Characterising ambient air quality over Gauteng using surface in-situ and remotely derived measurements

ZOZ Sibisi

orcid.org 0000-0003-2682-7551

Dissertation accepted in fulfilment of the requirements for the degree *Master of Science in Environmental Sciences* at the North-West University

Supervisor:	Dr B Language
Co-supervisor:	Dr A Ngie
Co-supervisor:	Prof RP Burger

Graduation May 2024

DEDICATION

I dedicate this Masters to my precious niece, ***Enelo Lujulil'uthando***, who was born on the same day as me (30 March) during the process of completing my Masters. Getting to share a birthday with you was the greatest gift and I will eternally be happy. Your birth was just the perfect epitome of God's grandeur. Thank you for giving me the extra push to continue to soldier on.

PREFACE

This dissertation is the original work of the author and has not been previously published at or submitted to any university institution. This dissertation was conducted between January 2021 to December 2023. It is written in accordance to the North-West University (NWU) rules and guidelines. The referencing style adopted in this study is the North-West University Harvard referencing style, found in the NWU 2020 referencing guide. The study was approved by the ethics committee to be of low risk (Ethics no: NWU-01281-21-A9).

During the course of this study period, the author has published and presented at the following conferences:

- Sibisi, ZOZ. 2021. Ambient air quality within urban communities of South Africa, in 3rd World Symposium of Sustainability and Science international conference, Brazil, 08 April 2021 - online.
- Sibisi, ZOZ, Language, B, Ngie, A, Burger, RL, Piketh, SJ. 2023. Validation of MODIS Aerosol Optical Depth Retrievals Using AERONET Data and Regression Modelling Techniques, in 54th Annual Conferences of the National Association for Clean Air, Polokwane, South Africa, 6-8 September 2023.
- Sibisi, ZOZ, Language, B, Ngie, A, Burger, RL, Piketh, SJ. 2023. Assessing Aerosol Optical Depth (AOD) over South Africa using an Assimilation Model and Satellite Data, in Society of Southern African Geographers Conference, University of Free State, Qwaqwa, South Africa, 2-4 October 2023.

ACKNOWLEDGEMENTS

I would like to thank everyone who contributed to this journey in completing my Masters. I would have not made it this far without all the help and patience I received.

- I would love to begin and thank my lovely parents, Collin Muziwandile Eugene Sibisi and Batseba Renolda Sibisi for being there for me. Your constant guidance and discipline shaped me and got me to where I am today. I am happy that I am able to continue raising the flag high for you. You are both my everything and I love you.
- To Nombuso and Calton, thank you for your love and support.
- To my heavenly grandparents, I love you all.
- I would like to thank all my supervisors, Brigitte Language, Adeline Ngie and Roelof Burger, for the constant help and support. And most importantly Brigitte, you have played a pivotal role in my academic career. Your supervision role from my Honours year (2020) till date (2023) has been impeccably played. Thank you.
- I would like to thank the National Research Fund (NRF) for funding me throughout my study period.
- I would like to express my gratitude to the Climatology Research Group (CRG) for also contributing to my funding.
- A special shoutout to Prof Piketh and Roelof for the endless opportunities I received in the CRG that aided in my academic advancement.
- I would like to thank the South African Weather Services (SAWS), more especially Mr Pieter Labuschangne for the valuable information that contributed towards this study as well as Dr Lynwill Martin.
- I would like to thank the South African National Space Agency (SANSA), more especially Dr Lerato Shikwambana, for providing assistance and input towards this study.
- I would like to acknowledge SAAQIS and the GES-DISC Interactive Online Visualization and Analysis Infrastructure (Giovanni) for the data used in this study.
- To the CRG ‘Peanut Gang’ (Felleng, Lerato, Gabby, Prince, Ribs & Monde), thank you for making this journey easier, fun times were had.
- And finally, to my high school Geography teachers (Mrs Bradford, Mrs Thiebout and Mrs Appelcryn), thank you for fostering the love for Geography in me.

ABSTRACT

Air quality studies using ground-based and satellite retrievals have gained popularity over the past two decades, offering many techniques and methods to explain atmospheric observations and contribute to climate predictions. Satellite retrievals assist us in capturing what ground-based measurements cannot; however, this is not to infer that one can be successfully employed as a proxy for the other. Ground-based measurements have presented major limitations on spatial coverage and data quality. This study aims to comprehensively assess air quality in the Gauteng Province, focusing on urban and industrial communities, by combining ground-based and remotely derived measurements. The first objective was to characterise ambient $PM_{2.5}$ levels in the Gauteng region using ground-based measurements by identifying and analysing temporal trends, diurnal variations, and seasonal fluctuations in ground-based $PM_{2.5}$ measurements. Secondary data for pollutant concentrations and meteorological data was used. The second objective is to evaluate the accuracy of satellite-derived data over the Gauteng region by validating the Aerosol Optical Depth (AOD) observations for $MODIS_{TERRA}$ and $MODIS_{AQUA}$ using the ground-based Aerosol Robotic Network (AERONET) observations. This objective used three retrieval algorithms, namely “Deep Blue (DB)”, “Dark Target (DT)”, and “Combined Deep Blue and Dark Target (DB & DT)”, to facilitate satellite AOD validations. The last objective was to characterise the aerosol loadings over Gauteng using satellite observations and modelled reanalysis data by comparing observations of AOD from $MODIS_{TERRA,}$ $MODIS_{AQUA}$, and the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) as well as characterising the aerosol type using the Ångström Exponent (α) for the monitored and unmonitored regions in Gauteng. ArcMap v10.8.1 was used for raster analysis and visualisation. The results show that $PM_{2.5}$ concentrations are still localised in low-income communities where major sources of pollution are biomass and waste burning, and domestic fuel burning, and industrial activities. When validating AOD, the “Combined DB & DB” algorithm performed best with the Machine Learning (ML) – SciKit-Learn model with a 90% model performance as opposed to the SLR and OLS statistical models. The retrieval algorithm performed better with satellite data from $MODIS_{AQUA}$. MERRA-2 was the most efficient in capturing aerosol species distribution. The findings highlight the importance of satellite validation studies as meteorological and topographic features impact different geographic areas. The findings suggest exploring different atmospheric measurement and monitoring techniques to fully understand the limitations in air quality studies and finding measures fully capture the $PM_{2.5}$ spatial distribution and AOD coverage.

Keywords: *Air quality, $PM_{2.5}$, MODIS-Terra, MODIS-Aqua, AERONET AOD, MERRA-2*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

ABBREVIATIONS & ACRONYMS

µm	-	Micrometre
µg/m^3	-	Micrograms per cubic metre
a.s.l	-	Above sea level
AAQ	-	Ambient Air Quality
AEL's	-	Atmospheric Emission Licenses
AERONET	-	Aerosol Robotic Network
AOD	-	Aerosol Optical Depth
APPA	-	Atmosphere Pollution Prevention Act
AQG	-	Air Quality Guideline
AQM	-	Air Quality Management
AQMP	-	Air Quality Management Plan
AQMS	-	Air Quality Management Systems
AQS	-	Air quality Standards
CF	-	Confidence Flags
CI	-	Confidence Interval
CoE	-	City of Ekurhuleni
CoJ	-	City of Johannesburg
CoT	-	City of Tshwane
CSIR	-	Council of Scientific and Industrial Research
CMF	-	Coarse-Mode Fraction
CTM	-	Chemical Transport Methods
CWV	-	Column Water Vapour
DB	-	Deep Blue
DRC	-	Democratic Republic of Congo
DT	-	Dark Target
DWMPC	-	Department of Waste Management and Pollution Control
EM	-	Environmental Management
EOF	-	Empirical Orthogonal Functions
EOSDIS	-	Earth Observing System Data and Information System
EU	-	European Union
FMF	-	Fine-Mode Fraction
GDP	-	Gross Domestic Product
HP	-	High Pressure
hPa	-	Hectopascal
HPA	-	Highveld Priority Area
IAP/HAP	-	Indoor Air Pollution/household Air Pollution
IDL	-	Instrument Detection Limit

IDP	-	Integrated Development Plan
IPCC	-	Intergovernmental Panel on Climate Change
IPWM	-	Integrated Pollution and Waste Management
L1	-	Level 1
L2	-	Level 2
L3	-	Level 3
LEO	-	Low Earth Orbits
LME	-	Linear Mixed Effect
LMIC	-	Low to Middle-Income Countries
LPG	-	Liquefied Petroleum Gas
LQA	-	Land Quality Assurance
LRM	-	Linear Regression Model
LULC	-	Land Use/ Land Cover
LUL	-	Land Use Regression
MAIAC	-	Multi-Angle Implementation of the Atmospheric Correction
METSI	-	Mooi River Ecosystem Trials for Scientific Investigations
MISR	-	Multi-Angle Imaging Spectroradiometer
ML	-	Machine Learning
MODIS	-	Moderate Imaging Spectro Radiometer
NAAQS	-	National Ambient Air Quality Standards
NaN	-	Not a Number
NASA	-	National Aeronautics Space Administration
NEMA	-	National Environmental Management Act
NEMAQA	-	National Environmental Management Air Quality Act
NDVI	-	Normalised Difference Vegetation Index
NIR	-	Near Infrared
NO_2	-	Nitrogen Dioxide
OLS	-	Ordinary Least Square
OQA	-	Ocean Quality Assurance
PA	-	Priority Area
PBL	-	Planetary Boundary Layer
PM	-	Particulate Matter
$PM_{2.4}$	-	Particulate Matter with aerodynamic size of ≤2.4
$PM_{2.5}$	-	Particulate Matter with aerodynamic size of ≤2.5
PM_{10}	-	Particulate Matter with aerodynamic size of ≤10
QA	-	Quality Assurance
RF	-	Rainfall
RFM	-	Random Forest Model
RH	-	Relative Humidity
RH	-	Relative Humidity

RMSE	-	Root Mean Square Error
SAAQIS	-	South African Air Quality Information System
SADC	-	Southern African Development Community
SAWS	-	South African Weather Service
SD	-	Standard Deviation
SDS	-	Scientific Data Set
S_e	-	Standard error
SFB	-	Solid Fuel Burning
SLR	-	Simple Linear Regression
SO_2	-	Sulphur Dioxide
SSA	-	Single Scattering Albedo
TSF	-	Tailing Storage Facility
Temp	-	Temperature
UAV	-	Unmanned Aerial Vehicle
UNFCCC	-	United Nations Framework Convention on Climate Change
UTC	-	Coordinated Universal Time
VIS	-	Visible light
VTAPA	-	Vaal Triangle Airshed was declared as a Priority Area
WBPA	-	Waterberg-Bojanala was declared as a Priority Area
WD	-	Wind Direction
WHO	-	World Health Organisation
WS	-	Wind Speed

1 INTRODUCTION & LITERATURE REVIEW

This chapter will provide an overview of the context and present a comprehensive literature review on ambient air quality. This chapter will emphasise particulate aerosols, legislation and policies pertaining to air quality and the application of remote sensing in this field. The conclusion of this chapter will encompass the problem statement, define the aim and objectives, the scientific importance, as well as the scope and limitations of the study. Additionally, a structure outline of the subsequent chapter's contents will be presented.

****

Air pollution is a subject which has been well studied, with studies evolving from the 20th century. Numerous studies have emerged, each providing results that add to a greater pool of knowledge. Clean air, as a basic human right, is challenged every day by degradational factors like urbanisation, motorisation, industrialisation, and globalisation (Gautam & Bolia, 2020). These factors contribute greatly to high levels of air pollution, which promote higher emission levels and exposure, ultimately resulting in increased negative health impacts.

There are global concerns arising surrounding the quality of air we breathe. These concerns are informed by the concentrations of harmful pollutants in the atmosphere, the declining quality of environmental health caused by air pollution, and, most significantly, the contribution that air pollution has to increased mortality rates (Tilt, 2019). Harmful pollutants, as declared by the World Health Organisation (WHO), include Sulphur Dioxide (SO_2), Carbon Monoxide (CO), Carbon Dioxide (CO_2), Particulate Matter (PM), Nitrogen Oxides (NO_x) and ground-level Ozone (O_3) (Rodríguez-Urrego$_a$ & Rodríguez-Urrego$_b$, 2020). Cohen et al. (2017) explain that exposure to excessive concentrations of these atmospheric pollutants contributes to early mortality rates. Studies have shown the congruence between poor air quality and its negative impact on human health (Rodríguez-Urrego$_a$ & Rodríguez-Urrego$_b$, 2020). Health impacts may include respiratory disorders and cardiovascular dysfunctions, which pre-existing health illnesses may exacerbate. Due to these negative impacts, ambient PM pollution has gained a rank as the fourth leading cause of premature death attributable to cardiovascular diseases (Vaduganathan et al., 2022).

Understanding ambient air quality can assist in informing many preventative measures being put in place to curb its harmful impacts. Traditionally, the use of ground monitoring networks has been a method used to understand ground-level air quality. This method has been useful in establishing local sources and estimating the air quality near the ground stations. However, this method does not fully account for changes in air quality in the upper atmosphere or planetary boundary layer (PBL). To account for this and get a full synoptic picture of a region's air quality, satellite remote sensing has been used (Engel-Cox et al., 2004$_a$). This method is a good way of remotely acquiring

data to establish aerosol concentration patterns in the atmosphere/ boundary layer, which atmospheric ground-monitoring stations cannot measure. This is advantageous as it can increase the scientific understanding of air quality as well can increase the scientific understanding of air quality and enable better policy decisions.

Air pollution is a global problem that has proven to have dire impacts on the environment, economy, and, most importantly, the health of people (Landrigan et al. 2017). Air pollution has been classified as the number one major contributor to severe health issues, as asserted in "*The Global Burden of Cardiovascular Diseases and Risk: A Compass for Future Health*" journal (Vaduganathan et al., 2022). This could be a result of high levels of exposure people have to indoor and ambient pollution (Leung, 2015). This is considered one of the driving forces in promulgating air pollution laws and policies to limit, if not eradicate, exceedances.

Though air pollution is a global problem, it mostly impacts developing countries like China and India. Zhang and Cao (2015) indicate this is attributed to increasing population rates and rapid economic expansion. With China having the world's largest population of almost 1.5 billion, it is bound to produce higher pollution levels and experience harsher impacts (Peng, 2011). Similar patterns can also be observed for local regions regarding high pollution levels.

Local regions like South Africa also experience high levels of air pollution (Henneman et al. 2016). This results from inadequate waste management practices, increases in industrial activities and the reliance on solid fuel burning as an energy source. All these activities lead to harmful health impacts, which increase susceptibility to chronic illnesses (Nkosi et al., 2018). Other impacts of air pollution affect the environment and the economy.

1.1 Impacts of air pollution (environmental, human and climate change)

The environment is rich in minerals and resources. However, misuse or exploitation of these minerals and resources prompts unsustainable energy uses, leading to air pollution. Air pollution impacts the environment in numerous ways which can lead to water contamination, soil degradation, damage to terrestrial ecosystems and vegetation (De Marco et al., 2019). Sources that influence and alter environmental systems are plenty, including indiscriminate waste disposal and the oil mining industry. These sources can impact the soil quality, contaminate underground water sources and negatively impact our air.

The impacts of air pollution on health are more dire in developing countries with a higher dependence on solid fuels. These health impacts tend to be exacerbated when a person has pre-existing chronic respiratory illnesses, which may ultimately lead to mortality (Pope et al., 2015). Factors that contribute to high air pollution exposure are living conditions, poorly ventilated spaces, small confined spaces, and indoor building materials (Mbandi, 2020). The long-term

effects of air pollution include cardiopulmonary diseases like asthma and lung cancer, and the short-term effects include heart rate variability, coughing and sneezing (Mocumbi et al., 2019).

Air pollution and climate change have a close relationship. Climate change encompasses various definitions. The Intergovernmental Panel on Climate Change (IPCC) defines it as "*any change in climate attributed to human activity or natural variability*" (IPCC, 2018), and the United Nations Framework Convention on Climate Change (UNFCCC) defines it as "*any change in climate that alters the composition of the atmosphere which may be attributed to direct and indirect human activity*" (IPCC, 2018). Human activities like burning fossil fuels industries and vehicle emissions release harmful pollutants and greenhouse gases into the atmosphere, leading to global warming. This ultimately leads to climate change, where natural (biodiversity) and human (health) systems are negatively impacted. Furthermore, due to environmental air pollution, climate change poses severe threats through the geographical distribution and alteration of infectious diseases (Manisalidis et al., 2020). This increases the triad risk relationship between hazards, vulnerability, and exposure.

1.2 Particulate matter ($PM_{2.5}$)

The classification of particulate matter is based on its physical and chemical characteristics. The physical characteristics are determined by the aerodynamic particle size, shape, and surface area; the chemical characteristics are determined by the particles' chemical mass, combustibility and hygroscopicity. These particles are aerosols which are suspended in the atmosphere, with a size fraction of ≤ 2.5 μm.

1.2.1 Sources of $PM_{2.5}$

Sources contributing to $PM_{2.5}$ include, but are not limited to coal power stations, wood heaters, traffic emissions, non-road diesel and domestic fuel burning (Chang et al., 2019). These abovementioned sources are mentioned because of their harmful health and environmental impacts when exposed to them. The health impacts include severe respiratory illnesses like asthma and cardiovascular diseases and the environmental impact can negatively impact crop yield and acidify bodies of water, creating an imbalance in aquatic ecosystems (Manisalidis et al., 2020).

Sources contributing to $PM_{2.5}$ need to be unpacked to fully comprehend the interplay between them and the extent of their impacts. Sources can be natural and anthropogenic (human-induced). Natural sources contributing to $PM_{2.5}$ include veld fires and decomposing waste materials, whereas contributory anthropogenic sources include solid fuel burning (SFB), industries and vehicular emissions (Adesina et al., 2020). The pollutant concentrations emitted from these

sources, particularly from anthropogenic sources, have harmful impacts on health when produced in excessive amounts.

Vehicle emissions (exhaust and non-exhaust) greatly harm human health, the economy, and the environment. This threat increases more with automobile ownership (Zhang et al., 2019). Traffic congestion has been the primary reason for increased vehicular emissions. Furthermore, traffic congestion is exacerbated by vehicle ownership, urban sprawl and commuting during peak hours (Cullen, 2019). Fuels like diesel, petrol and kerosene are mainly used in the transport sector, resulting in increased environmental pollution. Vehicular emissions can be categorised as exhaust and non-exhaust emissions. The former entails the combustion of diesel and gasoline that contribute to $PM_{2.5}$ emission concentrations. The latter entails emission from tire and brake wear, abrasion, and road particle suspension. In order to reduce the abovementioned emissions, factors like fuel type, vehicle type, engine type, and even road quality need to be considered.

Domestic solid fuel burning (SFB) is a leading source of increased indoor/household (IAP/HAP) $PM_{2.5}$ in developing countries. Domestic SFB has increasingly become a major energy source both locally and globally (Lee et al., 2015). Solid fuels like wood, coal, paraffin, and kerosene, also referred to as ‘dirty fuels’, are responsible for approximately 90% of household energy reliance in Sub-Saharan Africa and 75% for people residing in the rural areas of India (Faizan & Thakur, 2019) respectively. Exposure to IAP/HAP increases when people who live in poorly ventilated households/ areas use inefficient cooking and heating practices. Inefficient combustion and access to ‘clean fuels’ negatively impact human health and the environment. In this context, ‘clean’ fuels are defined as “energy sources used for cooking, heating, and lighting which aim to achieve a level of sustainability like the use solar electricity and liquefied petroleum gas (LPG)” and ‘dirty fuels’ are defined as “fuels that have high emission levels and a direct impact on human health” (Karakara & Osabuohien, 2020). Continuous use of ‘dirty fuels’ ultimately impacts and decreases air quality.

Burning materials accumulated from municipal waste (industrial, commercial, farm and domestic) is a process that poses harmful impacts on the environment and human health. Because municipal waste is highly heterogeneous, materials may contain organic and inorganic properties and when incinerated, they release harmful gases, toxins and pollutants like $PM_{2.5}$ into the atmosphere (Krecl et al., 2021). However, waste incineration has been seen as advantageous in reducing the demand for landfill sites and waste volume.

Mining has become an economic powerhouse for countries like South Africa, the sixth largest contributor to its Gross Domestic Product (GDP) (Cole & Broadhaurst, 2020). Having had economic advantages like employment and an increased GDP, the mining industry still poses deleterious impacts on the environment. The mining operations industry is categorised into

opencast mining and underground mining. The former deals with larger-scale operations, and the latter deals with small-scale operations. Opencast mining, or open-pit mining, extracts minerals from an 'open-air pit' through hauling, blasting, drilling, and loading (Ozdemir & Kumral, 2019). All these processes contribute to high $PM_{2.5}$, which negatively impacts human health. Underground mining is extracting minerals from below the Earth's surface. Similarly to opencast mining, excavation in underground mining occurs through blasting, grinding, crushing, and drilling, producing dust and fumes, with $PM_{2.5}$ being the dominant emitted pollutant.

1.3 History of air quality legislation in South Africa

Air quality has been a field of interest over many years in South Africa. Poor air quality has harmful impacts on human health and requires attention to combat this problem (Department of Environmental Affairs, 2018). Several strategies have been implemented to ensure poor air quality is curbed, one of them being the introduction of air quality management plans and environmental legislation (Department of Environmental Affairs, 2005). This will assist in ensuring that the air quality in South Africa is of a breathable quality.

Air quality management (AQM) has evolved drastically in South Africa. Severe health, environmental and socio-economic factors have been a driving force in managing the air quality (Department of Environmental Affairs, 2018). Managing air quality is an intricate process that needs to be handled delicately to ensure optimum success. Previously, strict air quality control and management practices were hindered because of a great need in legislation and cooperative governance (Scorgie, 2012). This was ameliorated when legal policies and frameworks emanated and were put in place (Department of Environmental Affairs, 2014).

AQM policies have been strategically put in place to combat poor air quality and have been amended over the years. Figure 1-1 shows the chronology of legislation that have been promulgated in South Africa pertaining to the protection, preservation and management of the environment and air quality, which are as follows: the atmosphere pollution prevention act (1965), The constitution of the republic of South Africa (1996), the national environmental management act (1998), the white paper integrated pollution and waste management for South Africa (2000), the national environmental management: air quality act (2004/2005) and lastly, the national framework for air quality management in South Africa.

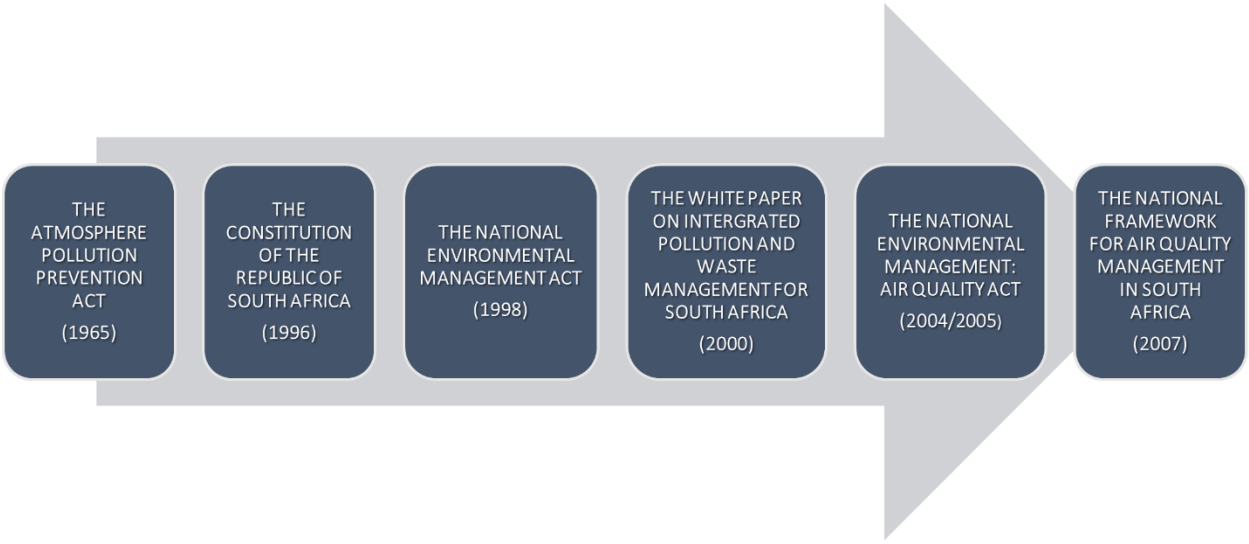


Figure 1-1: Timeline showing the process of the air quality legislation in South Africa (Source: Adopted from Naiker et al., 2012).

1.3.1 The Atmosphere Pollution Prevention Act (Act 45 of 1965)

The Atmosphere Pollution Prevention Act (APPA) was promulgated in 1965 as a means to regulate air pollution levels. As an approach to air pollution control, it focused primarily on industrial sources and gave little attention to vehicular emissions, noise, smoke, and dust control (Naiker et al., 2012). The pollution control approach of the APPA was flawed, but it also proved to be inadequate as it only devolved AQM among government and industry stakeholders. This meant that the local government minimal control in AQM juxtapose to the current Local Government: Municipal Systems Act, 2000 (Act 32 of 2000) which involves the local government.

The APPA has been ineffective in controlling air pollution for many years. This is because emission and source quantification, Air Quality Management Plans (AQMP's), ambient and emission standards, awareness, public consultation, compliance, and enforcement were not sought after (Tshehla & Wright, 2019). This can be seen in the emergence of air pollution hotspots all over the country. This was not prevented under the jurisdiction of the APPA and therefore, emergence of 'hotspots' are seen as a consequence of the APPA. The 2005 state of air report (2009) outlines that these emerging hotspots were because the APPA only focused on individual source emissions and did not recognise the accumulative impacts thereof. However, these impacts were ameliorated through the National Environmental Management: Air Quality Act (2004).

1.3.2 The Constitution of the Republic of South Africa (1996)

A shift in AQM was seen with the emergence of the South African constitution. The Constitution of the Republic of South Africa (Act 108 of 1996) emerged in 1996 and served as a legal framework for imminent acts as shown in Figure 1-1. This environmental legislation is found in

the South African constitution and was amended post the South African colonial system. Numerous acts emerged subsequently and the prominent one was the National Environmental Management Act (NEMA) in 1998, which was later amended in 2014 (NEMA, 2014). Therefrom, was the promulgation of the National Environmental Management: Air Quality Act (Act 39 of 2004). These legal frameworks and policies introduced new legislation that declared emission inventories, fuel parameters and emission standards. This paved a pathway to acquiring greater knowledge and amelioration to a new evolution of AQM in South Africa.

1.3.3 The National Environmental Management Act (Act 107 of 1998)

The National Environment Management Act came into action in 1998 where its legislation brought numerous policies into effect. Juxtapose to the APPA, NEMA stood firmly for social, economic, and sustainable development. Its policies included pollution prevention, environmental justice, waste minimization and avoidance. The NEMA legislation is there to enforce Section 24 of the Constitution of the Republic of South Africa (Act 108 of 1996) which states that "*everyone has the right to a clean a safe environment that does not impede on their health and well-being*". One of the NEMA principles stipulates that as part of environmental management (EM), "*people's social, psychological, physical and cultural interests need to take precedence*". This is promoted in their policies through legal conflict resolution, public participation, transparency, and co-operation among relevant governmental stakeholders.

1.3.4 The White paper on Integrated Pollution and Waste Management (IPWM) for South Africa

The white paper aims to curb poor air quality by integrating pollution and waste management as a holistic system. The primary objective of the white paper highlights that environmental degradation can be evaded through pollution and waste prevention. This paper is further described as "*a policy on pollution prevention, waste minimisation, impact management and remediation*". Instead of focusing on reducing and managing one aspect only, the IPWM sought to reduce both waste and pollution through integrative collaborative involvement of authorities and cooperative governance.

1.3.5 The National Environmental Management: Air Quality Act (2004/2005)

The National Environmental Management Air Quality Act (NEMAQA) has been a more comprehensive legislation that led to a significant transition in air pollution control. It has heralded new, improved, and sustainable AQM strategies. The main aim of this act is to protect the environment and give effect to Section $24_{(b)}$ of the Constitution of the republic of South Africa (Act 108 of 1996) which states that "everyone is entitled to a healthy and safe environment whereby their health and well-being is not compromised". The NEMAQA legislation has followed specific

strategies to ensure sustainable environmental development. These strategies include establishing national, provincial, and local standards, ambient air quality (AAQ) emission standards, AQMP's, declaring Priority Areas (PA's) and control measures through atmospheric emission licenses (AEL's). The NEMAQA has also made decentralization of AQM among relevant organs of state. These strategies signal an intense yearning to environmental quality that was not achieved in the APPA period.

1.3.6 The national framework for Air Quality Management in South Africa (2007)

The national framework serves as a blueprint to implement and achieve the objectives of AQM as declared in the NEMAQA (Department of Environmental Affairs, 2018). All technical aspects are liaised through the national framework whereby it provides the norms and standards and medium to long-term plans required for achieving AQM. The national framework is required by the NEMAQA to provide procedures in attaining compliance with AAQ standards, give effect to international agreements, set up national norms and standards, AQ monitoring, AQM planning and control emission from point and non-point sources. This framework was promulgated in 2007 and was governed by the guiding principles of Section 2 of the NEMA.

An Air Quality Management Plan (AQMP) is simply defined as a guideline used to develop, plan, manage and achieve an effective sound basis for air quality standards in an area of interest. This can be achieved by appointing reasonable measures in reducing air pollution and enhancing the air quality through protecting the environment (Air Quality Act of 2004) as this is one of the objectives of the National Environmental Management Act: Air Quality Act 39 of 2004.

To ensure that an area adheres to the air quality standards, plans need to be put in place in which one of them is an Integrated Development Plan (IDP). An IDP is a comprehensive framework between local and governmental stakeholders. This plan aims to develop, uplift, and improve the socioeconomic status of the residents as well as their quality of life. An IDP is crucial for municipalities so that local communities have access to basic needs. This is promulgated in Municipal Systems Act (Act 32 of 2000).

A sustainable practice in protecting the environment and surrounding communities is the establishment of by-laws (Department of Environmental Affairs, 2018). By-laws are regulations established by the municipality and relevant governmental stakeholders in which they need to ensure that they align with the Local Government: Municipal Systems Act, 2000 (Act 32 of 2000). According to Section $156_{(3)}$ of the Republic of South Africa, 1996, "*A by-law that conflicts with national or provincial legislation is invalid. If there is a conflict between a by-law and national or provincial legislation that is inoperative, the by-law must be regarded as valid for as long as that legislation is inoperative*".

1.4 Ambient air quality priority areas

Priority areas (PA) can be regarded as areas of concern. The PAs are determined from the observations made from air quality monitoring data obtained from monitoring stations. Some geographical regions exceeded the SAAQS and this was determined through emission inventories, area visibility and health related studies to air pollution (Moreoane et al., 2021).

These areas were then stated as National Priority Areas. In 2006, the Vaal Triangle Airshed was declared as a PA (VTAPA); in 2007, the Highveld was declared as a PA (HPA); and lastly, in 2012, Waterberg-Bojanala was declared as a PA (WBPA) (Feig et al., 2019; Department of Environmental Affairs, 2012).

As seen in Figure 1-2, the VTAPA covers parts of the Gauteng and the Free State province, the HPA covers some eastern parts of Gauteng and western parts of Mpumalanga, and the WBPA spans across some parts of the Limpopo and North West province. PAs are areas where there are high emissions of criteria pollutants due to industrial and non-industrial activities like domestic and biomass fuel burning, agricultural activities or transport emissions. These activities lead to harmful health impacts and decreased air quality in those areas. Because of PA's characteristics, this has birthed the urgent need for AQM in South Africa.

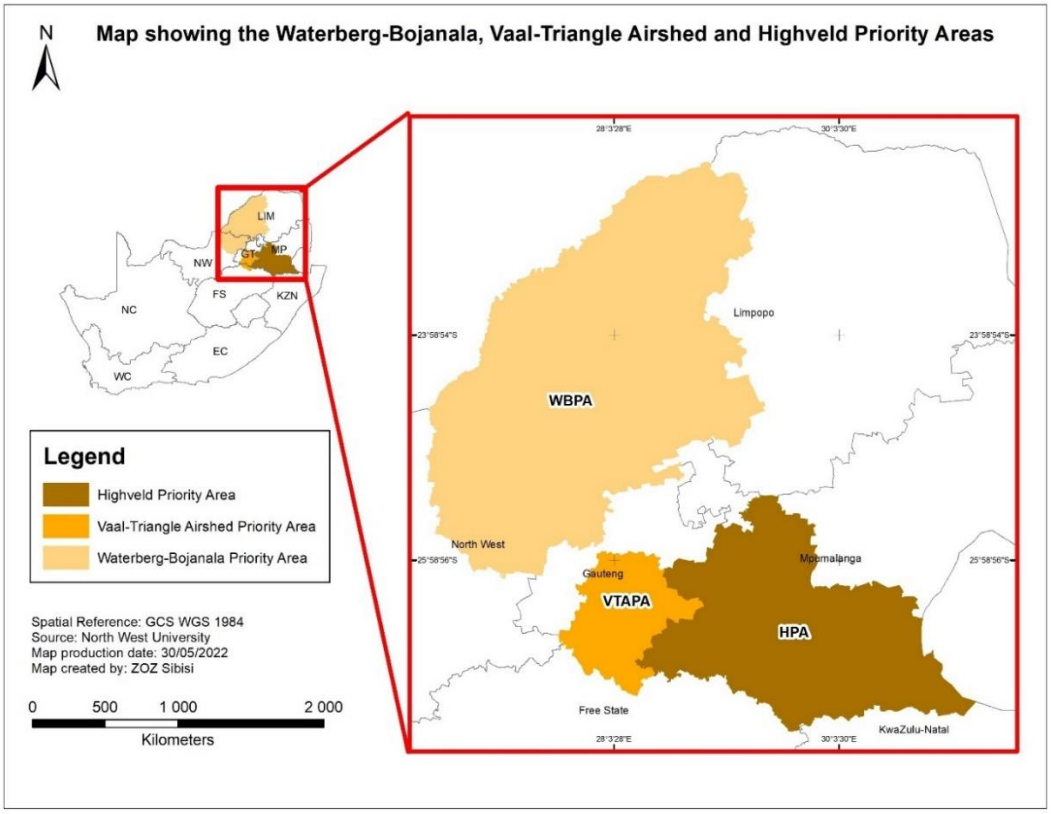


Figure 1-2: Priority areas in South Africa namely the HPA, VTAPA and WBPA.

Table 1-1 shows the WHO air quality guidelines (AQG). These worldwide guidelines were declared with the intention of combating global air pollution. These guidelines serve as a

benchmark for countries to set their AQS. The WHO AQGs offer guidance in combating the health and environmental impacts of air pollution. As seen in Table 1-1, the $PM_{2.5}$ annual standards are 5 $\mu g.m^{-3}$, and for 24H, the standards are 15 $\mu g.m^{-3}$. Compared to standards, guidelines are not legally binding as there are no legal consequences for countries that infringe on the AQGs. However, if countries do not set standards according to the WHO guidelines, they would be interfering with the basic health requirement for clean air because these guidelines assist policymakers assistance in setting appropriate targets.

Table 1-1: WHO air quality guidelines for $PM_{2.5}$ for averaging periods of a year and 24H (Source: Adopted from WHO global air quality guidelines).

Pollutant	Averaging period	Concentration
$PM_{2.5}$	1 year	5 $\mu g.m^{-3}$
	24 H	15 $\mu g.m^{-3}$

Table 1-2 shows the South African National Ambient Air Quality Standards (NAAQS). These standards are put in place to monitor and reduce the amount of air pollution in the atmosphere. Standards are there to limit criteria pollutant emissions, which ultimately should lessen the harmful impacts of air pollution on human health. The annual (1 year) and 24H concentration standards have changed with the compliance date starting from 1 January 2016 to 31 December 2029. $PM_{2.5}$ standards have decreased from 25 $\mu g.m^{-3}$ to 20 $\mu g.m^{-3}$ and from 65 $\mu g.m^{-3}$ to 40 $\mu g.m^{-3}$, respectively (DEA, 2012). The table below also shows the tolerable frequency of exceedances for a year to be zero and for 24H to be four. These tolerable exceedances are the recommended targets for $PM_{2.5}$.

Table 1-2: South African NAAQS) for $PM_{2.5}$ averaging periods of a year and 24H (Source: DEA, 2012).

Pollutant	Averaging period	Concentration	Tolerable frequency of exceedances
$PM_{2.5}$	1 year	20 $\mu g.m^{-3}$	0
	24 H	40 $\mu g.m^{-3}$	4

1.5 Meteorological parameters

Meteorological parameters are an important aspect to consider in air quality studies (Kayes et al., 2019). This is due to the fact that pollutant concentrations or transportation is driven by meteorology and understanding its impacts can assist in developing good air quality management systems. Meteorological parameters included in this study are temperature, relative humidity and rainfall. Wind speed and wind direction has been omitted from the analysis of this study, however, is included in this chapter and the discussion section in ***chapter 3***.

1.5.1 Temperature

Temperature plays a crucial role in the chemical reactions that occur in the atmosphere and this has made it an important meteorological parameter in air quality studies (Lui et al., 2020). Although some studies by, but not limited to Birim et al. (2023) and Olutola and Wichmann (2021) have shown temperature to not have a direct impact on PM, chemical reactions like photochemical smog may form (Leban et al., 2018).

1.5.2 Relative Humidity

Relative humidity is the amount of water vapour molecules present in the atmosphere. A high percentage of humidity in the atmosphere will increase the moisture content and a low percentage of humidity in the atmosphere means there will be less moisture in the atmosphere. The physical characteristics of dry aerosol particles like PM can be altered due the humidity in the atmosphere. The water vapour may attach to the particle which can increase its size. However, a local study conducted by (Olutola and Wichmann, 2021) proved that relative humidity has a negative correlation with PM.

1.5.3 Rainfall

Rainfall as a meteorological factor has an important role in the transport and dispersion of a pollutant. This is because it can remove PM and dissolve gaseous substances in the atmosphere but also deposit PM (Kayes et al., 2019). Although the latter may lead to the formation of acid rain, the former is beneficial in improving visibility.

1.5.4 Wind speed and wind direction

Wind moves in a horizontal direction and has a major impact on the air quality and also affect the pollutant concentration, dispersion and transportation (Garsa et al., 2023). A study by Janhäll (2015) shows that high wind speeds result in lower pollutant concentrations. This assists in reducing high concentration levels of a pollutant in the atmosphere by dispersing it and reducing concentration build-up of a pollutant in a certain area (Garsa et al., 2023). Wind direction is also an important aspect in air pollution transport. Analysis of seasonal wind patterns helps industrial planners to locate sources of air pollution in optimal locations in order to minimize their effect upon surrounding communities or the environment (Beelen et al., 2010).

1.6 Remote sensing

Remote sensing is the science and technology of acquiring geospatial data through nonphysical contact (Duan et al., 2020). It is also an assistive technological tool in providing spatiotemporal observations for vegetation, physical and atmospheric parameters, and everything pertaining to object space. It gathers object space information by measuring electromagnetic radiation and

interaction (Zhu et al., 2017). Collection of data is achieved through different platforms and needs to be pre-processed, visually analysed and interpreted prior to distribution.

Remote sensing platforms can be divided into three sub-categorises: ground-based, airborne and space-borne (Wójtowicz et al., 2016). Ground-based sensors may be handheld or ground/stationary instruments. The advantages of ground-based sensors are that they provide accurate measurements and have better spatial, temporal, and spectral resolution. The disadvantages are that it only operates on a small scale and does not account for Aerosol Optical Depth (AOD) properties found in the planetary boundary layer (PBL)/ mixing layer (Verma et al., 2019)

Airborne remote sensing is a non-ground-based platform that usually uses aeroplanes/ aircraft, and these aircraft are slowly being replaced with Unmanned Aerial Vehicles (UAVs), also known as drones (Aasen et al., 2018). UAVs are quite advantageous when it comes to flexibility. Their flexibility includes timing of measurements, height at which it flies and quick deployment. Images are also obtained with a high resolution. A disadvantage against UAVs is weather conditions and possible software malfunctions. Satellite remote sensing is acquiring geospatial data through space-borne sensors. The advantages are that it operates on a large scale and has a high spatial resolution. Satellite imagery is also advantageous because it can provide a complete micro-scale overview of a city or study area of interest. Satellite imagery can provide a comprehensive survey of the major sources of air pollution, the distribution pattern thereof, and key focus areas in reducing air pollution levels. Depending on the platform used, orbital times may be a disadvantage in satellite remote sensing.

There are four types of resolutions, which are spatial, spectral, temporal, and radiometric resolution, as outlined in Table 1-3.

Table 1-3: MODIS satellite resolution characteristics (spatial, spectral, temporal and radiometric resolution).

Type of resolution	
Spatial Resolution	250m (1 – 2 bands); 500m (3 – 7 bands);1000m (8 – 36 bands)
Spectral Resolution	36 spectral bands Band 1 – 2 (land/ cloud/ aerosol boundaries) Band 3 – 7 (land/ cloud/ aerosol properties) Band 8 – 36 (atmospheric water vapour/ cloud temperature/ cloud properties) Wavelength range (0.4 µm – 14.4 µm)
Temporal Resolution	1-2 days
Radiometric Resolution	12 bits

1.6.1 Spatial resolution

Spatial resolution is represents the Earth's surface in the form of pixels or cells. This representation shows the pixel size in relation to the ground representation; however, this representation is not a true geographical object. The appropriate spatial resolution is dependent on the pixel size. For example, the pixel size of an area may be defined as 1 m x 1 m or 250 m x 250 m. Therefore, the smaller the pixel size, the finer the spatial resolution, the greater the pixel size, the coarser the spatial resolution.

MODIS shows good spatial resolution capabilities. As shown in Table 1-3, MODIS has spatial resolutions of 250m, 500m and 1000m, indicating its ability to encompass finer spatial resolutions. A recent study by Gupta (2022) showed that MODIS had better spatial capabilities in conducting long-term trends in AOD across the globe. Another recent and local study by Yakubu and Chetty (2022) that aimed to assess the climatology of aerosol properties in South Africa over a decadal period agrees that MODIS possesses better spatial capabilities.

1.6.2 Spectral resolution

Spectral resolution is the sensor's ability to delineate fine wavelength intervals (Blue, green, red, NIR) on an electromagnetic spectrum. Electromagnetic waves have an impact on the spectral resolution. The longer the wavelength, the lower the spectral resolution; the shorter the wavelength, the higher the spectral resolution. This is also dependent on the number of bands for a particular wavelength. The more bands there are, the higher the spectral resolution (use of a colour film); the fewer bands there are, the lower the spectral resolution (use of a black and white film. Considerations of appropriate wavelength observations are crucial in air quality satellite studies mainly because of the satellite's ability to retrieve spectral data at different wavelengths. For example, if retrievals need to be made in the ozone, some aspects need to be taken into cognisance, like the ozone being comprised of four different spectral bands: Hartley (200 nm-310 nm), Huggins (310 nm-350 nm), Chappuis (380 nm-850 nm) and Wulf (750 nm-950 nm).

1.6.3 Temporal resolution

Temporal resolution is the time a satellite takes to make a complete orbit cycle. The temporal resolution also dependent on the type of sensor used, as MODIS has a short orbit cycle (1-2 days), and the MISR sensor has a longer orbit cycle (9 days). In air quality studies, it is more useful to use sensors that have a shorter orbit cycle. Local studies by Muyemeki et al. (2020) and Padayachi (2016) have used the MODIS sensor to assess and map PM concentrations in the Vaal Priority Area and Western Cape, respectively, because of their temporal resolution. MODIS has a shorter orbit cycle and can re-visit an area of interest almost daily, which serves as an

advantage as it provides more accurate diurnal patterns and improves ambient air quality monitoring.

1.6.4 Radiometric resolution

Radiometric resolution is the amount of data stored on a pixel in which a pixel represents the energy recorded. The imagery data in a pixel is stored in bits, whereby values to a selected power of 2 are assigned to them. For example, a 6-bit resolution has an assigned value of 26. This indicates that the sensor can store up to 64 digital values. Therefore, the greater the number of bits, the higher the radiometric resolution. MODIS sensor showed a radiometric resolution of 12-bits, which indicates a high radiometric resolution.

When conducting remote measurements and assessments of PM, specific criteria needs to be followed to ensure correct PM monitoring. Figure 1-3 shows this process, which can act as a guide to choosing the appropriate sensor. It is important to know which aerosol optical depth (AOD) property you want to measure. This is because Satellite AOD has been used for almost two decades as a measure/ proxy, considering the limitations thereof, for long-term ground-level PM estimations and has grown exponentially over the years (Xiao, 2018). Knowing which AOD property to use will govern the appropriate choice of sensor. With sensors having their strengths and limitations, the latter can be combated through the use of appropriate algorithms. After the algorithm for a specific sensor is chosen, dataset retrieval methods must be considered. Once this has been done, pre-processing and modelling methods must be considered.

1.6.5 Aerosol Optical Depth

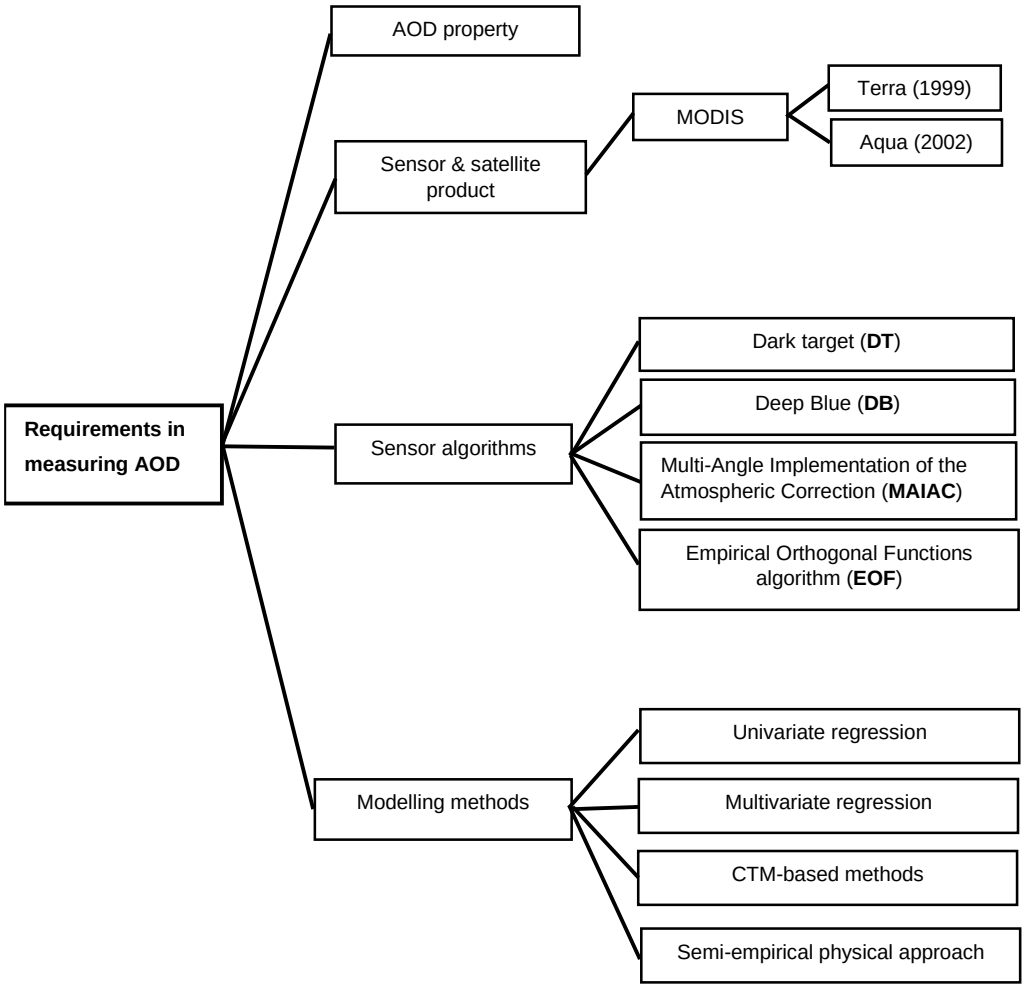


Figure 1-3: Conceptual diagram showing the requirements when measuring Aerosol Optical Depth using the MODIS satellite. (Source: Created by ZOZ Sibisi)

When observing PM, the Aerosol Optical Property is what is sought after. Table 1-4 shows the Aerosol Optical Properties that can be considered with observing PM. The chemical or physical property that needs to be observed will determine which aerosol optical property to use. For example, if aerosol mass or aerosol size distribution needs to be measured, AOD and spectral AOD can be chosen. However, a disadvantage in choosing this aerosol property is its hygroscopic nature (water vapour uptake). If the hygroscopic growth of an aerosol exceeds 80% of the relative humidity (RH), the true size, distribution, and aerosol density will be affected.

Table 1-4: Aerosol Optical Property of MODIS outlining its chemical/ physical properties.

Aerosol Optical Property	Chemical/ Physical property
Aerosol Optical Depth (AOD)	Aerosol mass
Spectral AOD	Aerosol size distribution
Ångström Exponent (α)	Fine/ coarse aerosol
Single scattering albedo	Aerosol chemical composition
Imaginary Refractive Index	Aerosol type
Scattering phase function	Particle shape

Van Donkelaar et al. (2006) describe AOD as measuring the light extinction by aerosol in an atmospheric column during the satellite-specific pass-over time. In contrast with the latest definition, Gui et al. (2021) describe it as the fundamental measurement of solar radiation attenuation by atmospheric aerosols along an entire atmospheric column. Despite the quindecinnial gap between the two definitions, they are both grounded on indispensable assumptions, that need to be made to avoid uncertainties in retrievals.

An early study by Tsay et al. (1991) and a recent study by Su et al. (2017) outlined two aerosol assumptions when working with AOD derivations. Tsay et al. (1991) maintain the assumptions of particle sphericity and aerosol external mixing. The former contends that particles are uniformly spherical and, the latter contends that aerosols are externally mixed without any interaction. Su et al. (2017) outline the assumption of confinement and homogeneity. The former contends that aerosols are vertically confined in a column at the top of the PBL, and the latter further extends that homogeneous mixing occurs there. These assumptions have been proposed to maintain aerosol characteristics as they constantly undergo physical and chemical alterations.

When assessing air quality, knowing which pollutant of interest will be measured or observed is required. Table 1-5 summarises the common platforms used to observe PM. It also provides a short description of the platforms.

Table 1-5: MODIS description, strengths and limitations.

Pollutant	**Satellite**
PM (AOD)	Moderate Resolution Imaging Spectroradiometer (MODIS - Aqua/ Terra)
Description	Measures properties of aerosols or solid particles in the atmosphere.
Strengths	Coverage, resolution, accuracy, calibration
Limitations	Bright surfaces, Cloud coverage

1.6.6 MODIS Terra and Aqua

Over the years, satellites have made improvements in deriving AOD. However, one specific satellite thrust its way into being the first satellite to successfully derive AOD measurements, MODIS Terra and Aqua (Chu et al., 2002). Su et al. (2017) and Wei et al. (2020) further elaborate that it is the most advanced instrument sensor for making long-term AOD measurements on a global scale. Table 1-3 outlines that the MODIS sensor measures aerosol properties in the atmosphere. Furthermore, its advantageous capabilities are that the sensor orbits the entire planet every 1 to 2 days at a distance of 705km from the earth's surface. MODIS is a sensor that is space-borne by satellites Terra (formally known as EOS AM-1) and Aqua (formally known as EOS PM-1), operated by the National Aeronautics Space Administration (NASA). The former was launched on the 18th of December in 1999 and has an equatorial pass-over time of 10:30, and the latter was launched a few years later on the 4th of May 2002 and has an equatorial pass-over time of 13:30.

Satellite platforms encompass many strengths and limitations; however, technological advancements over the years have enabled researchers to combat and overcome certain satellite retrieval limitations (Mutanga et al. 2016). Table 1-5 outlines the strengths and limitations of the MODIS. Strengths of the MODIS sensor include good coverage, resolution, accuracy and calibration. Certain algorithms must be applied to combat atmospheric, aerosol and surface parameters (Hashimoto & Nakajima, 2017). To combat the limitations outlined in Table 1-5, innovative ways of combating uncertainties and limitations are using an appropriate retrieval algorithm for better and more efficient data processing. These retrieval algorithms will be discussed in depth in the subsequent paragraphs.

For the MODIS sensor, the Dark Target (DT), Deep Blue algorithm (DB) and Combined Deep Blue and Dark Target (Combined DB & DT) is used to counteract surface parameters. The DT is used to retrieve AOD over dark surfaces/ vegetated areas, and the DB is used to retrieve AOD over bright land and ocean surfaces. The DT is normally applied if the area of interest is densely vegetated and to reduce the biases (Hu et al., 2021). This algorithm can be utilised for accurate AOD retrieval, by detecting the surface reflectance in the NIR band to approximate the surface reflectance in the visible band (Zhang et al., 2021). For the DT algorithm to be successfully employed, the sensor must have a wavelength of 2.1 µm to 2.2 µm. The DB algorithm is highly advantageous in enhancing the spatial continuity of AOD retrievals across space as it is used to retrieve AOD over bright surfaces. Frequently, the DT algorithm does not accurately retrieve AOD; therefore, the DB algorithm acts as a bridge. The Multi-Angle Implementation of the Atmospheric Correction (MAIAC) is another new image processing algorithm for the MODIS sensor that can retrieve AOD dark vegetated areas and bright surfaces at a high 1 km resolution (Pu & Yoo, 2021). It works by making simultaneous surface reflectance and aerosol retrievals. The Combined

DT & DB is an algorithm that is based on the strongest capabilities of DT and DB as theses MODIS products are continuously being improved to combat issues related to uncertainties (Filonchyk & Hurynovich, 2020).

Several aerosol products have been established in practice due to the rapid development of satellite platforms, sensors, and retrieval algorithms. There are now 21 aerosol databases that give data on various aerosol characteristics. MODIS, VIIRS, SeaWiFS, and Envisat/MERIS are among the 16 datasets offering sensor AOD data. Datasets containing aerosol physical and optical properties, such as the Ångström (α) Exponent (9 datasets) and Absorbing Aerosol Index (5 datasets), have been created in addition to the regularly used AOD products. Despite the importance of AOD and fine-mode fraction (FMF) in mapping fine PM, only datasets from MODIS and VIIRS offer FMF because of the difficulty in retrieving these two metrics.

Univariate regression is employed searching for a linear relationship between two variables, such as, the correlation between PM mass concentrations and AOD satellite observations. This can be achieved through linear (least squares) and non-linear approaches (quadratic methods). Juxtaposing the univariate regression modelling, multivariate regression modelling uses of a number of variables to better understand the true relationship between PM and AOD and solve the complexities thereof. The considered variables may be geographical properties, meteorological data or emissions data. Chemical Transport Methods (CTM) are useful in establishing near-surface PM concentrations. The observed AOD and simulated PM data are assumed to have a constant scale factor. Lastly, the semi-empirical physical approach theoretically makes empirical derivations of PM from AOD. This is achieved through understanding the empirical relationship between AOD optical properties and PM concentrations.

1.7 Applications of remote sensing in air quality studies

The application of remote sensing has proven to be a great tool in mapping or displaying the spatial and temporal distribution of air pollutants. However, satellite measurements alone have a role in air quality studies but cannot operate independently as an observing system. Applications such as event identification, transport and atmospheric composition determination are strengths of satellites.

Remote sensing has been used extensively in air quality studies. However, it has a major setback in yielding quantitative information in air quality studies instead of ground-based measurements. The integration of remote and ground measurements is useful in enabling better policy decisions to be made.

In South Africa, the applicability of remote sensing to air quality studies has come a long way. Choosing the appropriate requirements for making AOD observations is important as it eliminates

the probability of overestimations or underestimations. A study by Hersey et al. (2015) aimed to assess the air quality in South Africa through the use of both ground-based and remote measurements. To combat overestimations and underestimations, this study made use of more than one sensor (MODIS, MISR, TOMS), used the Ångström Exponent (α) of 550 nm - 865 nm and 470 nm - 660 nm for the $MODIS_{(AQUA,\ TERRA)}$ platforms, respectively. This study also used multivariate regression modelling methods to understand the impact meteorological parameters would have on AOD observations and PM concentrations. Furthermore, a study by Shikwambana and Tsoeleng (2020) further extended the study by investigating how population growth and land use impact air quality.

Some local studies incorporate spatio-temporal and meteorological parameters impacting the air quality in South Africa. A plethora of local studies using satellite-based data have been conducted by Saucy et al. (2018), Muyemeki et al. (2020), Zhang et al. (2021) and Mahesh et al. (2022).

Saucy et al. (2018) made use of land use regression modelling in $PM_{2.5}$ estimations in the Western Cape province which showed a poor model performance of 0.29 which is expressed to be a result of the variations found in the spatial determinants of particles. Luckson et al. (2020) conducted a similar study to Hersey et al. (2015) which was to evaluate the suitability of using satellite data as a proxy to ground-based $PM_{2.5}$ in the VTAPA. Results indicate a poor agreement between satellite retrieval and ground-based $PM_{2.5}$ data.

Zhang et al. (2021) made use of $MODIS_{(TERRA,\ AQUA)}$ satellite data to estimate $PM_{2.5}$ concentrations in the South African industrialised highveld region. The study made use of a machine learning model to achieve its study objective. Zhang et al. (2021) made use of a Random Forest Model (RFM) which performed extremely well. A recent study that explored the relationship between satellite AOD with PM, in Durban was conducted by Mahesh et al. (2022) with the aid of statistical modelling. The study made use of the Linear Mixed Effect (LME) Model which considered influential parameters (distribution of aerosols in a vertical column, optical aerosol properties and meteorology). The model results showed a good model performance for $MODIS_{(TERRA,\ AQUA)}$.

1.8 Problem statement

The justification of this research study is that satellite retrievals have their shortcomings, whereby discrepancies are often found between ground-based and satellite data (Yu et al., 2020). Relying only on ground-based measurements is not sufficient to make inferences on the status of air quality as the quality of air at higher atmospheric levels needs to be accounted for. Using satellite observations only also presents a short-coming when ground-level air quality needs to be accounted for as well. To overcome these short-comings, ground-based measurements will be used in tandem with satellite observations. Balsamo et al. (2018) also contend that this technique is beneficial in providing more understanding on the complex Earth/ climatological behaviours.

Such a technique ensures any spatial or temporal short-comings from either ground-based or satellite observations overcome (Arabi et al., 2020).

Routine air quality monitoring is often not possible with some sensors because of their spatial and spectral capabilities as well as the biases of the pollutants when using certain satellite products. The MODIS sensor is capable of covering a big geographical area while producing good images. This is because of its good spatial coverage and the 36 channels it encompasses, as most sensors have 4 to 8 channels.

1.9 Aim and objectives

This study aims to comprehensively assess air quality in the Gauteng Province, focusing on both urban and industrial communities, by utilizing a combination of ground-based (air quality monitoring stations) and remotely-derived (MODIS and AERONET) measurements.

The aim will be fulfilled through three objectives, namely:

(i) Quantify and characterise ambient $PM_{2.5}$ levels in Gauteng region using ground-based measurements, by:

   a. identifying and analysing temporal trends, diurnal variations, and seasonal fluctuations in ground-based $PM_{2.5}$ measurements.

   b. investigating and characterising the spatial distribution of ground-based $PM_{2.5}$ measurements.

(ii) Evaluate the accuracy of satellite-derived PM data in understanding the spatiotemporal variations over the Gauteng region, by:

   a. validating the AOD observations for $MODIS_{TERRA}$ and $MODIS_{AQUA}$ using AERONET observations.

(iii) Evaluate the aerosol loadings over Gauteng using satellite observations and modelled reanalysis data, by:

   a. Comparing observations of AOD from $MODIS_{TERRA}$, $MODIS_{AQUA}$, and MERRA-2 for the monitored and unmonitored regions in Gauteng.

1.10 Significance of the study

The findings obtained from this study will contribute to the scientific pool of knowledge by:

Expanding our knowledge on the spatio-temporal distribution of aerosols over the Gauteng Province. Improving knowledge and understanding on the interaction of $PM_{2.5}$ and meteorology, highlighting the possible confounders. The study will outline significant satellite limitations that are

assumed to true to the local context of South Africa – Gauteng and also present how different satellites perform and using appropriate satellite products to overcome the posed limitations.

1.11 Scope, assumptions and limitations of the study

1.11.1 There are assumptions that were considered and maintained in this study which are:

The stability of pollution sources assumes that sources of pollution, particularly $PM_{2.5}$, remain relatively constant during the study period. To maintain consistency in meteorological data this study assumes meteorological data from different stations are comparable and consistent over time. Another assumption is the confounding effect of time on the association between PM and meteorology. The representativeness of satellite data assumes that there is homogeneity in aerosol properties implying homogeneous aerosol mixing in the vertical atmospheric column. The assumption maintained for this study is that there is a certain degree of homogeneity in the aerosol properties across the studied region. Another maintained assumption is that aerosol particles in a vertical atmospheric column embody a spherical shape.

1.11.2 The limitations encountered in this study are:

There was limited consideration of all environmental variables affecting $PM_{2.5}$ and aerosol distribution (e.g., topography, urban heat island effects). There was some exclusion of certain types of pollutants and environmental factors not covered in the study. Secondary particulates were also not accounted for in the study. The spatial and temporal coverage of data was based on satellite and ground-based data. The satellite passover times were also a limitation and generalisations/ applicability of validation results needed to be made with utmost caution. The accuracy of satellite observations was limited due to the discrepancies between satellite-derived AOD measurements and ground-based observations and the potential for plumes in the middle parts of the troposphere. There were challenges in generalising findings to other regions or different climatic conditions. Data availability was an issue which presented with inconsistencies and gaps in data availability, especially from remote sensing sources. Modelling and statistical methods also have limitations inherent in the modelling and statistical methods used for data analysis.

The study mainly focuses on evaluating ground-based and remote-based measurements to better understand the ambient air quality in Gauteng and the correlation thereof. The choice of air quality monitoring stations was not based on specific criteria but rather on data availability. The choice of the sensor (MODIS) was based on its spatial, spectral and temporal capabilities. These capabilities assist in making AOD retrievals easier as appropriate algorithms and calibrations will be put in place. Moreover, the results and interpretations made in this study should only be

considered to be that of the study and not a generalisation outside the study bound as different areas have dissimilar inherent conditions.

1.11.3 Ground-based data limitations:

$PM_{2.5}$ will be the main pollutant of focus in this study because of harmful health impacts. $PM_{2.5}$ shows the concentrations at ground level, not fully capturing concentrations at the upper atmospheric boundary layer, in which AOD efficiently does. Measurement methods (use of instrumentation) are not uniform across all monitoring stations. Data and concentration values recorded by different instruments present their own limitations as the same instrument is not utilised for all monitoring stations in the study area, furthermore, a common instrument detection limit is set for all instruments. Spatial resolution of the stations present limitations for the statistical analyses in this study. For the meteorological data, wind speed and wind direction were omitted from the data analysis (windroses) however a heatmap with temperature, relative humidity, rainfall and wind was included. The sole purpose of including wind as a meteorological parameter is to assess if there is a direct link with $PM_{2.5}$ and not to see the general wind conditions as they would have not been a true representation of my results seeing as this study is covering 17 ground-based monitoring stations over a five year period. Averaged wind data for all 17 stations would have not provided the study with the scientific justification.

1.11.4 Remotely-derived measurements limitations:

There is a likelihood of vertical inhomogeneity in the aerosol column properties; hence certain assumptions have been adopted to combat this. Aerosol column properties are not entirely true representations of ground-based $PM_{2.5}$ concentrations. Satellite temporal resolution might be a limitation as the sensor has specific local pass-over times (Terra – 10:30 LT; Aqua – 13:30 LT). These pass-over times do not coincide with wintertime PM peak concentration times as concentrations reach minima and maxima in the morning and evening, and the sensor does not observe these. Only Machine Learning (ML) algorithms commonly used in literature have been applied to this study. ML is a branch of artificial intelligence and uses algorithms in computer science to conduct statistical analysis and make informed references from them. In ***chapter 4,*** this study uses AERONET sites (Pretoria_CSIR_DPSS and Pretoria_CSIR_EC) to fulfil objective ii. However, these AERONET sites are considered background sites in which from making general inferences on the air quality in residential-based areas should be refrained unless HYSPLIT trajectory modelling is conducted to discern the direction, transportation, and dispersion of air pollutants. The AERONET sites used to fulfil objective ii only have four years worth of data (Pretoria_CSIR_DPSS: 2017 to 2018 and Pretoria_CSIR_EC: 2020 to 2021), so the year 2019 was excluded in the analysis conducted in ***chapter 4.***

1.12 Outline of chapters

This dissertation is made up of six chapters, which are outlined as follows:

(a) Chapter 1: Overview

- This chapter provided the introductory overview of this research project and the literature review.

(b) Chapter 2: Study area description, data acquisition and methodology

- This chapter will provide information on the methods used, data quality control and analysis.

(c) Chapter 3: Characterising ground-based $PM_{2.5}$ concentrations in the Gauteng Province

- Results will be based on the $PM_{2.5}$ concentrations, observed at selected monitoring station, in the Gauteng province. A discussion on the results will be provided.

(d) Chapter 4: Evaluating Satellite Observations in the Gauteng Region

- Results will be based from evaluating the Aerosol Optical Depth (AOD) from both satellite observations to ground-based measurements at two sites in Gauteng. This chapter will make use of remote data ($MODIS_{TERRA,\ AQUA}$) and AERONET. This chapter also aims to use three different retrieval algorithms, further expand, and evaluate the suitability of using statistical and Machine Learning models for areas with sparse ground observations.

(e) Chapter 5: Evaluating aerosols over Gauteng using satellite and re-analysis model data.

- Results will be obtained using satellite imagery by providing a visual analysis of how aerosol loading in unmonitored regions within Gauteng look like.

(f) Chapter 6: Summary and conclusion

- It will outline the study assumptions, a summary of the findings for objectives i, ii, and iii, the study recommendations, and the contribution to the current body of knowledge.

****

This chapter provided the background and literature review on air quality from a global and local perspective. The prevalent problems encountered in air quality studies using ground-based and remote measurements were provided, the aim and objectives, the study significance and limitations were provided within the context of this research study. The next chapter will provide the methods undertaken in this study.

2 STUDY AREA CHARACTERISATION, DATA ACQUISITION AND ANALYSIS

This chapter summarises the methodologies applied in this research study to achieve the set objectives. A comprehensive outline of the research design, study area, data collection, data quality control and data analysis will be provided. Arithmetic formulae and calculations are accounted for and explained in detail.

****

2.1 Study area characterisation

This section discusses the study area characterisation in which Figure 2-1 provides a detailed conceptual diagram that outlines the structure of the research process undertaken. South Africa comprises nine distinct provinces, Gauteng, with the smallest land area of 18,178 km^2 (Stats SA, 2016). Gauteng is also the only landlocked province with no foreign border surrounding it. Despite it being the smallest province, it is considered South Africa's political capital and Africa's economic powerhouse as it contributes to 34.4% of the country's national Gross Domestic Product (GDP) (Moeletsi & Tongwane, 2020). Furthermore, the country's economic sector (storage, transport and communication) contributes 35.6% of the national total. The Gauteng province is home to the country's largest city, Johannesburg, and the administrative capital, Tshwane (CoJ report, 2007). This province also houses mining headquarters companies, a major contributor to increased air pollution levels.

The Gauteng province is made up of three Metropolitan areas, namely the City of Tshwane (CoT), the City of Johannesburg (CoJ) and the City of Ekurhuleni (CoE). The largest metropolitan (CoT), as shown in Figure 2-2 forms Gauteng's local government. The province is further divided into two district municipalities: the West Rand and Sedibeng. The local municipalities found in the former are Merafong City Local, Mogale City and Rand West City, while the latter includes Emfuleni-, Lesedi- and Midvaal- local municipalities.

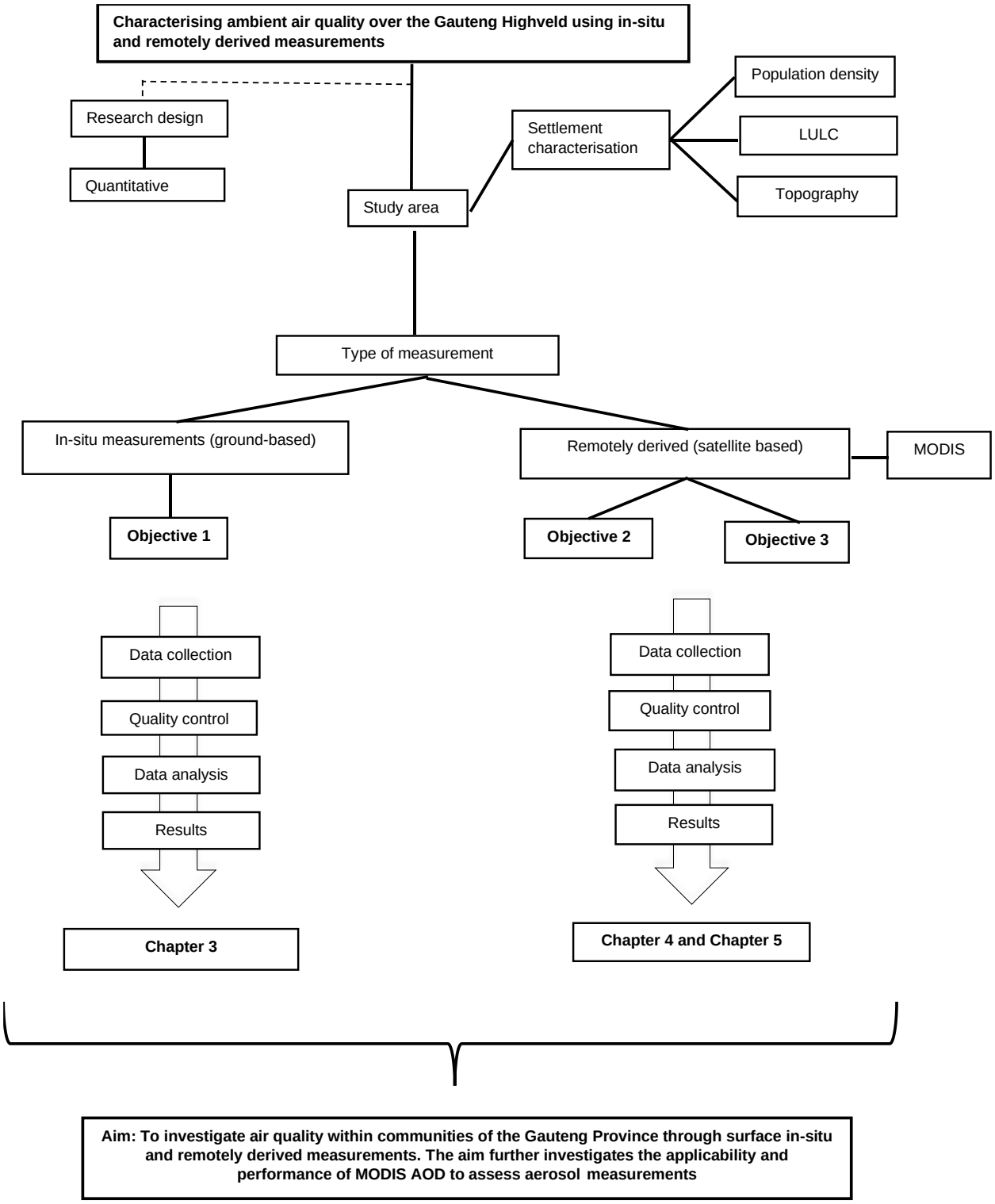


Figure 2-1: Conceptual model outlining the research process undertaken in this study.

Of the three metropolitan municipalities, CoT is one of the largest metropolitans in the northern part of Gauteng, as shown in Figure 2-2 , and forms the largest part of Gauteng's local government (Wright et al., 2011). Prominent air pollution-contributing activities in the CoT are industrial processes, power generation, vehicle emissions and household solid fuel burning (City of Tshwane, 2009). With CoT being the largest metropolitan area, its air quality is mainly impacted

by its proximity to Johannesburg industries (Matyeni, 2021). This is not the only prevalent problem because parts of the CoT are bordered by the Highveld Priority Area (HPA), Waterberg-Bojanala Priority Area (WPBA) and the Vaal-Triangle Air-shed Priority Area (VTAPA), which have an impact on its air quality.

CoJ is situated in the centre of Gauteng. It is a highly industrialised area due to manufacturing, construction, petrochemical, steel, brick and mining activities (Lourens, 2012). The CoJ also has an intricate network/ road system, which results in increased road and traffic emissions. This network system consists of national roads, freeways and other arterial roads, contributing to almost 80% of CO_2 from traffic emissions (Moyo et al., 2020).

The CoE consists of low-income, medium to upper-income residential areas, urban, and roadside traffic areas and is located in the central-eastern part of Gauteng. Air pollution sources stem from industries' domestic fuel burning, industrial fuel burning, mine dumping, mine tailings, and vehicle emissions.

Sedibeng district is located in the southern parts of Gauteng. Its pollution levels are expected to be high as parts of Sedibeng (Emfuleni local municipality and Midvaal local municipality) fall under the declared priority area, VTAPA (Moreoane et al., 2021). With parts of Sedibeng falling under the VTAPA, emission sources are most likely to emanate from collieries, industries, coal-producing power stations and quarries (Moreoane et al., 2021). Local municipalities like Emfuleni and Midvaal do not fall under the VTAPA; however, there is a high conglomeration of low-income settlements that use solid fuels as a source of energy for space heating and cooking. Such contributes to air quality, and high population densities exacerbate air pollution's health and environmental impacts.

The West Rand, consisting of three local municipalities, namely Mogale City, Rand West City and Merafong City, is situated in the western part of the Gauteng province. Air pollution sources identified in the West Rand are textile manufacturing, mining, mine tailings, waste dumping and household solid fuel burning (Ginindza & Muzenda, 2016; West Rand District Municipality AQMP, 2010). These sources contribute to air and dust pollution, affecting air quality in the West Rand.

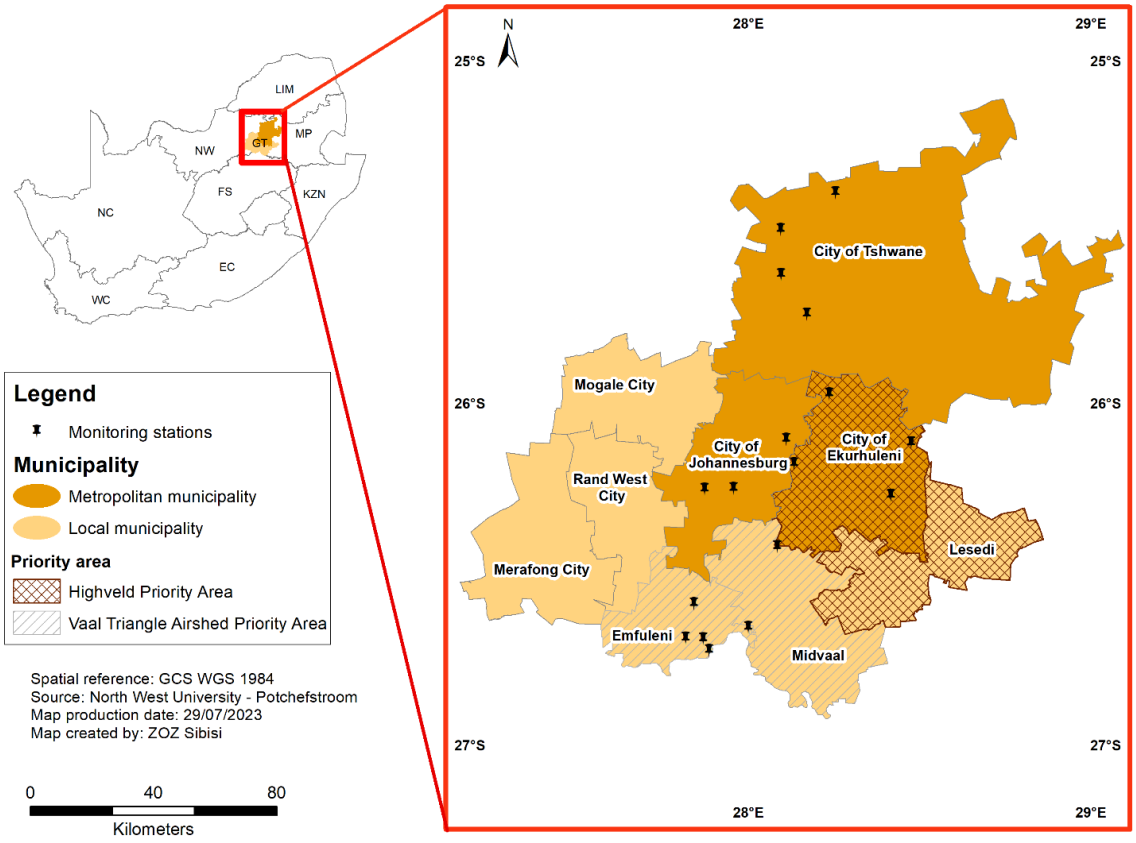


Figure 2-2: Study area map with location pins indicating different $PM_{2.5}$ monitoring stations. The brown shaded area represents part of the Highveld Priority Area, and the grey shaded area represents the Vaal Triangle Airshed Priority Area.

2.2 Emission sources in Gauteng

For a small area like Gauteng, the epicentre of Africa's economic hub, all these factors, especially urbanisation, feature strongly in this regard. There are numerous identified emission sources in the Gauteng province including: vehicle emissions, domestic fuel burning, biomass burning, mining, industries, quarry activities, factories and fugitive dust. Waste disposal is also a major contributor to air pollution in Gauteng.

2.2.1 Vehicle emissions

Traffic and road emissions have been shown to produce approximately 17% of the global greenhouse gas (GHG) footprint and 21% of global energy consumption (Contestabile et al., 2017). These numbers are not far off for South Africa as the transport sector contributes almost 13% of the country's GHGs. Private vehicle ownership has increased as urbanisation and socio-economic development heighten, contributing to 41% of the country's total road emissions (Moeletsi & Tongwane, 2020). In South Africa, the Gauteng province is the largest contributor to road and vehicle emissions. Tongwane et al. (2015) asserted that Gauteng is responsible for 33%

of the country's road emissions. This is because the public road network system was historically designed for buses, taxis and train networks (Bubeck et al., 2014). A study by Tomaschek et al. (2012) showed that the Gauteng province produced 92% of total emissions in 2007. In 2009, motorcars alone contributed to 47% of road emissions, increasing air pollution levels (Tongwane et al., 2015).

2.2.2 Mining

The mining industry in Gauteng has had a paramount impact on South Africa's national economy as it was and is still responsible for creating a labour-intensive market. One prominent contributor to Gauteng's economic upliftment was the gold mining in the Witwatersrand basin (Grab & Knight, 2015). Gold mining at the Witwatersrand basin became a cornerstone in forming the country's metropolis and a prominent goldfield (Tucker et al., 2016). Subsequent to the discovery of the Witwatersrand basin, more mines like the West and East Rand Goldfields and Durban Roodepoort Deep mine were discovered (Tucker et al., 2016). Despite a reported contribution of 6.8% to South Africa's GDP, the mining industry in Gauteng has also managed to increase ambient air pollution levels and exacerbate respiratory illnesses (Mpanza et al., 2020).

The mining basins also pose negative environmental and health impacts. Environmental impacts like flooding can result in acid mine water accumulation, which can flow into water and soil systems and, if in contact with or consumed by humans, can result in damage to the optimum functioning of the human nervous system (Njinga & Tshivhase, 2017; McCarthy, 2011). Furthermore, the mining industry has experienced some disadvantages, such as mine tailings, a by-product of mining activity. Mine tailings have a major contribution to the resuspension of dust, which can harm residents living nearby. Nkosi et al. (2017) also outline that schools near mine tailings have impacted school-going children, especially those prone to respiratory illnesses like asthma.

2.2.3 Industrial activities like factories

Air quality has deteriorated over the years due to industrial activities, which have been acknowledged to be the second most important contributing activity in Gauteng, according to a study by Lourens (2015). Areas like Tshwane and Johannesburg are the most affected because of their proximity to industrial areas. Mohlala (2020) also asserts that CoJ's air pollution sources stem from industrial activities. Furthermore, a report by Kaziboni et al. (2015) focused on three industrial nodes in Gauteng and their contribution to air pollution. These industrial nodes were Wynberg, Industria West and Aeroton. These nodes are close to all three Metros, namely, CoT, CoJ, and CoE, which greatly contribute to air pollution. Industrial activities in Gauteng include

power generation, petrochemical, steel industries, brick manufacturing, and mining. Parts of the HPA and VTAPA are also highly industrialised, contributing to Gauteng's overall air pollution.

2.2.4 Domestic fuel burning

With Gauteng housing the country's largest population, domestic fuel burning contributes majorly to Gauteng's air pollution (Chiwewe & Ditsela, 2016). With an increasing population, the Gauteng province has struggled with housing space, leading to informal and illegal dwellings. The proliferation of informal settlements is, more commonly than not, a phenomenon that alters the social and economic system as it continually changes (Lekonyane & Disoloane. 2013). The 2019 general household survey indicated that almost 17% of recorded households in CoT, CoJ and CoE fall under the class category of informal settlement. The percentage distribution of informal dwellings in CoT, CoJ and CoE is 16.4%, 19.1% and 18.4%, respectively. These dwellings are often unelectrified and use solid fuels like paraffin, wood or coal as primary energy sources for cooking and space heating, contributing to air pollution. Similar patterns are seen in low-income communities where domestic fuel burning still precedes electricity use (Xulu et al., 2020).

Surface in-situ measurements (*hereafter referred to as ground-based measurements*), having their strengths and limitations, have been used for decades in air quality studies. Its strength is the ability to give local point sources surrounding the area, measure the concentration of pollutants emitted, and if they adhere to the AQ standards set. The major limitation of ground-based measurements is their inability to capture full spatial coverage. The mixing layer/ vertical profile is where AOD observations are made, which ground-based stations cannot achieve. This is why these measurements are often coupled with remote-based measurements to acquire a holistic picture of the air quality at different atmospheric levels because, often, based on a geographical region of study, ground-based stations may not be able to offer the full spatial variability of $PM_{2.5}$ concentrations.

Ground-based measurements are often acquired at reputable sites. Some provide quality-controlled data, and others do not. Common requirements for using quality ground data are to use hourly resolutions to perform diurnal variations, remove all negative and null values, and consider stations with 70% data retention. A local study by Muyemeki (2020) used ground stations that have been operational since 2007. This unfolds a question in the literature about how air quality studies were conducted before to 2007 using ground station data. Another mindful aspect to consider in South African air quality studies using ground-based stations is that the ground stations are often placed in areas with high pollution levels (priority areas) or densely populated areas, not in rural areas.

2.3 Population and demographics for the province of Gauteng

2.3.1 Gauteng's population

Gauteng houses 26% of the country's population, which is 15,5 million, through which its population increased by 52% from 1993 to 2019 (Katumba & Everatt, 2022). This population increase may be attributed to migration seeking better economic opportunities and better educational and health services. CoT has a population of 3.6 million as opposed to CoJ, which has a population of 5.7 million, as shown in Table 2-1. CoE has a relatively high population of 3.8 million. West Rand and Sedibeng have lower population statistics than the metropolitan municipalities. The West Rand has a population of just below 950 000, and Sedibeng's population is just above 950 000. A possible reason for the high population density in CoJ is the urban sprawl experienced during the democratic period. This urban sprawl poses environmental fragmentation, which contributes to increased air pollution.

Table 2-1: Population statistics outlined for all metropolitan- and district municipalities in the province of Gauteng, along with the area in km^2 of each metropolitan and district municipality (Stats SA, 2011).

Municipality	Population	Area (km2)
City of Johannesburg	5 738 536	1 644
City of Ekurhuleni	3 888 873	6 297
City of Tshwane	3 649 053	1 975
Sedibeng	959 520	4 172
West Rand	940 135	4 087
Total	15 176 117	18 175

Figure 2-3 shows how the population is distributed per municipality, which tally with the information in Table 2-1. CoT makes up 24% of the population, CoJ 36%, CoE 26%, while Sedibeng and West Rand both make up 7% of Gauteng's population. Data was retrieved from the 2011 National Census Statistics of South Africa (Stats SA, 2011).

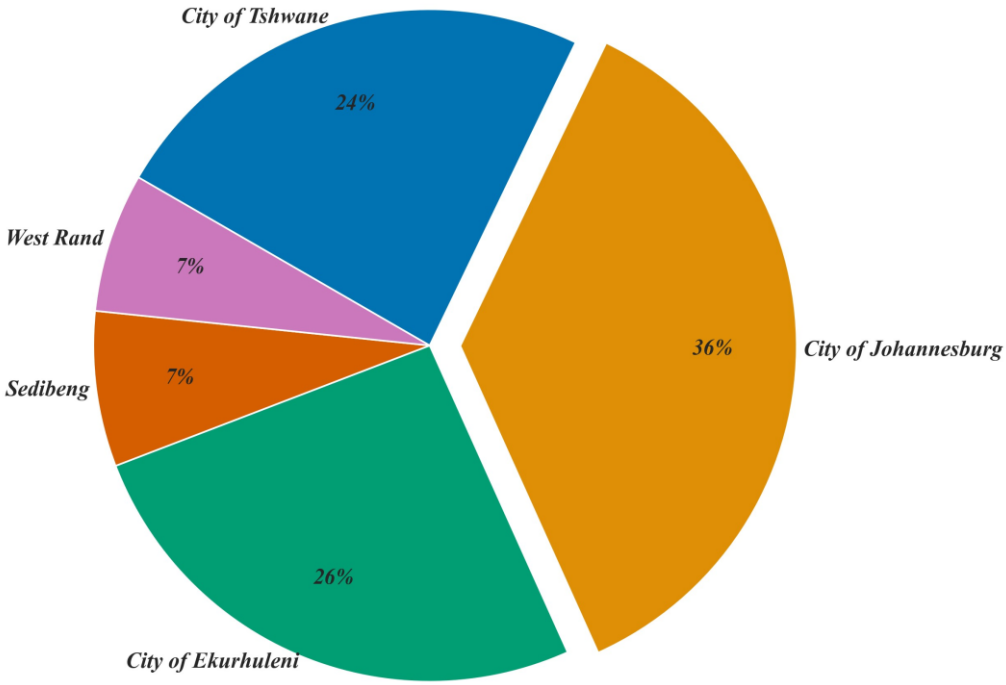


Figure 2-3: Population distribution of the Gauteng province per municipality shown in percentage (%) (Source: Stats SA, 2011).

2.3.2 Population density map

Population density is a measurement of how persons are distributed across a geographical area. Figure 2-4 shows a population density map of people per km^2 for Gauteng province with an outline of all municipalities. Figure 2-4 shows five classes describing the population density with ≤ 50 km^2 indicating a low population, ≤ 150 km^2 indicating a moderately low population, ≤ 250 km^2 indicating a moderate population, ≤ 1000 km^2 indicating a moderately high population and ≥ 2124 km^2 indicating a high population density. With data acquired from the census conducted in 2011 by Statistics South Africa, CoT shows a low to moderately high population, CoJ shows a very high population density, and CoE also has a high population density. With numerous industrial activities like mining, it shows people may have flocked there in search of better economic opportunities. Sedibeng and West Rand do not have a high population, considering they are only district municipalities. However, local municipalities like the West Rand and Emfuleni have a relatively high population.

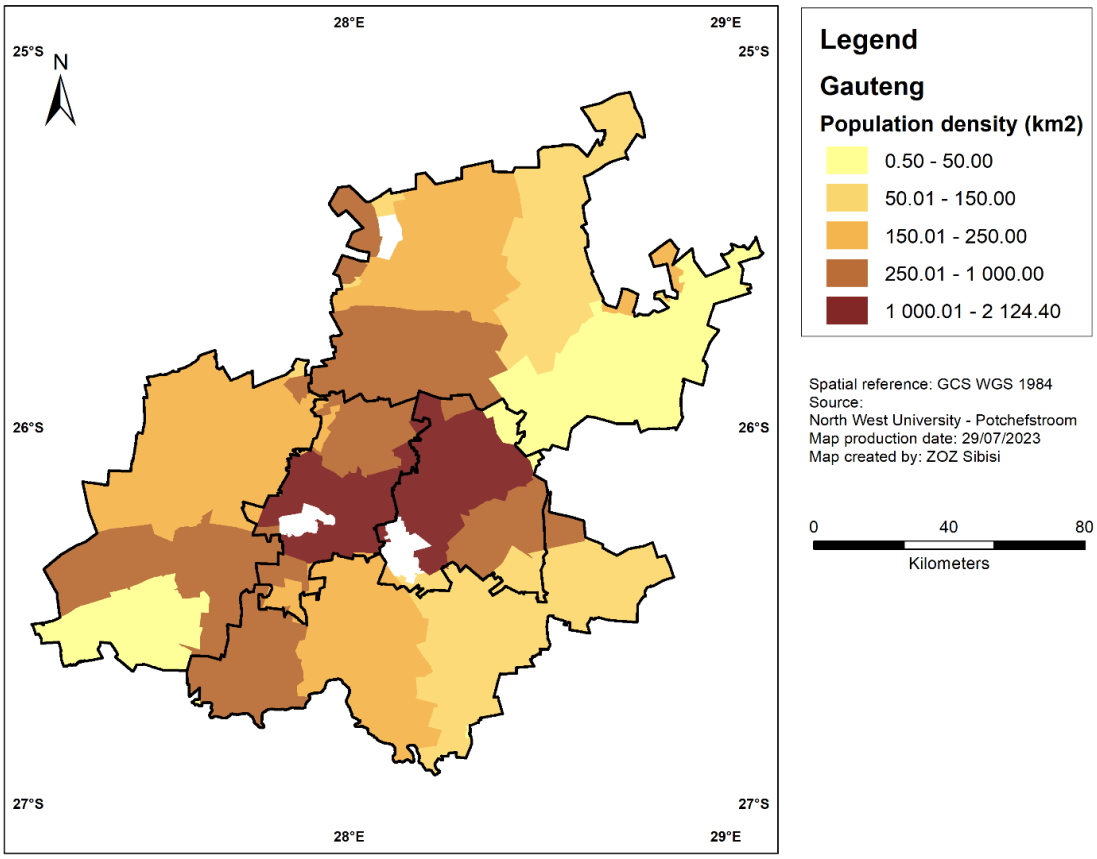
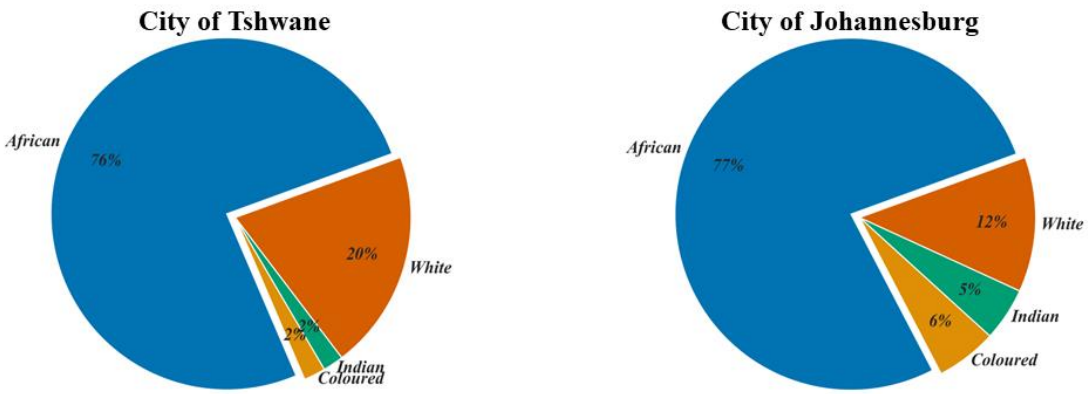


Figure 2-4: Map showing the population density of people per km² for the Gauteng province with an outline of the municipalities (Source: NWU).

2.3.3 Cultural demographics

The cultural demographics depicted in Figure 2-5 are separated according to each municipality. It is evident from Figure 2-5 that the African population is the dominating culture in all municipalities in Gauteng. Looking at the overall cultural population distribution in Gauteng, it is observed that the African, White, Coloured and Indian population makes up 79%, 16%, 3% and 2% of the entire province population, respectively. These statistics are retrieved from Statistics South Africa, 2011 (Stats SA, 2011).



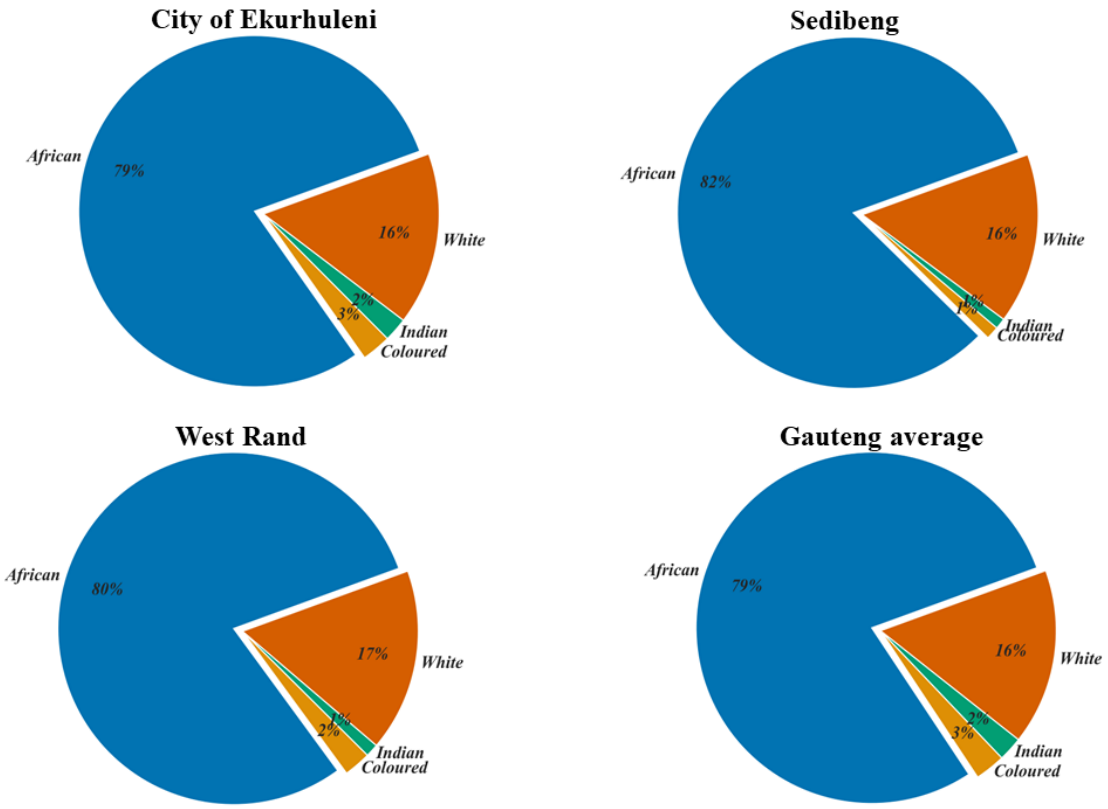


Figure 2-5: Pie charts indicating cultural demographics of Gauteng per municipality in percentage (%) (Source: Stats SA, 2011).

2.3.4 Language demographics

South Africa had 11 declared official languages (Bostock, 2018; Use of Official Languages Act, 2017). In 2022, section 6 of the constitution of the Republic of South Africa was amended to have South African Sign language as an official language (Department of Justice and Constitutional Development, 2022). This amendment was approved in 2023, making South Africa the only country with 12 official languages, namely English, Isizulu, IsiXhosa, Setswana, Sepedi, Sesotho, IsiNdebele, Xitsonga, Siswati, Tshivenda, Afrikaans and South African sign language (Seethal, 2023). Figure 2-6 depicts the distribution of first spoken language demographics for the entire Gauteng province.

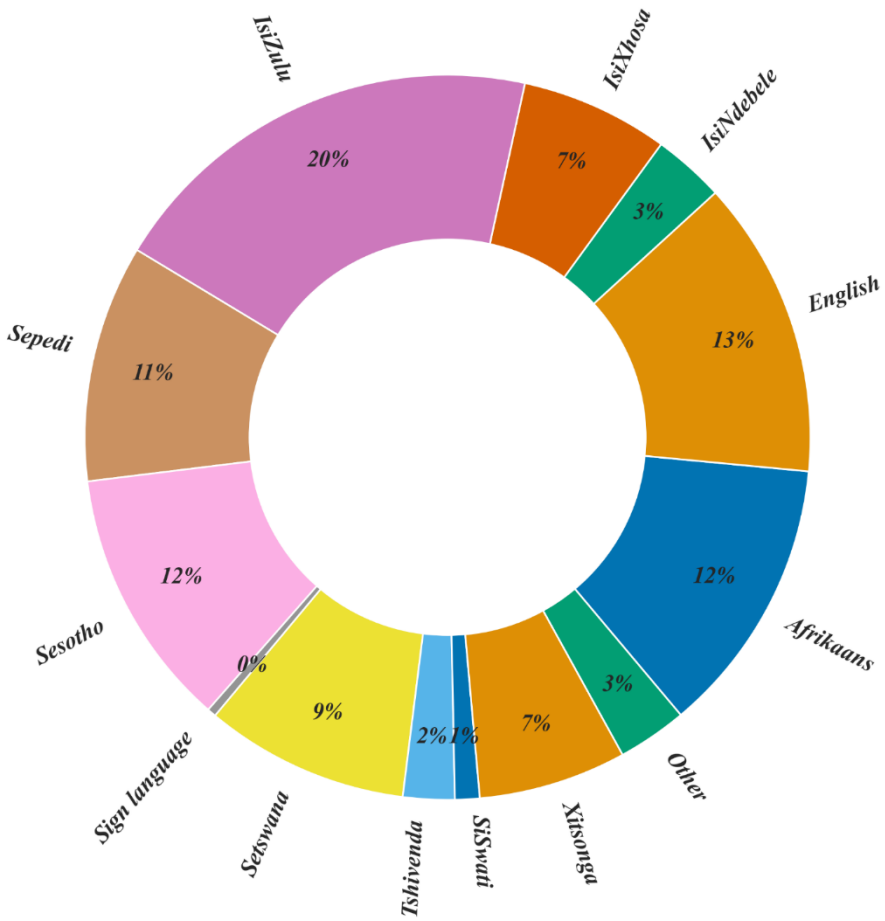


Figure 2-6: Pie chart indicating the percentage (%) distribution of first spoken languages in Gauteng (Source: Stats SA, 2011).

In Gauteng, IsiZulu is the most spoken language, followed by English, Sesotho, Afrikaans and Sepedi with 20%, 13%, 12%, 12% and 11%, respectively. First spoken languages below 10% in Gauteng are Setswana, IsiXhosa, Xitsonga, IsiNdebele, Tshivenda and SiSwati, with 9%, 7%, 7%, 3%, 2% and 1%. Sign language has a percentage of 0%, which is not entirely a true representation of the language in Gauteng as the statistics are for 2011. The category for 'other' has 3%, which could be a language spoken amongst community members that is not declared as an official language. For example, the language dominant in areas like Soweto was the 'Flaaitaal', also known as the 'Tsotsi-taal' (Molamu, 1995). This language was formed by apolitical township gangs that remain spoken today despite its dilution over the years (Hurst, 2015; Matima, 2010).

2.4 Topography

Figure 2-7 illustrates varying elevations in Gauteng, with the lowest point at 952 m.a.s.l (CoT) and the highest at 1913 m.a.s.l (Midvaal). South of Gauteng, low ridges are found in the Witwatersrand basin. Hills and low parallel ridges are found between Johannesburg and Tshwane. The Vaal River, at 1241 m a.s.l, separates Gauteng from the Free State. The lowest point of elevation is found in the CoT, and the highest is found in the district of Sedibeng. Elevation also plays a crucial

role in land use and population density. Land use activities are more prevalent in areas of low elevation. The high population also contributes to more land use activities, which can alter the geographic landscape.

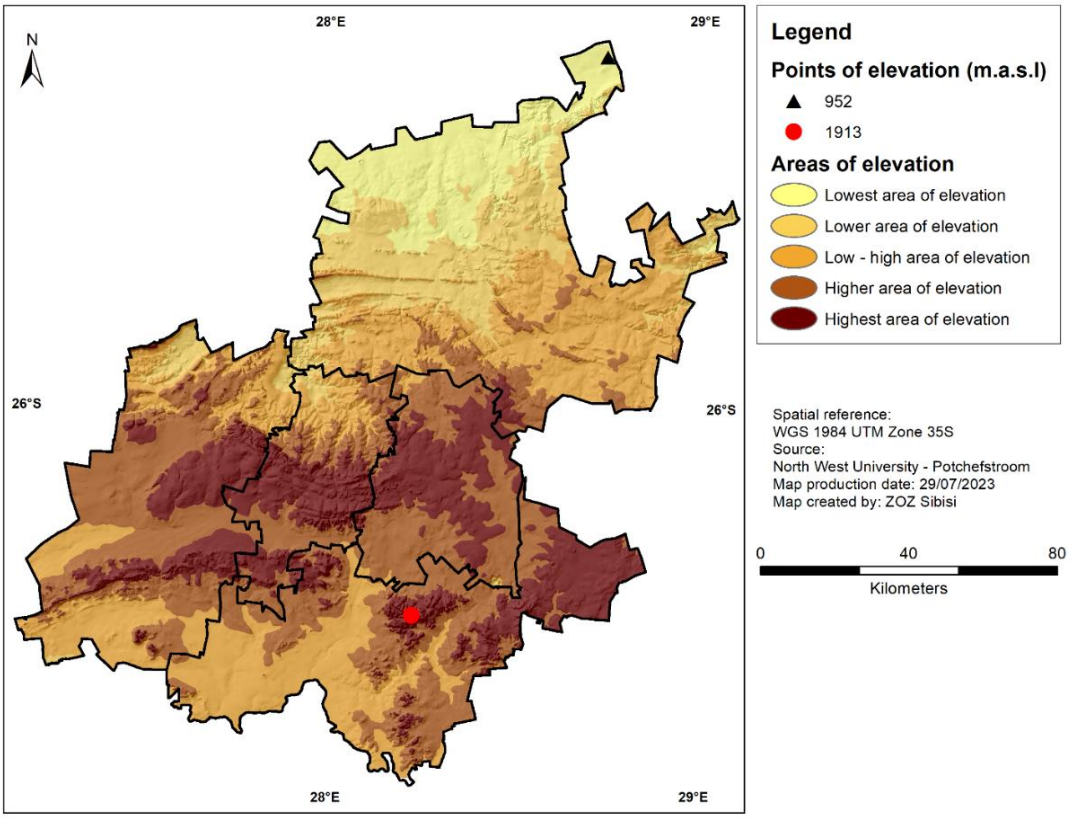


Figure 2-7: Map showing the topography of Gauteng depicting the highest and lowest points of elevation (Source: NWU).

2.5 Land use and land cover (LULC)

Land use activities of a geographical space are usually impacted by its topography. Land use is expected to be less in areas of high elevation, and this is seen in Figure 2-8 , where areas of high elevation under the Midvaal have land cover of grasslands, natural wooded land and natural grassland.

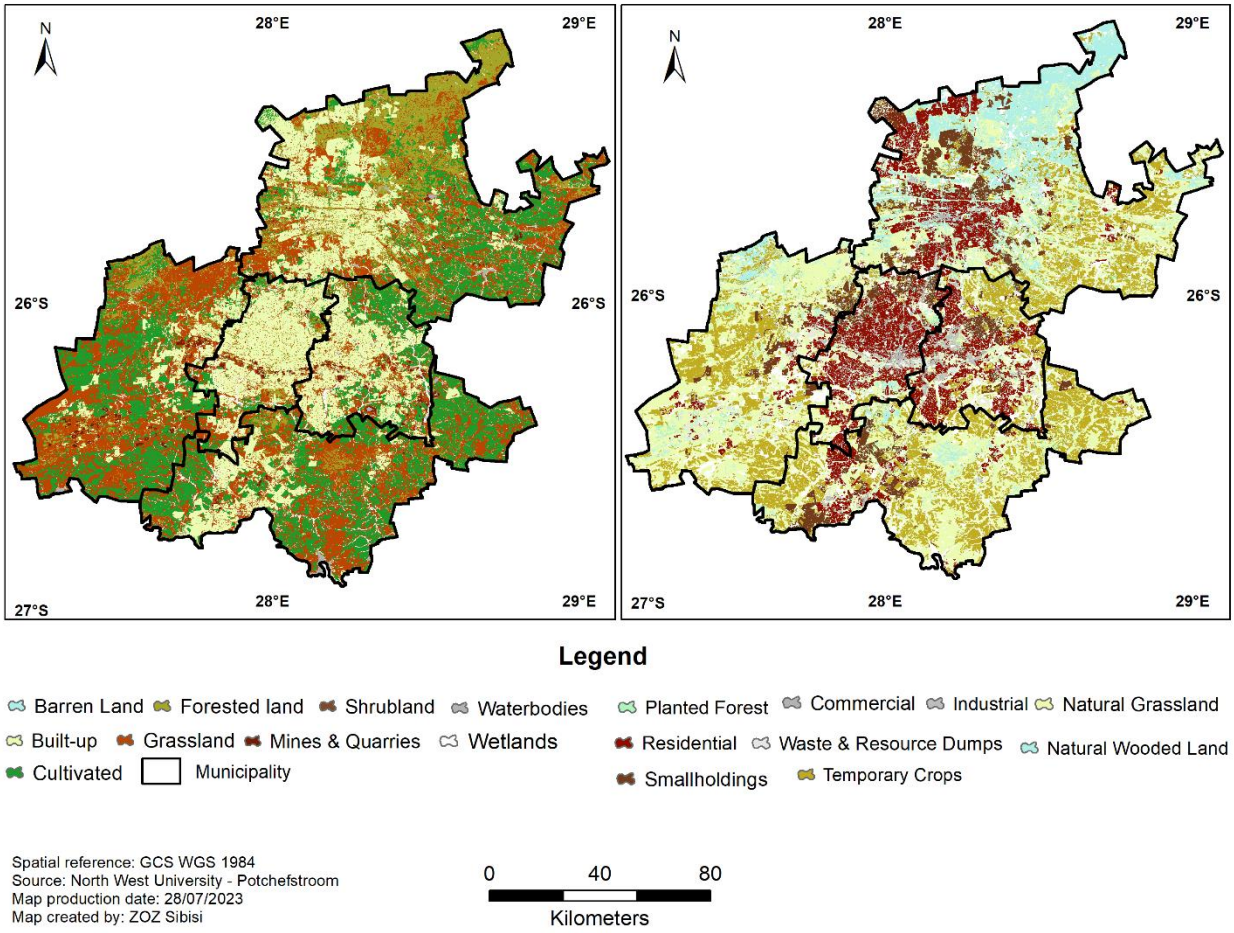


Figure 2-8: Land use and land cover map of Gauteng showing an outline of the municipalities (Source: NWU).

The land use in Gauteng includes agricultural activities, which are low to high intensity, as well as residential, commercial, and industrial activities. It is imperative to know and understand LULC to better comprehend and manage geographical transformation and to further incorporate spatial planning, environmental management and environmental (Obaid et al., 2023). Geographical transformations impact the physical environment and social aspects as these changes may lead to environmental vulnerability like droughts and floods and environmental degradation like soil erosion (Mawasha & Britz, 2023; Fairbanks et al., 2000). Figure 2-8 shows that grasslands, cultivated lands, residential areas, commercial, and industrial areas mainly occupy the land cover in Gauteng. Furthermore, most of the land use in Gauteng contributes massively to ambient air pollution. This includes mining, industries, waste burning and vehicle emissions. Considering rapid urbanisation, such land use may impact the urban geospatial distribution (Wray & Cheruiyot, 2015). For example, the need for new transport links or expansion of residential areas may alter the LULC of an area.

2.6 Ground-based data methods

2.6.1 South African Air Quality System Network Monitoring Stations

These data was compiled from the South African Air Quality System (SAAQIS). Five years of data, from 2017 to 2021 (January 1 to December 30), were downloaded. Table 2-2 shows the data availability of all 17 stations during the study period. The table further shows the percentage of data available from the raw and quality-controlled dataset. Data were downloaded at an hourly (1H) resolution for all 17 stations. The choice to use 17 of Gauteng's 39 stations was based on the availability of $PM_{2.5}$ data at the sites Seventeen AQM stations (*hereafter referred to as stations*) are had data retention of 60% and above (spatial distribution shown in Figure 2-9). It is regarded as 'good practice' to only use data sets with more than 60% data availability for averaged data sets (Garland et al., 2017).

Table 2-2: Summary the verified available data from the raw and quality-controlled (bold) dataset for all stations.

Station Name	**Verified data availability**				
	2017	***2018***	***2019***	***2020***	***2021***
Alexandra	*3%* **3%**	*4.7%* **4.7%**	*18.3%* **18.3%**	*54.3%* **54.3%**	*89.3%* **89.1%**
Bedfordview	---	---	*74.5%* **74.1%**	*89.1%* **88.9%**	*98.3%* **98.1%**
Bodibeng	---	*59.8%* **59.6%**	*34.4%* **34.2%**	*79.2%* **79%**	*42.4%* **42.4%**
Diepkloof	*84%* **83.7%**	*84%* **83.8%**	*73.3%* **73%**	*69.4%* **69.3%**	*85.8%* **85.6%**
Etwatwa	---	*27.4%* **27.2%**	*85.4%* **85.2%**	*73.5%* **73.2%**	*77.6%* **77.4%**
Hammanskraal	*26%* **25.7%**	*84.2%* **83.8%**	*72.2%* **71.9%**	*93.5%* **78.8%**	*59%* **49.1%**
Jabavu	---	*28.2%* **28%**	*78.8%* **78.6%**	*82.5%* **82.3%**	*82.9%* **82.7%**
Kliprivier	*54.1%* **53.6%**	*81.6%* **81.5%**	*52.4%* **52.3%**	*60.7%* **60.1%**	*85.3%* **85.1%**
NWU Vaal	---	---	*17.4%* **17.1%**	*99.2%* **98.9%**	*98.6%* **98.4%**
Olifantsfontein	---	*6.8%* **6.8%**	*81%* **79.6%**	*89.5%* **89.5%**	*79.3%* **79%**
Rosslyn	---	---	*44.8%* **44.5%**	*98.7%* **98.4%**	*88.2%* **87.9%**
Sebokeng	*51.6%*	*77.6%*	*43%*	*44.9%*	*61.8%*

	51.3%	**77.6%**	**42.8%**	**44.7%**	**61.8%**
Sharpeville	*57.5%* **57.4%**	*79.3%* **79.1%**	*39.8%* **39.6%**	*3%* **2.9%**	*13.7%* **13.5%**
Springs	---	*20.9%* **20.9%**	*71.8%* **71.3%**	*89.7%* **89.4%**	*53.5%* **53.3%**
Three Rivers	*48.3%* **48%**	*84.5%* **84.3%**	*19.5%* **19.4%**	*77.1%* **76.9%**	*60.9%* **60.9%**
Tshwane	---	---	---	*73.4%* **73.4%**	*28.9%* **28.7%**
Vanderbijlpark	---	---	*60.8%* **60.8%**	*81.7%* **81.4%**	*78.4%* **78.4%**

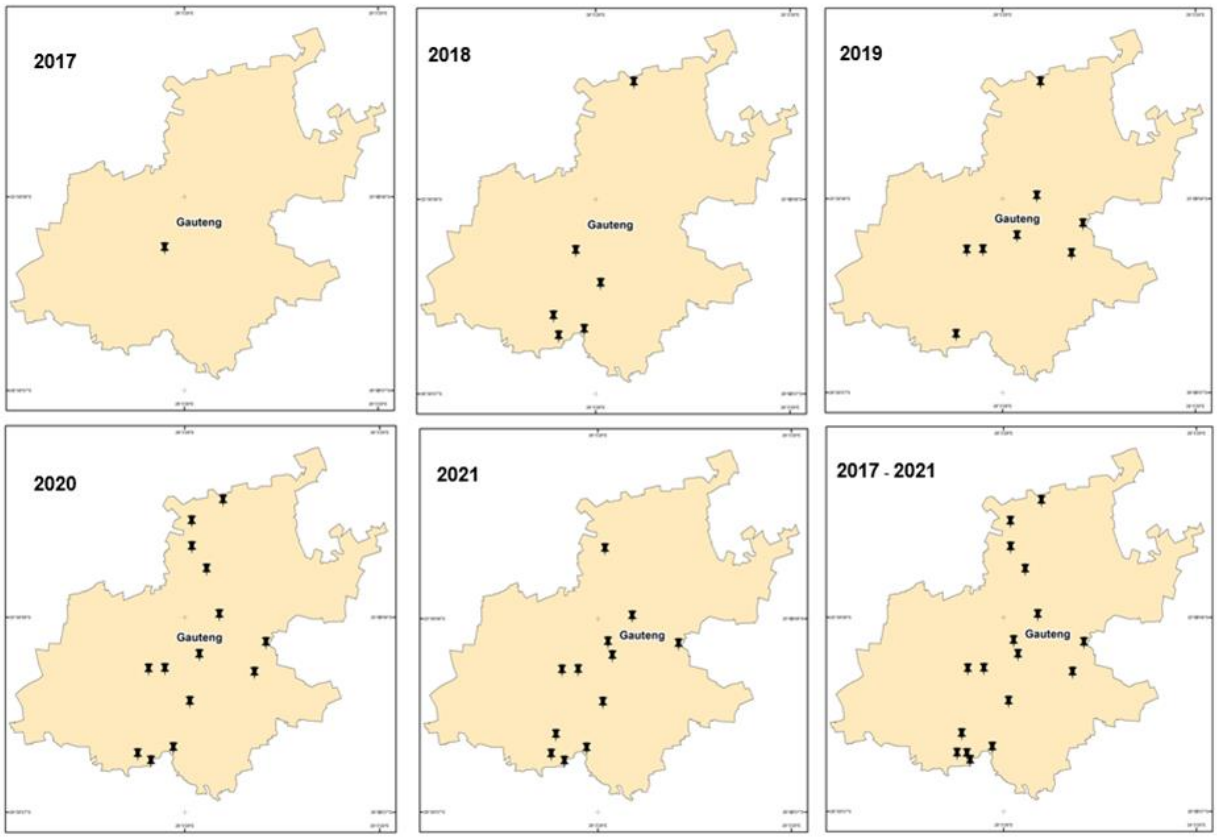


Figure 2-9: Maps showing the spatial distribution of the stations in the Gauteng province with a 60% and above data retention.

Most monitoring stations are located near low-income settlements, which assist in providing estimations of the air quality and pollution levels around those areas. Ambient air quality monitoring stations (AQMSs) in CoT are located close to high traffic-, residential-, and industrial activities (Wright et al., 2011).

Most air quality monitoring stations in the CoJ are fixed, with one being mobile. CoJ experiences high traffic emissions, and stations like Jabavu, Kliprivier and Diepkloof are located near the intricate road networks, which makes monitoring better. Bordering the townships of Jabavu,

Kliprivier, and Diepkloof are more arterial routes and national roads, which are the N1 and N12. The national roads are 86 km, 69 km and 99 km away from the townships, respectively. Surrounding townships are well-established route systems/ transport node operations known as 'Indingilizi' (Nyanda, 2019). This transport operation consists of the Metrobus, PUTCO, and DORLJOTA stakeholders, some of whose operation networks pass through the three townships, contributing to vehicle emissions.

For the CoE, the monitoring station in Olifantsfontein is located near the OR Tambo International Airport, which can contribute to poor visibility. $PM_{2.5}$ is the primary criteria pollutant emitted in observed airports alongside carbon and greenhouse gases (Kim et al., 2015). High pollutant concentrations are also governed by the geography and meteorology of an environment. Most monitoring stations in Sedibeng are found in the Vaal Triangle area, Vanderbijlpark and Vereeniging. These stations provide a better understanding of the air quality in the Vaal Triangle, which is still experiencing pressing issues regarding air pollution (Rabaji, 2019); for the West Rand, monitoring stations include Randfontein station and Mogale City station. However, these stations do not measure $PM_{2.5}$, hence their exclusion in this study.

2.6.2 Data quality control and assurance

After downloading the ground data, the data was quality-controlled and checked for instrumental errors. Table 2-4 shows the instruments used in each station. Before working with the data, data was converted to the most suitable timestamp with an appropriate time zone, Coordinated Universal Time (UTC). Data with extreme outliers were removed cautiously to avoid any data distortion. The data quality control process required adding and combining all the data into one final spreadsheet. Table 2-3 showed the upper and lower limits applied to the data.

Table 2-3: Data limits set for variables $PM_{2.5}$, temperature and relative humidity

Variable	**Data limits**	
	Values less than zero	**Data limit range**
$PM_{2.5}$	Not a Number (NaN)*	0 $\mu g.m^{-3}$ to 1000 $\mu g.m^{-3}$
Temperature (Temp)	Not a Number (NaN)# values less than 10 degrees were flagged as Nan's	-10°C to 50°C
Relative humidity (RH)	Not a Number (NaN)	0% to 100%
** Indicates all values that were below zero (0) were treated as missing value/s* *# Since the temperature minimum limit is -10°C, temperature values of less than -10°C were treated as missing values*		

Table 2-4: Summary of district, station name, station owner, co-ordinates, height and pollutant information for all stations in Gauteng (2017 to 2021).

District	Station name	Station owner	Type	Latitude	Longitude	Height (m.a.s.l)	Pollutant	
							$PM_{2.5}$	Meteorology
CoJ*[b]	Bedfordview	CoE	Traffic - Roadside	28.1331	-26.1787	1632	Teledyne API model T640	RM Young – Anemometer, RH%AT, rain-tipping bucket
CoE*[c]	Etwatwa	CoE	Residential - Low Income	28.4754	-26.1166	1618	Teledyne API T640	RM Young – Anemometer, RH%AT, rain-tipping bucket
CoE	Olifantsfontein	CoE	Urban	28.236	-25.9737	1535	Teledyne API T640	RM Young – Anemometer, RH%AT, rain-tipping bucket
CoE	Springs-new	CoE	Residential - Medium/Upper income	28.416	-26.2704	1636	Teledyne API T640	RM Young – Anemometer, RH%AT, rain-tipping bucket
CoJ	Alexandra	CoJ	Residential - Low Income	28.1103	-26.107	---	Teledyne API model T640 & Thermo TEOM 1405-DF	RM Young – Anemometer, RH%AT, rain-tipping bucket
CoJ	Diepkloof	SAWS	Urban	27.9564	-26.2507	1714	Thermo FH62C14	---
CoJ	Jabavu	CoJ	Residential - Low Income	27.8721	-26.2526	---	Teledyne API T640	RM Young – Anemometer, RH%AT, rain-tipping bucket
CoJ	Kliprivier	SAWS*[d]	Industrial	28.0841	-26.4202	1492	Thermo FH62C14	---
CoT*[a]	Bodibeng	CoT	Residential - Medium/Upper income	28.0937	-25.4928	1211	Thermo 5014i	RM Young – Anemometer, RH%AT, rain-tipping bucket
CoT	Hammanskraal	CoT	Residential - Low Income	28.2545	-25.3851	---	Thermo TEOM 1405-DF	RM Young – Anemometer, RH%AT, rain-tipping bucket
CoT	Rosslyn	CoT	Industrial	28.0947	-25.6251	1264	Thermo TEOM 1405-DF	RM Young – Anemometer, RH%AT, rain-tipping bucket

District	Station name	Station owner	Type	Latitude	Longitude	Height (m.a.s.l)	Pollutant	
							$PM_{2.5}$	Meteorology
CoT	Tshwane	CoT	Traffic Roadside	28.1703	-25.7402	---	---	---
Sedibeng	NWU_Vaal	NWU*[e]	Residential - Low Income	27.8848	-26.7247	---	---	---
Sedibeng	Sebokeng	SAWS	Residential - Medium/Upper income	27.841	-26.5873	1529	Thermo FH62C14	---
Sedibeng	Sharpeville	SAWS	Residential - Low Income	27.8678	-26.6898	1478	Thermo FH62C14	RM Young – Anemometer, RH%AT, rain-tipping bucket
Sedibeng	Three Rivers	SAWS	Urban	27.9995	-26.657	1448	---	---
Sedibeng	Vanderbijlpark	Sedibeng	Urban	27.8167	-26.6885	---	Teledyne API T640	RM Young – Anemometer, RH%AT, rain-tipping bucket

**[a]City of Tshwane *[b]City of Johannesburg *[c]City of Ekurhuleni *[d]South African Weather Service *[e]North-West University*

2.6.3 Data limit justifications

For $PM_{2.5}$, only values between 0 and 1000 $\mu g.m^{-3}$ were included in the data set.

Rainfall (RF) was recorded with values between 0 and 900 mm. 900 mm was chosen because this is the limit for the largest hydrological catchment area in Gauteng, Jukskei (Mawasha & Britz, 2021). This is asserted by CoJ (2008) as the annual rainfall ranges in South Africa are generally from 650 mm to 900 mm. Furthermore, elevated areas in Gauteng, like the Witwatersrand, receive approximately 700 mm of precipitation annually (Dyson et al., 2015).

RH values considered were values between 0 and 100%. Gauteng's temperature readings differ from the other provinces. Temperature includes values from -10°C and 50°C. Therefore, a benchmark of 50°C was used to account for extreme temperature conditions. According to a temperature-humidity index, heat waves are likely to occur in extremely high temperatures between 45°C and 55°C.

2.7 Statistical analysis techniques

2.7.1 Descriptive and inferential statistics

Descriptive and inferential statics analysis is the underlying foundation of quantitative research (Kaur et al*.,* 2018). Franzese and Luliano (2018) describe that when carrying out an investigation/ experiment, descriptive statistics precedes inferential statistics because the former serves to articulate summarised information from the data in a concise and comprehendible manner, and the latter aims to make inferences/ conclusions about the sampled data at hand (Subanji et al., 2021). Descriptive statistics analysis is useful for summarising and communicating data in a succinct manner. Before making inferences, data needs to be condensed and visualised, which may be in the form of charts, box and whisker plots, numbers, graphs, and tables (Ekoh, 2020). This visualised information makes it easier to make inferences about the air quality in the Gauteng region from the data obtained at the air quality monitoring stations.

For this study, measures of central tendency (mean, median, minimum, maximum), measures of variation (standard of deviation), and measures of position (percentile and quantile ranks) were conducted. The mean was calculated as follows in Equation 2-1:

$$Mean\ (\bar{X}) = \frac{\sum x}{n}$$

Equation 2-1: Measure of central tendency: Mean calculation

$$Median = \frac{(n+1)}{2}$$

Equation 2-2: Measure of central tendency: Median calculation

where x denotes the sum of all data points divided by the number of data points (n). Outliers were removed in this instance to avoid distortion on the central tendency of the dataset. The median is the middle value in the dataset distribution, and in the case of equal distribution, Equation 2-2 is applied. Equation 2-3 shows how to calculate the standard deviation (SD), which measures how spread-out values are from the mean (Mishra et al., 2019); the following formula was used:

$$\sigma=(x-x)2n$$

Equation 2-3: Measure of variation: Standard deviation calculation

where σ denotes the population standard deviation, x denotes each value from the population, x denotes the population mean, and n denotes the size of the population. To calculate the measures of position, Equation 2-4, Equation 2-5, Equation 2-6 and Equation 2-7 were followed where the 1st quartile (Q1), 2nd quartile (Q2), 3rd quartile (Q3), Inter-quartile range (IQR), lower bound limit and upper bound limit were conducted. Q1 represents the lowest values in the first quarter (25%), Q2 represents the middle values (median), and Q3 represents the values in the last quarter (75%). The IQR provides a steadier representation of the dataset as it is less influenced by outliers or extreme data values. Lower and upper bounds are limits set to exclude anything below or above the values in the dataset. Calculations are conducted as follows:

$$Q1 = (n+1)\left(\frac{1}{4}\right)$$

Equation 2-4: Measure of variation: First quartile calculation

$$Q2 = (n+1)\left(\frac{2}{4}\right)$$

Equation 2-5: Measure of variation: Second quartile calculation

$$Q3 = (n+1)\left(\frac{3}{4}\right)$$

Equation 2-6: Measure of variation: Third quartile calculation

$$IQR = Q3 - Q1$$

Equation 2-7: Measure of variation: Interquartile range calculation

To increase the validity of the data, Confidence Intervals (CI) were conducted using Equation 2-8. CI are an estimate of how far away the sample mean from the actual population is. CI also accounts for discrepancies between the sample mean, and the population mean. The equation is as follows:

$$CI = \bar{x} \pm a\frac{\sigma}{\sqrt{n}}$$

Equation 2-8: Confidence Interval calculation

Variables needed when calculating the Confidence Interval (CI) are the sample size/mean n , S.D σ, and alpha value (α). For this study, the confidence level was 95%; therefore, the significance value was 0.05 (1- α). $\bar{x}$ denotes the population mean, α denotes the significance value, σ represents the SD and $\sqrt{n}$ represents the population size.

2.7.2 Descriptive tables

Descriptive tables were used to describe the central tendency of the data, as they are an important component of statistical analysis (Language, 2020). All descriptive tables outline the count, mean $\bar{x}$, standard deviation (SD), minimum (min), 1st quartile (25%), 2nd quartile (50%), 3rd quartile (75%), 99th percentile and maximum (max).

2.7.3 Boxplots

Boxplots visually represent the measures of position or where the tendencies lie (Kaur et al., 2022). The whiskers on the boxplot indicate the minimum (1st percentile) and maximum (99th percentile) data values. The box's top, middle, and bottom parts represent the 25th, 50th (median) and 75th percentiles.

2.7.4 Time variation graphs

Time variation graphs are crucial in visually depicting different spatial and temporal patterns. Spatial patterns are generally represented on a diurnal, weekly, seasonal, and yearly time scale. Diurnal variations show concentration patterns over a 24H (day) time scale. Weekly variations show the concentration patterns from Monday to Sunday, and seasonal variations show the pollutant variation across the summer, winter, autumn and spring.

2.8 Aerosol Robotic Network (AERONET) – Sun photometers

Sun photometers provide continuous temporal data and have been doing so for almost two decades. They are directed towards the sun and take measurements every ±15 minutes during the day, with the sensor's head being 1° from the sun (Adesina et al., 2019). Like many other satellites, they directly measure aerosol light extinction in a vertical column. They take measurements across eight spectral bands of 1640 nm, 1020 nm, 870 nm, 675 nm, 500 nm, 440 nm, 380 nm and 340 nm. This allowed further improvements for acquiring other optical properties in the vertical column.

Remote data was collected from the Aerosol Robotic Network (AERONET) site (https://AERONET.gsfc.nasa. gov/) for four ground stations. AERONET is a ground-based data network that provides optical properties in a vertical column while stationed on the ground, not in space. AERONET data is comprised of three data levels: level 1.0 (unscreened), level 1.5 (cloud-

screened) and level 2.0 (quality controlled). The collected data for this study was level 2.0 data, cloud-screened and quality-controlled with pre- and post-field calibrations. Level 2.0 data allows for easier satellite data analysis without making conversions and pre-calculations before working with the data (Toledano et al., 2007). Data were downloaded and stored as CSV files.

2.8.1 AERONET station information

Figure 2-10 shows all the information on the $AOD_{AERONET}$ ground-based stations that are within 100 km of the study area. Using ArcMap v10.8.1 to map their spatial distribution, only two stations showed to be the best choice, which is Pretoria_CSIR_DPSS (Site 1) and Pretoria_CSIR_EC (Site 2), which are shown as one point on the map due to their close proximity to each other.

Table 2-5: Table indicating $AOD_{AERONET}$ stations that are within 100 km from the study area, including the AERONET site name, latitude, longitude, height (m), period of available data and the variables (AOD and AE)

AERONET site	No.	Latitude	Longitude	Height (m)	Period	Variables	
						AOD	AE
Metsi	705	-26.4908	27.13241	1404	2020 - 2021	✓	✓
Potchefstroom	705	-27.0970	26.7145	1355	2018 - 2020	✓	✓
Pretoria_CSIR_DPSS	659	-25.7643	28.2673	1449	2017 - 2018	✓	✓
Pretoria_CSIR_EC	659	-25.7501	28.2784	1403	2020 - 2021	✓	✓
Welgegund	705	-26.5694	26.93917	1480	2018 - 2020	✓	✓

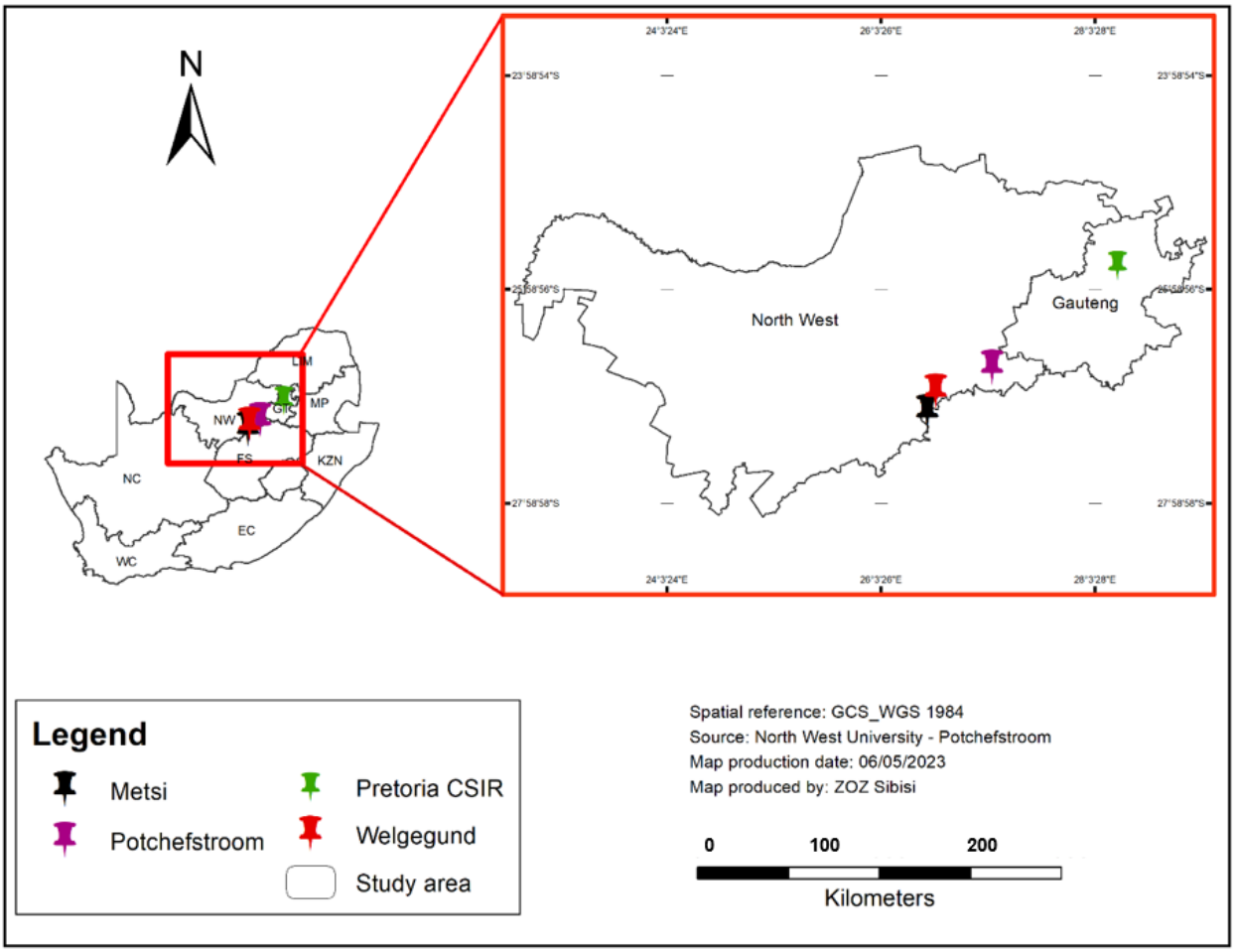


Figure 2-10: Study area map with location pins indicating different AOD monitoring AERONET sites. (Source: ZOZ Sibisi).

Table 2-5 shows the site name, site instrument number, coordinates, height, available data periods and retrieved variables. As seen from the table, the data available for each site is limited and does not cover the five-year monitoring period. However, despite data availability being a limitation, substantial results were achieved.

The data obtained contained data at different AOD spectral bands ranging from 1640 nm to 340 nm and different variables including AOD, angstrom exponent and precipitable water. For this study, $AOD_{AERONET}$ at a wavelength (λ) of 550 nm is required; however, AERONET data does not provide measurements in that wavelength. Many techniques have been employed to combat this shortcoming, including the spatiotemporal interpolation and 'power law' techniques to obtain $AOD_{AERONET}$ measurements of 550 nm. $AOD_{AERONET}$ of 500 nm, 675 nm and an Ångström (α) of 440 nm~675 nm are interpolated together, and autocorrelation will be achieved through the following formula:

$$\tau_{550nm} = \frac{\tau_{500nm}}{\exp\left(-\alpha^{*} In\left(\frac{\lambda_{500nm}}{\lambda_{550nm}}\right)\right)}$$

Equation 2-9: Spatiotemporal interpolation conversion technique

Where $\tau_{\mathbf{550}nm}$ and $\tau_{\mathbf{500}nm}$ are AOD values found at $\lambda_{\mathbf{500}nm}$ [λ_1 = 500 nm] and $\lambda_{\mathbf{550}nm}$ [λ_2 = 675 nm]. It is also expressed as the 'power law' where AOD_{MODIS} and $AOD_{AERONET}$ are interpolated to obtain a common wavelength, 550 nm, raising it to the power of the Ångström. This is the technique employed in this study. The formula is as follows:

$$Aeronet\_AOD_{\mathbf{550}} = AOD_{\mathbf{500}}(\frac{550_{nm}}{500_{nm}})^{-\alpha}$$

Equation 2-10: 'Power law' conversion technique

where 550 nm is the AOD_{MODIS} wavelength, 500 nm is the $AOD_{AERONET}$ wavelength, and α is the Ångström at 440 nm – 675 nm.

Furthermore, $MODIS_{(TERRA, AQUA)}$ AOD data for the specific stations were downloaded using the coordinates to facilitate easier data collocation. AOD_{MODIS}, relative humidity (precipitable water) and the α variables have been downloaded. $MODIS_{TERRA}$ α data has been excluded for Potchefstroom in the correlation matrix heatmap due to no available data.

2.8.2 Data quality control

Despite AERONET data being screened, -999 values were found in the data, and these need to be corrected before any analysis. These values were flagged and replaced as missing values. This was a crucial step prior to conducting descriptive statistics.

2.8.3 Data processing

This section's data processing comprised data merging, refinement and transformation. Before any analysis, $AOD_{AERONET}$ data were merged with MODIS data using Python's 'merge' function. This function was specified to merge the MODIS data according to the available AERONET data, as it was inconsistent with the amount of MODIS data. The data refinement process, where the data cleaning process was done and missing values were dealt with, has been discussed in the data quality control section. Transforming the data included changing the features in the dataset into float and numerical data to allow for effortless computation. All this was conducted using advanced statistical libraries found in Python.

2.9 Remote sensing data methods

A process is always followed when retrieving AOD satellite data, and Figure 2-11 provides a conceptual model for the process to retrieve AOD for the MODIS platform. When retrieving data, the type of measurement, instruments, spatial resolution, temporal resolution, wavelength band, data period, and parameters are considered. Furthermore, statistical analysis will be conducted

to visually demonstrate and compare results to the ground station data and see if a trend can be established.

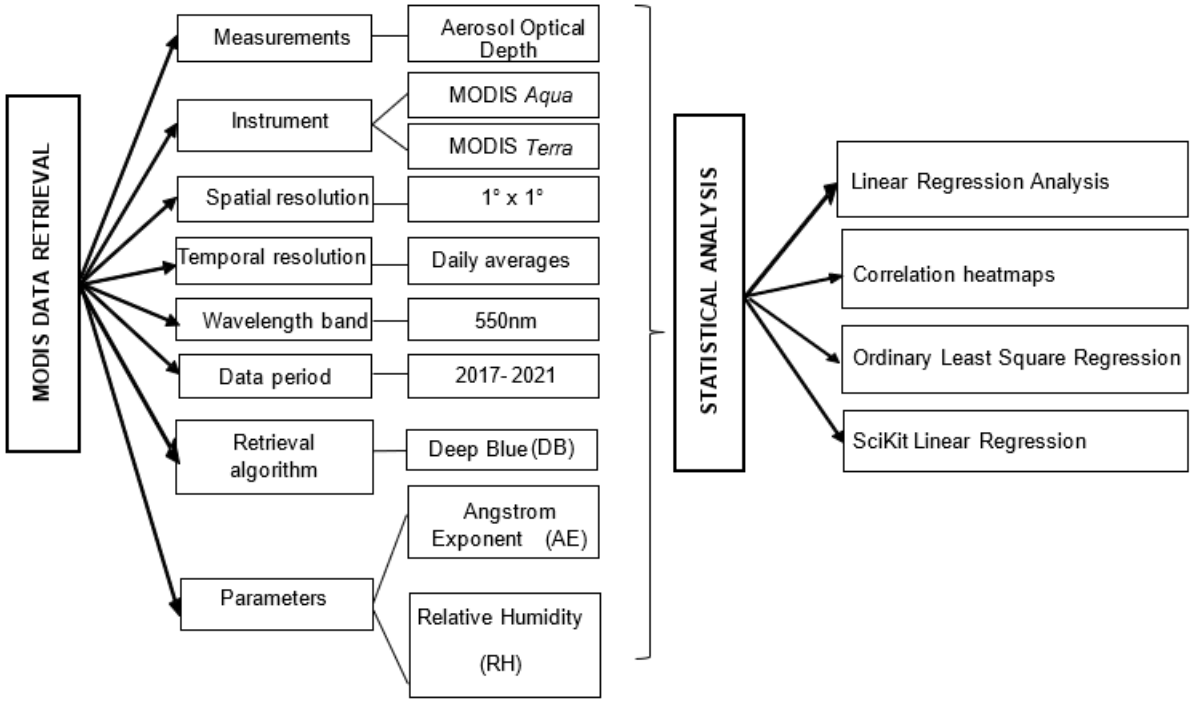


Figure 2-11: Conceptual diagram outlining the MODIS data retrieval along with the statistical analysis employed in this study.

2.9.1 AOD retrieval

Retrieving aerosol properties using satellites is often challenging and needs to be counteracted using the appropriate retrieval algorithms. The retrieval of AOD_{MODIS} presents surface reflectance constraints that must be accounted for. This influences the AOD-$PM_{2.5}$ relationship. Furthermore, climate conditions, particle compositions, and geography impact the AOD-$PM_{2.5}$ relationship (Hersey et al., 2015). The Deep Blue (DB), Deep Target (DT) and Combined DB & DT products accounted for this constraint.

MODIS has 36 spectral band channels and provides spectral information from VIS to IR wavelengths. Furthermore, AOD_{MODIS} is reported from seven spectral wavelengths, which are 470 nm, 550 nm, 660 nm, 870 nm, 1200 nm, 1600 nm, and 2100 nm (Anderson et al., 2012), in which the 550 nm wavelength is congruent with frequent AOD studies and has been used in this study. AOD data was retrieved for both $MODIS_{(TERRA,\ AQUA),}$ with both having a local satellite Passover time of 10:30 and 13:30, respectively.

2.9.2 MODIS data collection

The satellite data is from the National Aeronautics Space Administration (NASA) Geospatial Interactive Online Visualisation and Analysis Infrastructure (GIOVANNI) website (https://giovanni.gsfc.nasa. gov/giovanni/). The process of acquiring satellite data includes plot selection, appropriate date range, variables, variable-specific algorithms, platform/instrument, spatial resolution, temporal resolution, and wavelength.

For this study, the plot chosen was an area-averaged time series. The date range had to correspond with the ground-based $PM_{2.5}$ data from 2017 to 2021 (1 January - 31 December) and is given in UTC. Due to the small geographic scale of this study, the bounding box option could not be utilised when choosing a specific study area; therefore, each specific coordinate of each station was inputted to acquire the data. The following parameters were considered when choosing Aerosol Optical Depth (AOD) and Ångström Exponent (α). After choosing the variables, deciding on the appropriate algorithm is important. For MODIS $AOD_{(TERRA,\ AQUA)}$ and MODIS $AE_{(TERRA,\ AQUA)}$, the DB, DT and Combined DB & DT algorithms were chosen because they facilitate easier data retrieval over densely vegetated areas and bright surfaces (Butt et al., 2017). For $MODIS_{TERRA}$, the *Aerosol Optical Depth 550 nm (Deep Blue, Land-only) (MOD08_M3 v6.1)* product was used, and for $MODIS_{AQUA}$, the *Aerosol Optical Depth 550 nm (Deep Blue, Land-only) (MYD08_M3 v6.1)* product was used. Both datasets have data available from 1 February 2000 to date. The spatial resolution of $MODIS_{(TERRA,\ AQUA)}$ AOD and AE is 1° x 1°.

The Ångström (α) exponent is often used when there is a need to know the aerosol size, aerosol distribution and wavelength dependence (Falah et al., 2021). The α is usually used to measure aerosol size (Schuster et al., 2006). Table 2-6 describes the value characterisation attached to an α value. A lower α value indicates the presence of non-absorbing aerosols, which indicates clean atmospheric conditions. In contrast, a higher α value indicates the presence of absorbing aerosols, indicating polluted atmospheric conditions. An index of 1.5 is used in this study as a qualitative indicator of particle size. Values less than 1.5 indicate the presence of large aerosols like dust and sea salt. Values that are more than 1.5 indicate the combustion of by-products. The amount of light attenuation of the aerosols in the vertical column layer further governs these values.

Table 2-6: Ångström (α) index characterisation outlining the numerical value, variable degree, aerosol concentration, light attenuation, atmospheric conditions and aerosol absorption associated with the index (Source: Adopted from Falah et al., 2021).

Numerical value	Variable degree	Aerosol concentration	Light attenuation	Atmospheric conditions	Aerosol absorption
< 1.5	Low α	Low	Low	Clean	Non-absorbing aerosols EG: Sea salt/ dust
> 1.5	High α	High	Significant	Polluted	Absorbing aerosols EG: Black Carbon, Sulphates & Smoke

2.9.3 Satellite data processing levels

The Earth Observing System Data and Information System (EOSDIS) are products processed at different levels ranging from the lowest processing data level, Level 0, to the highest processing level, Level 4. Products at the lowest processing data level are at their rawest form, where the data is at full instrument resolution, and the highest processing data level provides model outputs from analyses of lower-level data.

The processing levels in data retrieval, as shown in Figure 2-12, are Level 0 (L0), Level 1 (L1), Level 2 (L2), Level 3 (L3) and Level 4 (L4). L0 data includes reconstructed and unprocessed instrument data. Level 2 (L2) data files, which operate similarly to L1, data is reconstructed, raw, full resolution, instrumental data that is referenced in time and annotated with information including geo-referenced. L1B data is L1A data processed to the sensor units, and instrumental calibrations are applied. L1C is L2B data that includes new variables to describe the spectra. L2 data files, which operate similarly to L1 data, obtain Earth's physical information at the same spatiotemporal resolution as L1 (Hubank et al., 2015).

Furthermore, L2 data is available as 6 min granules along the orbit trajectory. L2A data contains information derived from the geolocated sensor data, for example, elevation. L2B data are L2A data that are processed by the sensor unit. L3 data is L2 data projected onto a uniform spatiotemporal coverage and defined period. These data is gridded and quality-controlled. L3A data are periodic summaries of Level 2 data products. L3 monthly and daily data are aggregated on a 1° x 1° horizontal network. L3 daily data is generated from L2 data. Aggregation uses only quality assurance filtered data retrievals.

Additionally, aggregation includes only cells measured on the day of interest. At least three such cells are required for a given grid cell to be valid on a given day. Monthly data statistics are based on the arithmetic mean (mean, standard deviation) derived from L3 data. To remove poorly sampled grid items, you need at least three days of valid data to fill the monthly grid items. L4

data products are uncomplicated to use as the products are derived variables. L4 models outputs from all the other data processing levels.

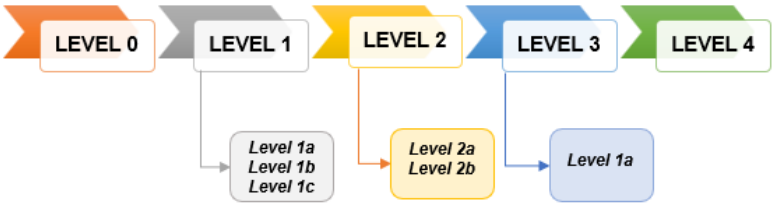


Figure 2-12: Data processing levels involved when retrieving satellite data.

Figure 2-13 shows the elements that affect how the remote signals between the 0.4 µm and 2.5 µm range are received. The figure shows that tropospheric aerosols, trace gases, Rayleigh scattering and stratospheric aerosols impact the remote signals. These elements may impact retrievals; therefore, quality control needs to be conducted on the data.

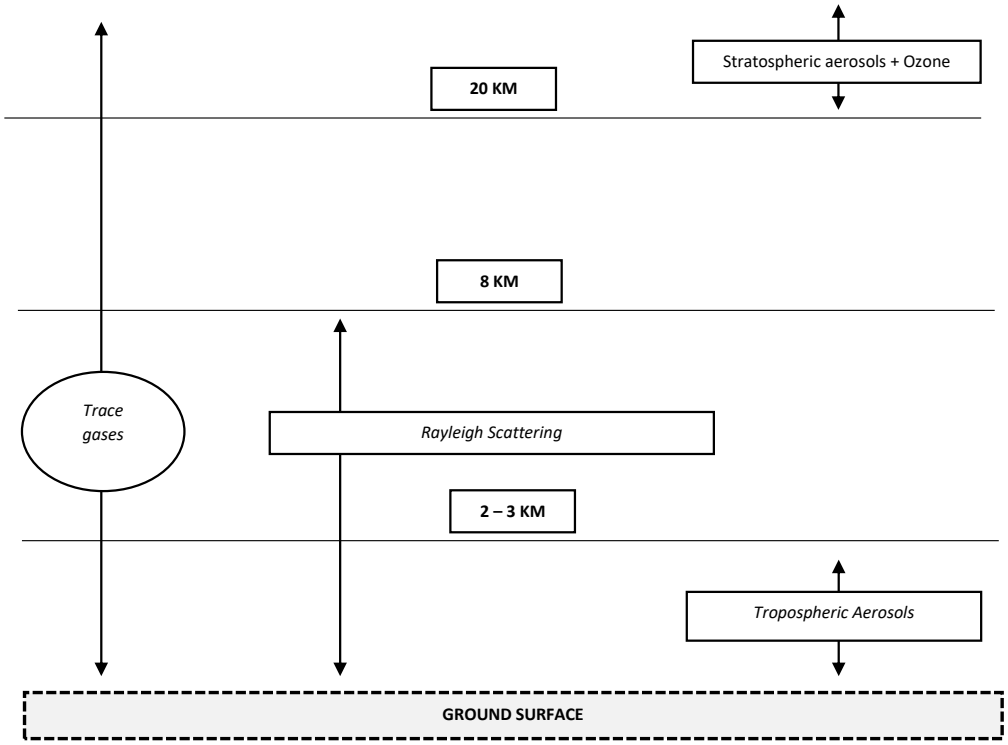


Figure 2-13: Figure depicting the elements impacting remote signals in the 0.4 µm -2.5 µm range (Source: Image re-created from Hubanks et al., 2015 - MODIS Atmosphere QA Plan for collection 006).

2.9.4 Quality Assurance flags (QA)

Quality Assurance (QA) Flags, or Confidence Flags (CF), are important for validating satellite acquisition data as they are used to identify problems during the acquisition process (Hubanks et al., 2008). The QA score in Table 2-7, derived from Hubanks et al. (2015), is expressed in integer

format from 1 to 3, with 1 being the lowest (unsatisfactory) and 3 being the highest (satisfactory). If a query is indicated by a flag value of 0 (zero), the query was not run for that particular pixel, proving it to be invalid. A flag value of 2 or 3 is highly recommended for scientific applications as it can provide monthly and daily data aggregation. Using data with a flag value of 1 means that the end user considered more data coverage than data accuracy.

QA flags are important when working with satellite datasets. They can facilitate aggregation or reduction when moving from L2 orbital data to L3 global gridded data (Hubanks et al., 2015). Additionally, QA flags can provide more detail into the input parameters of the L2 granules to produce improved global L3 gridded parameters.

Hubanks et al. (2015) outlined the aspects required during QA which are as follows:

1. The quality or confidence level of the acquired physical parameters.
2. Processing path of the retrieval algorithm
3. Acquisition method
4. Input data source
5. Search spectral bands used
6. Search outcomes
7. Dataset characteristics
8. Input source metadata

The initial bit value indicates the effectiveness of a particular parameter, followed by 2 or 3 bits indicating validity. Figure 2-14 shows that numbers are stored from 0 to 7, but the computer software package used for L3 data is intended to accommodate only 4 levels of validity that can be stored well in just 2 bits (Hubanks et al., 2008).

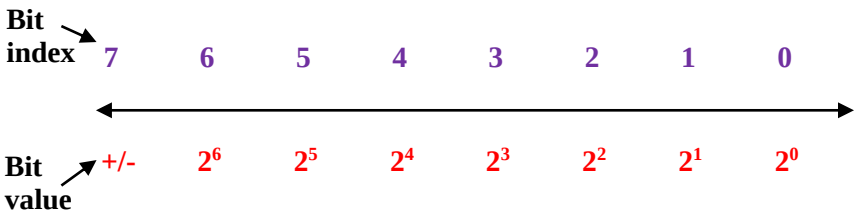


Figure 2-14: Bit index and value indicating effectiveness and validity of satellite data (Source: Image re-created from Hubanks et al., 2015 - MODIS Atmosphere QA Plan for collection 006).

The effectiveness of a flag should always be read in conjunction with the validity flags to distinguish valid data from missing data (Platnick et al., 2015). The MODIS L3 computer software package uses L2 effective and validity flags by producing L3 descriptive statistics (weighted

average and standard deviation). Table 2-7 and Table 2-8 show that QA weighting is performed by assigning a weight to each L2 input pixel and flagging its validity. This provides the invalid, unsatisfactory, substantial and satisfactory data with a weight of 0-, 1-, 2- and 3-within a L3 1° grid box.

Table 2-7: Table indicating baseline values applicable in weighting of quality assurance flags, which comprise the flag name, bit value, bit value definition and applicability in level 3 QA (Source: Hubanks et al., 2015).

Flag Name	Bit Value	Bit Value Definition	L3 QA-Weighting
Parameter e.g.	0	Ineffective	Not applied in L3
Effectiveness flag	1	Effective	Applied in L3
Parameter e.g.	0	Invalid	0 -
Validity flag	1	Unsatisfactory	1 -
	2	Substantial	2 -
	3	Satisfactory	3 -

MODIS Terra and Aqua QA flags are stored as two SDS,: Land Quality Assurance (LQA) and Ocean Quality Assurance (OQA) (Hubanks et al., 2015). LQA and OQA are 5-byte SDSs whereby the former and latter both contain numerous (input and processing) flags required for land and ocean retrievals.

Table 2-8: Table indicating the weightings assigned to all the MODIS QA flag names, which include the flag name, number of bits, bit value and bit value definition (Source: Hubanks et al., 2015).

Flag name	Number of Bits	Bit values	Bit Value definitions
0.47 µm Aerosol Optical Thickness	1	0	Ineffective
Effectiveness Flag		1	Effective
0.47 µm Aerosol Optical Thickness	3	0	Invalid
Validity Flag		1	Unsatisfactory
		2	Substantial
		3	Satisfactory
0.66 µm Aerosol Optical Thickness	1	0	Ineffective
Effectiveness Flag		1	Effective
0.66 µm Aerosol Optical Thickness	3	0	Invalid
Validity Flag		1	Unsatisfactory
		2	Substantial
		3	Satisfactory
Deep Blue Aerosol	1	0	Ineffective

Effectiveness Flag		1	Effective
Deep Blue Aerosol	2	0	Invalid
Validity Flag		1	Unsatisfactory
		2	Substantial
		3	Satisfactory
Deep Blue Aerosol type	2	0	Mixed
		2	Dust
		3	Smoke
		4	Sulphate

2.10 Re-analysis data methods

The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) is a highly advanced re-analysis satellite tool that has proven its efficiency in many climatological and air quality satellite studies (Campos et al., 2022). This tool operates on an assimilation basis where aerosol data and meteorological observations are assimilated into one system. This simplifies the work for the end user as this tool can integrate aerosol optical properties, aerosol constituents and concentrations with meteorological observations to facilitate effortless aerosol-climate studies (Khan et al., 2021). MERRA-2 was developed by the Global Modelling and Assimilation Office (GMAO) at NASA and is an updated version 1 of MERRA (Navinya et al., 2020). MERRA-2 v5.12.14 uses the updated version of The Goddard Earth Observing System Model, version 5 (GEOS-5), a modelling system that assimilates observations to produce gridded data. This assimilation tool improves spatiotemporal analysis of aerosols like AOD by integrating them into models. This re-analysis tool is highly advantageous because this joint assimilation provides aerosol measurements like AOD from a retrieval algorithm found in the data products under Collection Five (C5) satellites $MODIS_{(TERRA,\ AQUA)}$, MISR and Advanced Very High-Resolution Radiometer (AVHRR) over ocean (Campos et al., 2022). The re-analysis tool has a horizontal grid of 0.5° (latitude) x 0.65° (longitude) and a vertical grid of 72 layers (Gelaro et al., 2017; Reinecker et al., 2011).

2.11 MODIS validation using $AOD_{AERONET}$

The validation process is central in corroborating satellite data with ground-based AERONET data, and AOD_{MODIS} and $AOD_{AERONET}$ data must be collocated in space and time. To ensure a successful validation process without any hindrances, daily retrievals from AOD_{MODIS} and ground-based $AOD_{AERONET}$ data need to be collocated in space and time (Chu et al., 2002). Following this process, algorithms and statistical analysis need to be employed to facilitate this process. Algorithm testing and verification are important as they provide confidence in retrieval algorithms, model simulations, and comparisons between techniques and software programs.

2.12 Statistical analysis

For this study, two-regression methodologies will be employed; Simple Linear Regression and Ordinary Least Square Regression, to determine if there is an association between the variables of interest: AOD_{MODIS}, $AOD_{MERRA\text{-}2}$ and $AOD_{AERONET}$. Regression analysis will be conducted to investigate the relationship between the regressor and regressand variables. This is a common method used in statistical analysis. Prior to any statistical analysis, it is important to establish if variables of interest have some correlation among them. As part of data analysis, pairwise plot was first conducted for the dataset, however, it was not included in the study. This was helpful in visually depicting the correlation between the variables and what the correlation coefficient is.

Correlational analyses are quite helpful in determining the strength of an assumable linear relationship between variables governed by the dimensionless absolute (r) value (Janse et al., 2021). The r value ranges from -1 to 1, in which 1 represents a perfect/ very strong relationship, and 0 indicates no relationship present (Schober et al., 2018). Considering that the r values range from -1 to 1, the negative and positive signs indicate the direction of the association. The former implies that values closer to -1 indicate a perfect/ strong negative association, and the latter indicates that values closer to 1, as seen in Table 2-9, indicate a perfect/ strong positive association (Akoglu, 2018).

Table 2-9: Summary of the absolute r value and associated strength of the relationship.

The absolute value of r	The strength of the relationship
r < 0.3	Very weak
0.3 < r < 0.5	Weak
0.5 < r < 0.7	Moderate
r > 0.7	Strong

2.12.1 Simple Linear Regression

A Simple Linear Regression (SLR) was conducted as part of the statistical analysis process to investigate the additive and linear relationship between numeric variables. R-squared (R^2), Mean Absolute Error (MAE), Mean Square Error (MSE) and Relative Mean Bias (RMB) to see how efficiently the regression model performed. A SLR is a parametric method used to determine the regressand variable from a provided set of regressor variables. Regressand and regressor are used interchangeably for dependent (y) and independent (x) variables. Herein referred to as regressand and regressor variables. When conducting a SLR, several assumptions need to be maintained and these were maintained for the study as well. The assumptions are:

- **Linearity:** The relationship between the regressor and regressand must be additive and linear.

- **Numeric:** Variables of interest must be continuous numeric values, not categorical.
- **Independence of errors:** Error-values are statistically independent.
- **Normality of errors:** Error terms need to be normally distributed for and provided value of x
- **Equal variance (homoscedasticity):** Error terms must have a constant variance.
- Regressor variables should not be correlated

Therefore, the SLR was calculated as follows:

$$y = \beta_0 + \beta_1 * (X) + \in$$

Equation 2-11: Simple Linear Regression equation

Where y is the regressand variable (Dependent variable); β_0 is the intercept; β_1 is the gradient; X is the regressor (Independent variable), and $\in$ is the error term. This equation is also known as $y = mx + c$, where m (gradient) is measured by estimating the change in y when x increases by a unit of 1 and c (intercept) is the estimated value of the regressand when the regressor is equal to 0 (Pierce et al., 2020).

2.12.2 Ordinary Least Square Regression Model

The OLS model is a common statistical method used for establishing linear regression coefficients. It also consists of two betas (β), which are selected to reduce the square distance (minimise the square of error) between the regressor and regressand variable (Lakshmi et al., 2021). It is a branch of statistics from the generalised linear regression. This statistical model was used due to its computational simplicity in calculating statistical parameters. It is considered a constrained model, known for performing better even when noise is present. It proves efficient and provides the best approximation for a true regression line. Furthermore, they provide rudimentary results which are easy to understand.

2.12.3 Scikit Linear Regression Model

Scikit Linear Regression Model, also referred to as the Sklearn regression model, is a Machine Learning (ML) library used in Python. This approach to linear regression model building was included in this study only for two AERONET stations (Pretoria_CSIR_DPSS and Pretoria_CSIR_EC) to compare how its performance is to the other two models and if there is space in future studies to adopt this modelling method as part of satellite-based AOD studies. A linear regression was conducted using ML. The difference between this model and the SLR and OLS is that the former uses an ML library – Scikit learn and the latter two models use a computational statistical library – stats model in Python. The stats model library efficiently incorporates NumPy and Pandas libraries, is more inclined to statistical approaches, and has

built-in functionalities for conducting statistical tests. Scikit learn is an ML approach where a regression model is built, trained, tested and validated. ML regression modelling mainly focuses on acquiring and processing data that needs to be fed into the model (Karimian et al., 2023). Therefore, data needs to be separated into training and testing datasets. The undertaken process is shown in Figure 2-15 below and is as follows:

Figure 2-15: Figure shows the process undertaken in building, training and testing a linear regression model.

Datasets were split into training and testing data, a subset of the whole dataset. The training data for this study contained 70% of the data, and the testing data contained 30% of the data. The testing data is used to establish how well the training data performed. This is achieved through this sequential algorithm: ***X_train, X_test, y_train, y_test = train_test_split (X, y, test_size = 0.3, random_state=0).*** As mentioned above, the test size used in this study is 30%, expressed as 0.3. The random state is assigned an integer to avoid random shuffling, which means the same dataset will be split every time the test is run. A validation set is also created and used during the training to validate the model's accuracy. This validation is extracted from the training set using the ***train_test_split()*** method. When conducting this ***train_test_split***, data must be in a 2^D array and not 1^D.

Building a linear regression model does not end here, as checking if the model works is essential. The model must be trained using Python's ***LinearRegression()*** ML library. A line must be fitted into the model, and the gradient and intercept must be examined. After that, predictions must be made on the testing data and compared with the actual data. Verifying/ evaluating the model will assist in determining its applicability. In this study, the R-squared (R^2), Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) were used as evaluation benchmarks to determine the model's capabilities. Furthermore, a correlation coefficient of (R^2 ≥0.5) was set in this study to benchmark whether $AOD_{AERONET}$ can serve as a proxy to $PM_{2.5}$ estimations.

When calculating, the MAE entails subtracting the predicted values from the actual values, which will result in an error sum of absolute values where their mean is calculated. The MAE was calculated as follows:

$$mae = \left(\frac{1}{n}\right)\sum_{i=1}^{n}(Actual - Predicted)$$

Equation 2-12: Measure of Mean Absolute Error (MAE) equation

The MSE squares the absolute values of the error. The closer the value to 0, the better. The MSE was calculated as follows:

$$mse = \sum_{i=1}^{n}(Actual - Predicted)^2$$

Equation 2-13: Measure of Mean Square Error (MSE) equation

The RMSE is used as a performance indicator in regression modelling. It squares the value obtained by the MSE. The RMSE was calculated as follows:

$$rmse = \sqrt{\frac{\sum_{i=1}^{n}(Actual - Predicted)^2}{n}}$$

Equation 2-14: Measure of Root Mean Square Error (RMSE) equation

The RMB is used as a performance indicator in regression modelling that is calculated by averaging the deviation of two different datasets. The RMB was calculated as follows:

$$rmb = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}$$

Equation 2-15: Measure of Relative Mean Bias (RMB) equation

2.13 Spatially Mapping MODIS satellite and MERRA-2 re-analysis data

Objective iii will use satellite and re-analysis data to show how AOD presents in the unmonitored regions as well as characterise the different aerosol particles. To achieve this objective, this section will show how AOD is represented throughout the province of Gauteng.

2.13.1 $MODIS_{TERRA}$, $MODIS_{AQUA}$ and MERRA-2 remote data

Data was collected on the GIOVANNI website, a web-based system developed by NASA's Goddard Earth Sciences (GES) Data and Information Services Centre (DISC). This web-based system offers a plethora of functions that enable the end user to access different kinds of satellite data and reconnoitre a geographic area and visualise it (Campos et al., 2022). Table 2-10 shows how the data download process was undertaken.

Table 2-10: Table indicating the dataset name, satellite, variable, measurement parameter, spatial resolution, temporal resolution and the date range for $MODIS_{(TERRA\ AND\ AQUA)}$ and MERRA-2.

Dataset name	Satellite	Variable	Parameter	Spatial Resolution	Temporal Resolution	Date Range
MOD08_M3_v6.1	$MODIS_{TERRA}$	AOD	DB, DT, Combined DT & DB AOD at 550 nm	1° x 1°	Monthly	2017 - 2021
MYD08_M3_v6.1	$MODIS_{AQUA}$	AOD	DB, DT, Combined DT & DB AOD at 550 nm	1° x 1°	Monthly	2017 - 2021
M2TMNXAER v5.12.14	MERRA-2	AOD	Total Aerosol Extinction AOD at 550nm	0.65° x 0.5°	Monthly	2017 - 2021

For this objective, Collection 6.1 (C6.1) $MODIS_{(TERRA,\ AQUA)}$ AOD data were downloaded at a monthly temporal resolution, at a spatial resolution of 1° x 1°. Both retrieval algorithms for MODIS, Dark Target (DT) and Deep Blue (DB) were used to ensure quality AOD data. For the bounding region, Southern Africa was used. For $MODIS_{TERRA,}$ the parameter used was *DB, DT* and the *Combined DT and DB AOD at 550nm for land and ocean (MOD08_M3 v6.1)*, and for $MODIS_{AQUA}$, the parameter used is also *Combined DT and DB AOD at 0.55 micron for land and ocean (MYD08_M3 v6.1)*. L3 data was downloaded, which facilitated easier satellite analysis for map making. MERRA-2 data was also downloaded and used in this chapter. MERRA-2 data were also downloaded at a monthly temporal resolution; however, its spatial resolution is 0.5° x 0.65°. The parameter used was the *Total Aerosol Extinction AOD at 550nm (M2TMNXAER v5.12.14)*. For data visualisation, PNG and Network Common Data Form (NetCDF) files, which are in raster format, were downloaded, and maps were created on ArcGIS using ArcMap version 10.8.1.

Data for both $MODIS_{(TERRA,\ AQUA)}$ and $AOD_{MERRA\text{-}2}$ were downloaded on an annual and seasonal basis to assess the spatiotemporal variability of AOD. MERRA-2 has a different spatial resolution than MODIS and was, therefore resampled to have the same spatial resolution as that of MODIS. Resampling/ re-gridding of MERRA-2 data was conducted on ArcGIS v10.8.1 using a bilinear interpolation technique, which uses the four nearest input points in a grid to obtain one cell centre. Furthermore, a resampling methodology similar to an international study by Nguyen et al. (2015) was adopted to suit South Africa's spatial characteristics. Therefore, both MODIS and MERRA- 2 data had to be resampled to a finer resolution of 0.01° (latitude) x 0.01° (longitude), which is equivalent to a resolution of 1 km x 1 km to obtain optimal high-quality raster images. This process was undertaken because when working with raster data, data needs to be at its finest resolution. This technique utilises the raster processing tools found in ArcGIS. This resampling technique was chosen as it is quite useful for continuous datasets. The satellite images downloaded from

GIOVANNI are displayed in the ‘smoothed’ and ‘non-smoothed’ options. This is termed ‘smoothing’, but this operation is not an actual mathematical smoothing calculation.

GIOVANNI uses Matplotlib's filled contour algorithm to create contoured plot images. This algorithm removes the line boundaries between data pixels. The scaling was linear, and the projection was equidistant cylindrical. For the monthly NetCDF data files, seasons with missing months were discarded. Furthermore, seasonal data were downloaded in the following format: DJF = summer, MAM = Autumn, JJA = Winter and SON = Spring. Summer, Autumn, Winter, and Spring will be referenced throughout ***chapter 5***, including figures, results, and discussion.

****

This chapter has provided the research design undertaken in this research project, accompanied by a conceptual diagram of how this research will be conducted. A quantitative approach was used, and various methods have been undertaken to achieve the aim and objectives in ***chapter 1, section 1.9***. Subsequent data quality control and analysis, only reliable and quality data was used. The next chapter will characterise ground-based $PM_{2.5}$ concentrations in the Gauteng province.

3 CHARACTERISING GROUND-BASED $PM_{2.5}$ CONCENTRATIONS IN THE GAUTENG PROVINCE

This chapter presents the findings derived from the characterization of ambient air quality in Gauteng. It encompasses an exploration of the region's meteorological parameters and an analysis of hourly, daily, and annual concentrations of ambient $PM_{2.5}$. Additionally, a comprehensive discussion section synthesises these results, offering a cohesive understanding of the interrelationships among the various factors examined.

****

The meteorological parameters considered for this study are temperature, relative humidity and rainfall. This consideration contributes to assessing how meteorological parameters influence emission concentrations in the Gauteng province. Chen et al. (2018) study proved that temperature, relative humidity and wind speed are great impactors in $PM_{2.5}$ emission concentrations. With temperature data, an inverse relationship with $PM_{2.5}$ is expected to be seen as concentrations generally show an increase during the winter and a decrease in the summer (Lourens, 2012). Relative humidity plays a role in particle growth due to its hygroscopic nature, which can extend its presence in the atmosphere. Rainfall was also a considered parameter as it contributes to the decrease of $PM_{2.5}$ concentrations, which will assist in seeing where $PM_{2.5}$ concentrations decrease due to rainfall (Pérez et al., 2020). Wind speed is a great disperser of pollutant concentrations but also acts as a deposition agent. Understanding wind speed changes will assist in establishing if $PM_{2.5}$ has dispersed over an area. These parameters were chosen as candidate parameters to understand the short-term variability patterns (Chen et al., 2013).

3.1 Temperature

Table 3-1 depicts the average temperature of all monitoring stations of interest. Due to the geographical nature of the study area, temperature values above 35°C are considered high levels and values below 0°C are low.

The highest hourly average recorded temperature (47°C) is in Vanderbijlpark. Monitoring stations with high hourly temperature readings are Tshwane, Rosslyn, Bodibeng, Hammanskraal, Diepkloof, Kliprivier and Three Rivers, with readings of 46°C, 39°C, 38°C, 38°C, 37°C, 37°C and 37°C, respectively. These temperatures indicate hourly averaged readings at each station. Olifantsfontein reached a maximum of only 24°C through the study period (2017 to 2021), which does not fall in the category of high temperature as the stations mentioned above.

The lowest hourly average recorded temperature was at Kliprivier and Three Rivers, with a reading of -9°C. The low temperatures, normally associated with low ambient daily temperatures,

were recorded between June and July. Subsequent monitoring stations to record lower hourly averaged temperatures are NWU Vaal, Diepkloof, Hammanskraal, Jabavu, Alexandra, Bedfordview and Vanderbijlpark, with readings of -7°C, -4°C, -3°C, -3°C, -2°C, -2°C and -2°C.

Table 3-1: Descriptive statistics for hourly temperature data, in °C for 2017 to 2019.

Temperature									
Station name	**Count**	**Min**	**25%**	**Mean**	**Std**	**50%**	**75%**	**99%**	**Max**
Alexandra	26828	-2	13	18	7	18	22	31	35
Bedfordview	28360	-2	14	17	6	18	21	29	33
Bodibeng	36335	3	17	21	6	21	25	34	38
Diepkloof	41569	-4	13	17	6	17	22	31	37
Etwatwa	740	10	15	19	5	18	23	28	29
Hammanskraal	31929	-3	15	19	7	19	24	34	38
Jabavu	35264	-3	14	18	6	18	22	30	34
Kliprivier	39956	-9	12	17	8	17	22	32	37
NWU Vaal	19223	-7	13	17	8	18	23	33	36
Olifantsfontein	826	1	7	12	6	12	18	23	24
Rosslyn	36998	2	16	21	6	21	25	34	39
Sedibeng	35362	-6	14	18	7	18	23	32	36
Sharpeville	29893	-6	14	18	7	18	23	33	35
Springs	--	--	--	--	--	--	--	--	--
Three rivers	37000	-9	12	17	8	17	22	32	37
Tshwane	13535	-1	15	19	7	19	24	34	46
Vanderbijlpark	35384	-2	15	19	6	19	23	33	47
Gauteng	43659	-5	14	18	6	18	23	32	36

A surface inversion occurs when temperatures from the Earth's surface increase with height. Since South Africa's climate and weather systems are dominated by subtropical highs (22°S and 33°S), temperature inversions are strongest and last longer during winter (Yin et al., 2022). Surface inversions normally occur around sunrise, where warm air rises and weakens throughout the day, resulting in the cooling down of air at ground level. The warm air acts like a "blanket" and settles above the cool air, creating an inversion. Small particles like dust, smoke, and $PM_{2.5}$ become trapped along with the cool air beneath it, preventing vertical mixing (Yin et al., 2022).

In Figure 3-1, diurnal patterns, trends are observed mostly in the early morning and afternoon hours. Low temperatures are observed in the early morning and last for about an hour, while high temperatures are observed in the afternoon, lasting for about two hours.

All stations follow a similar diurnal pattern: Summer has high-temperature peaks, and winter has low diurnal temperatures. The seasons of Autumn and Spring also follow a similar diurnal pattern

where Spring concentrations are higher than the Autumn concentrations. The averaged temperature readings throughout Summer, Autumn, Winter and Spring are 28°C at Bodibeng, 26°C at Rosslyn, 23°C at Bodibeng and 27°C at Tshwane, respectively. All high temperature concentrations seem to be reached during the afternoon hours of 12:00-14:00. Low temperatures observed in Summer, Autumn, Winter and Spring are 14°C at Etwatwa, 9°C at Kliprivier, 2°C at NWU Vaal and Three Rivers and 7°C at Olifantsfontein, respectively. These diurnal temperature dips were observed in the early morning hours of 04:00-05:00. Regarding the readings found at Olifantsfontein, these may not be conclusive readings because the station presented some instrumental errors.

Furthermore, stations with observed instrumental errors are Etwatwa and Olifantsfontein. The former has diurnal temperature readings for the summer only and the latter for the Winter and Spring. Furthermore, Springs also has no data available. All three stations mentioned above have the same owner and fall under the same Metropolitan municipality (CoE).

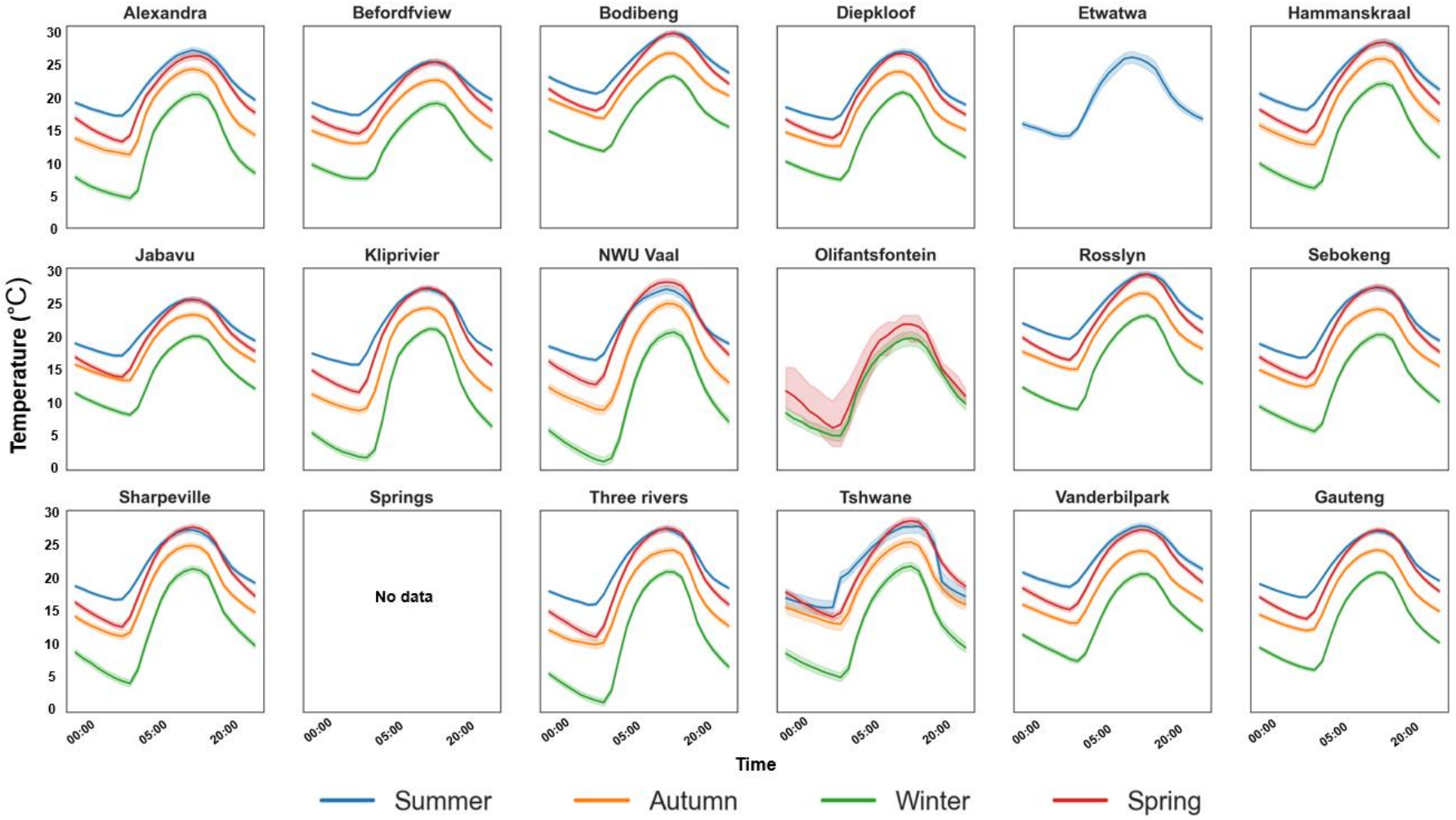


Figure 3-1: Diurnal plot from hourly averaged temperature (°C) data, for all stations. The shaded area represents a 95% confidence interval (CI).

3.2 Relative humidity

Table 3-2 shows that all stations monitored the relative humidity throughout the study period (2017 to 2021). A humidity level of 70% and above is considered high humidity, 50% is considered moderate humidity, and 30% and below is considered low humidity.

Table 3-2: Descriptive statistics for hourly relative humidity data, in %, for all stations.

Relative Humidity									
Station name	**Count**	**Min**	**25%**	**Mean**	**Std**	**50%**	**75%**	**99%**	**Max**
Alexandra	26289	4	25	57	36	58	78	96	99
Bedfordview	11264	1	25	52	31	51	73	98	100
Bodibeng	36286	2	22	47	30	45	63	95	100
Diepkloof	41570	3	21	47	30	47	65	84	88
Etwatwa	740	17	18	59	45	62	76	84	85
Hammanskraal	31928	3	23	52	34	52	70	95	100
Jabavu	34555	0	26	53	32	51	74	100	100
Kliprivier	39857	2	24	53	34	55	74	94	100
NWU Vaal	17146	5	27	59	35	60	84	100	100
Olifantsfontein	850	2	22	46	29	44	63	98	100
Rosslyn	36942	5	23	54	35	53	71	98	100
Sedibeng	35394	2	20	48	32	48	65	84	91
Sharpeville	29720	2	24	48	29	48	67	100	100
Springs	608	45	11	57	48	50	69	72	96
Three rivers	36920	2	23	52	34	54	72	92	100
Tshwane	13167	1	24	53	34	53	72	95	100
Vanderbijlpark	35290	2	25	52	31	50	73	98	100
Gauteng	43659	3	22	52	34	52	70	91	95

Most monitoring stations seem to have reached the maximum humidity level, 100%, at some point during the monitoring period, with two stations having a relatively high relative humidity observed at 85% and 88%. These monitoring sites are at Etwatwa and Diepkloof. All stations reached maximum humidity levels except for Alexandra, Diepkloof, Etwatwa, Sedibeng and Springs. Stations with humidity below the considered moderate level and above the considered low level are Bodibeng, Diepkloof, Olifantsfontein, Sedibeng and Sharpeville. ***Section 3.5*** will discuss the seasonal and diurnal patterns of humidity. The section will further discuss the relationship between relative humidity and ambient temperature, as the two variables have an inverse relationship.

Figure 3-2 shows high humidity levels for Summer, Autumn, Winter and Spring are all observed at the NWU Vaal monitoring station, where the readings are 90%, 90%, 88% and 80%, respectively. Peak humidity levels are observed in the early morning hours at 03:00-04:00. Low humidity levels for the Summer, Autumn, Winter and Spring are 39% at Etwatwa, 34% at Sharpeville, 19% at Sharpeville and 18% at Olifantsfontein. A decrease in humidity levels is observed in the late morning and afternoon hours of 15:00.

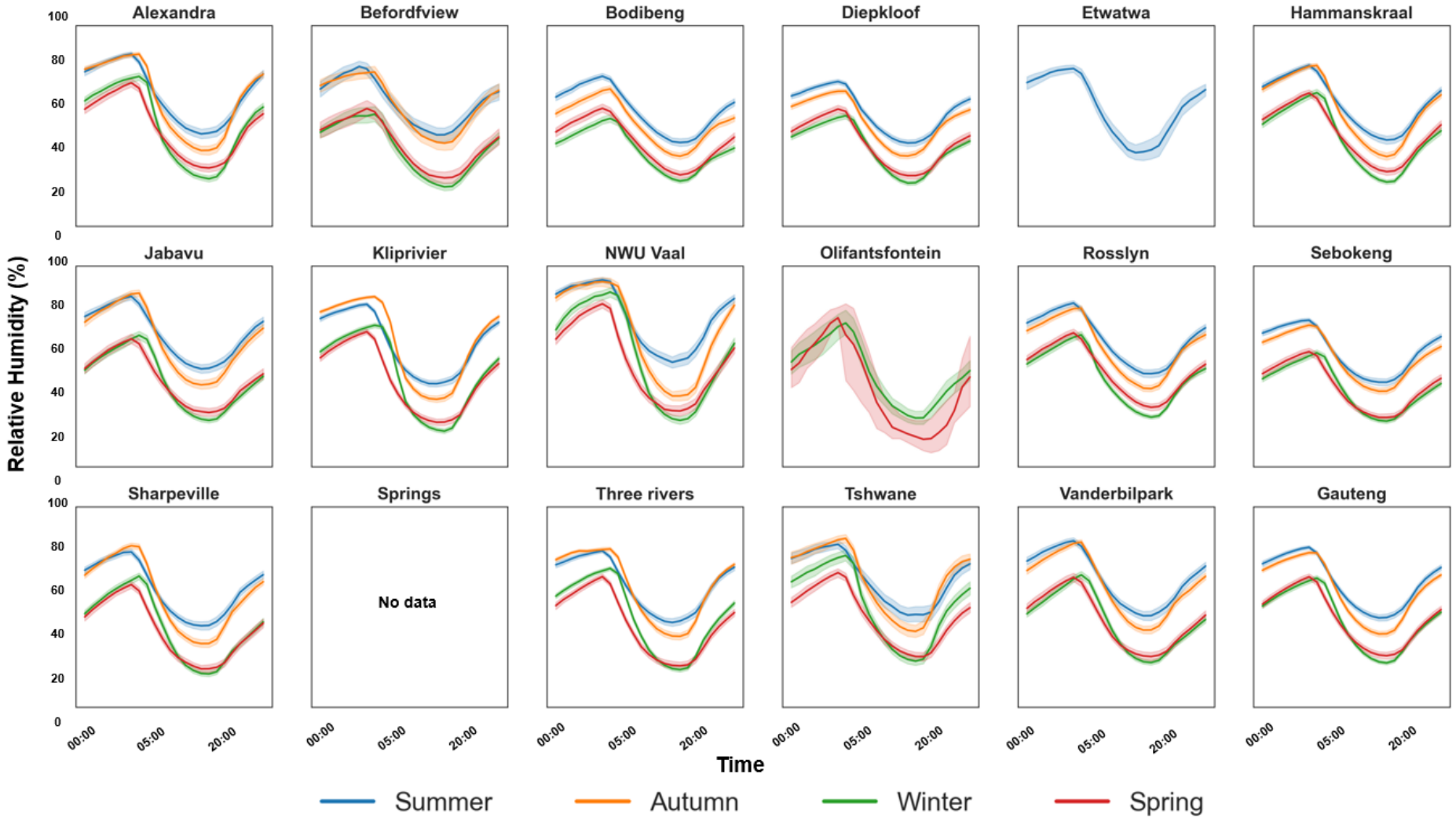


Figure 3-2: Diurnal plot from hourly averaged relative humidity (%). The shaded area represents a 95% confidence interval (CI).

3.3 Rainfall

All monitoring stations have a similar trend where rainfall is observed in the first four months of the year, January to April, except for Tshwane, where rainfall is observed in January and February. Furthermore, rainfall is observed in November and December at all monitoring stations. This phenomenon is congruent with the fact that Gauteng is a region that receives its rainfall during the summer (Fauchereau et al., 2003). High variability seems to be congruent for the monitoring station at Tshwane, where the variability in January, November and December are extremely high. This is attributed to the high summer rainfall events experienced in 2019. Since these rainfalls were episodic extreme weather events, the dataset becomes highly variable compared to the other monitoring stations.

The monthly rainfall over 12 months in the Gauteng province from 2017 to 2021 showed that the highest rainfall was received in November, and the lowest was received in July and August. There seems to be a trend where rainfall trends begin relatively high in January and gradually decrease until April. It is followed by almost no rain from May to September, after which it increases until December.

Table 3-3: Descriptive statistics of hourly rainfall, in mm.

Rainfall									
Station name	Count	Min	25%	Mean	Std	50%	75%	99%	Max
Alexandra	25381	0	1	0	0	0	0	2	27
Bedfordview	--	--	--	--	--	--	--	--	--
Bodibeng	32679	0	1	0	0	0	0	2	35
Diepkloof	33301	0	1	0	0	0	0	2	53
Etwatwa	--	--	--	--	--	--	--	--	--
Hammanskraal	29542	0	1	0	0	0	0	4	76
Jabavu	30970	0	1	0	0	0	0	2	35
Kliprivier	33968	0	1	0	0	0	0	2	24
NWU Vaal	--	--	--	--	--	--	--	--	--
Olifantsfontein	--	--	--	--	--	--	--	--	--
Rosslyn	29280	0	1	0	0	0	0	2	47
Sedibeng	24215	0	1	0	0	0	0	1	43
Sharpeville	25381	0	1	0	0	0	0	2	27
Springs	--	--	--	--	--	--	--	--	--
Three rivers	31347	0	1	0	0	0	0	1	43
Tshwane	12290	0	6	0	0	0	0	0	514
Vanderbijlpark	31797	0	1	0	0	0	0	2	72
Gauteng	43704	1	5	0	0	0	0	16	514

3.4 Ambient $PM_{2.5}$ air quality in Gauteng

This section will unpack the yearly, monthly, seasonal, weekly, weekday, weekend and daily variations of $PM_{2.5}$. Descriptive statistics are also provided to visually represent the data distribution to provide further depth to this discussion.

The following sub-sections will be discussed: hourly, diurnal, daily, weekday vs weekend, seasonal and annual $PM_{2.5}$ variations. The abovementioned temporal resolutions will be discussed to see how $PM_{2.5}$ varies over different timescales. Understanding the different impacts temporal resolutions have on $PM_{2.5}$ aids in having a deeper knowledge of weather feedback mechanisms and how aspects like climate change can contribute (Ruppert, 2016; Zhao et al., 2009). Temporal variations were beneficial to this study as they outlined when noticeable peaks and dips were experienced and what possibly could have been attributed to them.

Table 3-4 outlines the central tendency and count, minimum, 25th percentile, mean, SD, 50th, 75th, 99th percentiles and the maximum values. High and low mean values were observed for stations Etwatwa and NWU Vaal, with values of 41 and 22, respectively. High standard deviation values are seen at Hammanskraal and Etwatwa, with values of 64 and 63, respectively. The data variability at these stations is expected to be greater due to higher standard deviation values from the mean (Cooksey, 2020). Stations of low data variability are absent because all standard deviation values exceed the mean values.

Table 3-4: Descriptive statistics for hourly averaged $PM_{2.5}$, in $\mu g.m^{-3}$ for all stations.

Hourly $PM_{2.5}$ ($\mu g.m^{-3}$)									
Station Name	**Count**	**Min**	**25%**	**Mean**	**Std**	**50%**	**75%**	**99%**	**Max**
Alexandra	14161	2	16	34	28	26	44	139	402
Bedfordview	22871	0	14	27	20	21	34	100	345
Bodibeng	18861	0	12	29	26	21	36	132	416
Diepkloof	34650	0	11	23	19	19	30	86	683
Etwatwa	23040	0	12	41	63	21	41	317	996
Hammanskraal	27113	0	11	36	64	20	36	327	998
Jabavu	23792	1	12	37	50	21	39	264	864
Kliprivier	29182	0	15	34	33	26	44	151	945
NWU Vaal	18792	1	9	22	21	16	28	102	214
Olifantsfontein	22343	1	14	37	47	24	42	190	994
Rosslyn	20235	0	11	23	21	18	30	94	498
Sebokeng	24394	0	13	29	28	22	36	140	831
Sharpeville	16618	0	11	37	45	24	46	202	1000
Springs	20592	0	9	28	35	18	33	174	530
Three rivers	25381	0	12	25	20	21	33	91	384

Tshwane	8954	0	8	29	32	16	38	144	307
Vanderbijlpark	19335	1	13	32	31	23	39	148	761
Gauteng	43660	0	19	32	22	26	38	112	522

3.5 Temporal variations in $PM_{2.5}$ mass concentrations

3.5.1 Diurnal variations

Figure 3-3 shows the averaged $PM_{2.5}$ diurnal variations throughout all four seasons. High $PM_{2.5}$ concentrations for Summer, Autumn, Winter and Spring are at Hammanskraal (52 $\mu g.m^{-3}$), at Etwatwa (80 $\mu g.m^{-3}$), at Etwatwa (200 $\mu g.m^{-3}$) and again at Hammanskraal (50 $\mu g.m^{-3}$). In winter, high concentrations are observed at Alexandra (80 $\mu g.m^{-3}$), Hammanskraal (80 $\mu g.m^{-3}$), Jabavu (130 $\mu g.m^{-3}$), Olifantsfontein (80 $\mu g.m^{-3}$), Sharpeville (100 $\mu g.m^{-3}$), Springs (105 $\mu g.m^{-3}$) and Tshwane (75 $\mu g.m^{-3}$). These peaks are observed in the evening hours between 19:00-22:00. Morning winter peaks are observed in the early hours of 04:00-05:00. Low Summer, Autumn, Winter and Spring $PM_{2.5}$ concentrations are observed at Springs and Tshwane (20 $\mu g.m^{-3}$), Alexandra (25 $\mu g.m^{-3}$), NWU Vaal and Tshwane (20 $\mu g.m^{-3}$).

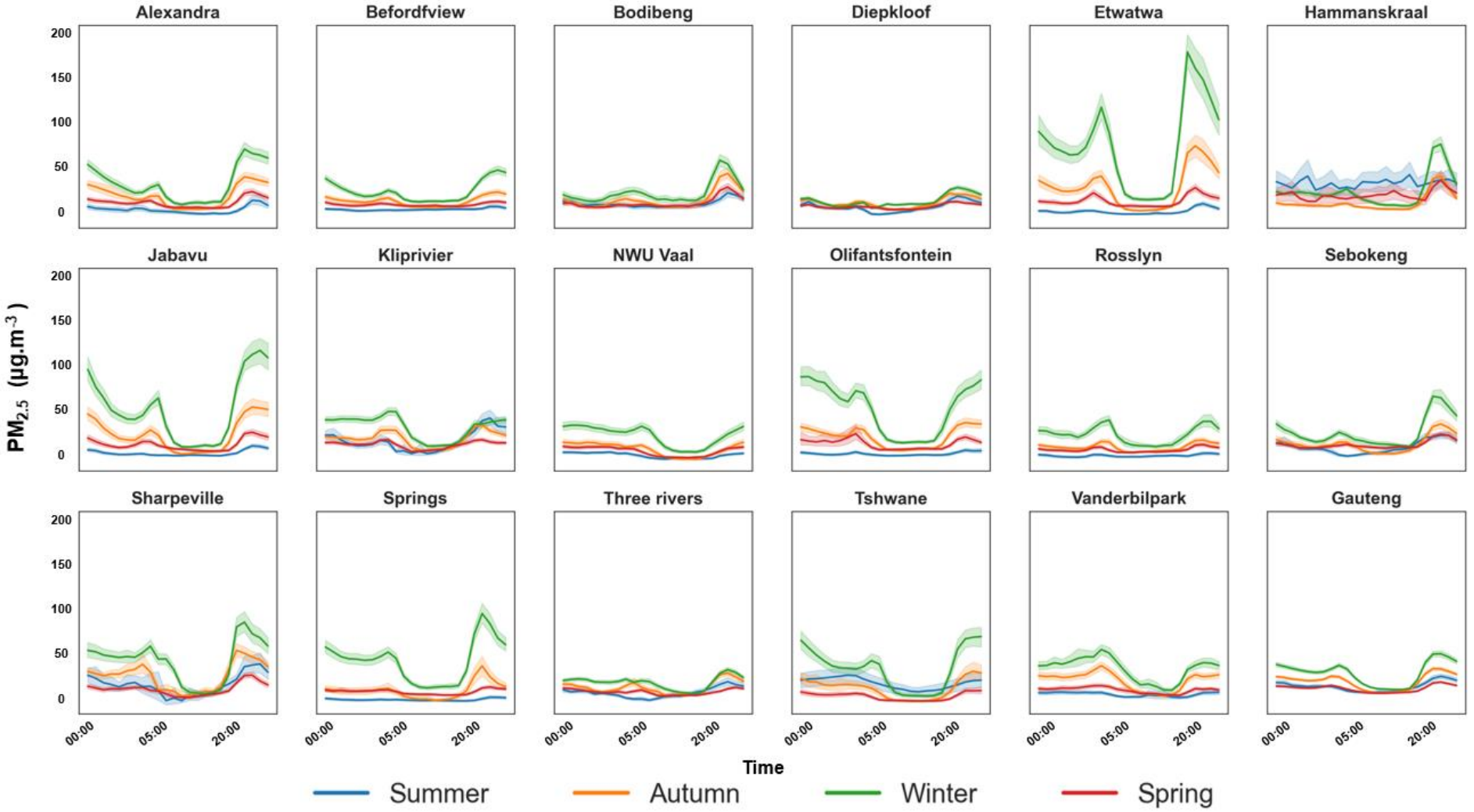


Figure 3-3: Diurnal plot from hourly averaged $PM_{2.5}$ data ($\mu g.m^{-3}$) for the 17 selected stations in Gauteng. The shaded area represents a 95% confidence interval (CI).

3.5.2 Seasonal variations

Figure 3-4 depicts the seasonal boxplots for the 17 AQMSs, which are hourly averaged. The boxplots will assist in evaluating how $PM_{2.5}$ concentrations vary across seasons. Data for all monitoring stations is positively distributed, as depicted in the figure. During summer, data is concentrated below 50 $\mu g.m^{-3}$, except for Tshwane. Autumn does not have high concentrations; however, Etwatwa, Jabavu, Kliprivier, Olifantsfontein and Sharpeville reach maximums of 50 $\mu g.m^{-3}$. High concentrations were observed during winter at Etwatwa with 120 $\mu g.m^{-3}$, Olifantsfontein with 80 $\mu g.m^{-3}$ and Sharpeville with 75 $\mu g.m^{-3}$. Similarly to Autumn, the concentrations for Spring are below 50 $\mu g.m^{-3}$.

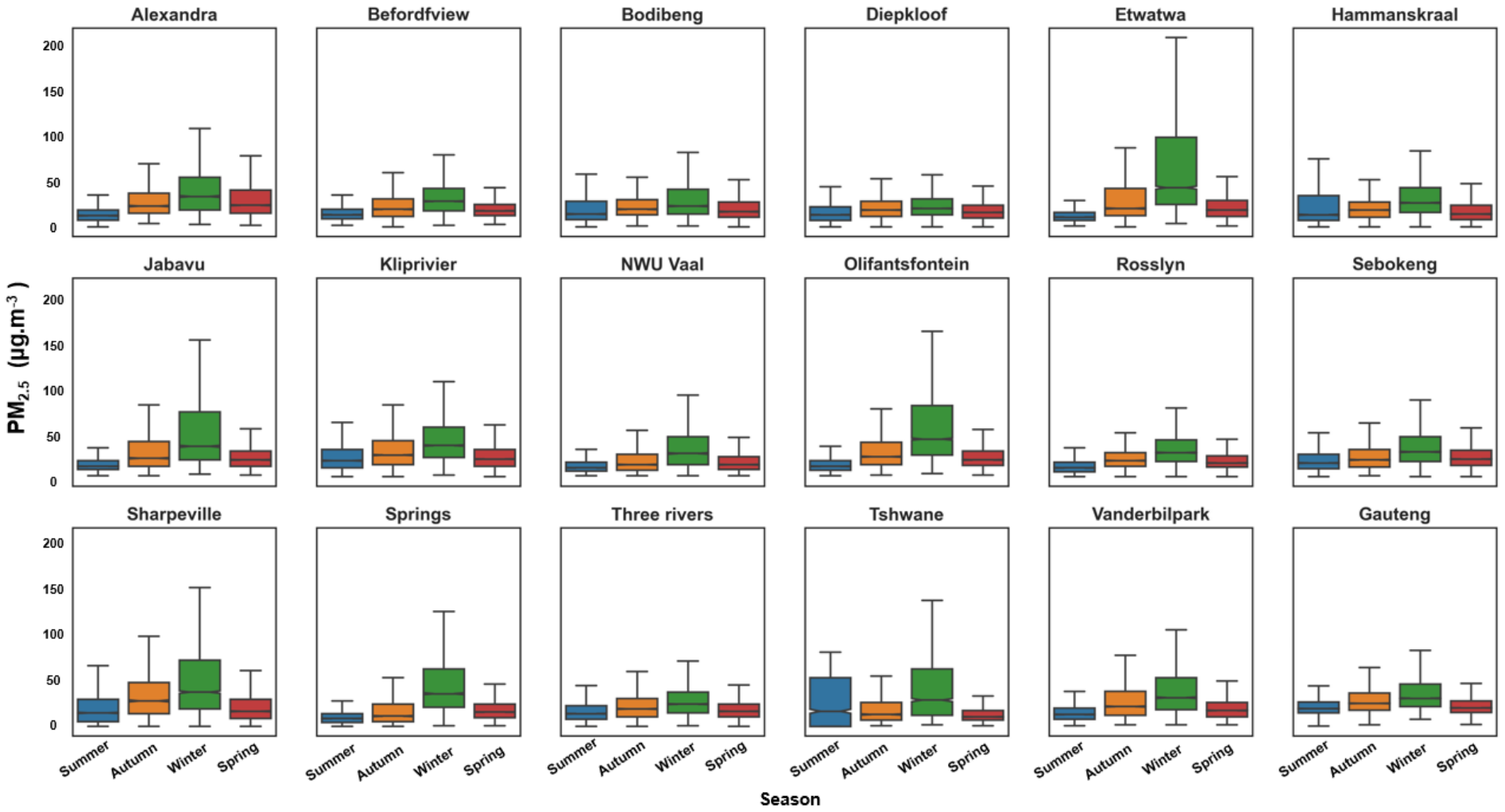


Figure 3-4: Seasonal boxplots (Summer, Autumn, Winter and Spring) from hourly averaged $PM_{2.5}$, in $\mu g.m^{-3}$for all stations.

There is a positive data distribution, for the hourly boxplots, as depicted in Figure 3-5. The months of May (52 µg.m^{-3}), June (64 µg.m^{-3}), and July (58 µg.m^{-3}) have higher $PM_{2.5}$ concentrations than the other months, respectively. June has the highest peak for both the hourly and daily data. February has an abrupt increase in data distribution comparable to January and March. There seems to be a gradual decrease in $PM_{2.5}$ concentrations from August to November, with a slight increase in December. These findings also tie in with the fact that domestic fuel burning is prominent in winter in low-income areas, contributing to the peaks experienced (Rabaji, 2019).

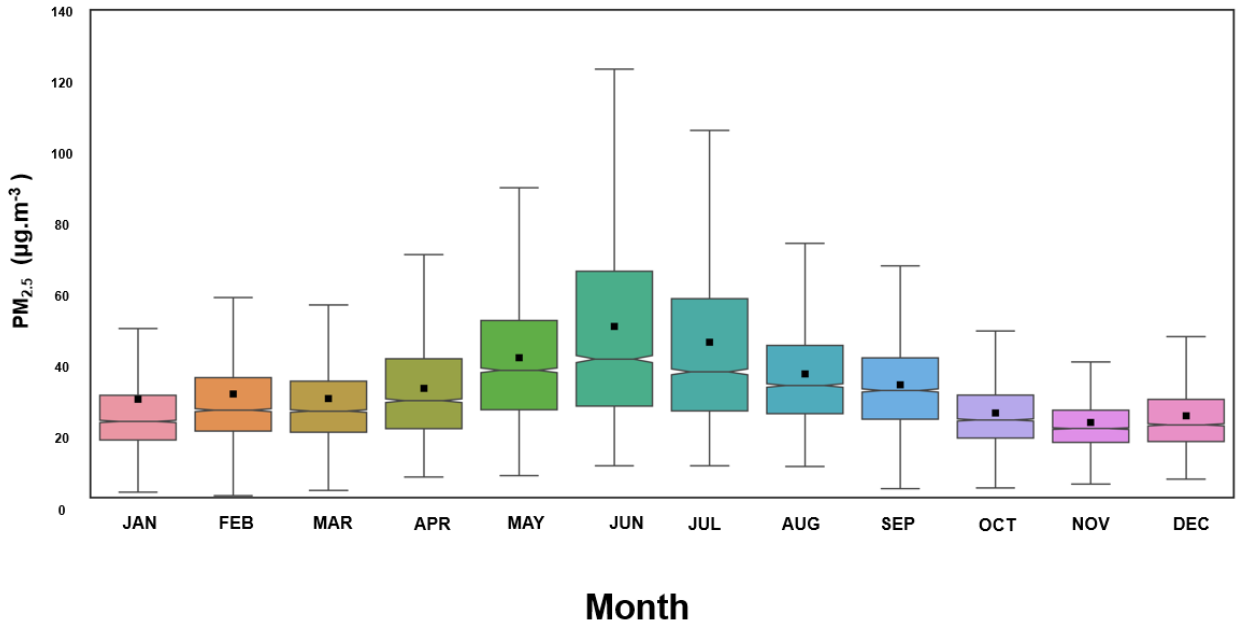


Figure 3-5: Monthly boxplot from January to December from hourly averaged $PM_{2.5}$ data in µg.m^{-3}. (Note: median = 50%; box = 25-75% ; and whiskers = min-max).

Like the preceding figures, Figure 3-6 shows a positive data distribution for all seasons. Furthermore, data for hourly and daily plots is found between the median (50th) and the third quartile (75th percentile). In the summer of 2017, there was an increase in concentrations compared to the other years. In Autumn, high concentrations were observed for hourly and daily plots in 2017, 2019 and 2021. During winter, there is a gradual increase in concentrations from 2017 to 2019 and a decrease in 2020 and 2021 for the hourly $PM_{2.5}$. For the hourly $PM_{2.5}$, 2019 has elevated concentrations. Spring has lower concentrations compared to Summer, Autumn and Winter. 2017 has the lowest concentrations, with a sharp increase is observed in 2018, a decrease in 2019 and 2020 and a slight increase again in 2021.

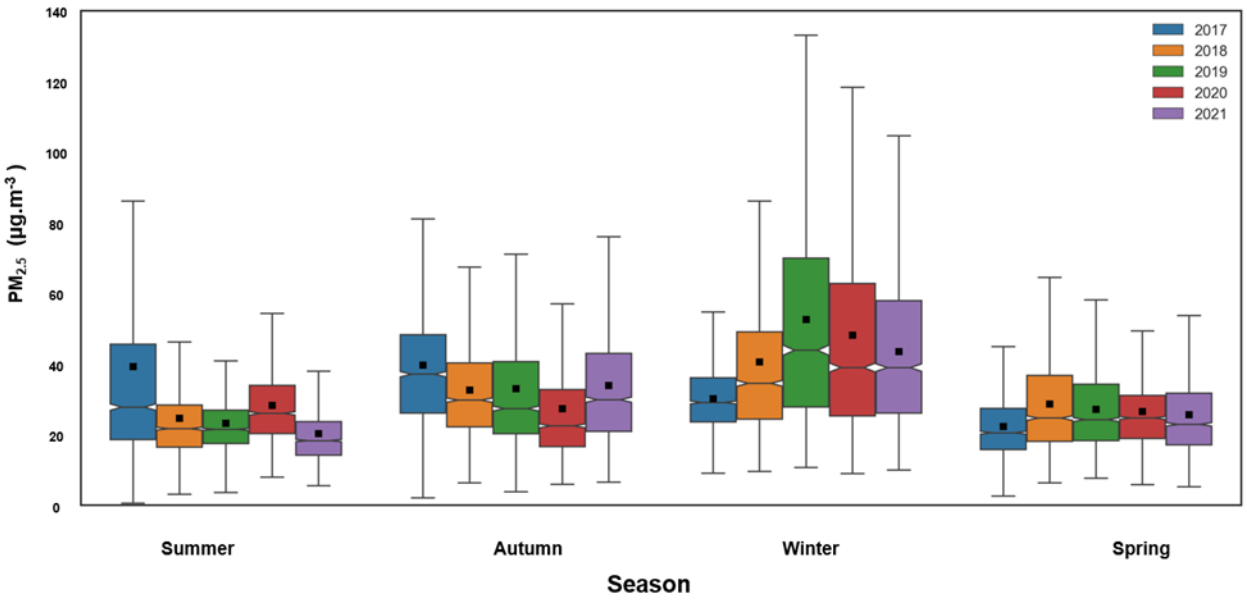


Figure 3-6: Seasonal boxplot (Summer, Autumn, Winter and Spring) from hourly PM2.5 data in µg.m-3. (Note: median = 50%; box = 25-75% ; and whiskers = min-max).

3.5.3 Annual variations

Figure 3-7 depicts the yearly $PM_{2.5}$ concentrations in boxplots. Most monitoring stations do not have data from 2017 to 2021 except for Alexandra, Diepkloof, Hammanskraal, Kliprivier, Sebokeng, Sharpeville and Three Rivers. Concentrations for the year 2017 are below 50 $\mu g.m^{-3}$ for all monitoring stations. The year 2018 depicts a similar trend to 2017, except for Alexandra. For 2018, data is concentrated in the 3rd quartile, with maximum concentrations reaching almost 350 $\mu g.m^{-3}$. For 2019, concentrations exceeding 50 $\mu g.m^{-3}$ were observed at Alexandra, Etwatwa and Sharpeville. For 2020, low concentrations are observed except for Tshwane, with concentrations of 60 $\mu g.m^{-3}$. 2021 also portrays similar patterns to 2020, and there seem to be lower concentrations.

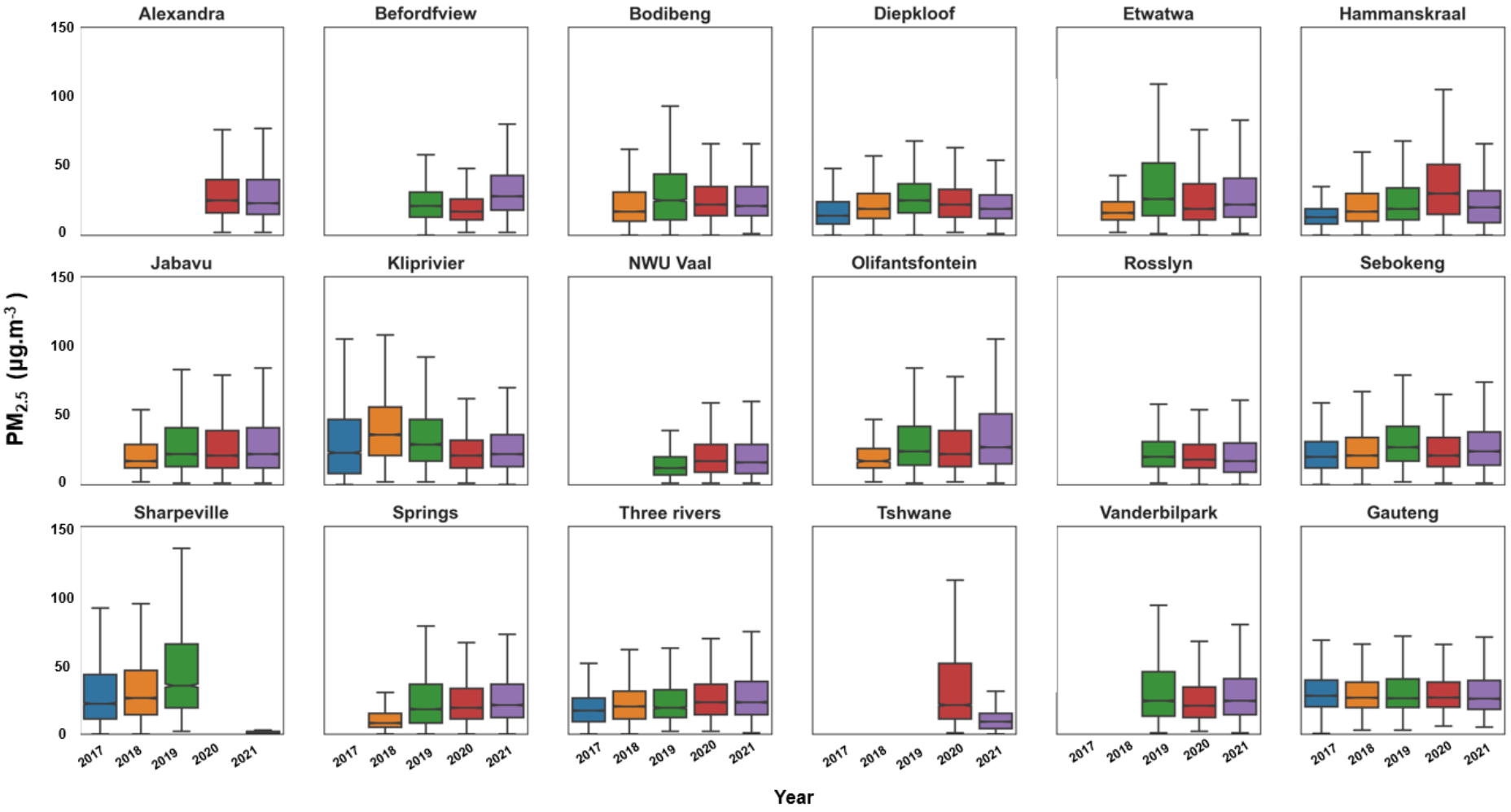


Figure 3-7: Yearly boxplots from hourly averaged data in µg.m^{-3}. (Note: median = 50%; box = 25-75%; and whiskers = min-max).

Table 3-5 outlines the descriptive statistics for daily averaged $PM_{2.5}$ concentrations. Furthermore, exceedances are included in the descriptive statistics table and will be compared to the NAAQS outlined in Table 1-2.

Maximum daily averaged concentrations were highest at Hammanskraal and Olifantsfontein stations with values of 877 $\mu g.m^{-3}$ and 855 $\mu g.m^{-3}$, respectively. Minimum daily averaged concentrations were lowest at Bedfordview and NWU Vaal with values of 90 $\mu g.m^{-3}$ and 93 $\mu g.m^{-3}$, respectively. High mean values are observed at Etwatwa and Hammanskraal, with a value of 41 $\mu g.m^{-3}$. The station with the highest standard deviation value of 73 $\mu g.m^{-3}$ is Hammanskraal, followed by Olifantsfontein and Etwatwa, with values of 47 and 41 $\mu g.m^{-3}$, respectively. A high standard deviation value at Hammanskraal could be attributed to data outliers. With data averaged to a daily resolution, different patterns are observed. Alexandra, Bedfordview, Bodibeng, Diepkloof, Jabavu, Kliprivier, NWU Vaal, Rosslyn, Sebokeng, Sharpeville, Springs, Three Rivers, Tshwane, and Vanderbijlpark have less variability in their data set as the mean values are exceeding the standard deviation values.

Table 3-5: Descriptive statistics for daily averaged $PM_{2.5}$, in $\mu g.m^{-3}$ for 2017 to 2021.

Daily $PM_{2.5}$										
Station Name	**Count**	**Min**	**25%**	**Mean**	**Std**	**50%**	**75%**	**99%**	**Max**	**Exceedances**
Alexandra	632	5	20	35	20	30	45	100	151	213
Bedfordview	963	5	15	27	15	24	35	72	90	173
Bodibeng	897	1	16	29	18	24	36	88	185	181
Diepkloof	1696	0	16	24	12	22	29	57	287	112
Etwatwa	1014	2	15	41	41	24	50	202	272	327
Hammanskraal	1277	0	14	41	73	23	37	352	877	286
Jabavu	1044	5	16	37	33	26	44	161	254	306
Kliprivier	1450	0	20	34	23	29	42	100	283	415
NWU Vaal	796	2	12	22	14	18	28	64	93	93
Olifantsfontein	974	4	16	38	47	27	50	130	855	326
Rosslyn	865	0	13	23	15	20	31	79	118	104
Sebokeng	1206	0	19	29	15	26	36	80	128	217
Sharpeville	825	0	19	37	29	30	46	140	322	259
Springs	905	1	11	28	24	20	37	111	164	200
Three rivers	1279	1	17	26	14	23	31	65	178	133
Tshwane	378	0	11	28	24	20	40	100	112	93
Vanderbijlpark	827	4	17	31	20	26	40	97	130	211
Gauteng	1826	7	22	32	15	29	38	80	173	404

The total number of exceedances for the 24H NAAQS is outlined in Table 3-5 (above), with a major limitation being data availability. All stations have exceedances over these five years. Permitted exceedances are based on the study period and the amount presented for each monitoring station. Depending on the monitoring station owner, most station owners like Eskom or Sasol assess their frequency of exceedances against a "*cumulative frequency of three years worth of data*" as this is part of their compliance with emission licencing (DEA, 2012). The current tolerable permitted exceedances per year for $PM_{2.5}$ with a 24H NAAQS of 40 $\mu g.m^{-3}$ and 20 $\mu g.m^{-3}$ is four (Edlund et al. 2021; Govender & Sivakumar, 2019).

All the stations have exceeded the 40 $\mu g.m^{-3}$ NAAQS. Kliprivier has an exceedance that is tenfold the daily NAAQS. Jabavu has an exceedance that is almost eightfold. These exceedances may be attributed to exacerbated emission levels. Furthermore, these two stations are situated in areas of high population density in which clean energy resources are limited, and the use of solid fuels is greater. Stations that have the lowest frequency of exceedances are NWU Vaal and Tshwane, with exceedances almost threefold than the NAAQS.

Considering that the allowed exceedance per year is four, Figure 3-8 shows aggregated data over five years; therefore, the limit seen on the graph is 20, representing the sum of the four exceedances over the five years. NWU Vaal did not exceed the limit of 20. This could be attributed to the fact that the station was not yet active in 2017 and 2018. It was only commissioned in late 2019. Similarly, neither Rosslyn nor Tshwane had above 20 exceedance, mainly because Rosslyn had no data during 2017 and 2018, and Tshwane had no data available during 2017, 2018, and 2019, with low data availability in 2021. All the other stations did surpass the 20 exceedance limit, with Kliprivier having the highest number of exceedances. 2020 had the highest number of exceedance, followed by 2021, 2019, 2018 and 2017. These depend not only on the emission concentration for that year but also on their data retention level. For 2017 and 2018, there was poor data retention, while 2019, 2020 and 2021 had good data retention (Table 2-2).

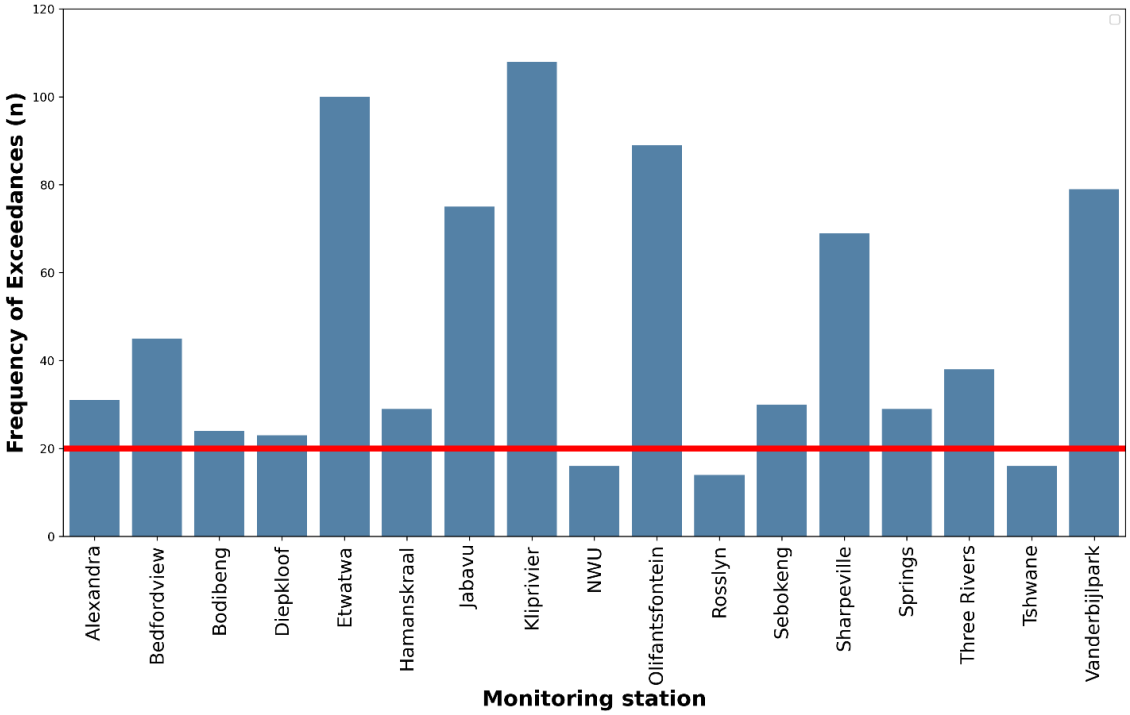


Figure 3-8: Bar plot showing the frequency of exceedances over an aggregated period of five years for all stations. Red line shows the tolerable exceedance of 20 over five years.

Figure 3-9 shows the seasonal exceedances per station per year. From the graph below, winter has the highest percentage of exceedances ranging from 0 to 88%, followed by Autumn (0 to 36%), Summer (0 to 32%) and Spring (0 to 20%). This indicates that during winter, exposed inhabitants of Gauteng may experience negative health impacts due to higher exposure to $PM_{2.5}$. Looking at the seasonal exceedances, all stations had the highest exceedance in winter, 12 stations in Autumn and five stations in summer. Meteorology plays a crucial role in the dispersion of pollutants, which may explain why exceedances may be associated with a specific season. Although a season like Autumn is associated with windy conditions, dispersed pollutants may be trapped in the atmosphere during a temperature inversion and linger during winter, resulting in more exceedances in winter.

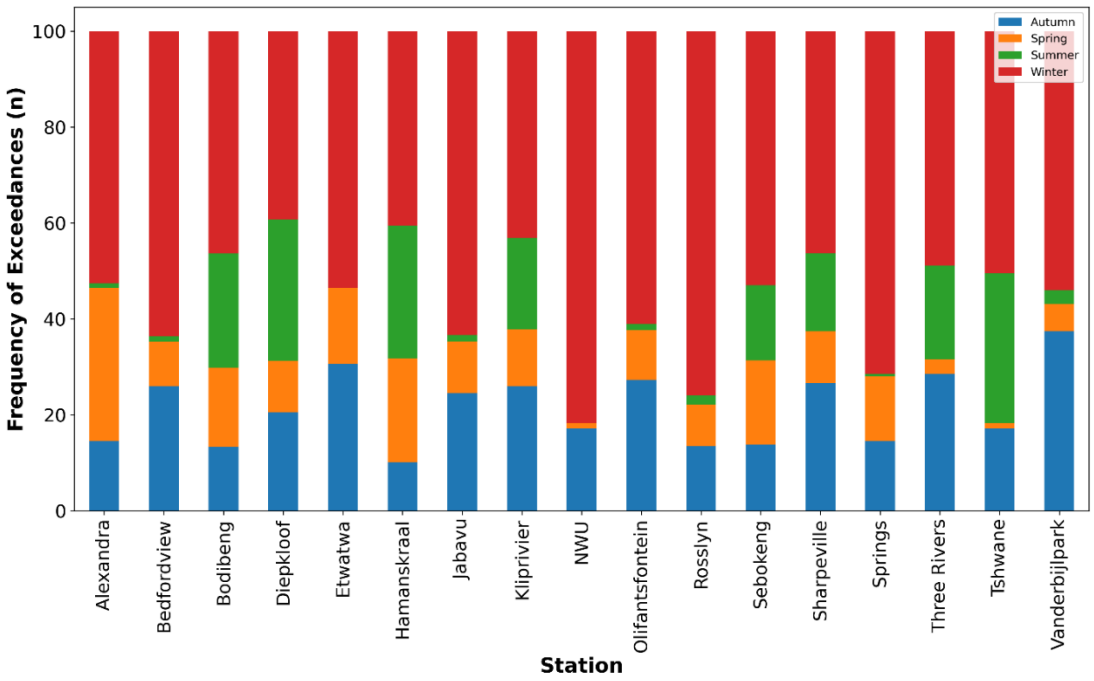


Figure 3-9: Stacked bar plot showing the frequency of exceedances per station by percentage (%) per season.

Figure 3-10 shows the frequency of exceedances per station per year. Only five stations (Diepkloof, Kliprivier, Sedibeng, Sharpeville and Three Rivers) had exceedance in 2017. Exceedances did not surpass 100. In 2018, there were nine stations that exceeded the NAAQS. The stations are Bodibeng, Diepkloof, Etwatwa, Jabavu, Hammanskraal, Kliprivier, Olifantsfontein, Sedibeng, Sharpeville, and Three Rivers. In 2019, all the stations, except NWU and Tshwane had exceedances. In 2020, all stations had exceedances, with Tshwane having its only exceedance in 2020. The increase in exceedances can be tied in with the global pandemic experienced in 2020 as a result of increase domestic fuel burning due to the “lockdown”. However, this is not solid justification for these exceedances and will need to be explored more. In 2021, all stations had exceedances except Sharpeville and Tshwane.

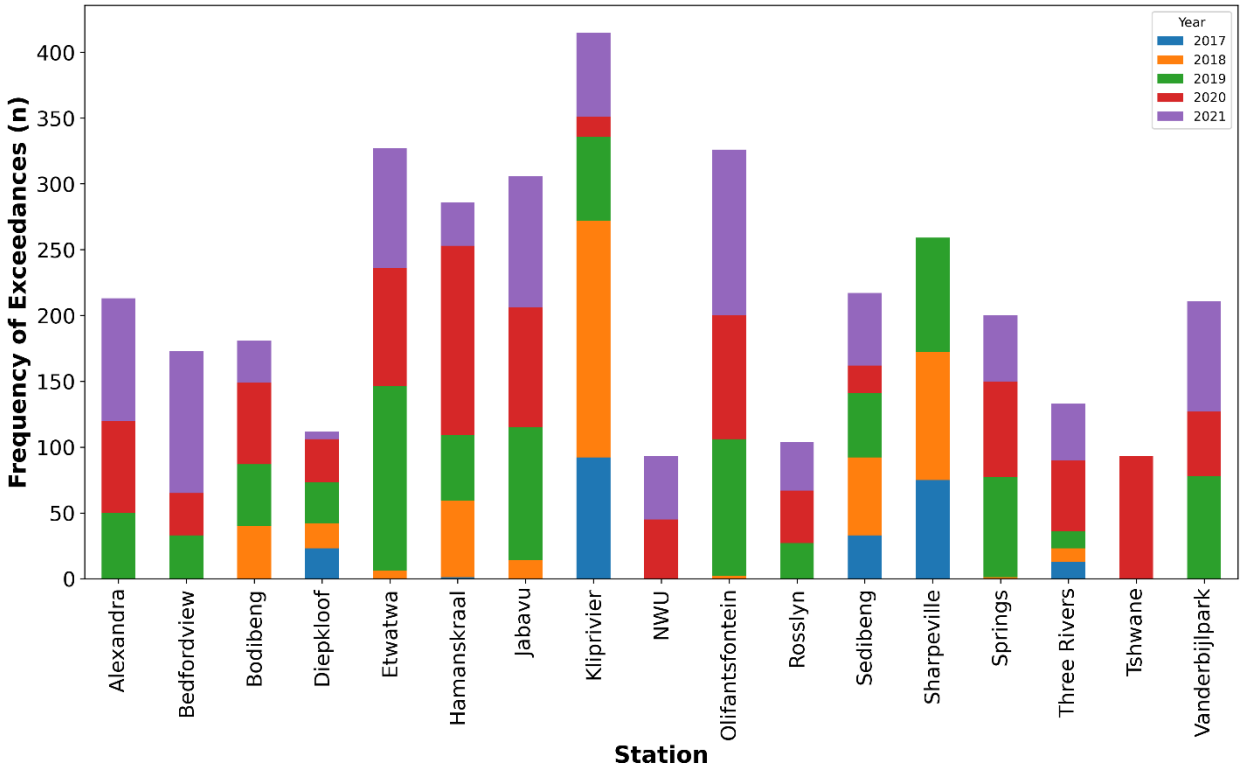


Figure 3-10: Stacked bar plot showing the frequency of exceedances per station per year.

3.6 Relationship between $PM_{2.5}$ and meteorological parameters

The correlation between $PM_{2.5}$ and the meteorological parameters like rainfall, relative humidity, temperature and wind speed are depicted in Figure 3-11. Although wind speed was not included in the analysis, it is included in the heatmap. The figure shows that there is no relationship between the observed variables. Rainfall has a negative correlation of 0.04 with $PM_{2.5}$. Relative humidity has a positive correlation of 0.11 with $PM_{2.5}$. Temperature has a negative correlation of 0.4 with $PM_{2.5}$. Wind speed also has a negative correlation of 0.33 with $PM_{2.5}$. A negative relationship implies an inverse relationship where one variable increases as the other decreases, and a positive relationship implies a relationship where as one variable increases, the other increases, considering the strength of the relationship is sufficient. It is also important to note that high spatial resolution plays a role in the interaction of $PM_{2.5}$ and meteorology. De Lange et al. (2021) outlines that there is a major gap in good quality pollution data at surface level, such that the study highlighted the significance of exploring meteorological air quality models.

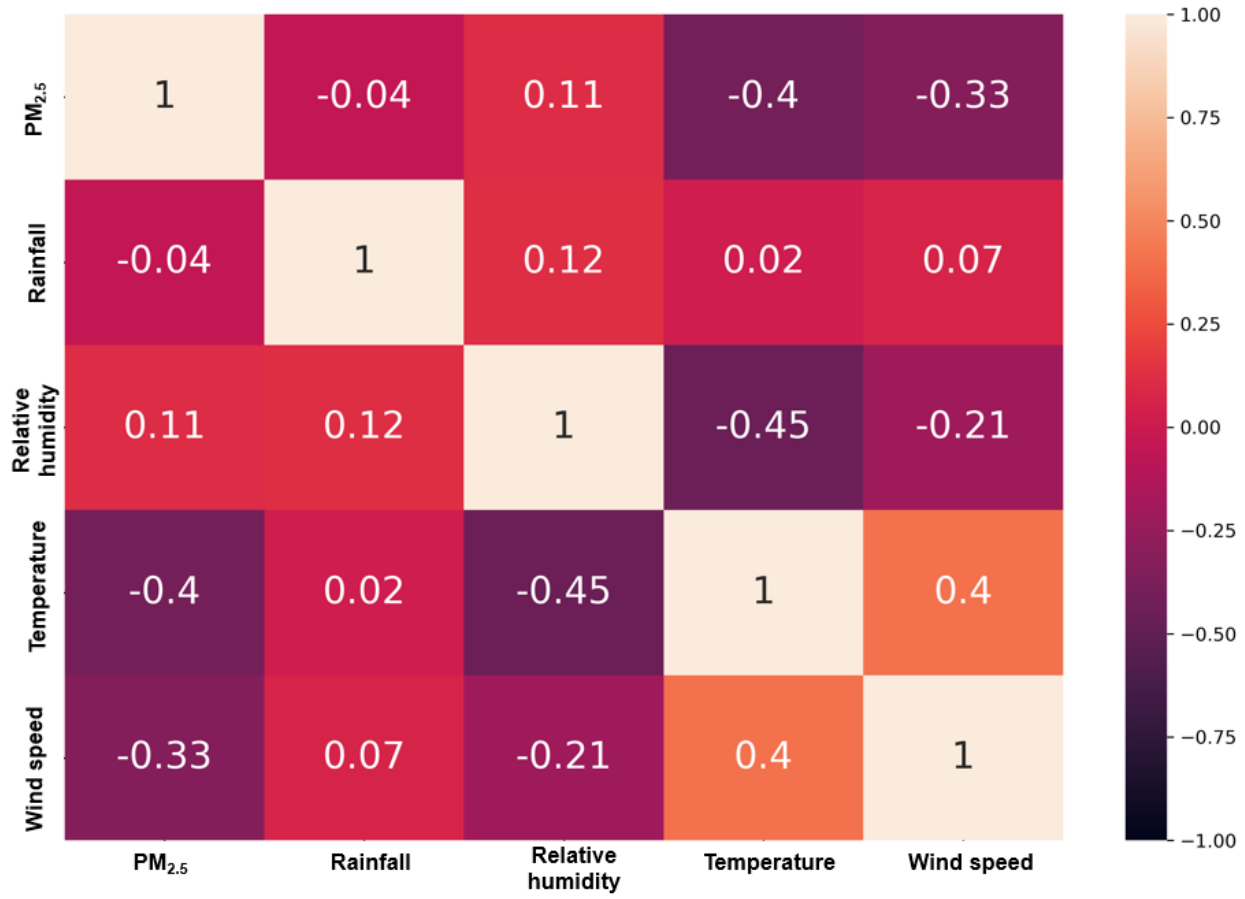


Figure 3-11: Correlation heatmap between $PM_{2.5}$ and meteorological parameters.

3.7 Discussion – South Africa's Meteorological parameters

South Africa's weather is dominated by anti-cyclonic atmospheric circulations, as illustrated in Figure 3-12, whereby in winter (June to August), High-Pressure Systems influence inland atmospheric conditions (Tyson & Preston-Whyte, 2000). In summer (November to March), excluding the Western Cape province, the country experiences seasonal summer rainfalls, in which rainy seasons are common over the country's interior region. Summer daytime temperatures range from 25°C to 32°C and winter daytime temperatures range from 15°C to 19°C.

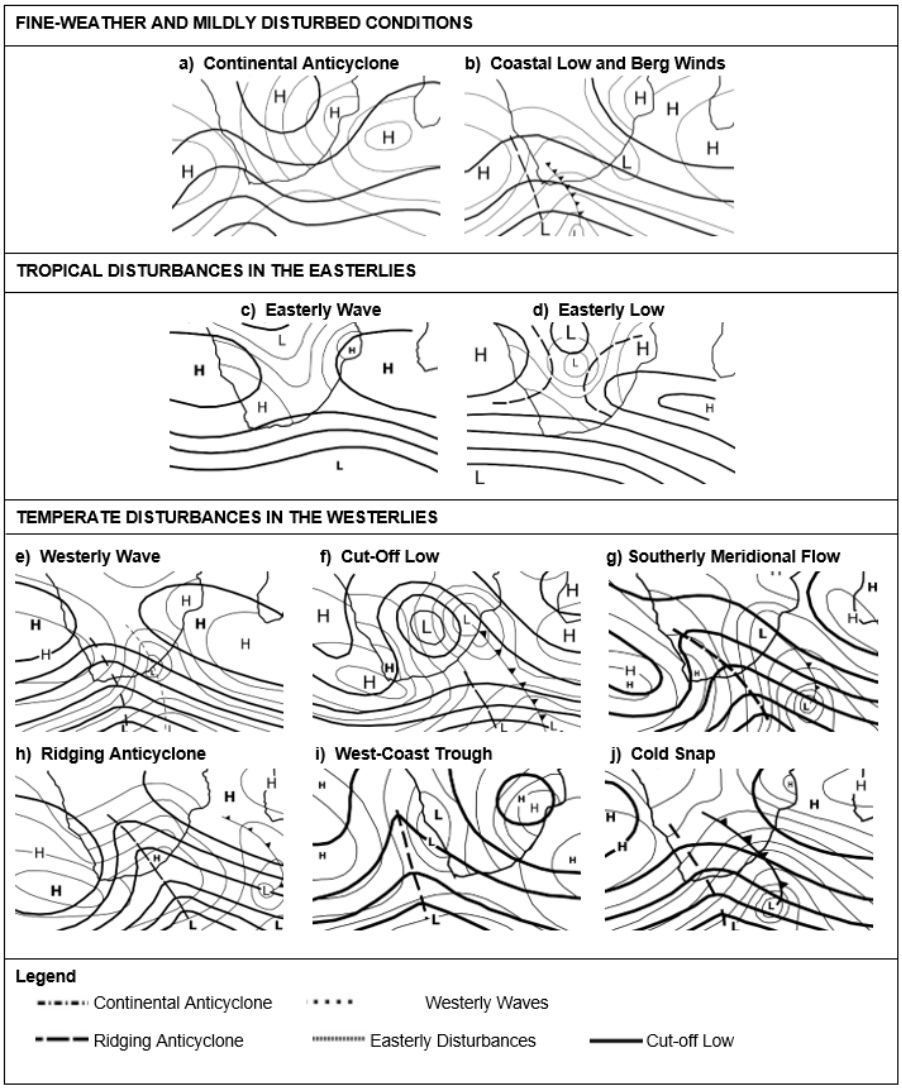


Figure 3-12: Atmospheric circulation and climate over South Africa (Source: Language, 2020, *In: Tyson et al., 1996; Tyson & Preston-Whyte, 2004*).

3.7.1 Relationship between temperature and $PM_{2.5}$

In Figure 3-3, $PM_{2.5}$ concentrations peak in the morning are contemporaneous with the predominance of surface temperature inversions in the early morning hours. $PM_{2.5}$ concentrations increase during the winter, and owing to the prior discussion that surface temperature inversions are strongest, Gauteng is set to experience a decline in air quality. There is a strong negative correlation of r= -0.4 between temperature and $PM_{2.5}$. This means that temperature has little to no direct influence on $PM_{2.5}$ concentrations in the atmosphere. A study by Olutola and Wichmann (2021) also had similar results to this study where their study intended to determine if temperature has a direct relationship with $PM_{2.5}$ and results showed a correlation of r= -0.477. A study by Adeyami et al. (2022) also shows that $PM_{2.5}$ was a correlation of r= -0.49. This phenomenon is explained as a negative response by Chen et al. (2020). This is due to aspects like evaporation loss of $PM_{2.5}$ and temperature related atmospheric convections (Chen et al., 2020).

A study conducted by Rabaji (2019) states that Sharpeville experiences numerous health impacts due to elevated pollution levels. industry. The study further contends that these impacts are exacerbated during the winter when temperatures drop and the need for energy and space heating increases. This decline in air quality will deteriorate the health (respiratory illnesses) and the environment (air contamination) state unless sustainable incentives for energy use are sought after.

3.7.2 Relationship between relative humidity and $PM_{2.5}$

Relative humidity is also an important meteorological parameter in air quality studies as it significantly impacts $PM_{2.5}$ concentrations. Studies have proven that relative humidity and temperature have an inverse relationship (Chidhindi, 2022). Relative humidity is a factor that needs to be considered about the particulate matter as it can change a particle's aerodynamic physical property due to its hygroscopic characteristic. Since humidity tends to be higher in the early morning and evenings, this may affect $PM_{2.5}$ concentration readings (Kapwata et al., 2018). The humidity level during those times tends to be high, as well as the $PM_{2.5}$ concentrations; this may alter the particle size, which may also lead to false instrumentation readings.

$PM_{2.5}$ readings may have been influenced by the humidity level where the particle size from the initial emission concentrations increased by size, which could have been mistaken or captured as a $PM_{2.5}$ reading. However, the results from this study show a positive weak correlation of 0.11 between relative humidity and $PM_{2.5}$. This can imply that the particles size did not increase due to the presence of humidity. Chen et al. (2020) explains this as a positive influence response mechanism. The study contends that low levels of humidity lead to increased evaporation of $PM_{2.5}$. Seasonal variation also plays a role in the relationship between humidity and $PM_{2.5}$. Relative humidity interacts differently with $PM_{2.5}$ differently across all four seasons due to atmospheric conditions presented with each season (Wang & Ogawa, 2015). A local study by Novela et al. (2020) assessed the relationship of relative humidity and $PM_{2.5}$, however in the upper northern parts of South Africa, Thohoyandou. The study yielded different results to this study with a correlation r= -0.26.

3.7.3 Relationship between rainfall and $PM_{2.5}$

Rainfall is a parameter that greatly impacts pollution concentrations and their lifetime in the atmosphere. It can wash out particulate matter and dissolve gaseous substances in the atmosphere. Although the latter may lead to the formation of acid rain, the former is beneficial in improving visibility. Results show a negative correlation of r= -0.04 between rainfall and $PM_{2.5}$. This shows that $PM_{2.5}$ concentrations in Gauteng are not impacted by rainfall. This may be influenced by the amount of precipitation received (Chen et al., 2020). Referring back to Table

3-3, the monitoring stations in the study showed to have recorded high amounts of rainfall in Tshwane. The other stations recorded low rainfall measurements. This shows that the magnitude of rainfall was not vast enough to have a direct impact on $PM_{2.5}$.

3.7.4 Relationship between wind and $PM_{2.5}$

Wind has a negative relationship of r =-0.33. This shows that there is little to no direct relationship with $PM_{2.5}$. This implies that based on these results, wind cannot be considered as a dominant disperser to $PM_{2.2}$ concentrations. A study by Novela et al. (2020) showed that wind speed had a correlation of r=0.18, however, the study was positively correlated with $PM_{2.5}$.

3.8 Ambient $PM_{2.5}$

Numerous studies have been conducted on $PM_{2.5}$; however, no study has brought forth results on all the stations measuring $PM_{2.5}$ in the province of Gauteng, and, in augmentation, very few have gone into detail to unpack the impacts of meteorology on $PM_{2.5}$ concentrations. Results and visual analysis on $PM_{2.5}$ concentrations been deliberated in a concise manner. Yearly, monthly and seasonal diurnal variations are provided in the ***section 3.5***. These will be discussed in the subsequent paragraphs.

In addition to the discussion, Table 3-6 outlines the electrical energy of all municipalities across Gauteng to better understand energy use. As per the data provided by Stats SA (2011), electricity is mostly used for lighting, cooking and heating across all municipalities. CoJ (90.8%) and Sedibeng (90.6%) use more electricity for lighting as opposed to CoT (88.6%) and CoE (82.2%). CoJ (87.4%) and Sedibeng (87.0%) also rank high in using more electricity for cooking, unlike CoE, which only uses 79.4%.

Table 3-6: Energy use table outlining the energy used during cooking, heating and lighting at the City of Tshwane, City of Johannesburg, City of Ekurhuleni and Sedibeng (Stats SA, 2011).

District	Energy use (%)		
	Cooking	Heating	Lighting
City of Tshwane	84.2	73.5	88.6
City of Johannesburg	87.4	82.1	90.8
City of Ekurhuleni	79.4	65.6	82.2
Sedibeng	87.0	79.4	90.6
West Rand	77.7	68.8	81.7

In **section 3.5**, three of the four mentioned stations with concentrations from 201 $\mu g.m^{-3}$ to 300 $\mu g.m^{-3}$ are found in the CoJ. Dominant sources of pollution in the CoJ are mine dumps,

unpaved roads, domestic biomass fuel burning, exhaust emissions, industrial/ mine tailings and illegal dumping (Selani, 2017). Monitoring areas like Jabavu have high emission concentrations attributed to household fuel burning as many homes in Jabavu still own a fireplace despite having access to electricity. Jabavu also has an arterial road that passes through the far east of Jabavu known as Elias Motsoaledi main road, formally known as Roodepoort Road (Nyanda, 2019). Commuters often use this network system to work, which contributes greatly to vehicle and traffic emissions. Rosslyn and Tshwane, both found in the CoT, display elevated $PM_{2.5}$ concentrations due to their geographical placement (Pretoria West) close to the Rooiwal and Pretoria West power stations. A study conducted by Matjyeni (2021) outlined that the dominant sources of air pollution in the CoT are vehicle emissions, domestic biomass burning, industrial processes and power generation, and this study further outlined the harmful health impacts experienced by Tshwane residents due to high exposure levels during the winter months as well as peak hours. The study recommended that "cleaner stoves and chimneys, broader electrification and "low-smoke fuels" be employed to ease the current air pollution problems experienced in Tshwane.

Specific diurnal peaks may be attributed to common activities in low-income areas, like domestic fuel burning for cooking and space heating (Language, 2020). These morning peaks are also attributable to vehicle and traffic emissions (Chidhindi, 2022). During the diurnal cycle, morning and evening peak concentrations can be attributed to solid fuel use. A decrease in $PM_{2.5}$ concentrations at the monitoring stations in Figure 3-3 is seen during mid-day hours, indicating less activity contributes to increased pollution concentrations. To substantiate the diurnal patterns observed in Figure 3-3, an international study by Dobson et al. (2021) conducted in Bangladesh, Dhaka, using $PM_{2.5}$ data to assess the air quality in a low-income area, portrayed similar results. Results were also in accordance with the fact that $PM_{2.5}$ concentrations are higher in the evening than in the afternoon.

****

From the results, it is shown that $PM_{2.5}$ concentrations are elevated during the winter which may be attributable to surface inversions and during the summer season, concentrations are less. During the study period, all monitoring stations, except three exceeded the tolerable frequency of exceedances. This is a strong indication that air quality management legislation should be enforced as there is still non-compliance to NAAQS. The study also indicated that there is no clear link between meteorology and $PM_{2.5}$. The next chapter will evaluate the use of satellite- and re-analysis data over the Gauteng region to validate AOD_{MODIS} satellite data using $AOD_{AERONET}$ data.

4 EVALUATING SATELLITE OBSERVATIONS AND MODELING PRODUCTS IN THE GAUTENG REGION

In this chapter, we delve into the assessment of satellite observations and modelling products in the Gauteng region, with a specific focus on the following aspect: the validation of Aerosol Optical Depth (AOD) observations from $MODIS_{TERRA}$, $MODIS_{AQUA}$ using AERONET observations. The objective is to assess the accuracy of these satellite-derived data in understanding atmospheric and environmental conditions in the Gauteng region.

****

To facilitate the study objective, multiple AOD datasets were scrutinised over the Gauteng region of South Africa. The datasets included are the $MODIS_{TERRA}$ and $MODIS_{AQUA}$ satellite observations, for which the Deep Blue (DB), Dark Target (DT), and Combined DB and DT (DB & DT) were considered.

The region had limited AERONET collocation sites; as such, there were two valid sites used, namely Pretoria CSIR_DRSS (Site 1) and Pretoria CSIR_EC (Site 2). The data from Site 1 was for the period 2017 and 2018, while Site 2 covered 2020 and 2021. None of the stations within the study area had valid data for 2019; thus, this year was excluded from further analysis in this chapter. The limitation to having few AERONET sites is that spatial coverage is not enough.

Table 4-1 and Table 4-2 below show a summary of $MODIS_{(TERRA, AQUA)}$, at two different sites (Site 1 and Site 2). Descriptive statistics are also included for all three retrieval algorithms, DB, DT and DB & DT. The results from these tables enlighten us on the possible aerosol type by assessing the AOD. High AOD values indicate significant light attenuation indicating aerosols are absorbing and a low AOD indicates low light attenuation of the aerosols and aerosols may be expected to be non-absorbing.

Table 4-1: Descriptive statistics summary of the daily $AOD_{MODIS\ (TERRA,\ AQUA)}$ for Site 1 (2017 and 2018).

Satellite product	Retrieval algorithm	Year	2017					2018				
		Season	Annual	Autumn	Spring	Summer	Winter	Annual	Autumn	Spring	Summer	Winter
MODIS Terra	DB	Mean	0.089	0.104	0.078	0.118	0.074	0.075	0.068	0.096	0.066	0.065
		±SD	0.082	0.1	0.057	0.136	0.04	0.093	0.062	0.149	0.056	0.048
		Median	0.065	0.064	0.065	0.069	0.064	0.048	0.041	0.053	0.049	0.046
	DT	Mean	0.169	0.235	0.152	0.179	0.115	0.13	0.146	0.152	0.107	0.112
		±SD	0.194	0.299	0.104	0.161	0.071	0.126	0.137	0.148	0.101	0.107
		Median	0.119	0.17	0.126	0.141	0.105	0.096	0.115	0.112	0.077	0.092
	DB & DT	Mean	0.153	0.235	0.118	0.179	0.098	0.118	0.146	0.132	0.107	0.092
		±SD	0.191	0.299	0.102	0.161	0.069	0.135	0.137	0.178	0.101	0.101
		Median	0.105	0.17	0.093	0.141	0.089	0.077	0.115	0.06	0.077	0.067
MODIS Aqua	DB	Mean	0.095	0.121	0.098	0.054	0.086	0.077	0.069	0.1	0.048	0.075
		±SD	0.096	0.141	0.102	0.033	0.05	0.088	0.059	0.129	0.026	0.071
		Median	0.062	0.062	0.063	0.047	0.071	0.047	0.039	0.061	0.041	0.049
	DT	Mean	0.142	0.193	0.137	0.134	0.108	0.126	0.138	0.16	0.105	0.1
		±SD	0.135	0.196	0.114	0.119	0.071	0.107	0.105	0.137	0.088	0.078
		Median	0.109	0.119	0.113	0.109	0.1	0.099	0.13	0.13	0.083	0.083
	DB & DT	Mean	0.13	0.193	0.113	0.134	0.097	0.115	0.138	0.141	0.105	0.081
		±SD	0.132	0.196	0.109	0.119	0.066	0.115	0.105	0.161	0.088	0.072
		Median	0.093	0.119	0.084	0.109	0.089	0.083	0.13	0.095	0.083	0.051

Table 4-2: Descriptive statistics summary of the daily $AOD_{MODIS\ (TERRA,\ AQUA)}$, for Site 2 (2020 and 2021).

Satellite product	Retrieval algorithm	Year	2020					2021				
		Season	Annual	Autumn	Spring	Summer	Winter	Annual	Autumn	Spring	Summer	Winter
MODIS Terra	DB	Mean	0.077	0.062	0.092	0.104	0.067	0.074	0.056	0.094	0.05	0.08
		±SD	0.059	0.051	0.073	0.078	0.031	0.08	0.041	0.111	0.03	0.083
		Median	0.057	0.045	0.064	0.089	0.058	0.046	0.038	0.048	0.041	0.057
	DT	Mean	0.136	0.127	0.164	0.17	0.11	0.124	0.126	0.13	0.124	0.12
		±SD	0.102	0.092	0.128	0.122	0.071	0.108	0.096	0.127	0.114	0.105
		Median	0.12	0.108	0.157	0.15	0.107	0.102	0.107	0.088	0.08	0.109
	DB & DT	Mean	0.123	0.127	0.129	0.17	0.092	0.117	0.126	0.113	0.124	0.111
		±SD	0.101	0.092	0.125	0.122	0.066	0.107	0.096	0.118	0.114	0.109
		Median	0.101	0.108	0.083	0.15	0.081	0.085	0.107	0.065	0.08	0.083
MODIS Aqua	DB	Mean	0.077	0.056	0.118	0.076	0.066	0.079	0.065	0.1	0.053	0.082
		±SD	0.071	0.043	0.112	0.073	0.029	0.093	0.084	0.12	0.041	0.086
		Median	0.056	0.043	0.09	0.048	0.058	0.046	0.035	0.05	0.036	0.05
	DT	Mean	0.123	0.123	0.153	0.154	0.077	0.127	0.111	0.154	0.144	0.117
		±SD	0.107	0.102	0.118	0.136	0.056	0.111	0.093	0.128	0.134	0.099
		Median	0.094	0.095	0.133	0.1	0.067	0.092	0.086	0.111	0.1	0.082
	DB & DT	Mean	0.114	0.123	0.135	0.154	0.069	0.119	0.111	0.14	0.144	0.102
		±SD	0.105	0.102	0.117	0.136	0.05	0.114	0.093	0.142	0.134	0.096
		Median	0.083	0.095	0.11	0.1	0.06	0.082	0.086	0.091	0.1	0.072

4.1 Validation of AOD_{550nm} Observations

4.1.1 Simple Linear Regression (SLR)

Table 4-3 shows a summary of the SLR regression results for $MODIS_{(TERRA, AQUA)}$ using the Deep Blue (DB), Dark Target (DT) and the Combined Deep Blue and Dark Target (DB & DT) retrieval algorithms against two AERONET sites; site 1 (Pretoria_CSIR_DPSS) and site 2 (Pretoria_CSIR_EC).

Table 4-3: Summary of the N, slope, intercept, R-value, R^2-value, RMSE, MAE, RMB for the simple linear regression done against the $AOD_{AERONET}$ values at Site 1 and Site 2.

Location	Variables	$MODIS_{TERRA}$			$MODIS_{AQUA}$		
		DB	DT	DB & DT	DB	DT	DB & DT
Site 1	N	274	244	261	245	238	253
	Slope	1.03	0.77	0.78	0.80	0.59	0.59
	Intercept	0.03	0.04	0.04	0.04	0.06	0.06
	R-value	0.93	0.94	0.94	0.92	0.90	0.90
	R^2	0.86	0.88	0.88	0.85	0.80	0.81
	RMSE	0.03	0.04	0.04	0.03	0.06	0.05
	MAE	0.02	0.03	0.03	0.02	0.04	0.04
	RMB	-0.18	-0.45	-0.55	-0.26	-0.87	-0.84
Site 2	N	344	298	330	316	298	330
	Slope	0.83	0.59	0.60	0.36	0.89	0.90
	Intercept	0.03	0.06	0.05	0.06	0.03	0.03
	R-value	0.88	0.83	0.84	0.73	0.95	0.95
	R^2	0.78	0.69	0.71	0.54	0.90	0.90
	RMSE	0.03	0.06	0.06	0.05	0.03	0.03
	MAE	0.02	0.04	0.04	0.03	0.02	0.02
	RMB	-0.18	-0.61	-0.61	-0.39	-0.29	-0.26

The following figures show the SLR to visually depict the correlation between the regressor ($AOD_{AERONET}$) and regressand (AOD_{MODIS}). The results will be discussed for $satellites_{(TERRA, AQUA)}$. The results will touch on the validation of the model performance as validations are an important aspect to consider as satellite products have potential impacts such as topography, surface albedo, surface brightness and retrieval algorithms. Therefore, it is imperative to constantly validate satellite products to reduce uncertainties (Eibedingil et al., 2021). Therefore, the discussion section will include statistical parameters of a SLR which are intercept, gradient and correlation coefficient (absolute R^2), RMSE, MAE and RMB to elaborate on the model performance. These parameters are presented on a 95% confidence interval.

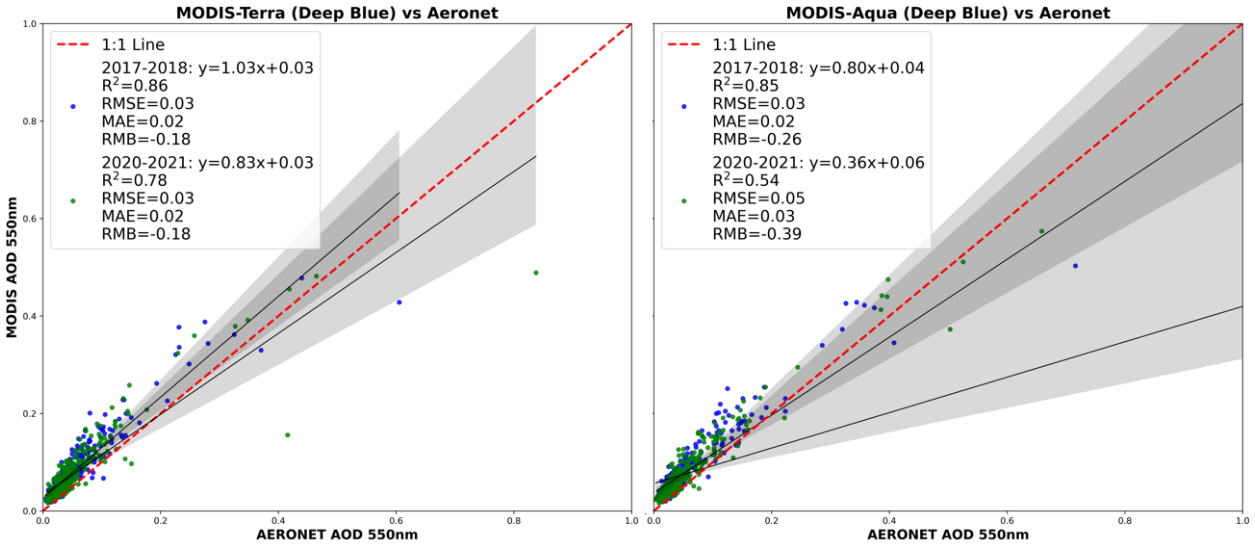


Figure 4-1: Regression plot depicting the relationship between "Blue Deep" AOD_{MODIS} and $AOD_{AERONET}$ for a) $MODIS_{TERRA}$ and b) $MODIS_{AQUA}$. Pretoria Site 1 (2017 to 2018) is shown in blue, while Pretoria Site 2 (2020 to 2021) is represented in green. Each plot includes the best-fit line and a 95% confidence interval.

Figure 4-1 shows a SLR relationship for $MODIS_{(TERRA, AQUA)}$ and $AOD_{AERONET}$ using the DB retrieval algorithm for both site 1 (2017 to 2018) and site 2 (2020 to 2021). For $MODIS_{TERRA}$ site 1, the results show a R^2 of 0.86, a RMSE of 0.03, a MAE of 0.02 and a RMB of -0.18. For site 2, the results show a R^2 of 0.78, a RMSE of 0.03, a MAE of 0.02 and RMB of -0.18. For $MODIS_{AQUA}$ site 1, results show a R^2 of 0.85, a RMSE of 0.03, a MAE of 0.02 and a RMB of -0.26. For site 2, results show an R^2 of 0.54, a RMSE of 0.05, a MAE of 0.03 and a RMB of -0.39.

Figure 4-2 shows a SLR relationship for $MODIS_{(TERRA, AQUA)}$ and $AOD_{AERONET}$ using the DT retrieval algorithm against two AERONET sites. For $MODIS_{TERRA}$ site 1, the results show a R^2 of 0.88, a RMSE of 0.04, a MAE of 0.03 and a RMB of -0.45. For site 2, the results show a R^2 of 0.69, a RMSE of 0.06, a MAE of 0.04 and RMB of - 0.61. For $MODIS_{AQUA}$ site 1, results show a R^2 of 0.80, a RMSE of 0.06, a MAE of 0.04 and a RMB of -0.87. For site 2, results show an R^2 of 0.90, a RMSE of 0.03, a MAE of 0.02 and a RMB of -0.29.

Figure 4-3 shows a SLR relationship for $MODIS_{(TERRA, AQUA)}$ and $AOD_{AERONET}$ using the Combined (DB & DT) retrieval algorithm against two AERONET sites. For $MODIS_{TERRA}$ site 1, the results show a R^2 of 0.88, a RMSE of 0.04, a MAE of 0.03 and a RMB of -0.55. For site 2, the results show a R^2 of 0.71, a RMSE of 0.06, a MAE of 0.04 and RMB of -0.61. For $MODIS_{AQUA}$ site 1, results show a R^2 of 0.81, a RMSE of 0.05, a MAE of 0.04 and a RMB of -0.84. For site 2, results show an R^2 of 0.90, a RMSE of 0.03, a MAE of 0.02 and a RMB of -0.26.

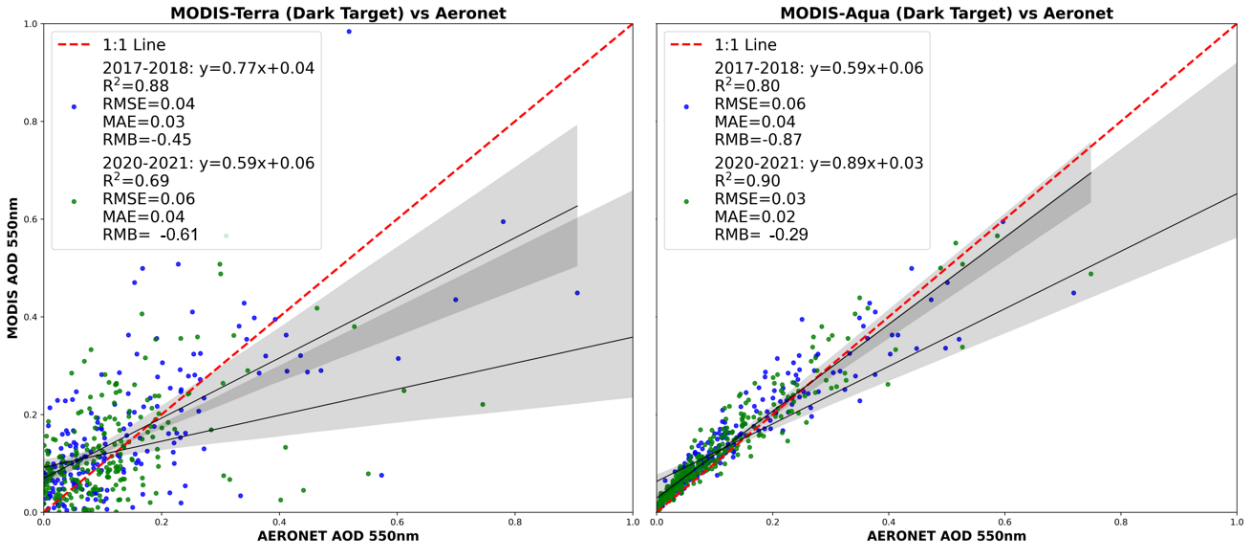


Figure 4-2: Regression plot depicting the relationship between "Dark Target" AOD_{MODIS} and $AOD_{AERONET}$ for a) $MODIS_{TERRA}$ and b) $MODIS_{AQUA}$. Pretoria Site 1 (2017 to 2018) is shown in blue, while Pretoria Site 2 (2020 to 2021) is represented in green. Each plot includes the best-fit line and a 95% confidence interval.

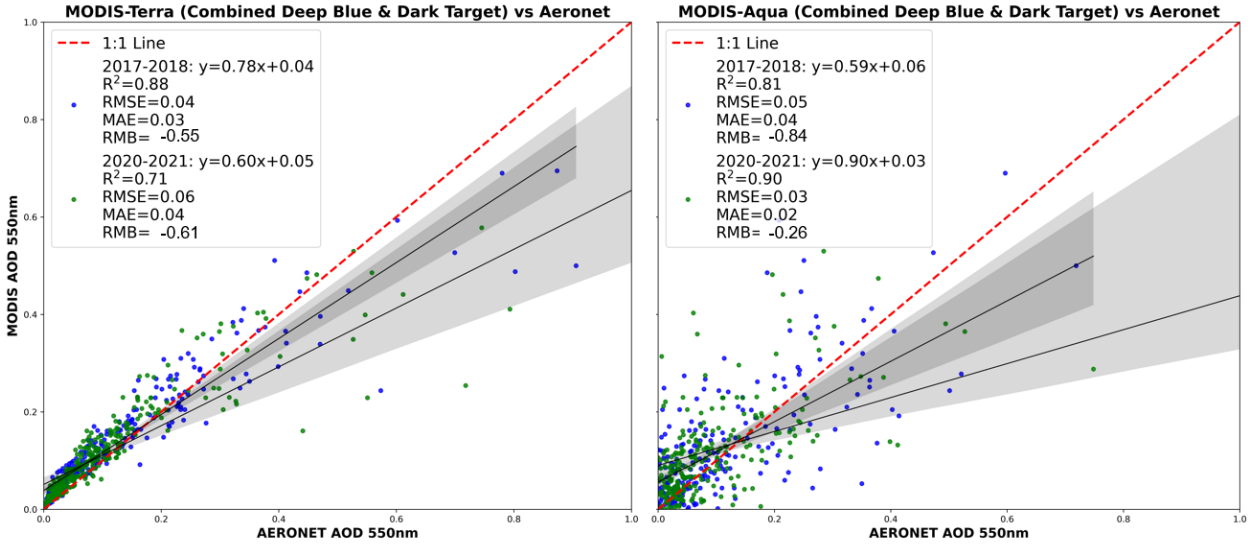


Figure 4-3: Regression plot depicting the relationship between "Combined Deep Blue and Dark Target" AOD_{MODIS} and $AOD_{AERONET}$ for a) $MODIS_{TERRA}$ and b) $MODIS_{AQUA}$. Pretoria Site 1 (2017 to 2018) is shown in blue, while Pretoria Site 2 (2020 to 2021) is represented in green. Each plot includes the best-fit line and a 95% confidence interval.

4.1.2 Ordinary Least Squared (OLS)

The Ordinary Least Squared regression was another regression analysis conducted to validate model performance. The discussion will include performance metrics: R-squared score (R^2), standard error (S_e) and p-value. For the R^2 value, a range of 0-1 will be used where a values that

are between ≥0 and <0.5 indicate poor model performance, values that are 0.5 indicate average performance and values ≥0.6 to 1 indicate good to exceptional model performance. With the standard error, we expect it to be smaller as it would indicate a better fit of the regression model to the data. The p-value will be used to establish a models statistical significance. If a p-value is <0.05, the model will be considered statistically significant meaning the null hypothesis is rejected, and if a p-value is >0.05 it indicates that the model is not statistically significant therefore the null hypothesis is acceptable.

Results from Table 4-4 show that for $MODIS_{TERRA}$ (DB) site 1, there is a good model performance, R^2 of 0.86. The regressor has an intercept of 0.027 and a positive bias (S_e) of 0.025. The p-value is <.001 which is less than the significant value of 0.05. For site 2, there is a good model performance, R^2 of 0.78. The regressor has an intercept of 0.032 and a positive bias (S_e) of 0.024. The p-value is <.001. For $MODIS_{AQUA}$ (DB) site 1, there is a good model performance, R^2 of 0.85. The regressor has an intercept of 0.038 and a positive bias (S_e) of 0.018. The p-value is <.001. For site 2, there is a weak average model performance, R^2 of 0.54. The regressor has an intercept of 0.056 and a positive bias (S_e) of 0.019. The p-value is <.001.

Results $MODIS_{TERRA}$ (DT) site 1 show that there is a good model performance, R^2 of 0.88. The regressor has an intercept of 0.042 and a positive bias (S_e) of 0.019. The p-value is <.001. For site 2, there is a moderate good model performance, R^2 of 0.69. The regressor has an intercept of 0.057 and a positive bias (S_e) of 0.023. The p-value is <.001. For $MODIS_{AQUA}$ (DT) site 1, there is a strong good model performance, R^2 of 0.80. The regressor has an intercept of 0.063 and a positive bias (S_e) of 0.019. The p-value is <.001. For site 2, there is a weak average model performance, R^2 of 0.54. The regressor has an intercept of 0.026 and a positive bias (S_e) of 0.043. The p-value is <.001.

Results $MODIS_{TERRA}$ (DB & DT) site 1 show that there is a strong good model performance, R^2 of 0.88. The regressor has an intercept of 0.038 and a positive bias (S_e) of 0.018. The p-value is <.001. For site 2, there is a strong good model performance, R^2 of 0.81. The regressor has an intercept of 0.058 and a positive bias (S_e) of 0.018. The p-value is <.001. For $MODIS_{AQUA}$ (DB & DT) site 1, there is a strong good model performance, R^2 of 0.81. The regressor has an intercept of 0.058 and a positive bias (S_e) of 0.018. The p-value is <.001. For site 2, there is a strong model performance, R^2 of 0.90. The regressor has an intercept of 0.026 and a positive bias (S_e) of 0.017. The p-value is <.001.

Table 4-4: Summary of the R^2, S_e, intercept and p-value, for the ordinary least square regression done against the $AOD_{AERONET}$ values at Site 1 and Site 2.

Location	Variables	$MODIS_{TERRA}$			$MODIS_{AQUA}$		
		DB	DT	DB & DT	DB	DT	DB & DT
Site 1	R^2	0.86	0.88	0.88	0.85	0.80	0.81
	Intercept	0.027	0.042	0.038	0.036	0.063	0.058
	SE	0.025	0.019	0.018	0.022	0.019	0.018
	p-value	< .001	< .001	< .001	< .001	< .001	< .001
Site 2	R^2	0.78	0.69	0.71	0.54	0.54	0.90
	Intercept	0.032	0.057	0.050	0.056	0.026	0.026
	SE	0.024	0.023	0.021	0.019	0.043	0.017
	p-value	< .001	< .001	< .001	< .001	< .001	< .001

4.1.3 Scikit Linear regression modelling

Table 4-5 shows a summary of the N, Train data, Test data, intercept, coefficient, R^2 (train), R^2 (test), MAE and RMSE for the scikit-learn regression done against the AOD_{MODIS} values at Site 1 and Site 2 using the Deep Blue (DB), Dark Target (DT) and the Combined Deep Blue and Dark Target (DB & DT) retrieval algorithms against two AERONET sites. The table below shows the calculated R^2 for both the training and testing data. This is imperative and beneficial to always compare training data performance against testing data performance to establish if there was overfitting or underfitting.

The model performance metrics will be discussed in the following manner: $MODIS_{(TERRA, AQUA)}$ "Deep Blue (DB)" for Site 1 and Site 2; $MODIS_{(TERRA, AQUA)}$ "Dark Target (DT)" for Site 1 and Site 2 and lastly, $MODIS_{(TERRA, AQUA)}$ "Combined Deep Blue and Dark Target (DB & DT)" for Site 1 and Site 2.

Table 4-5: Summary of the N, Train data, Test data, intercept, coefficient, R^2 (train), R^2 (test), MAE and RMSE for the scikit-learn regression done against the AOD_{MODIS} values at Site 1 and Site 2.

Location	Variables	$MODIS_{TERRA}$			$MODIS_{AQUA}$		
		DB	DT	DB & DT	DB	DT	DB & DT
Site 1	Train data	123	129	150	112	98	116
	Test data	54	56	64	49	42	51
	Intercept	0.01	0.04	0.03	0.01	0.02	0.02
	Coefficient	1.44	0.91	0.92	1.35	1.04	1.03
	R^2 (TRAIN)	0.88	0.93	0.93	0.90	0.88	0.91
	R^2 (TEST)	0.66	0.71	0.71	0.84	0.89	0.87
	MAE	0.02	0.03	0.03	0.01	0.02	0.02

	RMSE	0.02	0.05	0.05	0.02	0.03	0.03
Site 2	**Train data**	162	158	196	181	120	116
	Test data	70	69	70	78	52	51
	Intercept	0.02	0.04	0.04	0.02	0.02	0.02
	Coefficient	1.13	0.74	0.77	1.20	0.95	0.99
	R^2 (TRAIN)	0.74	0.74	0.74	0.86	0.90	0.92
	R^2 (TEST)	0.75	0.81	0.87	0.23	0.95	0.92
	MAE	0.01	0.02	0.02	0.02	0.02	0.02
	RMSE	0.02	0.03	0.02	0.04	0.03	0.03

Figure 4-4 shows the scatter plots for the Deep Blue retrieval algorithm for $MODIS_{(TERRA, AQUA)}$ and performance metrices to outline how the retrieval algorithm performed. For the training dataset, $MODIS_{TERRA}$ DB (Site 1 and Site 2) and $MODIS_{AQUA}$ DB (Site 1 and Site 2) have a model performance of (0.88 & 0.74) and (0.90 & 0.86), respectively. For the testing dataset. $MODIS_{TERRA}$ DB (Site 1 and Site 2) and $MODIS_{AQUA}$ DB (Site 1 and Site 2) have a performance of (0.66 & 0.75) and (0.87 & 0.23). Comparing the training and testing datasets, there is a clear distinction on the model performance. All models experience overfitting, with $MODIS_{TERRA}$ DB Site 1(0.88 & 0.66) and $MODIS_{AQUA}$ DB Site 2(0.86 & 0.23) showing the greatest model variability due overfitting on the training data.

The MAE and RMSE for the $MODIS_{TERRA}$ DB (Site 1 and Site 2) and $MODIS_{AQUA}$ DB (Site 1 and Site 2) is (0.02 & 0.02); (0.01 & 0.02) and (0.02 & 0.03); (0.02 & 0.03), respectively. The MSE and RMSE show to be closer to 0 which indicates that the model performed well.

Figure 4-5 shows the scatter plots for the Dark Target retrieval algorithm for $MODIS_{(TERRA, AQUA)}$ and performance metrices to outline how the retrieval algorithm performed. For the training dataset, $MODIS_{TERRA}$ DT (Site 1 and Site 2) and $MODIS_{AQUA}$ DT (Site 1 and Site 2) have a model performance of (0.93 & 0.74) and (0.88 & 0.86), respectively. For the testing dataset, $MODIS_{TERRA}$ DT (Site 1 and Site 2) and $MODIS_{AQUA}$ DT (Site 1 and Site 2) have a performance of (0.71 & 0.81) and (0.89 & 0.95). Only the $MODIS_{TERRA}$ DT (Site 1) presents with overfitting of the training dataset (0.93 & 0.71). The rest of the algorithms present with an underfitting of the training dataset with $MODIS_{TERRA}$ DT Site 2, $MODIS_{AQUA}$ DT Site 1 and $MODIS_{AQUA}$ DT Site 2 having an underfitting of (0.74 & 0.81), (0.88 & 0.89) and (0.90 & 0.95), respectively.

The MAE and RMSE for the $MODIS_{TERRA}$ DT (Site 1 and Site 2) and $MODIS_{AQUA}$ DT (Site 1 and Site 2) is (0.03 & 0.05); (0.02 & 0.03) and (0.02 & 0.03); (0.02 & 0.03), respectively. The model predictors indicate that the lower the error is, the better the model.

Figure 4-6 shows the scatter plots for the Combined DB and DT retrieval algorithm for $MODIS_{(TERRA, AQUA)}$. For the training dataset, $MODIS_{TERRA}$ DB & DT (Site 1 and Site 2) and $MODIS_{AQUA}$ DB & DT (Site 1 and Site 2) have a model performance of (0.93 & 0.74) and

(0.91 & 0.92), respectively. For the testing dataset, $MODIS_{TERRA}$ DB & DT (Site 1 and Site 2) and $MODIS_{AQUA}$ DB & DT (Site 1 and Site 2) have a performance of (0.71 & 0.87) and (0.87 & 0.92). Overfitting of the training dataset is observed for $MODIS_{TERRA}$ DB & DT (Site 1) and $MODIS_{AQUA}$ DB & DT (Site 1). Underfitting of the training dataset is observed for $MODIS_{TERRA}$ DB & DT (Site 2). $MODIS_{AQUA}$ DB & DT (Site 2) is the only satellite that uses a retrieval algorithm that did not result in over-underfitting because the R^2 for both the training and testing data is the same (0.92 & 0.92).

The MAE and RMSE for the $MODIS_{TERRA}$ DB & DT (Site 1 and Site 2) and $MODIS_{AQUA}$ DB & DT (Site 1 and Site 2) is (0.13 & 0.05); (0.02 & 0.02) and (0.02 & 0.03); (0.02 & 0.03), respectively. $MODIS_{TERRA}$ DB & DB (Site 1) has a slightly higher MAE than the other sites. This may be due to external influences that impact AOD retrieval (e.g. retrieval algorithms and surface albedo).

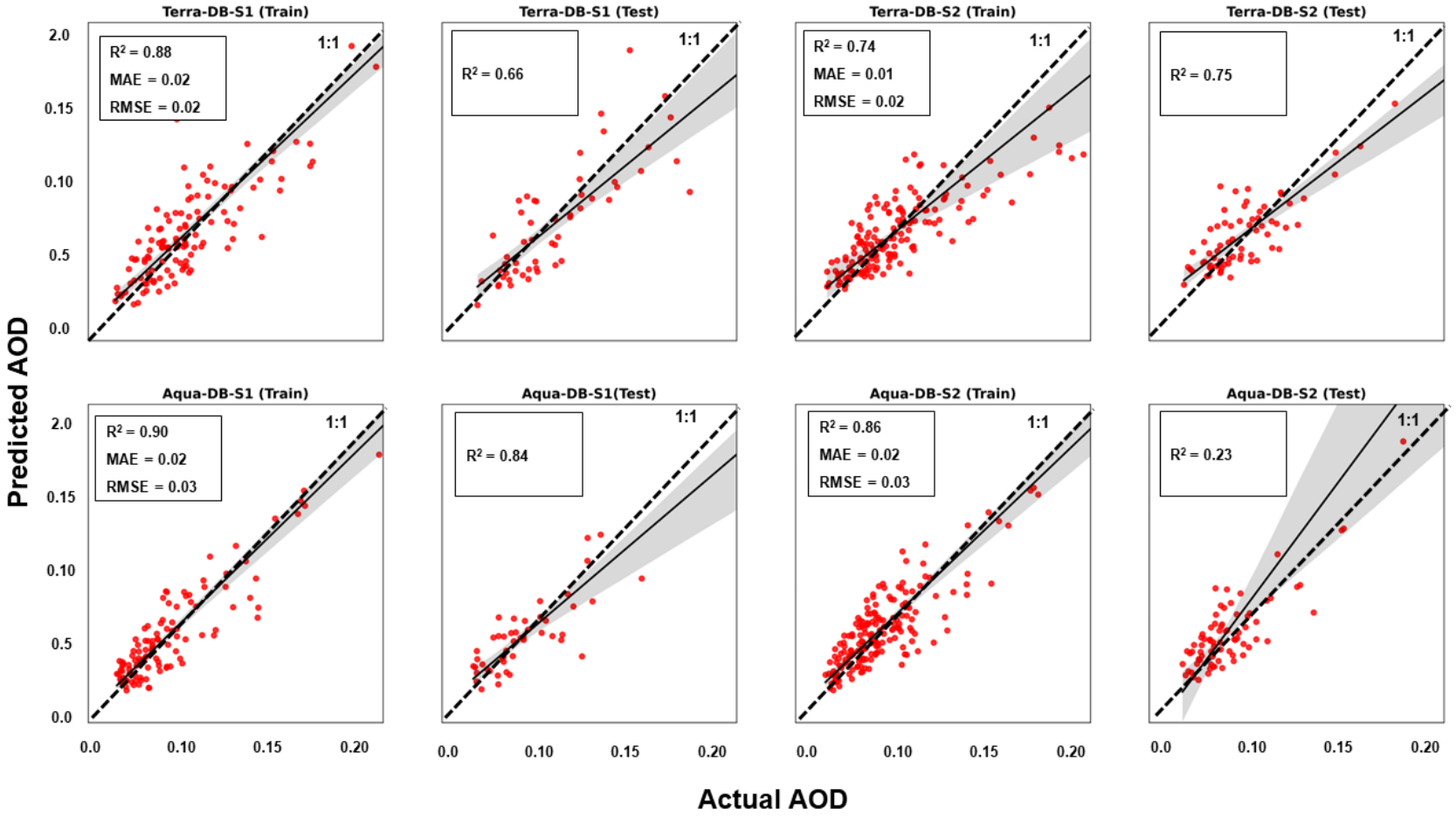


Figure 4-4: Scatter subplots depicting Scikit regression model performance for MODIS$_{(TERRA, AQUA)}$ "Deep Blue" for both Pretoria Site 1 (2017 to 2018) and Pretoria Site 2 (2020 to 2021). The graphs indicate the training and testing results, with the testing graphs only showing the R^2. Each plot includes the best-fit line and a 95% confidence interval.

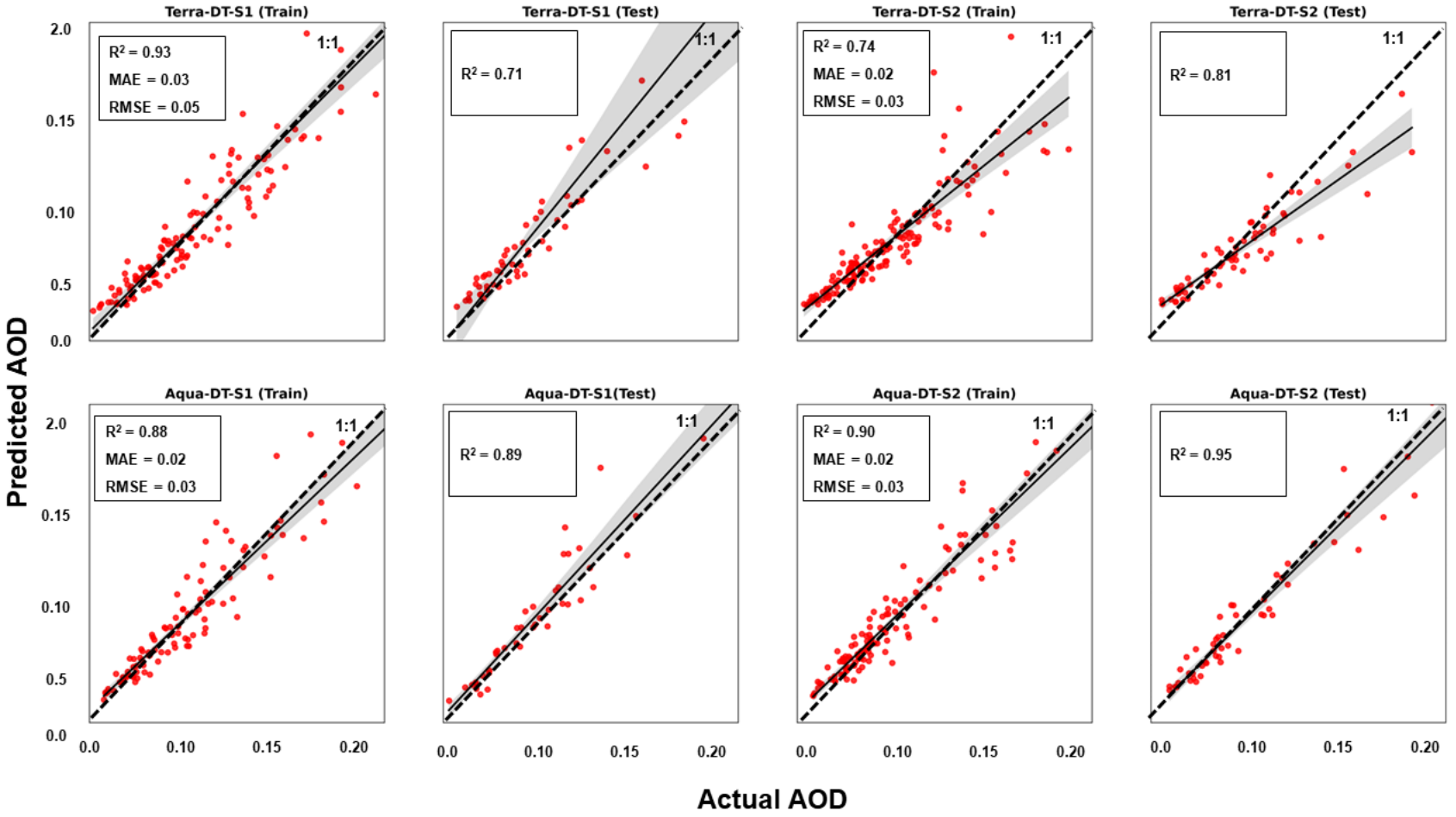


Figure 4-5: Scatter subplots depicting Scikit regression model performance for MODIS$_{(TERRA,\ AQUA)}$ "Dark Target" for both Pretoria Site 1 (2017 to 2018) and Pretoria Site 2 (2020 to 2021). The graphs indicate the training and testing results, with the testing graphs only showing the R^2. Each plot includes the best-fit line and a 95% confidence interval.

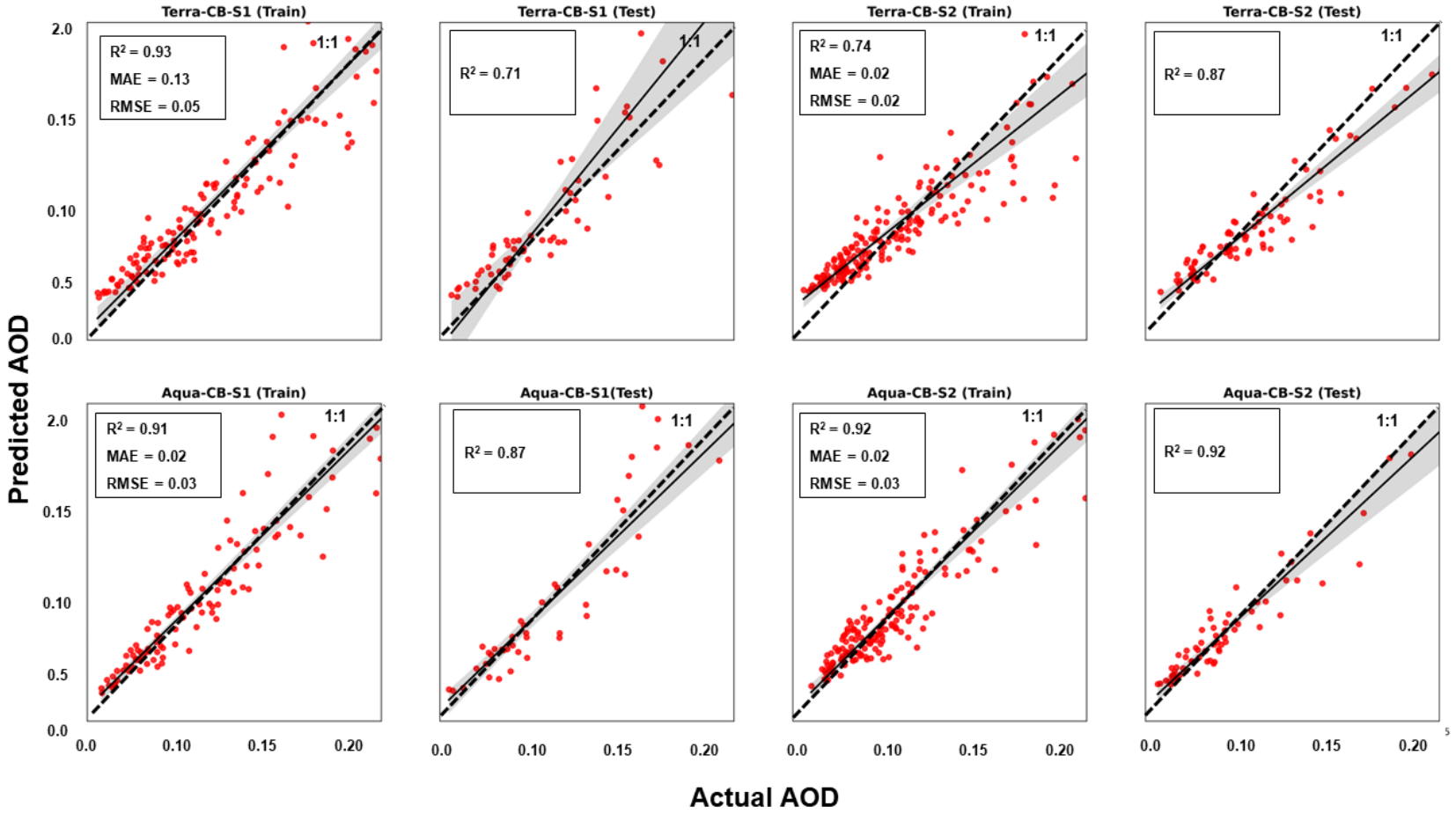


Figure 4-6: Scatter subplots depicting Scikit regression model performance for MODIS$_{(TERRA,\ AQUA)}$ "Combined Deep Blue & Dark Target" for both Pretoria Site 1 (2017 to 2018) and Pretoria Site 2 (2020 to 2021). The graphs indicate the training and testing results, with the testing graphs only showing the R^2. Each plot includes the best-fit line and a 95% confidence interval.

4.2 Satellite observations and atmospheric studies

Air quality studies using satellite data have been increasing in the space of scientific research. One of the earliest global satellite studies was by Holben et al. (1998), Wang & Christopher (2003), Engel-Cox et al. (2004) and Van Donkelaar et al. (2006). Holben et al. (1998) study leaned more towards aerosol characterisation using AERONET data. In these abovementioned studies, retrieval errors are highlighted to be attributed by bright snow surfaces and cloud cover. The studies concluded that careful application of retrieval algorithms are needed to combat the limitations in satellite studies and that AOD_{MODIS} can be quantitatively be applied in air quality studies which are undertaken in this study.

4.2.1 Assessing the accuracy of satellite AOD retrievals

Assessing the accuracy of AOD retrievals is important in air quality studies. Factors that greatly impact the accuracy of AOD retrievals are topography, surface albedo and retrieval algorithms.

Topography does have impacts on AOD accuracy as low a AOD loading is generally distributed in steep hilly (mountainous) areas and high AOD is dispersed among areas of low-lying grasslands (Yuan et al., 2023). Surface albedo highly impacts the accuracy of AOD. Due of bright surfaces AOD retrievals under moderate and low dust loading conditions are impede and the small upwelling radiance caused by aerosol loading relative to surface reflectance makes it challenging to distinguish between the aerosol signal and the surface contribution. That is why satellite AOD retrievals over desert regions are usually overestimated (Huang et al., 2021). However, for the purposes of this study, these abovementioned factors were not explored.

Different algorithms can impact how accurately AOD is retrieved. This study made use of all three AOD retrieval algorithms (Deep Blue, Dark Target and Combined Deep Blue and Dark Target) to assess the differences in validation performance.

The best performing retrieval algorithm throughout all regression techniques is the DT and Combined DB & DT (0.90 & 0.90), respectively for the SLR. The $MODIS_{AQUA}$ Combined DB & DT (0.90) also performed well with the OLS regression model against AERONET Site 2. This may be attributed to the location of the Site and the surrounding meteorology. Site 1 AOD retrievals could have been impacted by meteorology (Faniso et al., 2021). For the Scikit-Learn ML regression model, the Combined DB & DT model prediction for $MODIS_{AQUA}$ (Site 1 and Site 2) performed extremely well without presenting any over or-underfitting for the training dataset.

4.2.2 Model predictions

Model predictions have been a great way to predict a models efficiency and applicability, as well as to determine if AOD_{MODIS} can be scrupulously used to explain atmospheric and climatological

trends. A plethora of studies have used linear, non-linear and hybrid models, which all if not most have to undergo a process of training the model to see its performance. Using AOD measurements for the AERONET Site 1 (Pretoria_CSIR_DPSS) and Site 2 (Pretoria_CSIR_EC), for both $MODIS_{(TERRA, AQUA)}$ model performance was better using the Combined DB & DB retrieval algorithm as it is the only algorithm that did not present with overfitting and underfitting issues. The results from a study by Zhang et al. (2021) corroborated that the use of the Combined DB & DT algorithm improves AOD coverage. A study by Ranjun et al. (2021) highlighted the careful consideration of meteorological parameters because they can enhance the precision of AOD validation. This is a significant consideration that, if undertaken in this study would have yielded great model predictions.

4.2.3 Importance of satellite observations to atmospheric air quality studies

Satellite-based air quality studies have slowly been gaining popularity over the years which is beneficial for the growth of local/ regional studies. The importance of these studies is to provide better weather and climatological understandings and predictions, to make better and informed environmental decision, in the context of air quality management and environmental policy formulation. Local studies like that of Hersey et al. (2015), Zhang et al. (2021) use satellite-based observations in understanding atmospheric studies.

Hersey et al. (2015) highlighted that with a plethora of resources and information remotely sensed air quality studies, such studies contribute to AOD variability. The study highlighted the importance of satellite observations as they provide extensive information on the transportation of aerosol. Without satellite studies, it would be challenging to even understand transportation of aerosols aloft. Satellite observations have also been widely consulted in Machine Learning (ML) to modelling ground-based $PM_{2.5}$ concentrations. This study only explored the accuracy of these satellite-derived AOD data, however, a study by Zhang et al. (2021) computed satellite observations into a developed ML model (Random Forest Model) to estimate $PM_{2.5}$ in Gauteng on a daily scale. Such a study is import in atmospheric studies as it contributes to air quality management and environmental policy formulation.

Rigorous statistical analysis has validated the effectiveness of satellite observations against ground-based data. The performance metrics obtained from these analyses, such as R-squared values and standard errors, demonstrate a significant level of accuracy in satellite data, albeit with room for improvement (Shikwambana et al., 2019).

A key observation from this study is the variability in data availability across different satellite platforms and the impact of seasonal variations on data collection. This variability is an important consideration for researchers and policymakers who rely on satellite data for environmental monitoring and decision-making processes. Data insufficiency in AERONET measurements has

proven to be an impediment in significant trend analysis for validating AOD_{MODIS}. A study by Makokha et al. (2017) supports this notion as the study made use of $AOD_{AERONET}$ stations that had a minimum of three years' worth of data and the AERONET stations used in this study had much less data than that (2017 to 2018 and 2020 to 2021). Furthermore, the analysis has highlighted the influence of environmental factors such as topography, surface albedo, and retrieval algorithms on the accuracy of satellite AOD measurements. These factors are crucial in interpreting satellite data and must be considered in any comprehensive atmospheric study.

In conclusion, satellite observations provide a powerful tool for monitoring and understanding aerosol distributions on a large scale. However, they should be used with ground-based measurements for a more accurate and comprehensive understanding of atmospheric conditions. The findings from this chapter contribute to the ongoing effort to enhance the reliability and application of satellite data in atmospheric science, particularly in the context of air quality management and environmental policy formulation.

****

This chapter has presented a comprehensive evaluation of satellite-derived Aerosol Optical Depth (AOD) data in the Gauteng region, highlighting the importance of satellite observations in understanding atmospheric conditions and air quality. The findings indicate that while satellite data from $MODIS_{TERRA}$ and $MODIS_{AQUA}$ provides valuable insights into regional AOD patterns, there are notable discrepancies compared to AERONET observations. These discrepancies underscore the need for careful calibration and validation of satellite data for local-scale environmental assessments.

5 EVALUATING AEROSOLS OVER GAUTENG USING SATELLITE IMAGERY AND RE-ANALYSIS MODEL DATA

In this chapter, we employ remote sensing data and advanced modelling techniques to evaluate regions within Gauteng that lack active ground-based monitoring. Daily satellite observations of Aerosol Optical Depth (AOD) and Ångström Exponent (α) from $MODIS_{(TERRA,\ AQUA)}$ and the MERRA-2 re-analysis assimilation model are analysed, offering insights into the spatial and temporal variations of aerosol loading. Descriptive statistics, spatiotemporal patterns, and comprehensive discussions are outlined, with all maps visualized on ArcGIS using ArcMap v10.8.1. This approach enhances our understanding of air quality dynamics, especially in areas without direct monitoring, by showing the geographical distribution and identifying potential pollution hotspots. The generated outcomes contribute to an enhanced understanding of air quality across the Gauteng region.

****

Table 5-1 outlines the descriptive statistics of the AOD_{550nm}, based on the monthly data for each year for $MODIS_{(TERRA,\ AQUA)}$ and MERRA-2. The AOD_{550nm} values for $MODIS_{TERRA}$ are as follows: 2017 (0.131±0.009), 2018 (0.103±0.006), 2019 (0.094±0.006), 2020 (0.114±0.008) and 2021 (0.103±0.005). The values for $MODIS_{AQUA}$ are as follows: 2017 (0.118±0.008), 2018 (0.105±0.005), 2019 (0.090±0.006), 2020 (0.107±0.008) and 2021 (0.105±0.007). The values for MERRA-2 are as follows: 2017 (0.146±0.006), 2018 (0.152±0.004), 2019 (0.130±0.008), 2020 (0.145±0.005) and 2021 (0.144±0.005).

Table 5-1: Annual descriptive statistics of $MODIS_{(TERRA,\ AQUA)}$ and MERRA-2 showing the raster parameters (min, mean, standard deviation and max) for 2017 to 2021.

Raster parameter	2017	2018	2019	2020	2021
$MODIS_{TERRA}$					
Min	0.102	0.082	0.073	0.089	0.083
Mean	0.131	0.103	0.094	0.114	0.103
Std	0.009	0.006	0.006	0.008	0.005
Max	0.148	0.113	0.111	0.132	0.112
$MODIS_{AQUA}$					
Min	0.100	0.084	0.067	0.084	0.081
Mean	0.118	0.105	0.090	0.107	0.105
Std	0.008	0.005	0.006	0.008	0.007
Max	0.140	0.114	0.103	0.121	0.117

MERRA-2					
Min	0.127	0.136	0.109	0.127	0.128
Mean	0.146	0.152	0.130	0.145	0.144
Std	0.006	0.004	0.008	0.005	0.005
Max	0.154	0.158	0.143	0.151	0.150

For $MODIS_{(TERRA,\ AQUA),}$ both satellites used a combined correction retrieval algorithm of DB and DT from collection 6.1. Using a merged correction algorithm was the key resolution in accounting for densely vegetated and sun-glint-affected areas, as South Africa's landscape is quite diverse, intricate and inhomogeneous. Not only was this method undertaken in this study chapter but also in a study by Rizza et al. (2019) and Campos et al. (2022), both of which contended evidence to yield improved results with a reduced estimation bias. Pandey and Vinoj (2020) also contributed that using combined correction algorithms is advantageous as they can cover observations over a large area.

5.1 Annual variations

In Figure 5-1, a high aerosol loading is observed in the central moving to the upper parts of Gauteng, majorly covering the Metropolitan municipalities CoJ, CoE and CoE. District municipalities with high aerosol loading are Mogale City and Lesedi. The HPA also shows moderate to high aerosol loading, with CoT having high loadings. District municipalities with moderate aerosol loading are Midvaal and Rand West. Low aerosol loading is observed in Merafong City and Emfuleni district municipalities.

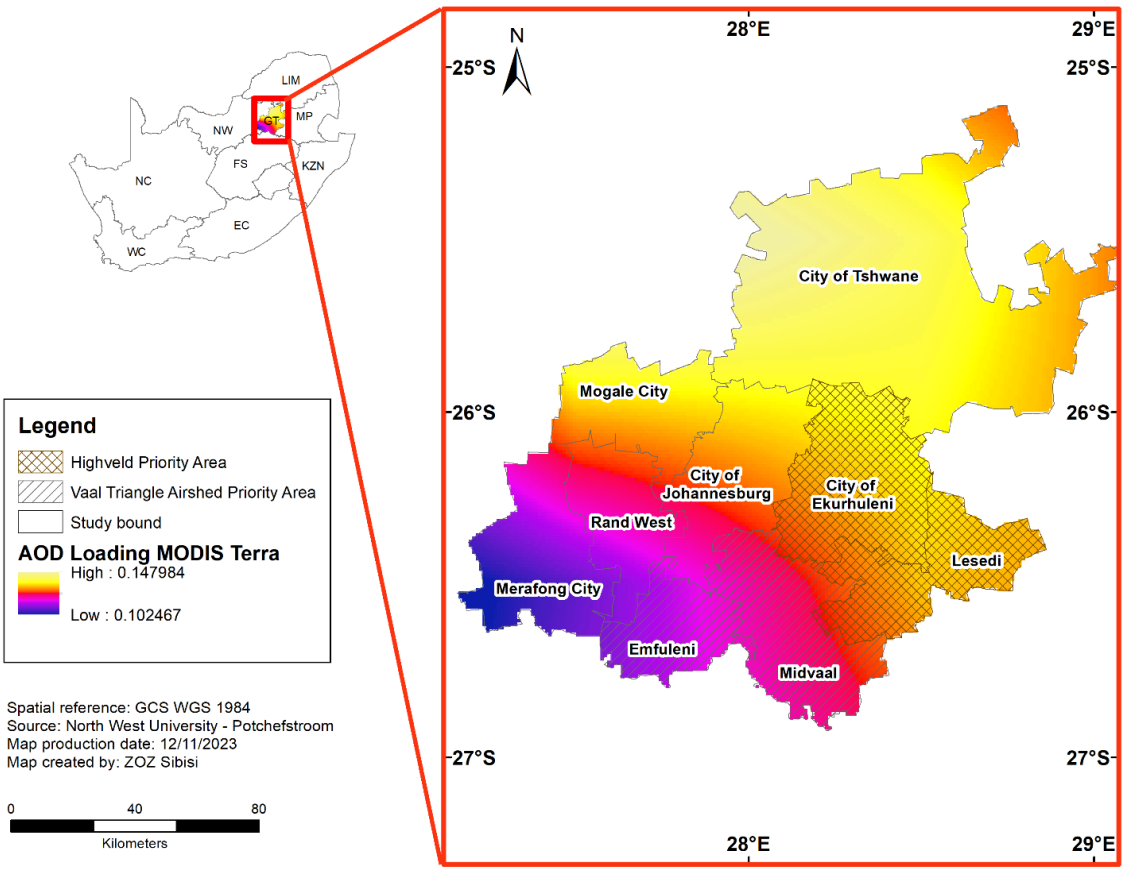


Figure 5-1: Map showing the multi-year averaged MODIS$_{(TERRA)}$ AOD at 550nm over Gauteng, South Africa from 2017 to 2022.

Figure 5-2 shows that MODIS$_{AQUA}$ captured a high aerosol loading over the north-eastern part of the province, mainly Mogale city and CoT. Making reference to Figure 2-2, all monitoring stations in the CoT (Bodibeng, Hammanskraal, Rosslyn and Tshwane) are located in the area where a high aerosol loading is observed. A low aerosol loading is observed in the south-western and eastern parts of the province. Merafong city, Emfuleni and Midvaal have the lowest aerosol loading.

Figure 5-1 and Figure 5-2 depict differences in spatial variability. MODIS$_{TERRA}$ showed to capture a higher AOD loading which covered the central to northern parts of the province as compared to MODSI$_{AQUA}$. MODIS$_{AQUA}$ showed to lower AOD loading than its counterpart satellite. These discrepancies can be alluded to their different overpass times (Arowosegbe et al., 2022). With the discrepancy of satellite of overpass times, differences in AOD loading can be attributed to this. A study by (Melchiorre et al., 2019) further elaborates that meteorological parameters like wind can also displace atmospheric species during overpass times. This explanation can be inferred to the results presented in Figure 5-1 and Figure 5-2.

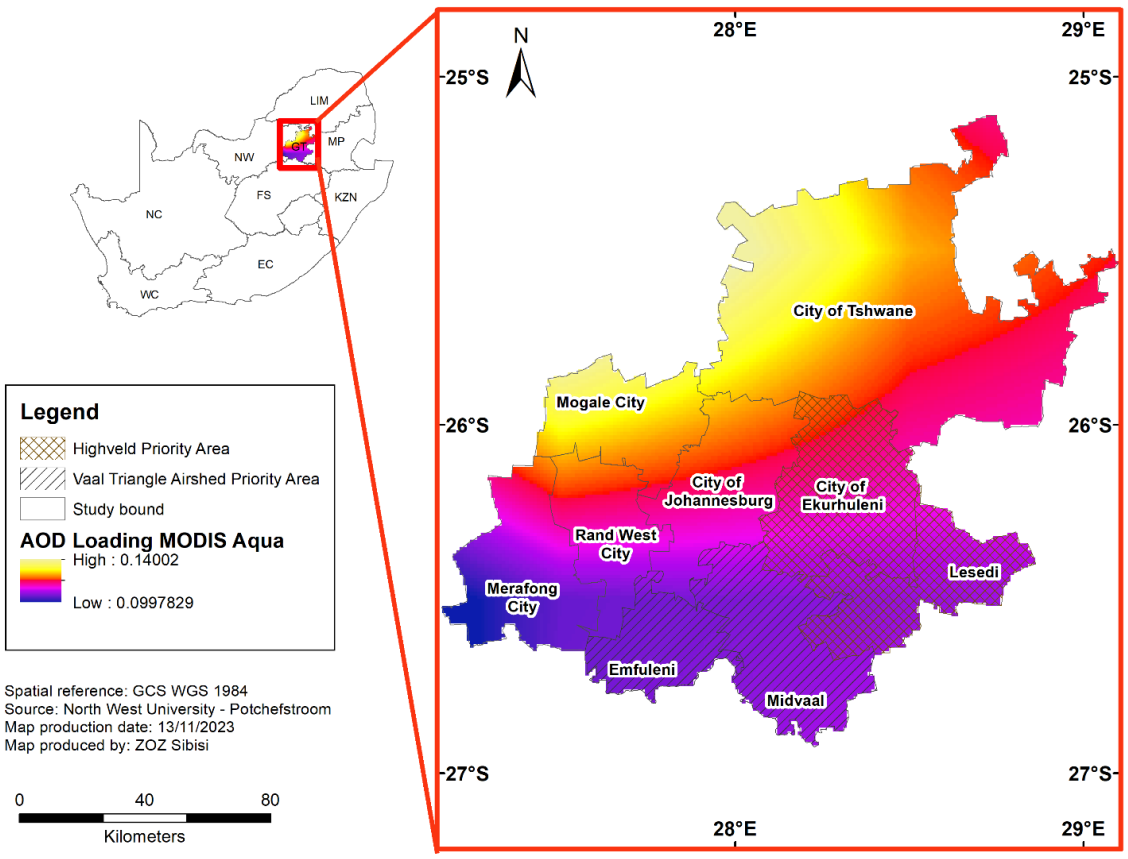


Figure 5-2: Map showing the multi-year averaged MODIS$_{(AQUA)}$ AOD at 550nm over Gauteng, South Africa from 2017 to 2022.

MERRA-2 was chosen to fully utilise the benefits of modelling and observations (Khoir et al., 2022). Regarding efficiency in capturing the region's aerosol loading, MERRA-2 outperformed both MODIS$_{(TERRA,\ AQUA)}$ as anticipated. This finding is further substantiated by Campos et al. (2022), who elucidate that MERRA-2 is extremely robust in areas that are high in elevation which are emblematic of Gauteng province in the Midvaal, as described in ***section 2.4***.

MERRA-2 data shows aerosol loading along the eastern (Figure 5-3), central and northern parts of the province. High aerosol loading is also observed over the HPA and VTAPA. Mogale City, Rand West City and Emfuleni have a medium to low aerosol loading, and Merafong City has a low aerosol loading. Areas showing high aerosol loading are situated near high pollution-contributing industries, covering some priority areas.

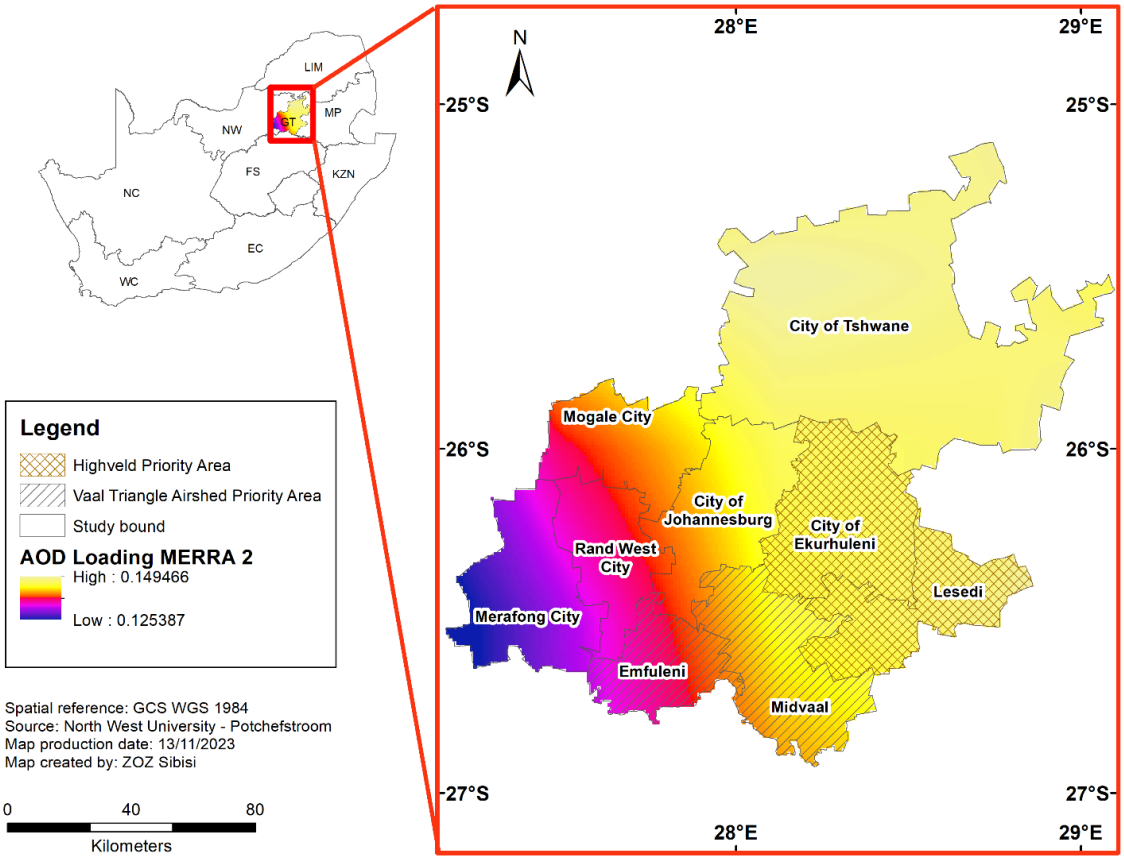


Figure 5-3: Map showing the multi-year averaged MERRA-2 reanalysis AOD at 550nm over Gauteng, South Africa from 2017 to 2022.

Figure 5-4 shows the aerosol loading for each year from 2017 to 2021, which will be discussed yearly. 2017 has a high aerosol loading at the City of Tshwane (CoT), where the monitoring station Bodibeng is located. 2018 has a high aerosol loading over CoT and the Highveld Priority Area (HPA). 2019 has a high aerosol loading in the west, moving to the northwest part of the province where the northern part of Mogale city and the western part of the CoT are impacted. The aerosol loading also decreases from the western to the central parts. In the CoT, Bodibeng and Rosslyn are the stations with high aerosol loading. 2020 has a high loading in the north-eastern and southwestern parts of the province. The former is the CoT (Bodibeng and Rosslyn), and the latter is at the HPA (Lesedi Local Municipality). 2021 has a high loading in the southeastern part of the province, covering areas in the VTAPA and HPA, the highly exposed local municipalities Emfuleni and the Midvaal. Throughout the years, low aerosol loading has been prominent in the southwestern (Rand West City) part, increasing in a northeasterly direction and southerly for 2019.

Figure 5-5 shows the aerosol loading for each year from 2017 to 2021 which will be discussed per year. In 2017, the aerosol loading increases from the southern region to the north-eastern part where you will find Hammanskraal, Bodibeng and Rosslyn stations. In 2018, there is a

decrease in the aerosol loading from centre towards the north and south-east. During this year, the HPA had a high aerosol loading. There is medium aerosol loading concentrations in 2019 from the centre extending to the north and southern parts of the province although loading continues to decrease towards the south-west. In 2020, aerosol loading is low from the south-west direction but increases northwards. A similar trend to 2020 is observed in 2021, where the HPA has a high loading. Interestingly, 2017 had a low aerosol loading over the VTAPA and parts of the HPA, which a discussion explaining such a phenomenon will be provided in ***section 5.3***.

Figure 5-6 showed a better depiction of aerosol loading than MODIS. 2017 has a low aerosol loading in the southwest, which increases in a north-westerly direction. 2018 has a high aerosol loading in the western part of the province, covering parts of the HPA and the City of Ekurhuleni (CoT). The areas that will be most affected are Etwatwa and Springs. In 2019, there was a high loading in the CoT and a medium-to-low loading from the central part, moving down to the southwestern parts. 2020 captured more high aerosols compared to the other years, with a high loading in the northern and western parts. 2021 has a high aerosol loading in the northern and south-western parts of the province.

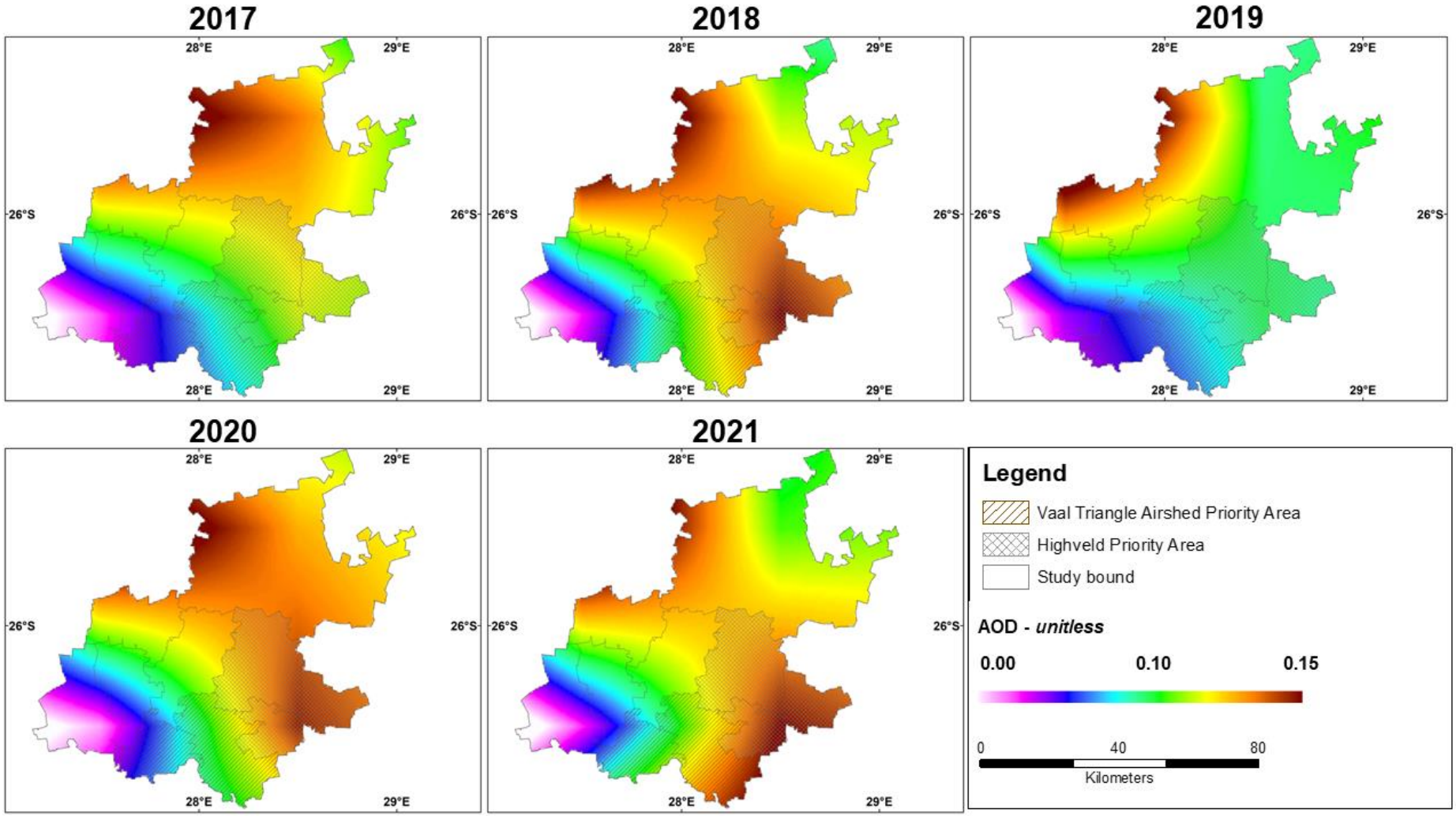


Figure 5-4: Spatial distribution of annual averaged MODIS$_{(TERRA)}$ AODs at 550nm in Gauteng, South Africa from 2017 to 2021.

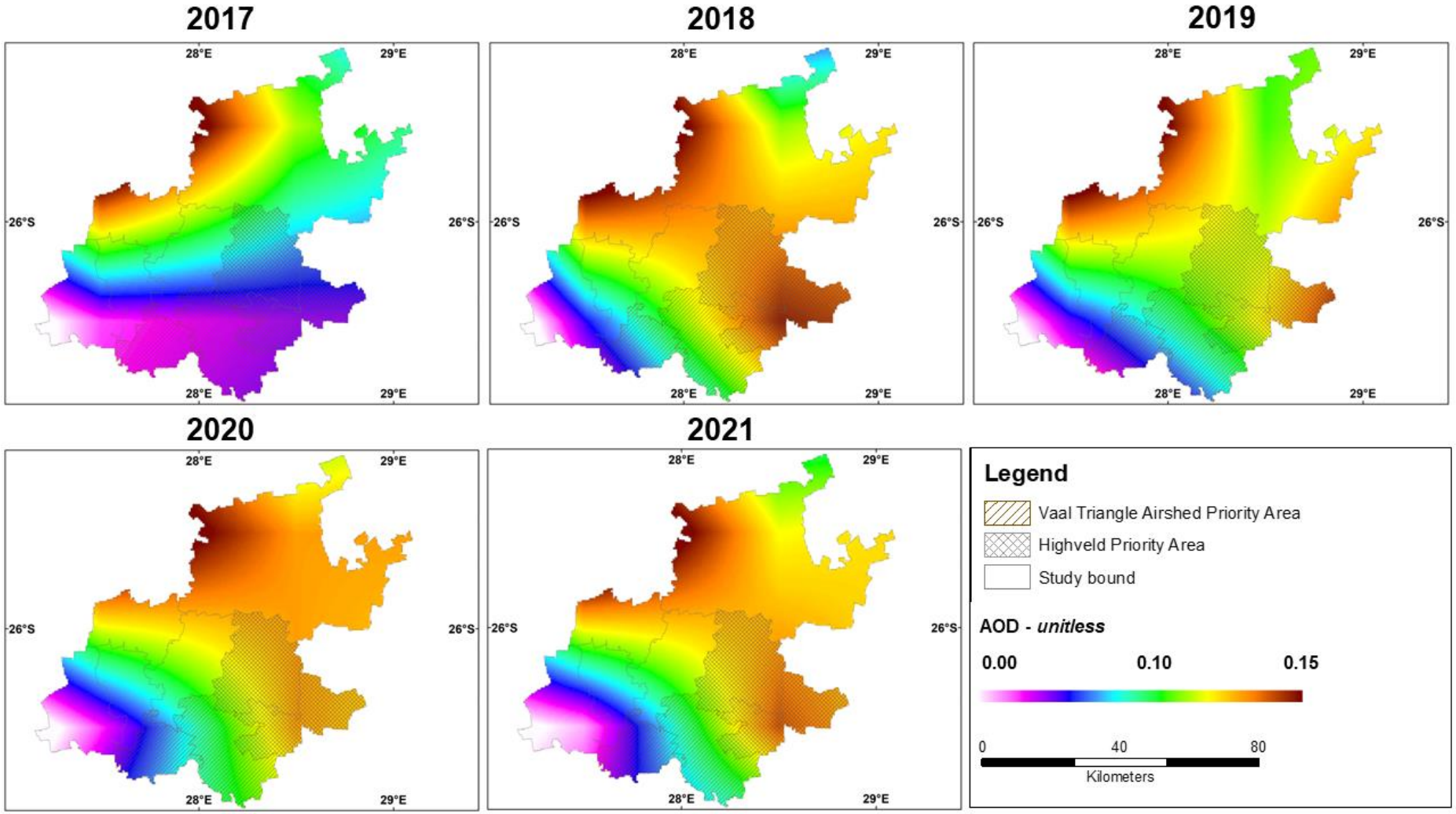


Figure 5-5: Spatial distribution of annual averaged MODIS$_{(AQUA)}$ AODs at 550nm in Gauteng, South Africa from 2017 to 2021.

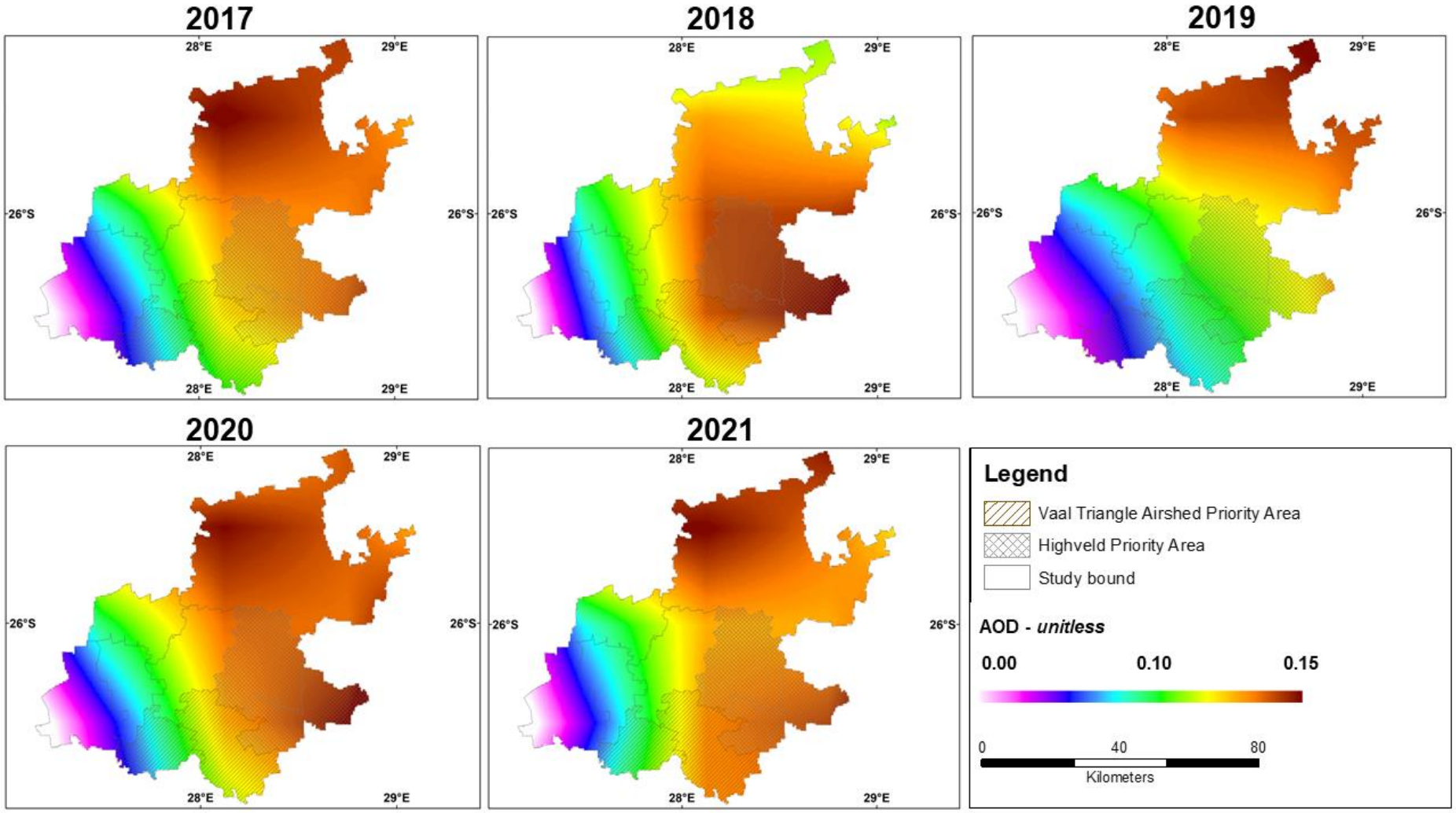


Figure 5-6: Spatial distribution of annual averaged MERRA-2 AODs at 550nm in Gauteng, South Africa.

5.2 Seasonal Variation

For $MODIS_{TERRA}$, the strongest seasonal variation is observed in the northern parts of the region during the autumn, and the weakest seasonal variation is observed during the spring. For $MODIS_{AQUA}$, a strong seasonal variation is observed during the summer and autumn and a weak seasonal variation in winter. For MERRA-2, a strong seasonal variation is observed in spring, and a weak variation is observed in winter. An in-depth analysis is provided in the subsequent paragraphs.

Seasonal variation is observed throughout the five-year monitoring period. As seen in Figure 5-7, each season had low and high aerosol loadings, with summer at (0.102 & 0.149), autumn at (0.100 & 0.150), winter at (0.078 & 0.107) and spring at (0.062 & 0.097). In summer, $MODIS_{TERRA}$ observed a high aerosol loading in the southeastern part covering the HPA. The extreme northern and southwestern areas have a low aerosol loading. Autumn seems to show a high aerosol loading in the CoT. Mogale City and CoT have high aerosol loadings, with low loadings in Merafong City. Midvaal shows notable/ relatively high loadings, which may indicate an event that might have occurred. Merafong City has consistently had low aerosol loadings throughout all the seasons, including spring. This may be attributed to low-contributing activities or strong winds that dissipate the aerosol prior to any satellite observation.

However, $MODIS_{AQUA}$ (Figure 5-8) has a seasonal variation similar to that of $MODIS_{TERRA}$, with a more elevated aerosol loading. The observed low and high aerosol loadings for all observed seasons are summer (0.101 & 0.143), autumn (0.084 & 0.133), winter (0.068 & 0.090) and spring (0.077 & 0.118). This association may likely be from transported aerosols from neighbouring countries like Botswana as outlined in a study by Freiman & Piketh (2003). Local municipalities like Lesedi (HPA) and Mogale City experienced a high aerosol loading during summer, in which the HPA is prominent for biomass burning and coal-producing power stations, and CoT has medium-to-low aerosol loading. During Autumn, CoT has a high loading, and Merafong City, Emfuleni and Midvaal have a low loading. Winter has a high loading in the north-western parts and a medium-to-high loading over the central parts. CoT and parts of the VTAPA and HPA also have a relatively high aerosol loading. Spring depicts similar trends to winter, with less aerosol loading observed in the VTAPA and HPA.

The observed low and high loadings for summer in Figure 5-9 are (0.134 & 0.165), autumn (0.100 & 0.128), winter (0.091 & 0.108) and spring (0.174 & 0.224). In summer, MERRA-2 observed a high aerosol loading in the Lesedi local municipality and Midvaal, which both fall under the HPA and VTAPA. This observation may be attributed to the emission activities in the HPA and VTAPA. CoT, Merafong, and Mogale City have a low aerosol loading. Medium-to-high loading is observed

by the satellite over the CoJ. In winter, the CoT had a high aerosol loading, and similarly, spring had a high loading in the CoT.

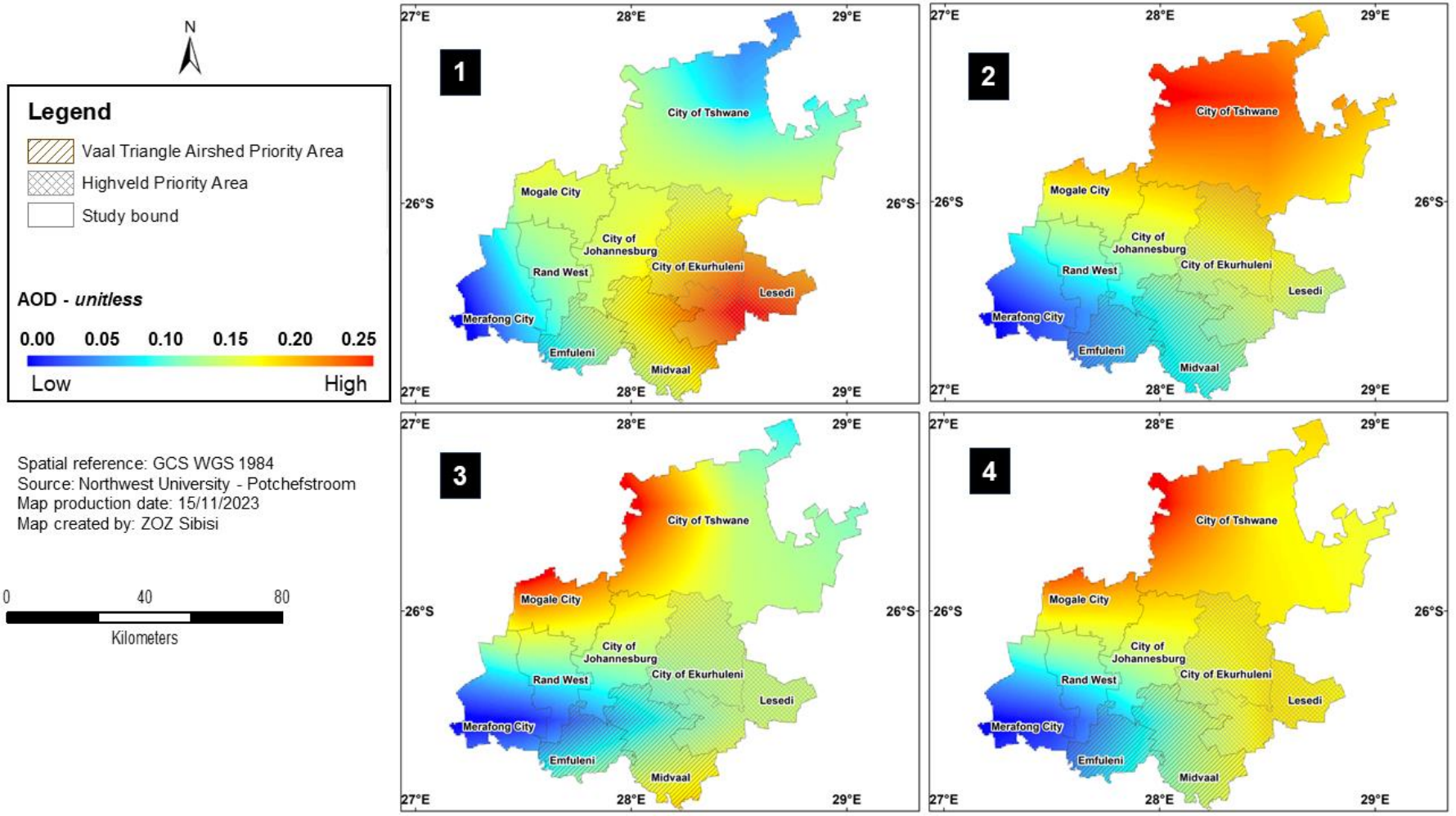


Figure 5-7: Spatial distribution of the seasonally averaged $MODIS_{TERRA}$ AODs at 550nm during the period 2017 to 2021 in Gauteng, South Africa, in (1) summer, (2) autumn, (3) winter, and (4) spring.

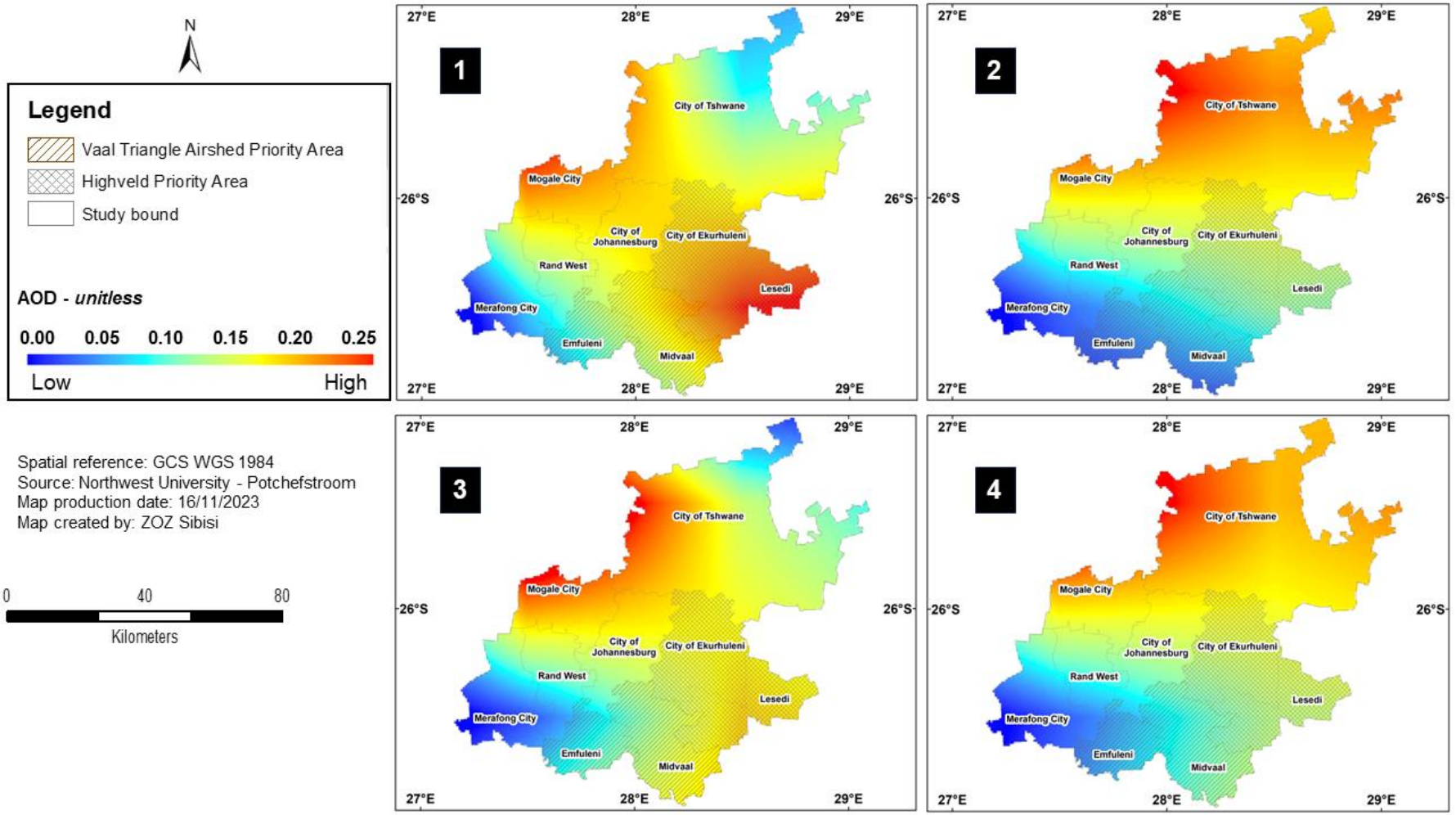


Figure 5-8: Spatial distribution of the seasonally averaged MODIS$_{AQUA}$ AODs at 550nm during the period 2017 to 2021 in Gauteng, South Africa, in (1) summer, (2) autumn, (3) winter, and (4) spring.

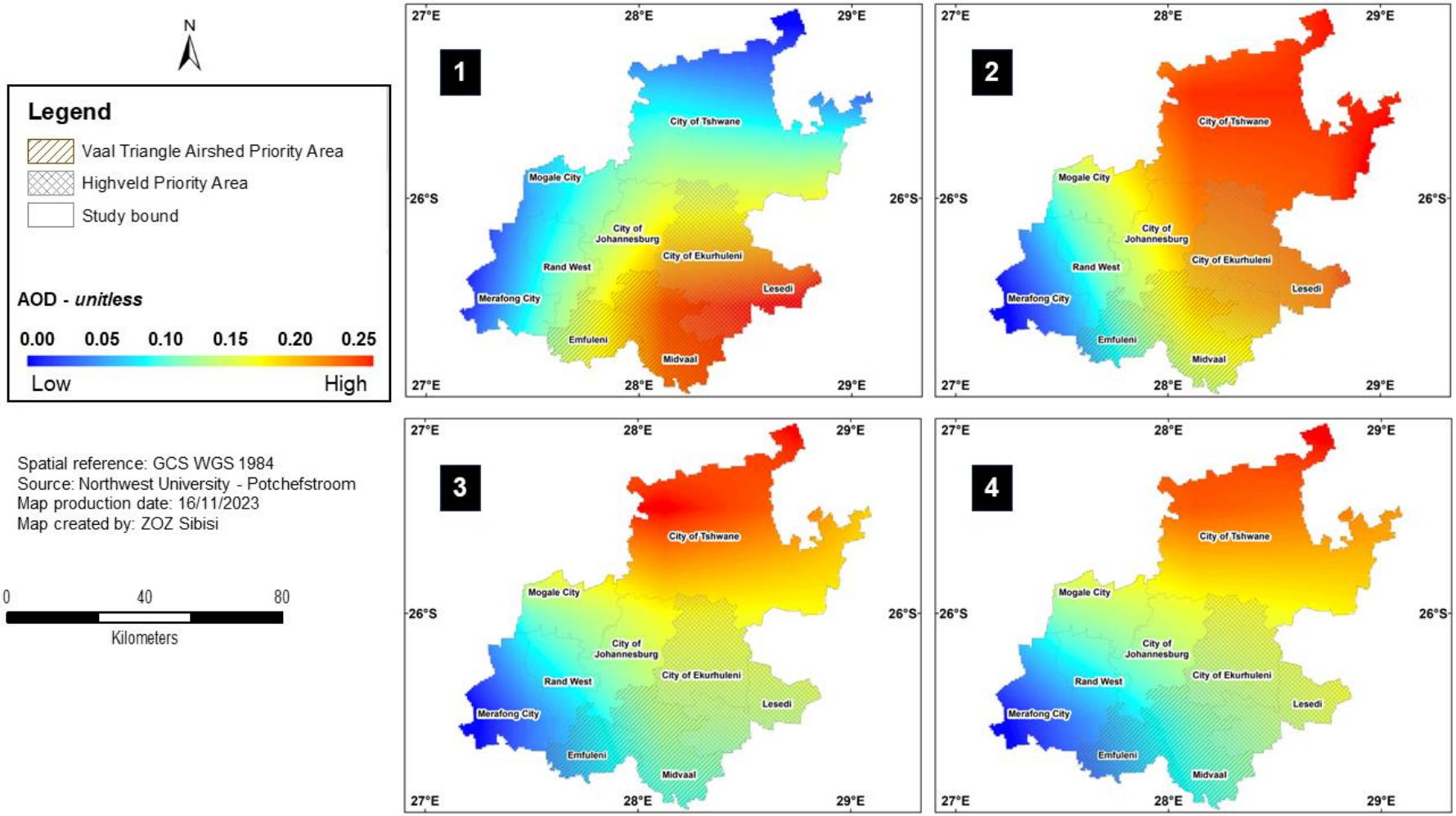


Figure 5-9: Spatial distribution of the seasonally averaged MERRA-2 reanalysis AODs at 550nm during the period 2017 to 2021 in Gauteng, South Africa, in (1) summer, (2) autumn, (3) winter, and (4) spring.

5.2.1 Aerosol types based on AOD and AE

Aerosol physical, chemical and optical qualities are impacted by space and time (Zheng et al., 2017). The α is a value assigned to an aerosols particle characteristic within an atmospheric column. This Ångström can be further described as an index used to classify whether an aerosol particle has absorbing or non-absorbing characteristics.

Light attenuation/ extinction is a crucial aspect that must be considered when working with satellite data as it can impact data retrieval. According to Beer-Lambert's law, it is referred to as the decline in the intensity of the light to travel through a medium (Kyriacou et al., 2019). This process is impacted by absorption, scattering, reflection or refraction (Kyriacou et al., 2019). AOD is a unitless variable; however, for context purposes, this section will be referring to AOD values with a range of 0.0 to 1.0. As seen in Figure 5-10, lower AOD values imply a weaker light attenuation, and higher AOD values imply a stronger light attenuation. And in reference to Table 2-6, aerosol having a weak or low light attenuation are classified as non-absorbing aerosols.

The classification of the aerosols present in Gauteng are adopted from a study by Ranaivombola et al. (2023) with most aerosol species in Gauteng are dust and biomass burning (Shikwambana et al., 2019). Although sea-salt is not dominant aerosol specie found in Gauteng, a study by Shikwambana et al. (2019) did show that meteorology and long range transportation can result in having traces of it inland.

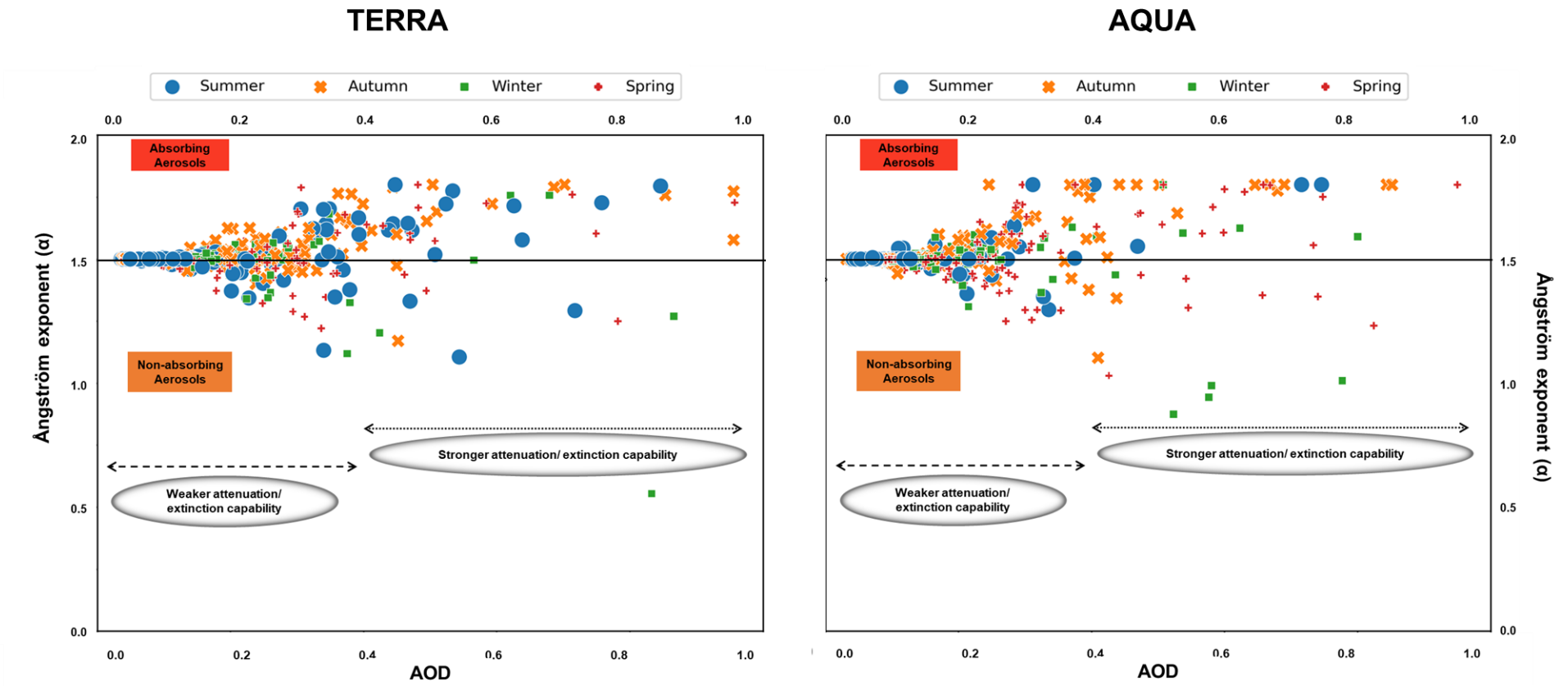


Figure 5-10: Ångström-aerosol classification diagram for MODIS$_{(TERRA, AQUA)}$ for all monitoring stations throughout the monitoring period. The aerosol classification diagrams show all four seasons- Summer, Autumn, Winter and Spring.

5.3 Discussion

The use of satellite data and re-analysis modelling has proven to be quite a useful route to take in air quality studies. In 2019, a study by Shikwambana et al. (2019) was one of the first regional studies to utilise satellite data (MODIS) and model simulations (MERRA-2) in investigating different aerosol species (AOD, BC, Sulphates and sea-salt). Prior to this, a study by Tsidu and Bililign (2018) used MODIS and MERRA-2 to compare diverse surfaces over the Southern parts of Africa (Botswana). An international study by Jiang et al. (2022) also provided scientific backing that MODIS and MERRA-2 provided overall great capability in capturing higher aerosol loadings in highly industrial areas and well as densely populated areas. Observing the annual aerosol loading throughout the monitoring period, MERRA-2 observed higher aerosol loading more than $MODIS_{(TERRA,\ AQUA)}$, which is attributable to satellite overpass times.

5.3.1 Spatiotemporal variability over Gauteng

Atmospheric and meteorological conditions have an impact on the increment and seasonality of aerosols as they are responsible for the transportation and deposition thereof (Garland et al. 2016). The interior of South Africa is dominated by a continental high pressure (HP) cell (Maritz et al., 2015; Tyson and Preston-Whyte, 2000). This impacts the way aerosols are transported and as McGill et al. (2020) outlined are transported “through re-circulation of air masses, westerly transport over the Indian Ocean (south of 20°S) and over the Atlantic Ocean (north of 20°S)”.

Aerosol transportation in the atmosphere has been studied locally over two decades with one of the earliest studies conducted by Piketh et al. (1999). Transportation of aerosols is quite inhomogeneous due to atmospheric conditions. Aerosols are mainly transported vertically and horizontally in which the former is restrained by the atmospheres stability and the latter is restrained by predominant local surface winds (Bigala, 2008). Cosjin and Tyson (1996) further corroborate that vertical aerosol transportation is owed to the presence of temperature inversions which result in stable conditions (no baroclinic westerly disturbances) that also constrain the vertical diffusion of aerosols. Cosjin and Tyson (1996) further elaborate that from the three prominent layers of stability outlined by Tyson et al. (1996), aerosol loading over the interior plateau is primarily found at ~500 hPa and over the coastal region at ~580 hPa.

Variations of aerosol loading are largely impacted by meteorology. For this study, summer is observed to have a high aerosol loading in the south-eastern parts of Gauteng (VTAPA, HPA and CoE) and the north-western parts (Mogale city and parts of CoT) which is an unexpected phenomenon based on South Africa’s atmospheric conditions as mentioned above. A more recent study by Khan et al. (2021) however, suggests that increased aerosol loading during the summer season may be attributed to long-range transport of other aerosol species. This is similar to the

findings of a study by Bigala (2008) where the researcher outlined that the main contributors to elevated aerosol loading in summer can be attributed to atmospheric dust.

Interestingly, aeolian dust and sulphate aerosols have also been proven to be evident during the summer (Piketh et al., 1999). Bearing in mind that some aerosol loading is located over priority areas, a study by Wernecke (2018) outlined the contributions of dust in two townships (KwaDela and KwaZamokuhle) found in the HPA. The study's findings show dust contribution to be from aeolian, wind-blown, soil and road. Aeolian dust showed to contribute 28% to dust emissions. Wind, soil and road dust are not considered in this discussion owing to the fact that they occur at surface level and are not carried aloft, making them insignificant to be observed by satellites. Findings from a study by Segakweng et al. (2022) conducted in the HPA for four townships (KwaZamokuhle, Kwadela, Jouberton and Zamdela) conceded to have high sulphate aerosols during the summer. In this instance, sulphate aerosol loading seem to be attributed by emissions from the coal power stations (Waterberg, Medupi and Matimba) and platinum smelters (Thabazimbi and Northam) located in the WBPA and the power stations located in the HPA (Hendrina). With aerosol loading attributed to areas out of the study area bound, the phenomenon of these aerosols in Gauteng is explained through the transport of aerosols which become deposited and observed in parts of Gauteng. These findings contribute in explaining why the summer has a high aerosol loading.

Aerosol loading during the autumn seems to be relatively high in the CoT, more especially when observed with MERRA-2. A study by Adesina et al. (2019) proved to have similar findings which explained this phenomenon to be alluded to backscattering, indicating the prevalence of coarse-fraction particles. However, this finding could not be inferred to this study as it was not presented in the data. This is also where mixing layer heights are high making satellite observation possible.

The winter shows to have a relatively low aerosol loading. Khan et al. (2021) summarises that this is due to the weak vertical mixing that is associated with stable atmospheric conditions. In winter, $MODIS_{TERRA}$ observed a relatively high aerosol loading in the Midvaal, which is part of the VTAPA which can be attributed to a possible event that occurred. The observed event may be dust or haze or because the close proximity of the coal fired power station in Zamdela, Sasolburg. The Midvaal area is also located near Three Rivers, Sharpeville and Sebokeng which in reference to ***section 3.5.2***, Sharpeville and Sebokeng showed extremely high PM concentrations during the winter season which are attributable to biomass burning. These concentrations are high enough to negatively impact human health, in which an earlier study by Komo (2013) outlined some of the health impacts resulting from biomass burning in the VTAPA are premature mortality and frequent hospitalisations due to cardiovascular and respiratory illnesses.

The spring shows an aerosol loading over the northern parts of the province. This is validated by Shikwambana et al. (2019) where the findings outline that elevated sulphate levels were observed over the eastern Limpopo province in which transportation thereof have settled in neighbouring provinces like Gauteng. Incidentally, a study by Freiman and Piketh (2003) further supports that South African neighbouring countries like Botswana may contribute to the increase in aerosol loading in central areas like Gauteng due to long range transport of pollutants.

$MODIS_{AQUA}$ observed a high aerosol loading for all seasons at parts covering Mogale City and the CoT. These two areas are adjacent to the Waterberg-Bojanala Priority Area (WBPA). With reference to ***section 1.4***, high aerosol loadings are observed in the WBPA and is associated more with emission contribution activities like industries and mining.

The south-western parts of Gauteng, especially Merafong city having shown consistent low aerosol loadings throughout all the seasons. Using satellite data was beneficial as there are no ground-based monitoring stations in the area. With the satellite data, it was observed that there was a consistency in low aerosol loadings, which is an interesting finding considering it is a gold-mining town which is prone to occasional mineral dust from Tailing Storage Facilities (TSF's) and harmful air pollutants as outlined by (Nealer, 2020; Kershaw et al., 2003). Merafong city, mainly Carletonville has been impacted by severe sinkholes due to the underlying dolomitic rock. It also surrounded by Tailing Storage Facilities (TSF's) which are major contributors to air pollution in the area (Mpanza et al. 2020). Considering low aerosol loading, this may be a result of the dust from TSF's not travelling far and high enough to be observed by a satellite within the vertical column.

****

Aerosol loading is prominent in the northern parts of the province as well as in areas promulgated as priority areas, with reference to ***chapter 1, section 1.4***. Interestingly, seasonal variation of aerosols in this study has proven to be impacted by meteorology (long-range transport of aerosols) as this is observed by the high aerosol loading in the VTAPA and HPA during summer. The next chapter will provide a summary of the inferences made from this research study, and further provide recommendations.

6 CONCLUSION AND RECOMMENDATIONS

This study aimed to investigate the air quality in Gauteng using ground-based and remotely-derived measurements. Furthermore, the study aimed to establish how the $MODIS_{(TERRA, AQUA)}$ platform performed in these measurements and it proved to be a good proxy to use in AOD-$PM_{2.5}$ studies, within the South African (Gauteng) context. The study has proven to successfully answer and fill gaps encountered. This chapter will outline the inferences made from this study. These inferences will stem from the outlined research objectives in ***chapter 1, section 1.9***. Not only will this chapter highlight the inferences made from each research objective, it will additionally provide some insight on the assumptions undertaken to facilitate the completion of this study as well as recommendations that could further improve not only this study but future studies pertaining to air quality.

****

6.1 Summary on the main findings of the study

6.1.1 Ground-based ambient $PM_{2.5}$ and local meteorology

Objective i: *Characterise ambient $PM_{2.5}$ levels and local meteorology in the Gauteng region using surface in-situ measurements.*

The study underscores the importance of understanding air quality, particularly focusing on $PM_{2.5}$, due to its significant health impacts and observed levels in Gauteng. The study finds a complex interplay between $PM_{2.5}$ concentrations and meteorological conditions such as temperature, humidity, rainfall, and wind speed. These findings are crucial for understanding how weather patterns affect air pollution levels. The study found that activities like biomass burning, and domestic fuel burning are still prevalent in low-income communities and greatly contribute to poor air quality. The study further highlights the significant temporal (hourly, daily, seasonal) and spatial variations in $PM_{2.5}$ concentrations across the Gauteng region. These variations are essential for designing targeted air quality management strategies.

- **Local meteorology**

Diurnal temperature variations were strongest during the winter rather for all monitoring stations due to a dominance of sub-tropical highs over South Africa. Diurnal relative humidity peaks were observed from 03:00 - 04:00 in the morning hours, and low humidity levels were observed during the noon hours, 12:00 - 15:00.

- **Ambient PM2.5 mass concentrations**

Maximum averaged $PM_{2.5}$ concentrations were 112 $\mu g.m^{-3}$ and 173 $\mu g.m^{-3}$ for hourly and daily averages. For hourly averaged $PM_{2.5}$ data, two out of four stations from the CoJ (Jabavu (864 $\mu g.m^{-3}$) and Kliprivier (945 $\mu g.m^{-3}$)) showed high $PM_{2.5}$ concentrations, two out of four stations from the CoE (Etwatwa (996 $\mu g.m^{-3}$) and Olifantsfontein (994 $\mu g.m^{-3}$)) showed high $PM_{2.5}$ concentrations, one out of four stations from the CoT (Hammanskraal (998 $\mu g.m^{-3}$)) and showed high $PM_{2.5}$ concentrations and three (Sebokeng (831 $\mu g.m^{-3}$), Sharpeville (1000 $\mu g.m^{-3}$) and Vanderbijlpark (761 $\mu g.m^{-3}$)) out of five stations from Sedibeng had high $PM_{2.5}$ concentrations. For daily averaged $PM_{2.5}$ data, four stations (Kliprivier (415 $\mu g.m^{-3}$), Olifantsfontein (326 $\mu g.m^{-3}$), Etwatwa (327 $\mu g.m^{-3}$) and Jabavu (306 $\mu g.m^{-3}$) had concentrations exceeding 300 $\mu g.m^{-3}$. All monitoring stations had daily NAAQS $PM_{2.5}$ exceedances. Kliprivier is shown to have the highest daily exceedance of 415, followed by Etwatwa (327), Olifantsfontein (326) and Jabavu (306).

- **Relationship between $PM_{2.5}$ and meteorological conditions**

My sentence is now constructed better to say, “Meteorological parameters have shown a very poor relationship with $PM_{2.5}$ concentrations. Based on the statistical results, there was quite less interaction with $PM_{2.5}$. So therefore, the study infers that meteorology did not play much of a role to the dispersal or accumulation of $PM_{2.5}$ concentrations and alteration of $PM_{2.5}$ particles. In the case of Gauteng, it was clear that temperature, relative humidity, rainfall and windspeed did not have a strong relationship with $PM_{2.5}$. The correlation coefficient (r) is as follows: -0.4, 0.11, -0.04 and -0.33. This poor relationship is a result of a positive response feedback. Another notable aspect that influenced the meteorology and $PM_{2.5}$ interaction is the magnitude of the meteorological conditions with $PM_{2.5}$ concentrations, simultaneously or within a short space of time. This could imply that meteorological conditions over Gauteng during the study period were not dominant enough to motivate/ stimulate for an on-going interaction.

6.1.2 Evaluation of satellite observations

Objective ii: *Evaluating Satellite Observations in the Gauteng Region*

The accuracy and limitations of satellite observations for Aerosol Optical Depth (AOD) reveals that while these observations provide valuable insights into regional air quality patterns, they have limitations and discrepancies compared to ground-based measurements. This emphasises the need for careful calibration and validation of satellite data. The study observed a lot of overfitting and underfitting which suggests that future, local based studies, which maintain similar assumptions to that of this study may implore the use of the Combined DB & DT retrieval algorithm

at Site 2 (Pretoria_CSIR_DPSS) using the $MODIS_{AQUA}$ satellite. This study proved that findings will be more valid and reliable.

Model validation presented with the following findings. The Simple Linear Regression presents the following: For DB, $MODIS_{TERRA}$ outperformed $MODIS_{AQUA}$ at both Site 1 and Site 2 (0.88 & 0.78). For DT, $MODIS_{TERRA}$ outperformed $MODIS_{AQUA}$ at Site 1 (0.88) and $MODIS_{AQUA}$ outperformed $MODIS_{TERRA}$ at Site 2 (0.90). For Combined DB & DT, $MODIS_{TERRA}$ performed better at Site 1 (0.88) and $MODIS_{AQUA}$ performed better at Site 2 (0.90). Ordinary Least Square Regression presents the following: For DB, $MODIS_{TERRA}$ performed better both at Site1 and Site 2 (0.86 & 0.78). For DT, $MODIS_{TERRA}$ outperformed $MODIS_{AQUA}$ at both Site 1 and Site 2 (0.88 & 0.69). For DB & DT, $MODIS_{TERRA}$ performed better at Site 1 (0.88) and $MODIS_{AQUA}$ performed better at Site 2 (0.90). Scikit-Learn model presents the following: DB presented with overfitting at site 1 and site 2 for $MODIS_{AQUA}$ and only at site 2 for $MODIS_{TERRA}$. The DT algorithm presented underfitting for both $MODIS_{AQUA}$ sites 1 and 2, and only site 2 for $MODIS_{TERRA}$. CB only showed for the algorithm to be applicable using Site 2 ($MODIS_{AQUA}$) indicating the model is fit for applicability studies.

Both the SLR and the OLS regression models produced incredibly similar results except for the DT algorithm for both Aqua satellites, in which the SLR had a better performance, making it the best preforming model against the SLR. SLR: For $MODIS_{TERRA}$, Site 1 performed better (100%) for all the retrieval algorithms (DB, DT and Combined DB & DT). For Site 2, only 1/3 (33.33%) of the retrieval algorithms (DB) preformed better. For $MODIS_{AQUA}$, there is 0/3 (0%) retrieval algorithms (DB, DT and Combined DB & DT) that performed better at Site 1. For Site 2, there are 2/3 (66.66%) retrieval algorithms (DT and Combined DB & DT) that performed better. OLS: For $MODIS_{TERRA}$, similarly to the SLR, all 3/3 (100%) retrieval algorithms (DB, DT and Combined DB & DT) for Site 1 performed better than Site 2. 2/3 (66.66%) retrieval algorithms (DB, DT) performed better. For $MODIS_{AQUA}$ there is no retrieval algorithms that performed better at Site 1, bringing it to 0/3 (0%) performance level. For Site 2, only 1/3 (33.33%) of the retrieval algorithms (Combined DB & DT algorithm) performed better.

6.1.3 Remotely derived aerosols over Gauteng

Objective iii: *Characterise the aerosol loadings over Gauteng using remotely derived measurements and modelled reanalysis data*

It is noteworthy to consider re-analysis tools in policy making to combat environmental issues experienced. The study successfully characterises aerosols in the Gauteng region using remote sensing data and re-analysis models. This approach is particularly valuable in areas lacking direct monitoring, providing insights into the spatial and temporal distribution of aerosols. The study suggests that effective air quality management in Gauteng requires an integrated approach that

considers the complex interactions between human activities, meteorological factors, and air pollutants. This approach should utilise both ground-based and satellite observations for comprehensive monitoring and analysis.

For $MODIS_{(TERRA, AQUA)}$, AOD loading was more concentrated in the HPA and VTAPA during the years 2018, 2020 and 2021. Long-range particle transportation has an influence on the aerosol loading where aerosol particles may travel from other distant areas in the country as well as neighbouring parts of the country. These aerosol particles may settle or become concentrated in an area far the source. A high aerosol loading in summer was driven by dust and biomass burning events. Low aerosol loading in winter was driven by weak vertical mixing that is associated with stable atmospheric conditions. MERRA-2 observed enhanced variability across Gauteng more than $MODIS_{(TERRA, AQUA)}$. Making inferences from literature, this study showed that aerosol species are likely to be from dust and-/or biomass burning.

6.1.4 Critical reflection on the study aim

The study aimed to thoroughly evaluate air quality in the Gauteng Province by combining ground- and remotely derived measurements. However, due to poor data retention and quality, the ground-based station faced limitations in offering a complete $PM_{2.5}$ distribution. While satellite data exhibited better spatial coverage from above, it did not account for atmospheric dynamics at the surface level. Utilising a re-analysis tool proved more efficient in capturing the spatiotemporal distribution of aerosols, yielding results like those observed by ground stations (specifically, aerosol concentration in the highveld priority areas). Consequently, the study proposes that effective air quality management in Gauteng necessitates an integrated approach considering the intricate interplay of human activities, meteorological factors, and air pollutants. This approach should incorporate ground-based and satellite observations for a comprehensive monitoring and analysis strategy.

****

This chapter has provided a succinct and thorough summary on the results obtained in this research study with confidence that research objectives outlined in ***chapter 1, section 1.9*** have been achieved.

REFERENCES

Aasen, H., Honkavaara, E., Lucieer, A. & Zarco-Tejada, P.J. 2018. Quantitative remote sensing at ultra-high resolution with UAV Spectroscopy: a review of sensor technology measurement procedures, and data correction workflows. *Remote Sensing*, 10(7), 1091. https://doi.org/10.33 90/rs10071091

Adesina, A.J. 2015. *Aerosol characteristics over different regions of Southern Africa - using sun photometer and satellite measurements*. Westville: University of KwaZulu-Natal. (Thesis – PhD). https://ukzn-dspace.ukzn.ac.za/handle/10413/12897

Adesina, J.A., Piketh, S.J., Formenti, P., Maggs-Kölling, G., Holben, B.N. & Sorokin, M.G. 2019. Aerosol optical properties and direct radiative effect over Gobabeb, Namibia. *Clean Air Journal*, 29(2):1-11. http://dx.doi.org/10.17159/caj/2019/29/2.7518

Adesina, J.A., Piketh, S.J., Qhekwana, M., Burger, R., Language, B. & Mkhatswa, G. 2020. Contrasting indoor and ambient particulate matter concentrations and thermal comfort in coal and non-burning households at South Africa Highveld. *Science of the Total Environment*, 699:1-9. https://doi.org/10.1016/j.scitotenv.2019.134403

Adeyemi, A., Molnar, P., Boman, J. & Wichmann, J. 2022. Particulate matter ($PM_{2.5}$) characterization, air quality level and origin of air masses in an urban background in Pretoria. *Archives of Environmental Contamination and Toxicology*, *83*(1), pp.77-94.

Air quality Act 39 of 2004.

Air quality management plan for the city of Tshwane Metropolitan municipality 2006-2008.

Akoglu, H. 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18(3):91-93.

Aldabash, M., Balcik, F.B. & Glantz, P. 2020. Validation of MODIS C6. 1 and MERRA-2 AOD using AERONET observations: a comparative study over Turkey. *Atmosphere*, 11(9), 905. https:// doi.org/10.3390/atmos11090905

Anderson, J.C., Wang, J., Zeng, J., Petrenko, M., Leptoukh, G.G, & Ichoku, C. 2012. Accuracy assessment of Aqua-MODIS aerosol optical depth over coastal regions: importance of quality flag and sea surface wind speed. *Atmospheric Measurements Technique Discussion*, 5:5205-5243. https://doi.org/10.5194/amtd-5-5205-2012

Arabi, B., Salama, M.S., Pitarch, J. & Verhoef, W. 2020. Integration of in-situ and multi-sensor satellite observations for long-term water quality monitoring in coastal areas. *Remote Sensing of Environment*, *239*, p.111632. https://doi.org/10.1016/j.rse.2020.111632

Arowosegbe, O.O., Röösli, M., Künzli, N., Saucy, A., Adebayo-Ojo, T.C., Schwartz, J., Kebalepile, M., Jeebhay, M.F., Dalvie, M.A. & De Hoogh, K. 2022. Ensemble averaging using remote sensing data to model spatiotemporal PM_{10} concentrations in sparsely monitored South Africa. *Environmental Pollution*, *310*, p.119883. https://doi.org/10.1016/j.envpol.2022.119883

Balsamo, G., Agusti-Parareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., … Buizza, R., 2018. Satellite and in situ observations for advancing global Earth surface modelling: a Review. *Remote Sensing*, *10*(12), p.2038. https://doi.org/10.3390/rs10122038

Beelen, R., Voogt, M., Duyzer, J., Zandveld, P. & Hoek, G. 2010. Comparison of the performances of land use regression modelling and dispersion modelling in estimating small-scale variations in long-term air pollution concentrations in a Dutch urban area. *Atmospheric environment*, *44*(36), pp.4614-4621. https://doi.org/10.1016/j.atmosenv.2010.08.005

Bigala, T.A. 2008. *Aerosol loading over the South African Highveld*. Johannesburg: University of Witwatersrand. (Dissertation – MSc).

Birim, N.G., Turhan, C., Atalay, A.S. & Gokcen Akkurt, G. 2023. The Influence of Meteorological Parameters on PM10: A Statistical Analysis of an Urban and Rural Environment in Izmir/Türkiye. *Atmosphere*, *14*(3), p.421. https://doi.org/10.3390/atmos14030421

Bostock, W.W. 2018. South Africa's evolving language policy: educational implications. *Journal of Curriculum and Teaching*, *7*(2):27-32. https://doi.org/10.5430/jct.v7n2p27

Bubeck, S., Tomaschek, J. & Fahl, U. 2014. Potential for mitigating greenhouse gases through expanding public transport services: a case study for Gauteng Province, South Africa. Transportation Research Part D: Transport and Environment, 32:57-69. https://doi.org/10.1016/j.t rd.2014.07.002

Butt, J.M., Assiri, M.E. & Ali, M.D. 2017. Assessment of AOD variability over Saudi Arabia using MODIS Deep Blue products. *Environmental Pollution,* 231:143-153.

Campos, P.M., Pires, J.C. & Leitão, A.A. 2022. Assessment of aerosols over five cities of Angola based on MERRA–2 reanalysis data. *Atmospheric Pollution Research*, 13(10), 101569. https:// doi.org/10.1016/j.apr.2022.101569

Carmona, J.M., Gupta, P., Lozano-García, D.F., Vanoye, A.Y., Yépez, F.D. & Mendoza, A. 2020. Spatial and temporal distribution of $PM_{2.5}$ pollution over north-eastern Mexico: Application of MERRA-2 reanalysis datasets. *Remote Sensing*, 12(14), 2286. https://doi.org/10.3390/rs1214 2286

Chang, L.T.C., Scorgie, Y., Duc, H.N., Monk, K., Fuchs, D. & Trieu, T. 2019. Major source contributions to ambient $PM_{2.5}$ and exposures within the new South Wales greater metropolitan region. *Atmosphere*, 10(3), 138. https://doi.org/10.3390/atmos10030138

Chidhindi, 2020. *Using dispersion models as a regulatory tool in South Africa*. Potchefstroom: North-West University. (Dissertation - MSc).

Chiwewe, T. M. & Ditsela, J. 2016. Machine learning based estimation of Ozone using spatio-temporal data from air quality monitoring stations. IEEE 14th International Conference on Industrial Informatics (INDIN). Poitiers: IEEE. pp. 58-63. doi:10.1109/INDIN.2016.7819134.

Chen, Z., Chen, D., Zhao, C., Kwan, M.P., Cai, J., Zhuang, Y., … Li, R. 2020. Influence of meteorological conditions on $PM_{2.5}$ concentrations across China: A review of methodology and mechanism. *Environment international*, *139*, p.105558. https://doi.org/10.1016/j.envint.2020.105558

Chen, Z., Xie, X., Cai, J., Chen, D., Gao, B., He, B., … Xu, B. 2018. Understanding meteorological influences on $PM_{2.5}$ concentrations across China: a temporal and spatial perspective. *Atmospheric Chemistry and Physics*, *18*(8), pp.5343-5358. https://doi.org/10.5194/a cp-18-5343-2018

Chu, D.A., Kaufman, Y.J., Ichoku, C., Remer, L. A, Tanre´, D. & Holben, B.N. 2002. Validation of MODIS aerosol optical depth retrieval over land. *Geophysical Research Letter*, 29(12), 1617. https://doi.org/10.1029/2001GL013205

City of Johannesburg Metropolitan Gauteng. 2007. Profile and analysis: district development model.

Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., … Forouzanfar, M.H. 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet*, 389(10082):1907-1918. https://doi.org/10.1016/S0140-6736(17)30505-6

Cole, M.J. & Broadhurst, J.L. 2020. Mapping and classification of mining host communities: a case study of South Africa. *The Extractive Industries and Society*, 7(3):954-964. https://doi.org/10. 1016/j.exis.2020.06.007

Constitution of the Republic of South Africa 1996.

Contestabile, M., Mohammed, A. & Almubarak, B. 2017. Will current electric vehicle policy lead to cost-effective electrification of passenger car transport?. *Energy Policy*, 110:20-30. https://doi. org/10.1016/j.enpol.2017.07.062

Cooksey, R.W. 2020. *Descriptive statistics for summarising data*. In: Cooksey, R, eds. Illustrating statistical procedures: finding meaning in quantitative data. Singapore: Springer. pp. 61-139. https://doi.org/10.1007/978-981-15-2537-7_5

Cosijn, C. & Tyson, P. 1996. Stable discontinuities in the atmosphere over South Africa. *South African journal of science*, 92(8):381-386. https://hdl.handle.net/10520/AJA00382353_7756

Cullen, S. 2019. *Causes and effects of suburban traffic dynamics - a case study in a municipality close to Munich.* Uppsala: Uppsala University. (Dissertation - MSc).

De Lange, A., Naidoo, M., Garland, R.M. & Dyson, L.L. 2021. The sensitivity of simulated surface-level pollution concentrations to WRF-ARW-model PBL parameterisation schemes over the Highveld of South Africa. *Atmospheric Research*, *254*, p.105517. https://doi.org/10.1016/j.atmosres.2021.105517

De Marco, A., Proietti, C., Anav, A., Ciancarella, L., D'Elia, I., Fares, S., … Leonardi, C. 2019. Impacts of air pollution on human and ecosystem health, and implications for the national emission ceilings directive: insights from Italy. *Environment International*, 125:320-333. https:// doi.org/10.1016/j.envint.2019.01.064

Department of Environmental Affairs (South Africa). 2018. National Environmental Management: Air Quality Act, 2004 (Act No. 39 of 2004): The 2017 National framework for air quality management in the republic of South Africa. (Notice 1144). *Government Gazette*, 41996, 26 Oct

Department of Environmental Affairs (South Africa). 2012. Highveld Priority Area Air Quality Management Plan.

Department of Environmental Affairs (South Africa). 2014. National Environmental Management: Air Quality Amendment Act, 2004 (Act No. 20 of 2014). (Notice 390). *Government Gazette*, 37666, 19 May

Department of Environmental Affairs (South Africa). 2012. National Environmental Management: Air Quality Act, 2004 (Act No. 39 of 2004): highveld priority area air quality management plan. (Notice 144). *Government Gazette*, 35072, 2 Mar.

Department of Environmental Affairs (South Africa). 2012. National Environmental Management: Air Quality Act, 2004 (Act No. 39 of 2004): National Ambient Air Quality Standard for Particulate Matter with Aerodynamic Diameter Less than 2.5 Micron Metres ($PM_{2.5}$). (Notice 486). *Government Gazette*, 354637, 29 Jun.

Department of Environmental Affairs (South Africa). 2005. National Environmental Management: Air Quality Act, 2004 (Act No. 39 of 2004). (Notice 163). *Government Gazette*, 27318, 24 Feb

Department of Justice and Constitutional Development. 2022. Invitation for public comments on the constitution of the republic of South Africa, 1996: amendment to section 6 of the constitution. (Notice 1156). *Government Gazette*, 47049:3, 19 Jul

Dobson, R., Siddiqi, K., Ferdous, T., Huque, R., Lesosky, M., Balmes, J. & Semple, S. 2021. Diurnal variability of fine-particulate pollution concentrations: data from 14 low- and middle-income countries. *The International Journal of Tuberculosis and Lung Disease*, 25(3):206-214. https://doi.org/10.5588/ijtld.20.0704

Duan, P., Wang, Y. & Yin, P. 2020. Remote sensing applications in monitoring of protected areas: a bibliometric analysis. *Remote Sensing*, 12(5), 772. https://doi.org/10.3390/rs12050772

Dyson, L.L., van Heerden, J. & Sumner, P.D. 2015. A baseline climatology of sounding-derived parameters associated with heavy rainfall over Gauteng, South Africa. *International Journal of Climatology*, 35(1):114-127. https://doi.org/10.1002/joc.3967

Edlund, K.K., Killman, F., Molnár, P., Boman, J., Stockfelt, L. & Wichmann, J. 2021. Health risk assessment of PM2. 5 and PM2. 5-bound trace elements in Thohoyandou, South Africa. *International journal of environmental research and public health*, *18*(3), p.1359. https://doi.org/10.3390/ijerph18031359

Eibedingil, I.G., Gill, T.E., Van Pelt, R.S. & Tong, D.Q. 2021. Comparison of Aerosol Optical Depth from MODIS Product Collection 6.1 and AERONET in the Western United States. *Remote Sensing*, 13(12), 2316. https://doi.org/10.3390/rs13122316

Ekoh, H.C. 2020. Spatial variation of air quality in Mpape area of Abuja, Nigeria. *World Scientific News*, 140:79-112.

Engel-Cox, J.A., Hoff, R.A. & Haymet, A.D.J. 2004. Recommendations on the use of satellite remote-sensing data for urban air quality. *Journal of the Air and Waste Management Association*, 54(11):1360-1371. https://doi.org/10.1080/10473289.2004.10471005

Fairbanks, D.H.K., Thompson, M.W., Vink, D.E., Newby, T.S., Van den Berg, H.M. & Everard, D. 2000. The South African land-cover characteristics database: a synopsis of the landscape. *South African Journal of Science*, 96(2):69-82. https://hdl.handle.net/10520/AJA00382353_7429

Faizan, M.A. & Thakur, R. 2019. Measuring the impact of household energy consumption on respiratory diseases in India. *Global Health Research and Policy*, 4(10):1-9. https://doi.org/10.118 6/s41256-019-0101-7

Falah, S., Mhawish, A., Sorek-Hamer, M., Lyapustin, A.I., Kloog, I., Banerjee, T., Kizel, F. & Broday, D.M. 2021. Impact of environmental attributes on the uncertainty in MAIAC/MODIS AOD retrievals: a comparative analysis. *Atmospheric Environment*, 262, 118659. https://doi.org/10.10 16/j.atmosenv.2021.118659

Fauchereau, N., Trzaska, S., Rouault, M. & Richard, Y. 2003. Rainfall variability and changes in southern Africa during the 20th century in the global warming context. *Natural hazards*, 29(2),139. https://doi.org/10.1023/A:1023630924100

Feig, G., Garland, R.M., Naidoo, S., Maluleke, A. & van der Merwe, M. 2019. Assessment of changes in concentrations of selected criteria pollutants in the Vaal and Highveld priority areas. *Clean Air Journal*, 29(2):75-87. https://doi.org/10.17159/caj/2019/29/2.7464

Filonchyk, M. & Hurynovich, V. 2020. Validation of MODIS aerosol products with AERONET measurements of different land cover types in areas over Eastern Europe and China. *Journal of Geovisualization and Spatial Analysis*, *4*, pp.1-11. https://doi.org/10.1007/s41651-020-00052-9

Franzese, M. & Luliano, A. 2018. Descriptive statistics. Encyclopaedia of Bioinformatics and *Computational Biology*. 1-13. doi:10.1016/B978-0-12-809633-8.20354-3

Freiman, M.T. & Piketh, S.J. 2003. Air Transport into and out of the industrial Highveld region of South Africa. *Applied Meteorology*, 42:994-1002. https://doi.org/10.1175/1520-0450(2003)042% 3C0994:ATIAOO%3E2.0.CO;2

Garland, R.M., Horowitz, H.M., Engelbrecht, C.J., Dedekind, Z., Bopape, M.J.M. & Engelbrecht, F.A. 2016. Representation of aerosol particles and associated transport pathways in regional climate modelling in Africa. Paper delivered at 32nd Annual conference of South African Society for Atmospheric Sciences (SASAS), Pretoria. http://www.csag.uct.ac.za/wp-content/uploads/2016/04/SASAS_2016_Conference_Proceedings_Final_18Nov_16.pdf Date of access:5 May 2022.

Garland, R.M., Naidoo, M., Sibiya, B. & Oosthuizen, R. 2017. Air quality indicators from the Environmental Performance Index: potential use and limitations in South Africa. *Clean Air Journal*, 27(1):33-41. http://dx.doi.org/10.17159/2410-972X/2017/v27n1a8

Garsa, K., Khan, A.A., Jindal, P., Middey, A., Luqman, N., Mohanty, H. & Tiwari, S, 2023. Assessment of meteorological parameters on air pollution variability over Delhi. *Environmental Monitoring and Assessment*, *195*(11), p.1315. https://doi.org/10.1007/s10661-023-11922-2

Gautam, D. & Boli, N.B. 2020. Air pollution: impact and interventions. *Air Quality, Atmosphere and Health*, 13:209-223. https://doi.org/10.1007/s11869-019-00784-8

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., … Zhao, B. 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of climate*, 30(14):5419-5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Geng, G., Meng, X., He, K. & Liu, Y. 2020. Random forest models for $PM_{2.5}$ speciation concentrations using MISR fractional AODs. *Environmental Research Letters*, 15(3), 034056. https://doi.org/10.1088/1748-9326/ab76df

Ginindza, B. & Muzenda, E. 2016. Waste management challenges to opportunities in the west rand district municipality, Gauteng, South Africa: initiatives.

Govender, K. & Sivakumar, V. 2019. A decadal analysis of particulate matter ($PM_{2.5}$) and surface ozone (O3) over Vaal Priority Area, South Africa. *Clean Air Journal*, 29(2). https://doi.org/10.171 59/caj/2019/29/2.7578

Grab, S. and Knight, J. 2015. *Landscapes and Landforms of South Africa*. London: Springer Cham. https://doi.org/10.1007/978-3-319-03560-4_1

Gu, Z. 2022. Complex heatmap visualisation. *iMeta*, 1(3), e43. https://doi.org/10.1002/imt2.43

Gui, K., Che, H., Li, L., Zheng, Y., Zhang, L., Zhao, H., … Zhang, X. 2021. The significant contribution of small-sized and spherical aerosol particles to the decreasing trend in total aerosol optical depth over land from 2003 to 2018. *Engineering*, 16:82-92. https://doi.org/10.1016/j.eng. 2021.05.017

Gupta, G., Ratnam, M.V., Madhavan, B.L. & Narayanamurthy, C.S. 2022. Long-term trends in Aerosol Optical Depth obtained across the globe using multi-satellite measurements. *Atmospheric Environment*, 273, 118953. https://doi.org/10.1016/j.atmosenv.2022.118953

Hashimoto, M. & Nakajima, T. 2017. Development of a remote sensing algorithm to retrieve atmospheric aerosol properties using multiwavelength and multipixel information. Journal of Geophysical Research: *Atmospheres*, 122(12):6347-6378. https://doi.org/10.1002/2016JD0256 98

Henneman, L.R.F., Rafaj, P., Annegarn, H.J. & Klausbruckner, C. 2016. Assessing emissions levels and costs associated with climate and air pollution policies in South Africa. *Energy Policy*, 89:160-170. https://doi.org/10.1016/j.enpol.2015.11.026

Hersey, S.P., Garland R.M., Crosbie, E., Shingler, T., Sorooshian, A., Piketh, S. & Burger, R. 2015. An overview of regional and local characteristics of aerosols in South Africa using satellite, ground, and modelling data. *Atmospheric Chemistry and Physics*, 15:4259-4278. https://doi.org/ 10.5194/acp-15-4259-2015

Hoff, R.M. & Christopher, S.A. 2009. Remote sensing of particulate pollution from space: Have we reached the promised land?. *Journal of the Air and Waste Management Association*, 59(6):645-675. https://doi.org/10.3155/1047-3289.59.6.645

Holben, B.N., Eck, T.F., Slutsker, I., Tanre´, D., Buis, J.P., Setzer, A., … Smirnov, A. 1998. AERONET - A federated instrument network and data archive for aerosol characterization. *Remote Sensing Environment*, 66(1):1-16. https://doi.org/10.1016/S0034-4257(98)00031-5

Hu, Z., Jin, Q., Ma, Y., Pu, B., Ji, Z., Wang, Y. & Dong, W, 2021. Temporal evolution of aerosols and their extreme events in polluted Asian regions during Terra's 20-year observations. *Remote Sensing of Environment*, 263, p.112541. https://doi.org/10.1016/j.rse.2021.112541

Hubanks, P., Platnick, S., King, M. & Ridgway, B. 2015. MODIS Atmosphere L3 gridded product algorithm theoretical basis document (atbd) & users guide. *ATBD reference number ATBD-MOD-30, NASA*, *125*, p.585.

Hubanks, P.A., King, M.D., Platnick, S. & Pincus, R. 2008. MODIS atmosphere L3 gridded product algorithm theoretical basis document. *ATBD Reference Number: ATBD-MOD-30*, *30*, p.96.

Hurst, E. 2015. Overview of the Tsotsitaal's of South Africa: Their different base languages and common core lexical items. *Youth language practices in Africa and beyond*, (105):169-184.

IPCC, 2018: Summary for Policymakers. In: *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 3-24. https://doi.org/10.1017/9781009157940.001

Janhäll, S., 2015. Review on urban vegetation and particle air pollution–Deposition and dispersion. *Atmospheric environment*, *105*, pp.130-137.

Janse, R.J., Hoekstra, T., Jager, K.J., Zoccali, C., Tripepi, G., Dekker, F.W. & van Diepen, M. 2021. Conducting correlation analysis: important limitations and pitfalls. *Clinical Kidney Journal*, 14(11):2332-2337. https://doi.org/10.1093/ckj/sfab085

Jiang, D., Wang, L., Yi, X., Su, X. & Zhang, M. 2022. Comprehensive evaluation of multisource aerosol optical depth gridded products over China. *Atmospheric Environment*, 278, 119088. https://doi.org/10.1016/j.atmosenv.2022.119088

Karakara, A.A. & Osabuohien, E.S. 2020. Clean versus dirty energy: empirical evidence from fuel adoption and usage by households in Ghana. *African Journal of Science, Technology, Innovation and Development*, 13(7):785-795. https://doi.org/10.1080/20421338.2020.1816266

Kapwata, T., Language, B., Piketh, S. & Wright, C.Y. 2018. Variation of indoor particulate matter concentrations and association with indoor/outdoor temperature: a case study in rural Limpopo, South Africa. *Atmosphere*, *9*(4), p.124. https://doi.org/10.3390/atmos9040124

Karimian, H., Li, Y., Chen, Y. & Wang, Z. 2023. Evaluation of different machine learning approaches and aerosol optical depth in $PM_{2.5}$ prediction. *Environmental Research*, 216, 114465. https://doi.org/10.1016/j.envres.2022.114465

Katumba, S. & Everatt, D. 2021. Urban sprawl and land cover in post-apartheid Johannesburg and the Gauteng City-Region, 1990-2018. *Environment and Urbanization ASIA*, 12(1_suppl):147S-164S. doi:10.1177/0975425321997973

Kaur, P., Stoltzfus, J. & Yellapu, V. 2018. Descriptive statistics. *International Journal of Academic Medicine*, 4(1):60-63. doi:10.4103/IJAM.IJAM_7_18

Kayes, I., Shahriar, S.A., Hasan, K., Akhter, M., Kabir, M.M. & Salam, M.A. 2019. The relationships between meteorological parameters and air pollutants in an urban environment. *Global Journal of Environmental Science and Management*, *5*(3), pp.265-278. doi:10.22034/gjesm.2019.03.01

Kaziboni, L., Mondliwa, P. & Robb, G. 2015. Towards an understanding of the economy of Johannesburg: Industrial Nodes Report. http://dx.doi.org/10.2139/ssrn.2716031

Kershaw, D., Cairncross, B., Freese, B. & De Vries, P. 2003. Secondary minerals from the Carletonville Gold Mines: Witwatersrand Goldfield, South Africa. *Rocks and Minerals*, *78*(6):390-399. http://dx.doi.org/10.1080/00357529.2003.9926753

Khan, R., Kumar, K.R., Zhao, T., Ullah, W. & de Leeuw, G. 2021. Interdecadal changes in aerosol optical depth over Pakistan based on the MERRA-2 reanalysis data during 1980–2018. *Remote Sensing*, 13(4):822-839. https://doi.org/10.3390/rs13040822

Kim, B.Y., Nakada, K., Wayson, R., Christie, S., Paling, C., Bennett, M., … Roof, C. 2015. *Understanding Airport Air Quality and Public Health Studies Related to Airports* (Vol. 135). Washington, DC: Transportation Research Board. DOI:10.13140/RG.2.1.1028.2086

Komo, A. 2013. *Towards clean energy options: a study of energy use patterns in Zamdela (Sasolburg, Free State province).* Pretoria: University of Pretoria. (Dissertation – MSc).

Krecla, P., de Limaa, C.H., Bosco, T.C.D., Targino, A.C., Hashimoto, E.M. & Oukawa, G.Y. Open waste burning causes fast and sharp changes in particulate concentrations in peripheral neighbourhoods. *Science of the Total Environment*, 765, 142736. https://doi.org/10.1016/j.scitotenv.2020.142736

Kruger, A.C., Retief, J.V. & Goliger, A.M. 2013. Strong winds in South Africa: part 2 mapping of updated statistics. *Journal of The South African Institution of Civil Engineering*, 55(2):46-58. https://hdl.handle.net/10520/EJC145418

Kyriacou, P., Budidha, K. & Abay, T.Y. 2019. Optical techniques for blood and tissue oxygenation. *Encyclopedia of Biomedical Engineering*, *3*, pp.461-472. https://doi.org/10.1016/B978-0-12-801238-3.10886-4

Lakshmi, K., Mahaboob, B., Rajaiah, M. & Narayana, C. 2021. Ordinary least squares estimation of parameters of linear model. *Journal of Mathematics and Computer Sciences*, 11:2015-2030.

Landrigan, P.J., Fuller, R., Acosta, N.J.R., Adeyi, O., Arnold, R., Basu, N., … Zhong, M. 2018. The Lancet Commission on pollution and health. *The Lancet Commissions*, 391(10119):462-512. https://doi.org/10.1016/S0140-6736(17)32345-0

Language, B. 2020. *Characterisation of respirable indoor particulate matter in South African low-income settlements*. Potchefstroom: North-West University. (Thesis - PhD).

Laban, T.L., Van Zyl, P.G., Beukes, J.P., Vakkari, V., Jaars, K., Borduas-Dedekind, N., … Laakso, L., 2018. Seasonal influences on surface ozone variability in continental South Africa and implications for air quality. *Atmospheric chemistry and physics*, *18*(20), pp.15491-15514.

Lee, A., Kinney, P., Chillrud, S. & Jack, J. 2015. A systematic review of innate immunomodulatory effects of household air pollution secondary to the burning of biomass fuels. *Annals of Global Health*, 81(3):368-374. https://doi.org/10.1016/j.aogh.2015.08.006

Lekonyane, B.C. & Disoloane, V.P.P. 2013. Determining strategies to manage informal settlements. Administratio Publica, 21(2):57-72. https://hdl.handle.net/10520/ejc-adminpub-v21-n2-a5

Local Government: Municipal Systems Act 32 of 2000.

Lourens, A.S.M. *Air quality in the Johannesburg-Pretoria megacity, its regional influence and identification of parameters that could mitigate pollution*. 2012. Potchefstroom: North-West University. (Dissertation - MSc). https://repository.nwu.ac.za/handle/10394/8760

Lozano, R., Fullman, N., Mumford, J.E., Knight, M., Barthelemy, C.M., Abbafati, C. et al. 2019. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020.

Liu, Y., Zhou, Y. & Lu, J. 2020. Exploring the relationship between air pollution and meteorological conditions in China under environmental governance. *Scientific reports*, *10*(1), p.14518. https://doi.org/10.1038/s41598-020-71338-7

Mahesh, B., Sivakumar, V., Kulkarni, P. & Sreekanth, V. 2022. Particulate air pollution in Durban: Characteristics and its relationship with 1 km resolution satellite aerosol optical depth. *Advances in Space Research*, 70:371-382.

Makokha, J.W. & Odhiambo, J.O. 2017. Trend analysis of aerosol optical depth and Angström exponent anomaly over East Africa. *Atmospheric and Climate Sciences*, 7:588-603. https://doi. org/10.4236/acs.2017.74043

Manisalidis, I., Stavropoulou, E., Stavropoulos, A. & Bezirtzoglou, E. 2020. Environmental and health impacts of air pollution: a review. *Frontiers in Public Health*, 8, 14. https://doi.org/10.3389/ fpubh.2020.00014

Maritz, P., Beukes, J.P., Van Zyl, P.G., Conradie, E.H., Liousse, C., Galy-Lacaux, C., … & and Pienaar, J.J. 2015. Spatial and temporal assessment of organic and black carbon at four sites in the interior of South Africa. *Clean Air Journal*, 25(1), 20-20. http://dx.doi.org/10.17159/2410-972X/2015/v25n1a1

Matima, C.M. 2010. Secondary stigma: a case study of people affected by HIV/Aids in white city Jabavu – Soweto. Johannesburg: University of the Witwatersrand. (Dissertation – MA).

Matyeni, A. 2021. *Study on sulphur dioxide (SO2) and Particulate Matter 10 (PM10) variations in the City of Tshwane, Gauteng*. Westville: University of KwaZulu-Natal. (Dissertation - MSc).

Mawasha, T.S. & Britz, W. 2021. Hydrological impacts of land use - land cover change on urban flood hazard: a case study of the Jukskei river in Alexandra Township, Johannesburg, South Africa. *South African Journal of Geomatics*, 10(2):139-162. doi:10.4314/sajg.v10i2.11

Mbandi, A.M. 2020. Air Pollution in Africa in the time of COVID-19: the air we breathe indoors and outdoors. *Clean Air Journal*, 30(1):1-3. http://dx.doi.org/10.17159/caj/2020/30/1.8227

McCarthy, T.S. 2011. The impact of acid mine drainage in South Africa. South African Journal of Science, 107(5-6):1-7. http://dx.doi.org/10.4102/sajs.v107i5/6.712

McGill, M.J., Swap, R.J., Yorks, J.E., Selmer, P.A. & Piketh, S.J. 2020. Observation and quantification of aerosol outflow from southern Africa using spaceborne lidar. *South African Journal of Science*, 116(3-4):1-6. http://dx.doi.org/10.17159/sajs.2020/6398

Melchiorre, A., Boschetti, L. & Roy, D.P. 2020. Global evaluation of the suitability of MODIS-Terra detected cloud cover as a proxy for Landsat 7 cloud conditions. *Remote Sensing*, *12*(2), p.202. https://doi.org/10.3390/rs12020202

Mishra, P., Pandey, C.M., Singh, U., Gupta, A., Sahu, C. & Keshri, A. 2019. Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, 22(1), 67. https://doi.org/10.4103%2Faca.ACA_157_18

Mocumbi, A.O., Stewart, S., Patel, S. & Al-Delaimy, W.K. 2019. Cardiovascular effects of indoor air pollution from solid fuel: relevance to Sub-Saharan Africa. *Current Environmental Health Reports,* 6:116-126. https://doi.org/10.1007/s40572-019-00234-8

Moeletsi, M.E. & Tongwane, M.I. 2020. Projected direct carbon dioxide emission reductions as a result of the adoption of electric vehicles in Gauteng province of South Africa. *Atmosphere*, 11(6), 591. https://doi.org/10.3390/atmos11060591

Mohammed, G., Karani, G. & Mitchell, D. Trace elemental composition in PM10 and $PM_{2.5}$ collected in Cardiff, Wales. *Energy Procedia*, 111:540-547. https://doi.org/10.1016/j.egypro. 2017. 03.216

Mohlala, K. 2020. *A critical evaluation of local government air quality management: the Gauteng experience*. Potchefstroom: North West University. (Dissertation – MSc)

Molamu, L. 1995. Wietie: The emergence and development of Tsotsitaal in South Africa. *Alternation*, 2(2):139-158. https://hdl.handle.net/10520/AJA10231757_69

Moreoane, L., Mukwevho, P. & Burger, R. 2021. The quality of the first and second Vaal Triangle Airshed Priority Area Air Quality Management Plans. *Clean Air Journal*, 31(2):1-14. http://dx.doi. org/10.17159/caj/2020/31/2.12178

Moyo, T., Mbatha, S., Aderibigbe, O.O., Gumbo, T. & Musonda, I. 2022. Assessing spatial variations of traffic congestion using traffic index data in a developing city: lessons from Johannesburg, South Africa. *Sustainability*, 14(14): p.8809. https://doi.org/10.3390/su141 48809

Mpanza, M. 2019. Dust deposition at a Gold Mine Village in the West Rand.

Mpanza, M., Adam, E. & Moola, R. 2020. Perceptions of external costs of dust fallout from gold mine tailings: West Wits Basin. *Clean Air Journal*, 30(1):1-12. https://doi.org/10.17159/caj/2020/ 30/1.7566

Mutanga, O., Dube, T. & Ahmed, F. 2016. Progress in remote sensing: vegetation monitoring in South Africa. *South African Geographical Journal*, 98(3):461-471. https://doi.org/10.1080/03736 245.2016.1208586

Muyemeki, L., Burger, R. & Piketh, S.J. 2020. Evaluating the potential of remote sensing imagery in mapping ground-level fine particulate matter ($PM_{2.5}$) for the Vaal Triangle Priority Area. *Clean Air Journal*, 30(1):1-7. http://dx.doi.org/10.17159/caj/2020/30/1.8066

Naiker, Y., Diab, R.D., Zunckel, M. & Hayes, E.T. 2012. Introduction of local Air Quality Management in South Africa: overview and challenges. *Environmental Science and Policy*, 17:62-71. https://doi.org/10.1016/j.envsci.2011.11.009

National Environment Management: Air Quality Act 39 of 2004.

Navinya, C.D., Vinoj, V. & Pandey, S.K. 2020. Evaluation of $PM_{2.5}$ surface concentrations simulated by NASA's MERRA version 2 aerosol reanalysis over India and its relation to the air quality index. *Aerosol and Air Quality Research*, 20(6):1329-1339. https://doi.org/10.4209/aaqr. 2019.12.0615

NEMA (South Africa). 2014. National Environment Act (Act 107 of 1998): implementation guidelines: sector guidelines for environmental impact assessment regulations. (Notice 654). *Government gazette*, 33333:4, 29 June.

Nealer, E.J. 2020. Geohydrological aspects of importance in the public management of basic water supply services in Merafong City Local Municipality (Carletonville area), South Africa. *TD: The Journal for Transdisciplinary Research in Southern Africa*, *16*(1):1-9. https://doi.org/10.4102/ td.v16i1.671

Nguyen, T.T., Bui, H.Q., Pham, H.V., Luu, H.V., Man, C.D., Pham, H.N., Le, H.T. & Nguyen, T.T. 2015. Particulate matter concentration mapping from MODIS satellite data: a Vietnamese case study. *Environmental Research Letters*, 10(9), p.095016. http://dx.doi.org/10.1088/1748-9326/10/9/095016

Njinga, R.L. & Tshivhase, V.M. 2017. The impact of mine tailings on the Witwatersrand and the surrounding water bodies in Gauteng Province, South Africa. *Mine Water and the Environment*, *36*:638-645. https://doi.org/10.1007/s10230-017-0469-x

Nkosi, N.C., Piketh, S.J. & Burger, R.P. 2018. Fine PM emission factors from residential burning of solid fuels using traditional cast-iron coal stoves. *Clean Air Journal*, 28(1):35-41. http://dx.doi. org/10.17159/2410-972x/2018/v28n1a10

Nkosi, V., Wichmann, J. & Voyi, K. 2017. Indoor and outdoor PM_{10} levels at schools located near mine dumps in Gauteng and North West Provinces, South Africa. *BMC public health*, 17(42):1-7. https://doi.org/10.1186/s12889-016-3950-8

Novela, R.J., Gitari, W.M., Chikoore, H., Molnar, P., Mudzielwana, R. & Wichmann, J. 2020. Chemical characterization of fine particulate matter, source apportionment and long-range transport clusters in Thohoyandou, South Africa. *Clean Air Journal*, *30*(2), pp.1-12. https://doi.org/10.17159/caj/2020/30/2.8735

Nyanda, N. 2019. *The Planning Process Around an Integrated Transport Node/Station: Indingilizi*. Johannesburg: University of Johannesburg. (Dissertation – MSc).

Obaid, A., Adam, E. & Ali, K.A. 2023. Land use and land cover change in the Vaal dam catchment, South Africa: a study based on remote sensing and time series analysis. *Geomatics*, 3(1):205-220. https://doi.org/10.3390/geomatics3010011

Olutola, B.G. & Wichmann, J. 2021. Does apparent temperature modify the effects of air pollution on respiratory disease hospital admissions in an industrial area of South Africa?. *Clean Air Journal*, *31*(2), pp.1-11. http://dx.doi.org/10.17159/caj/2021/31/2.11366

Ozdemir, B. & Kumral, M. 2019. A system-wide approach to minimize the operational cost of bench production in open-cast mining operations. *International Journal of Coal Science and Technology*, 6(1):84-94. https://doi.org/10.1007/s40789-018-0234-1

Padayachi. Y.R. 2016. *Satellite remote sensing of particulate matter and air quality assessment in the Western Cape, South Africa*. Westville: University of KwaZulu-Natal. (Dissertation – MSc).

Pandey, S.K. & Vinoj, V. 2020. Surprising Changes in Aerosol Loading over India amid COVID-19 Lockdown. *Aerosol and Air Quality Research*, 21(3), 200466. https://doi.org/10.4209/aaqr.2020.07.0466

Pérez, I.A., García, M.Á., Sánchez, M.L., Pardo, N. & Fernández-Duque, B. 2020. Key points in air pollution meteorology. *International Journal of Environmental Research and Public Health*, *17*(22), p.8349. https://doi.org/10.3390/ijerph17228349

Peng, X. 2011. China's demographic history and future challenges. *Science*, 333(6042):581-587. https://doi.org/10.1126/science.1209396

Pierce, R., Stacey, K. & Bardini, C. 2010. Linear functions: teaching strategies and students' conceptions associated with y= mx+ c. *Pedagogies: An International Journal*, *5*(3), pp.202-215. https://doi.org/10.1080/1554480X.2010.486151

Piketh, S.J., Annegarn, H.J. & Tyson, P.D. 1999. Lower tropospheric aerosol loadings over South Africa: The relative contribution of aeolian dust, industrial emissions, and biomass burning. *Journal of Geophysical Research: Atmospheres*, 104(D1):1597-1607. https://doi.org/10.1029/ 1998JD100014

Platnick, S., King, M.D., Meyer, K.G., Wind, G., Amarasinghe, N.A, Marchant, B., … Riedi, J. 2015. MODIS cloud optical properties: User guide for the Collection 6 Level-2 MOD06/MYD06 product and associated Level-3 Datasets. *Version*, *1*, p.145.

Pope, C.A., Muhlestein, J.B., Anderson, J.L., Cannon, J.B., Hales, N.M., Meredith, K.G., Le, V. & Horne, B.D. 2015. Short- term exposure to fine particulate matter air pollution is preferentially associated with the risk of ST-segment elevation acute coronary events. *Journal of the American heart association*, 4(12), e002506. https://doi.org/10.1161/JAHA.115.002506

Rabaji, O.P. 2019. *Waste dumping in Sharpeville (Emfuleni Municipality): An investigation of the characteristics and the potential impacts on air quality*. Potchefstroom: North West University. (Mini-Dissertation - MSc)

Ranjan, A.K., Patra, A.K. & Gorai, A.K. 2021. A review on estimation of particulate matter from satellite-based aerosol optical depth: data, methods, and challenges. *Asia-Pacific Journal of Atmospheric Sciences*, 57:679-699. https://doi.org/10.1007/s13143-020-00215-0

Rizza, U., Mancinelli, E., Morichetti, M., Passerini, G. & Virgili, S. 2019. Aerosol Optical Depth of the main aerosol species over Italian cities based on the NASA/MERRA-2 Model Reanalysis. *Atmosphere*, 10(11):709. https://doi.org/10.3390/atmos10110709

Robbins, L.J., Mänd, K., Planavsky, N.J., Alessi, D.S. & Konhauser K.O. 2020. *Trace Metals*. In: Gargaud M., et al. eds. Encyclopaedia of Astrobiology. Springer, Berlin: Heidelberg. pp 96-102. https://doi.org/10.1007/978-3-642-27833-4_5422-1

Rodríguez-Urrego, D. & Rodríguez-Urrego, L. 2020. Air quality during the COVID-19: $PM_{2.5}$ analyses in the 50 most polluted capital cities in the world. *Environmental Pollution*, 266, 115042. https://doi.org/10.1016/j.envpol.2020.115042

Ruppert Jr, J.H. 2016. Diurnal timescale feedbacks in the tropical cumulus regime. *Journal of Advances in Modelling Earth Systems*, *8*(3). pp.1483-1500. https://doi.org/10.1002/2016MS000713

Saucy, A., Röösli, M., Künzli, N., Tsai, M.Y., Sieber, C., Olaniyan, T., … de Hoogh, K. 2018. Land Use Regression Modelling of Outdoor NO2 and PM2.5 Concentrations in Three Low Income Areas in the Western Cape Province, South Africa. *International Journal of Environmental Research and Public Health,* 15(1452):1-14. doi:10.3390/ijerph15071452

Schober, P., Boer, C. & Schwarte, L.A. 2018. Correlation coefficients: appropriate use and interpretation. *Anaesthesia and Analgesia*, 126(5):1763-1768. https://doi.org/10.1213/ANE.0000000000002864

Schuster, G.L., Dubovik, O. & Holben, B.N. 2006. Ångström exponent and bimodal aerosol size distributions. *Journal of Geophysical Research*, 111, D7. https://doi.org/10.1029/2005JD006328

Scorgie, Y. 2012. *Urban air quality management and planning in South Africa*. Johannesburg: University of Johannesburg. (Thesis - PhD).

Seethal, C. 2023. The State of Languages in South Africa. In: Brunn, S.D., Kehrein, R. *eds*. Language, Society and the State in a Changing World. Springer, Cham. Date of access: 3 Aug 2023. https://doi.org/10.1007/978-3-031-18146-7_7

Segakweng, C.K., van Zyl, P.G., Liousse, C., Beukes, J.P., Swartz, J.S., Gardrat, E., … & Piketh, S.J., 2022. Measurement report: Size-resolved chemical characterisation of aerosols in low-income urban settlements in South Africa. *Atmospheric Chemistry and Physics*, 22(15):10291-10317. https://doi.org/10.5194/acp-22-10291-2022

Selani, L. 2017. *Mapping illegal dumping using a high-resolution remote sensing image. Case study: Soweto township in South Africa*. Johannesburg: University of the Witwatersrand. (Dissertation - MSc).

Shikwambana, L. & Sivakumar, V. 2019. Investigation of various aerosols over different locations in South Africa using satellite, model simulations and LIDAR. *Meteorological Applications*, 26(2):275-287. https://doi.org/10.1002/met.1761

Shikwambana, L. & Tsoeleng, L.T. 2020. Impacts of population growth and land use on air quality. A case study of Tshwane, Rustenburg and Emalahleni, South Africa. *South African Geographical Journal*, 102(2)209-222. https://doi.org/10.1080/03736245.2019.1670234

Shikwambana, L., Kganyago, M. & Mhangara, P. 2021. Temporal analysis of changes in anthropogenic emissions and urban heat islands during COVID-19 restrictions in Gauteng Province, *South Africa. Aerosol and Air Quality Research*, 21(9), 200437. https://doi.org/10.4209/ aaqr.200437

South African Air Quality Information System, 2010. http://saaqis.environment.gov.za/

State of Air Report 2005. A report on the state of air in South Africa. Department of Environmental Affairs. 2009.

Statistics South Africa. 2011. Census 2011 municipal report Gauteng. Pretoria.

Statistics South Africa. 2016. Community Survey 2016 in Brief. Pretoria.

Su, T., Li, J., Li, C., Lau, A.K., Yang, D. & Shen, C. 2017. An inter-comparison of AOD-converted $PM_{2.5}$ concentrations using different approaches for estimating aerosol vertical distribution. *Atmospheric Environment*, 166:531-542. https://doi.org/10.1016/j.atmosenv.2017.07.054

Subanji, S., Nusantara, T., Rahmatina, D. & Purnomo, H. 2021. The statistical creative framework in descriptive statistics activities. *International Journal of Instruction*, 14(2):591-608. https://doi.org /10.29333/iji.2021.14233a

Swart, A. 2016. *Assessment of the baseline meteorological and air quality conditions over Uubvlei, Oranjemund, Namibia*. Pretoria: University of Pretoria. (Dissertation - MSc).

Titi, A., Dweirj, M. & Tarawneh, K. 2015. Environmental effects of the open cast mining a case study: Irbid Area, North Jordan. *American Journal of Industrial and Business Management*, 5(6), 404. doi:10.4236/ajibm.2015.56041

Toledano, C., Cachorro, V.E., De Frutos, A.M., Sorribas, M., Prats, N. & De la Morena, B.A. 2007. Inventory of African desert dust events over the southwestern Iberian Peninsula in 2000–2005 with an AERONET Cimel Sun photometer. *Journal of Geophysical Research: Atmospheres*, 112(D21). doi:10.1029/2006JD008307

Tomaschek, J., Haasz, T., Dobbins, A., Fahl, U. & Annergarn, H. 2012. Energy related Greenhouse Gas Inventory and Energy Balance Gauteng: 2007--2009. Institute for Energy Economics and the Rational Use of Energy (IER), University of Stuttgart, Stuttgart, p.36.

Tongwane, M., Piketh, S., Stevens, L. & Ramotubei, T. 2015. Greenhouse gas emissions from road transport in South Africa and Lesotho between 2000 and 2009. *Transportation Research Part D: Transport and Environment*, 37:1-13. https://doi.org/10.1016/j.trd.2015.02.017

Tsay, S.C., Stephens, G.L., & Greenwald, T.J. 1991. An investigation of aerosol microstructure on visual air quality. *Atmospheric Environment*. 25(5-6):1039-1053. https://doi.org/10.1016/0960-1686(91)90146-X

Tshehla, C. & Wright. C.Y. 2019. 15 Years after the National Environmental Management Air Quality Act: is legislation failing to reduce air pollution in South Africa? *South African Journal of Science*, 115(9-10):1-4. http://dx.doi.org/10.17159/sajs.2019/6100

Tsidu, G.M. & Bililign, S. 2018. Comparison of MERRA-2 and Collection 6 Modis AODs at 0.55 micron over diverse surfaces of Southern Africa during 2000-2017. *In American Geophysical Union (AGU) Fall Meeting Abstracts*, 2018, A51J-2301.

Tucker, R.F., Viljoen, R.P. & Viljoen, M.J. 2016. A review of the Witwatersrand basin - the world's greatest goldfield. *Episodes*, 39(2):104-133. https://doi.org/10.18814/epiiugs/2016/v39i2/95771

Tyson, P.D. & Preston-Whyte, R.A. 2000. *The Weather and Climate of Southern Africa*. 2nd ed. Southern Africa: Oxford University Press.

Tyson, P.D., Garstang, M., Swap, R., Kallberg, P. and Edwards, M., 1996. An air transport climatology for subtropical southern Africa. *International journal of climatology*, *16*(3):265-291. https://doi.org/10.1002/(SICI)1097-0088(199603)16:3%3C265::AID-JOC8%3E3.0.CO;2-M

Use of Official Languages Act. 2017. Legal aid South Africa: language policy. (Notice 244). *Government Gazette*, 40733:130, 31 Mar.

Vaduganathan, M., Mensah, G.A., Turco, J.V., Fuster, V. & Roth, G.A. 2022. The global burden of cardiovascular diseases and risk: a compass for future health. *Journal of the American College of Cardiology*, 80(25):2361-2371. https://doi.org/10.1016/j.jacc.2022.11.005

Van Donkelaar, A., Martin, R.V. & Park, R.J. 2006. Estimating ground-level $PM_{2.5}$ using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research*, 111, D21. https://doi.org/10.1029/2005JD006996

Verma, S., Prakash, D., Soni, M. & Ram, K. 2019. Atmospheric aerosols monitoring: ground and satellite-based instruments. *Advances in Environmental Monitoring and Assessment*; IntechOpen: London, 67-80. http://dx.doi.org /10.5772/intechopen.80489

Wang, J. & Christopher, S.A. 2003. Intercomparison between satellite derived aerosol optical thickness and $PM_{2.5}$ mass: implications for air quality studies. *Geophysical Research Letter*, 30(21). https://doi.org/10.1029/2003GL018174

Wang, J. & Ogawa, S. 2015. Effects of meteorological conditions on $PM_{2.5}$ concentrations in Nagasaki, Japan. *International journal of environmental research and public health*, *12*(8), pp.9089-9101. doi: 10.3390/ijerph120809089

Wei, X., Chang, N., Bai, K. & Gao, W. 2020. Satellite remote sensing of aerosol optical depth: advances, challenges, and perspectives. *Critical Reviews in Environmental Science and Technology*, 50(16):1640-1725. https://doi.org/10.1080/10643389.2019.1665944

Wernecke, B. 2018. *Ambient and indoor particulate matter concentrations on the Mpumalanga Highveld*. Potchefstroom: North-West University. (Dissertation – MSc).

West rand district municipality air quality management plan, 2010.

Wójtowicz, M., Wójtowicz, A. & Piekarczyk, J. 2016. Application of remote sensing methods in agriculture. *Communications in Biometry and Crop Science*, 11(1) 31-50. http://agrobiol.sggw.waw.pl/CBCS

Wray, C. & Cheruiyot, K. 2015. Key challenges and potential urban modelling opportunities in South Africa, with specific reference to the Gauteng city-region. *South African Journal of Geomatics*, 4(1):14-35. doi:10.4314/sajg.v4i1.2

Wright, C., Oosthuizen, M.A., Mostert, J. & Van Niekerk, L. 2011. Investigating air quality and air-related complaints in the City of Tshwane, South Africa. *Clean Air Journal*, 20(2):3-12. https://hdl.handle.net/10520/EJC119371

Xiao, Q., Chang, H.H., Geng, G. & Liu, Y. 2018. An ensemble machine-learning model to predict historical $PM_{2.5}$ concentrations in China from satellite data. *Environmental Science and Technology*, 52:13260-13269. https://doi.org/10.1021/acs.est.8b02917

Xulu, N.A., Piketh, S.J., Feig, G.T., Lack, D.A. & Garland, R.M. 2020. Characterizing light-absorbing aerosols in a low-income settlement in South Africa. *Aerosol and Air Quality Research*, 20:1812-1832. http://hdl.handle.net/10204/11807

Yakubu, A.T. & Chetty, N. 2022. A decadal assessment of the climatology of aerosol and cloud properties over South Africa. *Atmospheric Chemistry and Physics*, 22(17):11065-11087. https://doi.org/10.5194/acp-2021-796

Yang, J. & Hu, M. 2018. Filling the missing data gaps of daily MODIS AOD using spatiotemporal interpolation. *Science of the Total Environment*, 633:677-683. https://doi.org/10.1016/j.scitotenv. 2018.03.202

Yin, P.Y., Chang, R.I., Day, R.F., Lin, Y.C. & Chang, C.Y. 2022. Improving $PM_{2.5}$ concentration forecast with the identification of temperature inversion. *Applied Science*, 12(1):71-88. https://doi.org/10.3390/ app12010071

Yue, Q., Jiang, J.H., Heymsfield, A., Liou, K.N., Gu, Y. & Sinha, A. 2020. Combining in situ and satellite observations to understand the vertical structure of tropical anvil cloud microphysical properties during the TC4 experiment. *Earth and Space Science*, *7*(4), p.e2020EA001147. doi: 10.1029/2020EA001147

Zhang, D., Du, L., Wang, W., Zhu, Q., Bi, J., Scovronick, N., … Liu, Y. 2021. A machine learning model to estimate ambient $PM_{2.5}$ concentrations in industrialized highveld region of South Africa. *Remote Sensing of Environment*, 266, 112713. https://doi.org/10.1016/j.rse.2021.112713

Zhang, X., Wang, Q., Qin, W. & Guo, L. 2019. Sustainable policy evaluation of vehicle exhaust control - empirical data from China's air pollution control. *Sustainability*, 12(1), 125. https://doi.org/ 10.3390/su12010125

Zhang, Y., Li, Z., Bai, K., Wei, Y., Xie, Y., Zhang, Y., …Li, D. 2021. Satellite remote sensing of atmospheric particulate matter mass concentration: advances, challenges, and perspectives. *Fundamental Research*, 1:240-258. https://doi.org/10.1016/j.fmre.2021.04.007

Zhang, Y.L. & Cao, F. 2015. Fine particulate matter ($PM_{2.5}$) in China at a city level. *Scientific Reports*, 5(1):1-12.

Zhao, X., Zhang, X., Xu, X., Xu, J., Meng, W. & Pu, W. 2009. Seasonal and diurnal variations of ambient PM2. 5 concentration in urban and rural environments in Beijing. *Atmospheric Environment*, *43*(18): pp.2893-2900. https://doi.org/10.1016/j.atmosenv.2009.03.009

Zhu, L., Suomalainen, J., Liu, J., Hyyppä, J., Kaartinen, H. & Haggren, H. 2018. A review: remote sensing sensors. *Multi-purposeful application of geospatial data*, 19-42. https://doi.org/10.5772/intechopen