# AMBIENT AIR POLLUTION MONITORING, MODELLING AND HUMAN HEALTH RISKS IN THULAMELA MUNICIPALITY, VHEMBE DISTRICT OF LIMPOPO PROVINCE, SOUTH AFRICA


**BY**

**ENITAN IBIRONKE TITILAYO**

**STUDENT NO: 18021500**


**A THESIS SUBMITTED TO THE DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL SCIENCES, FACULTY OF SCIENCE, ENGINEERING AND AGRICULTURE, UNIVERSITY OF VENDA, IN FULFILMENT OF THE REQUIREMENTS FOR A DOCTORAL DEGREE IN ENVIRONMENTAL SCIENCES**


**SUPERVISOR: PROF. EDOKPAYI J.N.**

**CO-SUPERVISOR: PROF. PIKETH S.J.**


**AUGUST 2025**


© University of Venda

# DECLARATION

I, Ibironke Titilayo Enitan, student number 18021500, hereby declare that this thesis titled - **AMBIENT AIR POLLUTION MONITORING, MODELLING AND HUMAN HEALTH RISKS IN THULAMELA MUNICIPALITY, VHEMBE DISTRICT OF LIMPOPO PROVINCE, SOUTH AFRICA -** submitted to the Department of Geography and Environmental Science, Faculty of Science, Engineering and Agriculture, University of Venda, for the PhD in Environmental Sciences, is my work and has not been previously submitted, in whole or in part, to any university for any degree; and all reference materials contained herein have been duly acknowledged.



20/08/2025

Student's signature

Date

**IT Enitan**



© University of Venda

# DEDICATION

This thesis is dedicated to the Jehovah God, the source of all wisdom, inspiration, and strength. His guidance has been my steadfast anchor, and His grace has provided the endurance needed. To God be the glory.



© University of Venda

# ACKNOWLEDGEMENTS

First and foremost, I extend my deepest gratitude to my Father and Friend, Jehovah God, and His Son, Jesus Christ, my Saviour, for granting me the knowledge, wisdom, strength, and perseverance to complete this Ph.D. program. Through Jehovah's unfailing love and support, I have achieved this significant milestone.

I want to thank my Promoter, Prof. J.N. Edokpayi, and Co-Promoter, Prof. S.J. Piketh, for their immense contribution to this study. Their exceptional guidance and mentorship have been invaluable, impacting my personal and professional growth. Their support and encouragement during challenging times motivated me to persevere - qualities of true academic leadership I will forever cherish. I am grateful for their understanding, patience, and confidence in my abilities. Their reliable guidance has been crucial in my pursuit of the doctoral degree. My sincere gratitude goes to the late Prof. J. Odiyo, whose belief in my potential provided me with the confidence to conduct research under his expert supervision.

My heartfelt appreciation goes to my guardians, Deacon S. & Deaconess B. Ipadeola, for their unwavering love, support, prayers, and encouragement. Their presence has been a source of strength and inspiration. I extend gratitude to my siblings, Engr & Mrs Temitayo Enitan, Dr Abimbola & Mr Ola-Oluwa Folami, for their confidence and steadfast love. My thanks go to Dr & Mrs Olusola Olubiyi, Ipadeola's family, Mr and Mrs Tosin Odeyale, Mr G. Akinola, Mr and Mrs O. Enikanoselu, Mrs A. Somoye, Ms Oluwaseun Ipadeola, Barrister and Mrs Adeyemi, Commander Dosunmu, Mr T. Omotayo, Mr and Mrs F. Oderinde, Mr B. Osunla, Mrs L. Oyovota, the Olubiyi's, Edobor's, Agboro's, Olatunji's, Adekanmi's and Adebayo's family, Pastors: Otolorin, Dipeolu, Akingbolu, uncles, aunties, cousins and family members for their prayers, encouragement, and support.

The financial and academic support from the Eskom Power Plant Engineering Institute (EPPEI), North-West University, and the University of Venda Research and Publications Committee (RPC) is gratefully acknowledged. My special thanks go to Mr Rickhein and the technician team, including Dr B. Language at North-West University, Potchefstroom, for their invaluable assistance in modifying the Dylos equipment, and to the technicians who installed and later removed the equipment. This research journey benefited greatly from the generous support and confidence of Mrs Gabi Mkhatshwa, Dr Makungo, Dr Le Grange, and other contributors. Several departments at the University of Venda, including Research and Innovation, Asset Management, Maintenance, Security, Finance, Student Administration, International Office, and the Disability Unit, provided



© University of Venda

essential administrative support throughout this academic process.

Special thanks to the families, homes, and offices where the Dylos DC1700 LCS equipment was installed for $PM_{2.5}$ monitoring in Thulamela Municipality. This includes the following individuals and organisations, though not limited to - Mr and Mrs V. Khorommbi, Mr and Mrs M. Luvhengo, Mrs M. Nemutshada, Mr and Mrs J. Siboiboi, Mr and Mrs V. Tshivhase, Mr and Mr C. Muyengwa, Mrs. V. Nembulu, Mrs M. Munyai, Mrs P. Matshotshi, Mr and Mrs E. Mphephu, Mrs B. Marima, Mr Maluma, Mr and Mrs Ligege, Mr. B. Matsena, Manager: M. Single, D. Grobler, and R. Botha. I also thank Khoroni Hotel Casino and Convention Resort, Thulamela Library, and Vhavenda Brick (Shayandima and Lwamondo offices). Without your support, this work would not have been possible. My appreciation also extends to individuals and institutions that provided invaluable data for this research. These include Manager Shandukani Nemauluma (Vhembe District Statistics South Africa), South African Weather Stations (SAWS), NATIS, drivers from the University of Venda, Mvusoluzo Mall, Thulamela, and Venda taxi ranks, Shayandima industrial areas, and private car owners who contributed to the emission activities data collection. I am grateful to the 30 households in Lunungwi and Gondeni villages who participated in the wood measurement study, and to Mrs. M. Makhesha and the Musandas - Mr. G. Mikosi (Gondeni) and Mr. M. Mamatho (Lunungwi) - for their cooperation.

Thanks to Dr Akintola for advice and proofreading this thesis. Appreciation goes to friends, colleagues and mentors for guidance and support during my study period: Mr and Mrs Ojelade, Dr and Mrs Durowoju, Mrs F. Ikotun, Dr Ekanade, Prof. J. Adeyemo, Ms R. Bassey, Mr O.Ogunfowote, Mr M. Augustine, Mr and Mrs Emmanuel, Mrs P. Akintola, Mr O. Bada, Mr B. Ajibola, Mr J. Adekeye, Dr Tambe, Dr Oyebamiji, Mrs K. Ramabubuda, Dr Unahumare, Dr Volenzo, Prof. B. Odhiambo, Miss Novela, Prof. Gitari, Dr E. Umejuru, Mr Ologundudu, Dr Adesina, Dr K. Amaechi, Dr Iwara, Mrs Sathekge, Madam Teresa, Dr Obijole, Prof. Adeyemo, Mrs Adebayo, Dr O. Bassey, Dr C. Edokpayi, Dr Olamiti, Mr A. Adeniran, Dr Y. Jimoh, Mr O. Ajao, Mrs P. Dickson, Prof Chikoore Dr. Oyebanjo, Dr Ayinde, and Dr Adeeyo. My gratitude extends to Mr Oladunjoye, Mrs I. Omowunmi, Miss F. Akpene, Mr A. Olanrewaju, Mr Oladokun, Miss Fungi, Mrs M. Etineh, Mr and Mrs M. Dry, Mrs Maringana, Mr S. Akinbola, Mr B. Ekanem, Mr Ogungbe, Mr I. Akinduyile, Mr A. Adeboye, Mr B. Awonusi, Mr H. Benjamen, Mr R. Ekoja, Mr J. Momoh, Miss D. Oni, Mrs Balqees, Mr Asiedu, Ms K. Ajayi, Mr Agunleti, Mrs W.Idris, Ms K. Kuye, Mrs K. Adele, Mr S. Onifade, Dr Adeyeye, Mr A. Adewunmi, Mr Morakinyo, Mr Hamidu, Dr Dondofema, Dr Oke, Dr Folaranmi, Mr Mathibela (KK), the Akola, Obisanya, Yehwenu, Toviho, and Lusunzi family.



© University of Venda

I also extend my thanks to all research assistants who selflessly supported me during data collection, including V. Sinthumule (Donald), N. Sinthumule, Z. Khorommbi, P. Ratshilumela, Madidi, P. Siboiboi, M. Mareya, F. Shana, M. Mulidzi, R. Mulidzi, P. Rabele, R. Makhado, L. Nngobeli, and P. Mikosi. My appreciation goes to those who contributed to the research analysis, including Prof. Roelof, Dr L. Munyai, P. Chidhindi, Dr M. Belelie, Dr Akinshipe, Pharm Ifedayo, Dr Morakinyo, Dr S. Sooktawee, Dr G. Pindihama, Dr Olawoyin, Mr D. Gumbo, Dr F. Moffat, Prof J. Adeniran, Dr Y. Aliyu, and Prof. Ikudayisi. I acknowledge the Air Quality Research Team at the University of Venda, including Mr K. Manyatsha, Dr P. Njoku, Dr R. Adeeyo, Dr T. Aniyikaiye, and P. Mudau, for their collaboration and insights.

My special thanks to the Murovhi, Muyengwa, Siboiboi, Simwanza, Mukansi, Munzhelele, Dowelani, Mmbengwa, Ngobeni, Ramulifo, Shana, Ndlovu, and Zhuwawo families, to name a few of my brothers and sisters from the Thohoyandou, Shayandima, Louis Trichardt, Durban and Jeffreys Bay congregations of Jehovah's Witnesses. Your sincere love, encouragement, hospitality, and spiritual support created a homely and nurturing environment for me throughout this journey. My heartfelt appreciation goes to Adebayo, Afolabi, Oke, Arowosegbe, Soyelu, Elegbede, Osadare, Omoroga, Ilori, Bamgbose, and Adabonyan families for believing in me.

I am grateful to the medical doctors, healthcare professionals, and nurses who provided me with essential healthcare support throughout this journey. This includes Dr. Khosa, Dr. N. Murovhi, Dr. M. Rudzani, Dr. Mogashou, Dr. Adeyinka, Dr. Magashule, Dr. Edet, as well as specialists, including neurosurgeons, epileptologists, psychiatrists, ophthalmologists, psychologists, endocrinologists, and pharmacists across the South African provinces.

Special thanks to my dear friend and mother, Mrs. B. Ipadeola, and Dr. Folami for their support, prayers, encouragement, and understanding. I am grateful to Dr. Folami, my beloved sister, for her guidance in the machine learning aspect of this thesis, her dedication to editorial and proofreading activities, and her boundless love and support throughout the submission process.

This PhD journey has been one of immense learning, perseverance, and growth. I am deeply thankful to everyone who played a role academically, professionally, financially, emotionally, and spiritually in making this achievement possible.

**Ndo livhuwa vhone vhothe (Thank you all)**



© University of Venda

## ABSTRACT

Ambient air pollution is a pressing global environmental concern, with pollutants from natural and anthropogenic sources contributing to deteriorating air quality. This research investigates the impact of sulfur dioxide ($SO_2$), nitrogen oxides ($NO_x$), carbon monoxide (CO), and particulate matter ($PM_{2.5}$ and $PM_{10}$) on human health and the environment. $PM_{2.5}$ in particular is a significant public health risk linked to respiratory and cardiovascular diseases. The study focuses on Thulamela Local Municipality (LM) in Limpopo Province, South Africa, a region with diverse semi-urban and rural environments, where air quality data is limited despite ongoing urbanisation and economic growth. Given the lack of comprehensive baseline data, this study aims to quantify, monitor, and model air pollution while assessing its potential health impacts. This research integrates the development of an emission inventory for criteria pollutants ($SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$), real-time $PM_{2.5}$ monitoring using Dylos DC1700 low-cost sensors (LCSs), $PM_{2.5}$ dispersion modelling with AERMOD, health risk assessment (HHRA), and machine learning-based predictive modelling. By evaluating pollution sources, exposure risks, and predictive trends, the study provides critical insights to inform air quality management strategies in semi-urban settings.

This study developed the first-ever 10-year vehicle emission inventory (VEI) for criteria pollutants ($SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$) in Thulamela Local Municipality (LM), South Africa. In addition, it produced the first comprehensive bottom-up emission inventory for the year 2021, quantifying total $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ emissions from two major anthropogenic sources: vehicular emissions and residential wood combustion (RWC). The results indicated annual emissions of 66.3, 1821.7, 21002.6, 2348.7, and 2649.3 tonnes, for $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$, respectively. RWC contributed over 96% of $PM_{2.5}$ and $PM_{10}$ emissions, while $NO_x$ (84%) and $SO_2$ (60%) originated primarily from vehicular sources. The spatial analysis revealed that emissions were highest in the southwest and northwest, where population density and human activities were most concentrated. A one-year $PM_{2.5}$ monitoring campaign (March 2021 – February 2022) was conducted at 14 sites, covering urban, suburban, and rural areas. The strong correlation ($R^2$ = 0.75) between Dylos and reference monitors (GRIMM EDM180) confirmed the reliability of LCS performance for $PM_{2.5}$ measurements. The study found that annual $PM_{2.5}$ concentrations exceeded both WHO guidelines and SA NAAQS, except at one site. Monthly trends showed higher $PM_{2.5}$ levels from May to September, with diurnal peaks from 6–8 am and 6–8 pm, primarily influenced by traffic emissions and residential wood combustion. The bivariate polar plot (BVP) analysis identified pollution hotspots near traffic intersections, residential areas, and commercial zones. The study identified significant spatial and temporal variations in $PM_{2.5}$ levels across different monitoring sites, with higher



© University of Venda

concentrations in suburban areas than rural ones due to vehicular emissions and wood-burning. VTV2 recorded the highest annual concentration (32.3 µg/m³), while RL4 had the lowest (19.1 µg/m³). The average seasonal $PM_{2.5}$ levels, despite differences, consistently exceeded air quality guidelines, except in summer for specific locations. The results also showed that evening $PM_{2.5}$ concentrations were generally higher than morning peaks, highlighting significant differences in pollution levels across the area.

Using AERMOD, $PM_{2.5}$ dispersion was modelled to assess spatial patterns and regulatory compliance. Results showed that pollution hotspots exceeded SA NAAQS (40 µg/m³ daily, 20 µg/m³ annually) and WHO standards (15 µg/m³ daily, 5 µg/m³ annually), with the highest daily $PM_{2.5}$ concentration (49.2 µg/m³) in densely populated areas. Model validation comparing modelled and monitored $PM_{2.5}$ showed systematic underestimation ($R^2 = 0.176$, negative fractional bias), indicating unaccounted sources such as biomass burning, unpaved roads, industrial activities, and agricultural emissions. At Lunungwi village (RL receptor group), the modelled hourly $PM_{2.5}$ concentration (47.12 µg/m³) exceeded the observed value (32.27 µg/m³, FAC2 = 0.25), indicating localised overestimation. Daily peaks aligned with morning and evening rush hours. At the same time, monthly trends from May to September reflected seasonal variations and increased residential wood combustion, while a notable decline in July 2021 was due to COVID-19 lockdown restrictions, highlighting the impact of reduced human activity on air quality. Although AERMOD has limitations in accounting for some emission sources, it effectively identified pollution hotspots, offering valuable insights for policy interventions and air quality management.

The study further assessed long-term exposure and health risks using the HHRA model, focusing on risk characterisation. The USEPA health risk assessment revealed that $PM_{2.5}$ exposure posed both non-carcinogenic and carcinogenic risks, with infants and toddlers being the most vulnerable. Hazard Quotient (HQ) values exceeded safe limits (1.10 in winter) due to high $PM_{2.5}$ levels in colder months. Excess Lifetime Cancer Risk (ELCR) values surpassed the acceptable threshold ($10^{-4}$) for all age groups, with adults facing the highest long-term cancer risks ($2.35 \times 10^{-1}$). These findings emphasise the urgent need for air pollution control measures, particularly in areas with high residential fuel combustion and vehicular emissions.

To enhance $PM_{2.5}$ predictions, machine learning models were applied using meteorological variables. Random Forest (RF) outperformed Multiple Linear Regression (MLR) and Polynomial Ridge Regression (PRR), with the highest $R^2$ (0.80) at VP4 for site-specific predictions and 0.604



© University of Venda

for area-wide estimates. Seasonal performance showed better RF accuracy in winter ($R^2$ = 0.58, RMSE = 2.10 µg/m³) than in summer ($R^2$ = 0.45, RMSE = 4.72 µg/m³), reflecting the impact of seasonal meteorology on $PM_{2.5}$ dispersion.

In conclusion, this research provides a holistic approach to quantifying air pollution sources, exposure risks, and predictive modelling in Thulamela Municipality. The study highlights urgent public health risks and the need for stronger air quality regulations, improved monitoring networks, and targeted mitigation strategies by establishing a clear link between air quality and human health. The findings underscore the need to expand emission inventories, refine meteorological inputs, and integrate multi-pollutant assessments to develop a comprehensive air quality management strategy. To address identified challenges, the study proposes mitigation strategies, including eco-friendly phytoremediation techniques, promotion of cleaner cooking and transport technologies, and urban planning measures to minimise exposure in densely populated areas. By combining scientific research with practical applications, this work aims to support sustainable air quality management, improve public health, and contribute to a healthier environment.

**Keywords:** Ambient air pollution, $PM_{2.5,}$ Emission inventory, Criteria air pollutants, Monitoring, Dispersion Modelling, Health Risk Assessment, Machine Learning, Thulamela Municipality, South Africa.



© University of Venda

## RESEARCH OUTPUTS

The following lists are research outputs originating from this PhD thesis, contributing to the scientific community and expanding the body of knowledge:

### A. Peer-reviewed journal article(s)

1. Enitan, I.T., Durowoju, O.S., Edokpayi J.N. and Odiyo J.O. (2022). A Review of Air Pollution Mitigation Approach Using Air Pollution Tolerance Index (APTI) and Anticipated Performance Index (API). *Atmosphere,* 13(3), pp.374.

**2. Ibironke T. Enitan**, Stuart J. Piketh, and Joshua N. Edokpayi. Development of Vehicular Emission Inventory of Criteria Air Pollutants for Sustainable Air Quality Management in the Thulamela Municipality, Vhembe District of South Africa. Target journal: Environmental Science; Advance (Submitted and under review).

3. **Ibironke T. Enitan**, Stuart J. Piketh, and Joshua N. Edokpayi. Assessing fuel wood consumption pattern and emission rate in the residential sector of Gondeni and Lunungwi villages in Thulamela Municipality of South Africa. Target journal: Environmental Pollution (75% complete in preparation).

4. **Ibironke T. Enitan**, Stuart J. Piketh, and Joshua N. Edokpayi. Quantifying the Contribution of Road Traffic to Ambient $PM_{2.5}$ in Thohoyandou and its Vicinity, Limpopo Province, South Africa. Target journal: Frontiers in Atmospheric Environmental Pollution (50% complete in preparation).

5. **Enitan, I. T**. *et al*. Machine Learning-Based Prediction of Traffic-Related Ambient $PM_{2.5}$ Concentrations in Suburban Areas of Thulamela LM in Limpopo Province, South Africa. Target journal: Atmospheric Environment (In preparation).

6. **Enitan, I. T.** *et al*. Predicting Ambient $PM_{2.5}$ Concentrations from Residential Wood Combustion in Rural Area of Limpopo Province, South Africa: Machine Learning Approach. Target journal: Environmental Pollution (In preparation).

### B. Conference Proceedings and Poster Presentations

1. **Enitan I.T**., Piketh S.J., and Edokpayi J.N. (2023). Contribution of Road Traffic to Ambient $PM_{2.5}$ in Thohoyandou and its Environment, Limpopo Province, South Africa. 2023 South African Young Academy of Science (SAYAS), Annual Young Scientist Conference, **Oral presentation** 12- 13 June 2023, Thohoyandou, University of Venda, South Africa. pp. 15.

**2. Ibironke T. Enitan**, Stuart J. Piketh, Tolulope E. Aniyikaiye, and Joshua N. Edokpayi. (2022). Spatiotemporal Variation of Ambient $PM_{2.5}$ Concentration in Thulamela Municipality, Limpopo Province, South Africa. 2022 Annual Conference of the National Association for Clean Air (NACA), 5 – 7 October 2022, Premier Hotel O.R. Tambo, Kempton Park, Gauteng, South Africa. pp. 35 (**Poster**).



© University of Venda

3. **Ibironke T. Enitan**, Stuart J. Piketh, and Joshua N. Edokpayi (2022). Temporal variability of ambient $PM_{2.5}$ in low-income settlements in South Africa. 2022 Annual Conference of the National Association for Clean Air (NACA), 5 – 7 October 2022, Premier Hotel O.R. Tambo, Kempton Park, Gauteng, South Africa. **Three-minute talk (3MT) Oral presentation**.

4. Tolulope E. Aniyikaiye, Stuart J. Piketh, **Ibironke T. Enitan**, and Joshua N. Edokpayi (2022). Assessment of Ambient Air Quality around University Residences in Thohoyandou, South Africa. 2022 Annual Conference of the National Association for Clean Air (NACA), 5 – 7 October 2022, Premier Hotel O.R. Tambo, Kempton Park, Gauteng, South Africa.

5. **Enitan I.T**., Edokpayi J.N., Piketh S.J., and Gabi Mkhatshwa (2021). An empirical analysis of residential fuelwood consumption rate and its pattern from villages in Thulamela municipality, South Africa. EPPEI Eight Student Workshop – A **Virtual** Event 11 & 12 November 2021, South Africa. pp. 70-71.

6. **Enitan I.T**., Odiyo J.O., Piketh S.J., and Edokpayi J.N. (2021). Vehicle Emission Inventory for Thulamela Municipality, Limpopo province, South Africa. 2021 Annual Conference Proceedings of the National Association for Clean Air, Virtual, 6 – 8 October 2021, South Africa. pp. 31-36.

7. **Enitan I.T**., Odiyo J.O., Piketh S.J., and Edokpayi J.N. (2021). Air Pollution Tolerance Index Level of Plants (APTI). 2021 Annual Conference Proceedings of the National Association for Clean Air, virtual, 6 – 8 October 2021, South Africa. Three-minute talk (3MT) Oral presentation.

8. **Enitan I.T**., Edokpayi J.N., Odiyo J.O., and Gabi Mkhatshwa (2020). Estimation of Vehicular Emission Inventory (VEI) within Thulamela municipality, Limpopo province. A case study of Thulamela Municipality, Vhembe District, South Africa. Seventh EPPEI Student Workshop – A Virtual Event, 12 & 13 November 2020, South Africa. pp. 57-58.

**9. Enitan I.T.,** Edokpayi J.N., Odiyo J.O., and Enitan A.M. (2018). Reduction of Vehicular Pollutants Using Phytoremediation Method: A Review. Proceedings of the first ICSMNR: 15-17 October 2018, Bolivia Lodge, Polokwane, South Africa, pp 247-253.



© University of Venda

University of Venda
Creating Future Leaders

# TABLE OF CONTENTS

*DECLARATION* *ii*
*DEDICATION* *iii*
*ACKNOWLEDGEMENTS* *iv*
*ABSTRACT* *vii*
*RESEARCH OUTPUTS* *x*
**A. Peer-reviewed journal article(s)** **x**
**B. Conference Proceedings and Poster Presentations** **x**
*TABLE OF CONTENTS* *xii*
*LIST OF FIGURES* *xvii*
*LIST OF TABLES* *xxii*
*CHAPTER ONE: INTRODUCTION* *i*
**1.1 Preamble** **1**
**1.2 Background to the Study** **1**
**1.3 Statement of Problem** **5**
**1.4 Motivation** **7**
**1.5 Aim and Objectives of Study** **9**
1.5.1 Aim of the study 9
1.5.2 Specific Objectives 9
**1.6 Research Hypothesis** **10**
**1.7 Description of the Study Area** **10**
**1.8 Thesis Outline and Research Methodology Framework** **11**
**References** **12**
*CHAPTER TWO: LITERATURE REVIEW* **19**
**2.1 Preamble** **19**
**2.2 Air Pollution** **19**
2.2.1 Overview and Health Impacts of Air Pollution 19
2.2.2 Major Air Pollutants and Sources 19
2.2.3 Global Air Pollution Trends and Monitoring 20
2.2.4 Policy and Technological Interventions 29
**2.3 Air Pollutants** **30**
2.3.1 Particulate matter (PM) 30
2.3.2 Carbon monoxide (CO) 35
2.3.3 Sulphur dioxide ($SO_2$) 36
2.3.4 Nitrogen dioxide ($NO_2$) 37
2.3.5 Ground-level Ozone (GLO) 37
2.3.6 Lead (Pb) 38
**2.4 Drivers of Pollution and Emission by Sectors** **39**
2.4.1 Residential Sector 40
2.4.2 Vehicle Emission 43
2.4.3 Industrial Emission 45
**2.5 Air Quality Management in South Africa** **48**
**2.6 Instruments for Monitoring Particulate Matter (PM) Pollutants.** **50**
2.6.1 GRIMM EDM180 Model 52



© University of Venda

University of Venda
Creating Future Leaders

2.6.2 BAM 1020 Monitoring System 52
2.6.3 DustTrak II 8530 Monitoring Instrument 53
2.6.4 Sidepak AM 510 Monitoring Instrument 53
2.6.5 DataRAM 4 (Model DR-4000) Monitoring Instrument 53
2.6.6 Dylos DC 1700 Sensor 54
**2.7 Dispersion Models for the estimation of pollutants 55**
2.7.1 Gaussian Plume Models 56
2.7.2 Gaussian Puff Models 59
2.7.3 Meteorological Data 63
**2.8 Air Pollution Health Risk Assessment (APHRA) 64**
2.8.1 Human Health Risk Assessment (HHRA) Model 69
**2.9 Machine Learning 70**
2.9.1 Types of Machine Learning 71
**2.10 Applications and Real-World Impact of Machine Learning in Predictive Modeling 75**
2.10.1 Prediction of Future Air Quality Using Machine Learning Models 77
**References 79**
***CHAPTER THREE 95***
***QUANTIFICATION AND INVENTORY OF EMISSIONS FROM ANTHROPOGENIC SOURCES: GASEOUS ($SO_2$, $NO_x$, CO) AND PARTICULATE MATTER ($PM_{2.5}$, $PM_{10}$) IN THULAMELA MUNICIPALITY OF VHEMBE DISTRICT, SOUTH AFRICA 95***
**Abstract 95**
**3.1 Introduction 95**
3.1.1 Study Domain 97
**3.2 Materials and Methods 98**
3.2.1 Sampling Sites and Data Collection 98
3.2.2 Development and Estimation of Sector-Based Emission Inventory 101
3.2.3 Ethical Considerations 106
3.2.4 Quality Assurance and Control (QA/QC) 106
**3.3 Results and Discussion 107**
3.3.1 Vehicle Sector 107
3.3.2 Residential Sector 121
**3.4 Source contributions to total annual emissions 140**
**3.5 Limitations of the study 140**
**3.6 Conclusion 141**
**References 142**
***CHAPTER FOUR 150***
***EVALUATION OF DYLOS DC1700 LOW-COST SENSORS (LCS) FOR AMBIENT $PM_{2.5}$ AIR QUALITY MONITORING IN THE THULAMELA MUNICIPALITY, VHEMBE DISTRICT OF SOUTH AFRICA 150***
**Abstract 150**
**4.1 Introduction 150**
**4.2 Materials and Methods 153**
4.2.1 Sampling Site 153
4.2.2 Experimental and Reference Instruments 155



© University of Venda

4.2.3 Modification of Dylos DC1700 to Measure Ambient Outdoor 155
4.2.4 Calibration procedure 156
4.2.5 Data Collection 158
4.2.5.1 Data logging 158
4.2.6 Quality Control of Data 159
4.2.7 Conversion of PNC to PMC and Statistical Analysis 160
**4.3 Results and Discussion 161**
4.3.1 Assessment of Dylos DC1700 LCS for Measuring Ambient $PM_{2.5}$ level 161
4.3.2 Spatial Distribution of Annual Ambient $PM_{2.5}$ Concentrations in Thulamela Local Municipality (LM) 165
4.3.3 Seasonal Variability of $PM_{2.5}$ Concentrations 168
4.3.4 Monthly Pattern of $PM_{2.5}$ Pollutant in the Study Domain 169
4.3.5 Diurnal and Weekdays Pattern of $PM_{2.5}$ Pollutant Concentration 171
4.3.6. Pollutants Exceedances 173
4.3.7. Assessment of Inter-Unit Variability of Dylos DC1700 Sensor 174
4.3.8 Identification of Potential Sources of Ambient $PM_{2.5}$ Air Pollutants using Bivariate Polar Plot (BVP) Analysis in the Study Area 175
**4.4 Limitations of the Study 179**
**4.5 Conclusion 180**
**References 181**
***CHAPTER FIVE 189***
***ATMOSPHERIC DISPERSION MODELLING OF PARTICULATE MATTER ($PM_{2.5}$) USING AERMOD IN THULAMELA MUNICIPALITY, LIMPOPO PROVINCE, SOUTH AFRICA 189***
**Abstract 189**
**5.1 Introduction 190**
**5.2: Materials and Methods 192**
5.2.1 Quantification of $PM_{2.5}$ Emissions in the study domain 192
5.2.2 AERMOD Modelling software, components and description 193
5.2.3 Data Collection and Preparation 195
5.2.4 AERMOD Model Configuration and Procedures 195
5.2.5 Model Validation 198
5.2.6. Data Analysis and Visualization 198
5.2.7. Ethical Considerations 199
**5.3 Results and Discussions 199**
5.3.1 Emissions estimations for the identified sources 199
5.3.2. Spatial Distribution of Ambient $PM_{2.5}$ in Thulamela LM 201
5.3.3 Ambient $PM_{2.5}$ Dispersion Modeling Results Validation 213
**5.4 Limitations of the Study 216**
**5.5 Conclusion 217**
**References 218**
***CHAPTER SIX 221***
***POTENTIAL HUMAN HEALTH RISKS OF AMBIENT $PM_{2.5}$ EXPOSURE VIA INHALATION PATHWAY: A CASE STUDY OF THULAMELA MUNICIPALITY, LIMPOPO PROVINCE, SOUTH AFRICA 221***



© University of Venda

University of Venda
Creating Future Leaders

**Abstract** **221**
**6.1 Introduction** **222**
**6.2 Methodology** **225**
6.2.1 Study Area And Sampling Sites 225
6.2.2 Data Analysis and Modeling 225
6.2.3 Human Health Risk Assessment (HHRA) Model of Inhalable Particulate Matter ($PM_{2.5}$) 225
**6.3 Results and Discussion** **229**
6.3.1 Spatio-Temporal Variability of Ambient $PM_{2.5}$ Levels Per Group in the Study Area 229
6.3.2 Ambient $PM_{2.5}$ Dose-response Assessment 232
6.3.3 Exposure Assessment and Health Risk Analysis of Ambient $PM_{2.5}$ Inhalation 232
**6.4 Limitations of the Study** **243**
**6.5 Conclusions** **244**
**Reference** **245**
***CHAPTER SEVEN*** ***254***
***PREDICTION OF AMBIENT $PM_{2.5}$ USING MACHINE LEARNING AND NATURE-BASED SOLUTIONS FOR SUSTAINABLE AIR QUALITY MANAGEMENT IN THULAMELA LOCAL MUNICIPALITY, LIMPOPO, SOUTH AFRICA*** ***254***
**Abstract** **254**
**7.1 Introduction** **255**
**7.2 Materials and Methods** **257**
7.2.1 Ambient $PM_{2.5}$ Concentration and Meteorological Data 257
7.2.2 Data Collection, Pre-Processing and Feature Engineering for Modeling 258
7.2.3 Development and Prediction of Ambient $PM_{2.5}$ using Machine Learning Algorithms 259
**7.3 Results and Discussion** **265**
7.3.1 Meteorological Variables and their Descriptive Statistics within Thulamela Municipality 265
7.3.2 Association between Changes in Meteorological Parameters within Thulamela Municipality 268
7.3.3 Sites-specific Relationship between Meteorological Variables and Particulate Matter ($PM_{2.5}$) using Spearman Approach 270
7.3.4 Unveiling Site-Specific $PM_{2.5}$ Dynamics with Machine Learning Models 273
7.3.5. Integration of Feature Importance with the Prediction of Average $PM_{2.5}$ Across the Study Areas 281
7.3.6. Modelling and Prediction of Seasonal Variations in $PM_{2.5}$ Pollution using Polynomial Ridge Regression and Random Forest Models 321
**7.4 Implications for Air Quality Management** **327**
**7.5 Policy Recommendations for Air Quality Improvement: Lessons from APTI and API Studies** **328**
**7.6 Limitations of the Study** **328**
**7.7 Conclusion** **329**
**References** **330**
***CHAPTER EIGHT*** ***336***
***CONCLUSIONS AND RECOMMENDATIONS*** ***336***



© University of Venda

University of Venda
Creating Future Leaders

**8.1 Conclusions** **336**
**8.2 Recommendations** **338**
**8.3 Contributions to Knowledge** **339**
**8.4 Areas for Further Research** **340**
***APPENDICES*** ***341***



© University of Venda

# LIST OF FIGURES

**Figure 1.1:** Study Area map of Thulamela Municipality. Data source: Limpopo Provincial Government in Surveyor General Department; map developed using ArcMap 10.8 (Esri software). ..... 11
**Figure 1.2:** Flowchart showing the approach and tools for the research. ..... 12

**Figure 2.1:** Major sources of anthropogenic air pollution.(Source:www.clarity.io/blog). ... 20
**Figure 2.2:** Illnesses and deaths across the world due to air pollution (Source: WHO, 2016a). ..... 21
**Figure 2.3:** Shows the percentage of air pollutants in the U.S from the year 1990 to 2018 (Source: US EPA, Air quality system 2020). ..... 22
**Figure 2.4:** The number of deaths attributable to household air pollution in Nigeria from 1990 to 2017 (Chasant, 2019). ..... 26
**Figure 2.5:** Overview of the relative sizes of $PM_{10}$ and $PM_{2.5}$ in relation to human hair (US EPA, 2008; American Lung Association, 2013). ..... 31
**Figure 2.6:** Estimated 2020 annual average $PM_{2.5}$ concentrations ($\mu g/m^3$) (Source: Institute for Health Metrics and Evaluation, 2025). ..... 32
**Figure 2.7:** Transportation sector emissions showing high CO from motor vehicles. Source: (DEA, 2012). ..... 36
**Figure 2.8**: Main types of energy used for cooking in South Africa (StatsSA, 2011). ..... 42
**Figure 2.9:** Main types of energy used for cooking in Vhembe district (StatsSA, 2011). .. 42
**Figure 2.10:** Schematic representation of the mechanisms of vehicular emissions and pollutant pathways into air, water, and soil environments (Winther and Slento, 2010). ..... 45
**Figure 2.11:** South Africa CO contribution to global emissions (Source: IEA, 2023). ..... 47
**Figure 2.12:** Health effects associated with air pollution (Source: WHO, 2016a). ..... 65
**Figure 2.13:**Types of machine learning. ..... 71
**Figure 2.14:** Illustration of the derivation of the residual sum of squares. ..... 72
**Figure 2.15:** Polynomial Regression ..... 74
**Figure 2.16**: Random Forest ..... 75

**Figure 3.1:** Study domain map of Thulamela Municipality. ..... 98
**Figure 3.2**: Sampling point where fuelwood usage was measured. ..... 99
**Figure 3.3:** Thulamela taxi rank as one of the locations where the questionnaire was administered. ..... 100
**Figure 3.4:** Collecting the amount of wood used at a) Gondeni village, and b) Lunungwi village. ..... 105
**Figure 3.5:** Crane scale used for wood measurement. ..... 105
**Figure 3.6:** Population of Vehicle by Category in Thulamela, from 2012 to 2021 (Source: eNaTIS, 2020). ..... 107
**Figure 3. 7:** Average kilometre travelled by all the vehicles. ..... 109
**Figure 3.8**: Linear regressions between average distance travelled and vehicle age. ..... 109
**Figure 3.9:** The total a) gaseous emissions and b) particulate emissions estimated for all vehicles during the study period (2012–2021). ..... 112
**Figure 3.10:** Total emission trend by each vehicle type for the study period. ..... 113



© University of Venda

**Figure 3.11:** Contribution of vehicle type to emissions of $SO_2$, $NO_x$, and CO from 2012 to 2021. ..... 114
**Figure 3.12:** Contribution of each vehicle type to $PM_{2.5}$ and $PM_{10}$ emissions from 2012 to 2021. ..... 115
**Figure 3.13:** Trend ratio of $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ emissions to GDP during the study period (2012-2021). ..... 118
**Figure 3.14:** Road network within the study domain. ..... 118
**Figure 3.15:** 2021 estimated annual emissions of a) $SO_2$, b) $NO_x$, c) CO, d) $PM_{2.5}$, and e) $PM_{10}$ from vehicles (all classes) for the domain in the 1 km per grid model. ..... 120
**Figure 3.16**: Percentages of the energy sources consumed by households for cooking and heating. ..... 125
**Figure 3.17:** Three-legged stove use by household for cooking and other domestic activities. ..... 125
**Figure 3.18:** Wood usage in different seasons recorded in this study. ..... 126
**Figure 3.19:** Percentage of responses on average daily hours of wood usage in the villages for summer and winter. ..... 127
**Figure 3.20:** Percentage of respondents' frequency for a) uses of different wood types in the villages b) Sources, c) Reasons for using specific wood types in the villages, and d) Hotspots where wood is sourced in the village. ..... 129
**Figure 3. 21:** Various sources of wood through, a and b) Harvest in the field, c) Buying of different types, and d) Buying of Mopani wood (Source: Fieldwork). ..... 129
**Figure 3.22:** Average amount of wood consumed in a) morning and evening, and b) summer and winter per household (HH) in the 2 villages. ..... 134
**Figure 3.23:** Average daily wood consumption per Household (HH) by season recorded in this study. ..... 135
**Figure 3.24:** Areas using wood for domestic activities within the study domain. ..... 136
**Figure 3.25:** Number of households using wood in 2021 within the study area. ..... 137
**Figure 3.26:** A) $SO_2$, $NO_x$, and CO gaseous emission, and B) $PM_{2.5}$ and $PM_{10}$ particulate emission by village in the study domain. ..... 138
**Figure 3.27**: Spatial distribution of annual A) $SO_2$, B) $NO_x$, and C) CO emissions (tpa) from Residential Wood Combustion (RWC) within the 1 km model domain. ..... 139
**Figure 3.28:** Spatial distribution of annual A) $PM_{2.5}$, and B) $PM_{10}$ emissions (tpa) from RWC in the 1 km model domain. ..... 139
**Figure 3.29:** Source contributions to the total annual emissions by pollutants for the study. ..... 141

**Figure 4.1:** Map of the study area showing the monitoring sites. ..... 154
**Figure 4.2:** Modified Dylos DC1700 with solar panel. ..... 156
**Figure 4.3:** Monitoring site for calibration (Source: Wernecke et al., 2021). ..... 157
**Figure 4.4:** Diagram showing (a) GRIMM EDM180 (Source: GRIMM EDM180 user manual) and (b) Dylos DC1700 and the reference instrument in same location. ..... 157
**Figure 4.5**: Data collection process; a) opening of the equipment using screwdriver, and b) downloading the data using Male and female cable, Dylos logger software and laptop. . 159
**Figure 4.6:** Scatterplots showing the relationship between GRIMM EDM180 and Dylos DC1700 particle number count (PNC) with the corresponding $R^2$ value. ..... 162



© University of Venda

**Figure 4.7:** Scatterplots of concentration for GRIMM EDM180 and Dylos DC1700, with the corresponding $R^2$ value using the corrected $PM_{2.5}$ (PMC) concentrations (µg/m$^3$)..... 163
**Figure 4.8:** The time series of $PM_{2.5}$ concentrations (µg/m$^3$) obtained from Grimm EDM180 and Dylos DC1700. ..... 165
**Figure 4.9:** Annual average $PM_{2.5}$ concentrations at different sites in the study area during the monitoring period. The red line indicates the annual $PM_{2.5}$ threshold limit of 20 µg/m³ set by South Africa's NAAQS. ..... 167
**Figure 4.10:** Spatial distribution map of annual average $PM_{2.5}$ concentrations of the study area during the monitoring period (March 2021 to February 2022). ..... 167
**Figure 4.11:** Seasonal mean concentrations of $PM_{2.5}$ (µg/m$^3$) at different groups across the monitoring area in Thulamela LM for the study period with red line indicating South Africa NAAQS annual $PM_{2.5}$ threshold limit of 20 µg/m³. ..... 168
**Figure 4.12:** Monthly variations of $PM_{2.5}$ concentrations, with the bold line indicating the average, in the Thulamela LM during the study period, with the red and blue bold lines indicating the SA NAAQS and WHO daily standard, respectively. ..... 170
**Figure 4.13:** Diurnal variations of $PM_{2.5}$ concentrations, with the bold line indicating the average levels in the study area from March 2021 to February 2022. ..... 172
**Figure 4.14:** Weekday variations in the daily mean of $PM_{2.5}$ concentrations across the study area in Thulamela LM..... 172
**Figure 4.15:** The correlation matrix heatmap for the daily average concentrations of $PM_{2.5}$ reported by 14 distinct Dylos DC1700 units. ..... 175
**Figure 4.16:** Bivariate polar plot (BVP) of $PM_{2.5}$ concentrations (µg/m³) per site. The color scale indicates $PM_{2.5}$ concentration, and the radial scales show wind speed, which increases radially from the centre. The site grouping is represented in a) for residential site in Gondeni (RG), b) residential site in Lunungwi (RL) and c) vehicular and residential areas (VA-VC) in the study area. ..... 178
**Figure 4.17:** Bivariate polar plot (BVP) of $PM_{2.5}$ concentrations (µg/m³) per group (RG, RL, VA, VB and VC). The color scale indicates $PM_{2.5}$ concentration and the radial scales show wind speed, which increases radially from the center. ..... 179

**Figure 5.1:** Flowchart process for AERMOD model (Adopted from Adeniran et al., 2024). ..... 192
**Figure 5.2:** AERMOD modelling system's data flow (Villalvazo, Davila & Reed G. 2007). ..... 193
**Figure 5.3**: Modelled Windrose generated for the study domain for 2021 to 2022. ..... 194
**Figure 5.4**: Study domain within Limpopo Province..... 196
**Figure 5.5:** The diurnal trend in domestic fuel combustion used during the simulation of concentrations from Thulamela LM (This study, 2021)..... 200
**Figure 5.6:** Spatial distribution of the daily average model $PM_{2.5}$ level within the study domain from all estimated sources..... 202
**Figure 5. 7:** The spatial distribution of the annual average model $PM_{2.5}$ within the study domain from all estimated sources..... 203
**Figure 5.8**: The spatial distribution of the daily average model $PM_{2.5}$ exceedance within the study domain from all estimated sources. ..... 203
**Figure 5.9**: The spatial distribution of the daily average model $PM_{2.5}$ within the study domain from the residential sources. ..... 205



© University of Venda

**Figure 5.10:** Spatial distribution of the annual and daily average model $PM_{2.5}$ within the study domain from the residential sources. ..... 205
**Figure 5.11:** Spatial distribution of daily average modelled $PM_{2.5}$ exceedances within the study domain from residential sources. ..... 206
**Figure 5.12:** The spatial distribution of the daily average model $PM_{2.5}$ within the study domain from the vehicular sources. ..... 207
***Figure 5.13:*** *The spatial distribution of the annual average model $PM_{2.5}$ within the study domain from the vehicular sources.* ..... 208
**Figure 5.14:** Map showing the locations used as discrete receptors in the study domain, a) monitoring sites simulated, and b) sensitive receptors. ..... 209
**Figure 5.15:** Box plots showing the modelled $PM_{2.5}$ concentration from all discrete receptors simulated at Thulamela. ..... 213
**Figure 5. 16**: Scatter plot of monitored against modelled daily average $PM_{2.5}$ concentrations for Thulamela LM, South Africa. ..... 216
**Figure 5.17:** Temporal variation comparing the modelled and monitored $PM_{2.5}$ concentration in the study area of South Africa. ..... 216

**Figure 6.1:** The flowchart of the human health risk assessment (HHRA) model (Source: This study). ..... 226
**Figure 6.2:** Hazard Quotient (HQ) for non-carcinogenic risks of $PM_{2.5}$ by land-use activity and across the study area. The blue bar represents the winter, the solid orange line is summer, and the dashed orange line represents the HQ standard limit for all age groups. ..... 238
**Figure 6.3:** Excess lifetime cancer risk (ELCR) for $PM_{2.5}$ exposure in Thulamela municipality. ..... 243

**Figure 7.1**: Flow chart used for the analysis and prediction of $PM_{2.5}$ in this study. ..... 258
**Figure 7.2**: Monthly mean relative humidity (%), temperature (°C), and rainfall (mm) for the study area. ..... 266
**Figure 7.3**: Windrose depicting the annual and seasonal wind speed, showing the prevailing wind patterns that affect $PM_{2.5}$ transport across the study area, a) annual, and b) by seasons. ..... 266
**Figure 7.4**: Spearman rank correlation between the $PM_{2.5}$ pollutant and the meteorological variables. ..... 271
**Figure 7.5:** Predictive accuracy and model fit metrics for each site using a multiple linear regression model. ..... 275
**Figure 7.6:** Predictive accuracy and goodness of fit metrics for each site using random forest model. ..... 279
**Figure 7.7:** Relationship between predicted and measured $PM_{2.5}$ concentration ($\mu g/m^3$) for some selected sites using random forest model a) RG3, b) RL3, c) VP4, d) VUI6, e) VMV8 and f) VTV9. ..... 279
**Figure 7.8:** Bar plot showing the model comparison feature importance of the machine learning models. ..... 282
**Figure 7.9:** Scatter plot visualizing the predicted vs. actual $PM_{2.5}$ in the study area using the three selected machine learning models. a) MLR, b) RF, and c) PRR models. ..... 290



© University of Venda

**Figure 7.10:** Scatter plot illustrating the comparison between actual $PM_{2.5}$ concentrations (yellow dots) and predictions made by the three machine learning models; MLR (blue dots), RF (green dots), and PRR (purple dots) over time. ..... 292
**Figure 7.11:** Actual $PM_{2.5}$values against the predicted values for the machine learning models. The yellow line represents the actual values, blue dashed line represents the MLR predictions, green dashed line represents the random forest prediction while 'purple colour represents the polynomial ridge regression model. ..... 292
**Figure 7.12:** Bar plot comparing the machine learning techniques across all metrics (a), box plot showing the distribution of the model metrics, and (b) visualization of cross-validation results for better insights into the stability and consistency of the models across key metrics. ..... 293
**Figure 7.13:** Residual analysis of a) MLR, b) random forest and c) purple colour represent the polynomial ridge regression. ..... 320
**Figure 7. 14:** Relationship between predicted and measured average $PM_{2.5}$ concentration during the (a) winter, and (b) summer using polynomial ridge regression model ..... 322
**Figure 7.15:** Relationship between predicted and measured average $PM_{2.5}$ concentration during the winter and summer seasons using RF Model, a) winter and b) summer prediction. ..... 325
**Figure 7.16:** Bar plot showing the winter and summer model comparison feature Importance for RF using average $PM_{2.5}$ concentration. ..... 326
**Figure 7.17:** Residual analysis for random forest model using average $PM_{2.5}$ concentration. ..... 327



© University of Venda

# LIST OF TABLES

**Table 2.1**: Emission estimates by sectors in China, 2010 (Gg/year) (Meng et al., 2017)..23
**Table 2.2:** South African ambient air quality standard for $PM_{2.5}$, $PM_{10}$, $SO_2$, $NO_x$, and CO (DEAT 2009 and Garland et al., 2021).....34
**Table 2.3:** Models for air pollution health impact assessment .....68

**Table 3.1:** Residential fuel use within Thulamela municipality (Data was derived during field study).....101
**Table 3.2:** Emission factors (g/km) for $SO_2$, $NO_X$, CO, $PM_{2.5}$, and $PM_{10}$ for calculation of vehicular emissions.....103
**Table 3.3:** Emission factors for residential fuelwood use.....106
**Table 3.4**: Mean VKT by different vehicle types in the study area .....110
**Table 3.5:** Pearson correlation matrix between GDP and the pollutants within the study area.....116
**Table 3.6:** VKT for the year 2021 in the study area. .....119
**Table 3.7:** Estimated emission (tons/year) of each vehicle type for 2021 in the study domain .....119
**Table 3.8:** Chi-square Cross-tabulation of Respondents' Socio-demographic Attributes by Village .....123
**Table 3.9:** Respondent frequency of different energy types mainly used for household activities with seasons as highlighted in the field study questionnaires.....124
**Table 3.10:** Preferred Tree species for fuelwood.....128
**Table 3.11:** Cross-tabulation between the choice of energy source used for cooking during the seasons using chi-squared test.....130
**Table 3.12:** Crosstab for village comparison versus respondent's perspective of reason, satisfaction, awareness, and condition.....132
**Table 3.13:** Mean, total and maximum amounts of wood burnt in the mornings and evenings by all 15 households in the villages for summer and winter 2020/2021. .....133
**Table 3.14:** Estimated RWC emission for 2021 ($10^{\wedge -4}$ tpa).....138
**Table 3.15:** 2021 emission inventories (ton per annum) per sector for the study domain in Thulamela municipality.....140

**Table 4.1:** Data completeness (%) per month at each $PM_{2.5}$ monitoring site from March 2021–February 2022. Sites with 70% and higher are bolded.....160
**Table 4.2:** Calibration of Dylos DC1700 with reference instruments in previous studies 164
**Table 4.3:** Summary statistics of Ambient $PM_{2.5}$ concentrations in Thulamela LM..... 166
**Table 4. 4:** Daily $PM_{2.5}$ pollutant exceedances of the NAAQS and WHO air quality standards.....173
**Table 5.1:** Quantified emissions rate of total $PM_{2.5}$ (ton per annum) per sector used by AERMOD model for Thulamela LM .....199
**Table 5.2:** Average annual $PM_{2.5}$ modelled level for discrete receptors in Thulamela LM.....212
**Table 5.3:** Comparison of measured and predicted average $PM_{2.5}$ and Model Performance Evaluation within Thulamela, South Africa.....215



© University of Venda

**Table 6.1:** Recommended parameter values for calculating exposure doses (ADD and LADD) and the health risks (Non-Carcinogenic and Carcinogenic) from ambient $PM_{2.5}$ via inhalation…..……229
**Table 6.2:** Seasonal mean of $PM_{2.5}$ Concentrations (µg/m³) by Group in the Study Area……231
**Table 6.3:** Average daily dose (ADD) µg/kg/day of $PM_{2.5}$ exposure via inhalation pathway in the study area……235
**Table 6.4:** Lifetime Average Daily Dose (LADD) µg/m³ for $PM_{2.5}$ exposure in the study area……240

**Table 7.1:** Descriptive statistical analysis of $PM_{2.5}$ and meteorological parameters……. 267
**Table 7.2:** Correlation coefficients between daily mean meteorological parameters in the study area from March 2021 to February 2022……269
**Table 7.3:** Multiple Regression (MLR) Model equations and the goodness of fit statistics for each study site……276
**Table 7.4:** Comparative analysis of model metrics in current and previous studies……. 295



© University of Venda

# CHAPTER ONE

# INTRODUCTION

## 1.1 Preamble

Presented in this chapter are the background, problem statement, motivation, aim, specific objectives, hypotheses, as well as a description of the study area.

## 1.2 Background to the Study

Globally, clean air is essential for the environment – the nexus of public health, however, an increase in industrialisation, urbanisation, and economic growth has led to the deterioration of air quality due to the discharge of pollutants from these numerous sources (natural and anthropogenic sources) into the environment (Phalen and Phalen, 2013; Chopra, 2016; Sun *et al*., 2019). Air pollution is one of the most alarming environmental problems in the world, due to its impacts on human health and climate change (WHO, 2018). Air pollution has been declared by the World Health Organisation (WHO) as “the silent killer”, also described by Chandrappa and Kulshrestha (2016) as “the world’s largest single environmental health risk”.

More than 80% of people living in urban areas, as well as 92% of low-income and high-income households, are exposed to air contamination levels that exceed the WHO standard limits of air pollution (WHO, 2016; Das and Horton, 2018). Studies indicate that ambient air pollution in sub-Saharan Africa is rising rapidly, aggravating the burden of human disease due to insufficient air quality monitoring, limited research, and population vulnerabilities (Egondi *et al*., 2016; Coker & Kizito, 2018). Heavily polluted urban areas like Lagos in Nigeria and Nairobi in Kenya experience heightened exposure to harmful pollutants from industrialisation and urbanisation, leading to increased cases of respiratory and cardiovascular diseases (Kiai *et al*., 2021; Laro & Raheem, 2022). The lack of adequate monitoring systems hinders accurate assessment and intervention, leaving populations exposed to pollutant levels exceeding regulatory standards, thereby making them vulnerable to severe health issues and premature death.



© University of Venda

Occurrences of volcanic eruptions, wind and soil erosion, forest fires, sandstorms (particularly in regions close to deserts), and dispersion of plant pollen are major natural sources of air pollution. Most atmospheric pollutants, however, are mainly introduced through anthropogenic activities, such as industrial emissions from mining activities, waste disposal, transportation, domestic fuel burning, landfill sites, agriculture, and brickmaking (Keuken *et al*., 2012; Bonjour *et al*., 2013; Aniyikaiye *et al*., 2023; Njoku *et al*., 2023). The burning of wood, coal, or paraffin for heating, cooking, and lighting purposes has been reported as part of the contributory sources of ambient and indoor air pollution (Paulin *et al*., 2013).

Common air pollutants include particulate matter (PM) and gaseous compounds such as nitrogen oxides, sulfur dioxide, carbon monoxide, and carbon dioxide, among others. The concentration of these pollutants depends on their sources, their distribution patterns, meteorological conditions, and topographical features of the area (Chattopadhyay, 2014; Sharma *et al.,* 2018). There have been reports that air pollution alters the ecosystem and negatively impacts plants by reducing photosynthetic pigment, stomata conductance, and net photosynthetic rate (Rai and Panda, 2014). Additionally, changes in meteorological parameters, such as wind speed and direction, can disperse these pollutants over long distances, causing transboundary contamination (NRC, 2010).

Several health problems have been linked to both ambient and indoor air pollution, with low-income countries reported to be more negatively affected by these pollutants (WHO, 2020). In addition, epidemiological studies have shown that air pollution can cause different human ill-health like pulmonary, cardiovascular, and neurological diseases (Koenig, 2000; Pope iii *et al.,* 2002; Kampa and Castanas, 2008); chronic respiratory symptoms, and diseases among elderly people worldwide (Nkosi *et al*., 2015). A consistent increase in, cardiac and respiratory diseases, lung cancer, and mortality in the world have been reported due to people's exposure to air pollution from different sources (Hamra *et al*., 2014; Franklin *et al*., 2015). It is estimated that atmospheric Particulate Matter ($PM_{2.5}$), which refers to tiny airborne particles that can be inhaled, typically measuring 2.5 micrometres or less in diameter, causes 3.3 million premature deaths annually, primarily in Asia, and they have had a variety of negative effects on human health (Lelieveld *et al*., 2015). In the United



© University of Venda

States, the Environmental Protection Agency (EPA) currently sets the maximum allowable concentration of microscopic particles at 150 μg/m$^3$ of air. The amount of pollutants in the air is expressed in terms of its mass/volume concentration, usually as micrograms of pollutants per cubic meter of air (μg/m$^3$).

Emission inventories (EI) are important for locating air pollutant sources, evaluating emissions, and their trends over time. Emission inventories provide a detailed and comprehensive account of the sources and quantities of air pollutant emissions within a specific geographic area, such as a city, state, or country (EEA, 2019). Over the years, scholars have developed reliable emission inventories in South Africa, and yet significant improvement in emission estimates is still ongoing. These emissions can be estimated in different sectors using the activity rate and emission factors (Scorgie *et al*., 2003; Blignaut *et al*., 2005). Recently, South Africa has been rated as one of the most air-polluted countries in the world due to poor urban ambient air quality in various areas of the country with an adverse effect of pollutants emitted from numerous sources, with a $PM_{2.5}$ rate of above 60 mg/m$^3$ (Meth, 2018). Awareness of air contamination, monitoring and control measures are inadequate, especially within Thulamela Municipality, Vhembe District, which is categorised as a low-income area of South Africa. Industrial, commercial, transportation, and domestic land-use activities are commonly present in the Vhembe District and are usually major sources of air pollution in the area. These detrimental air conditions have motivated scholars to continuously research the contribution of these land-use activities on ambient air pollution (Freiman & Piketh, 2003; Shabbir & Ahmad, 2010; Gordon *et al*., 2014).

Domestic coal and wood combustion have been recognised as one of the major sources of air pollution, especially within the lower-income settlements of South Africa. This has been shown to have the largest impact on human health because of the proximity of these air pollutant emissions to humans (Scorgie *et al*., 2003; Scorgie, 2012). Most of these low-income areas lack the finance, technology, and expertise to carry out comprehensive air quality monitoring in their area, and this has led to non-compliance by the manufacturing, industrial and domestic sectors in these areas. Monitoring of air quality is essential for all activities in the semi-urban and rural areas of Thulamela Municipality, in order to regulate



© University of Venda

the quality of air, assess emission sources and quantity, simulate the dispersion of the pollutants, and promote good human health and a healthy environment.

EI is frequently used as input data for air quality and dispersion modelling in air quality management (Maes *et al*., 2009). Air dispersion modelling is a steady-state Gaussian plume model that provides an effective method of assessing the impact of atmospheric emission from a wide variety of sources and measuring compliance with the standard air quality (EEA, 2013; Rood, 2014; Chidhindi, 2022). The South African government has published regulations concerning air dispersion modelling, for instance, in Government Gazette No. 37804 vol. 589; on 11 July 2014 (Bird and von Gruenewaldt, 2016).

Currently, the use of low-cost air monitoring sensors can enhance the process of air monitoring (Dacunto *et al*., 2015; Semple *et al*., 2015). This motivated the researcher to validate a low-cost air monitoring equipment with regulatory government standard equipment, in a South African context. The low-cost air monitoring equipment was then adopted to monitor the ambient air around the Vhembe District. This approach was intended to eventually help local households and commercial establishments to monitor air quality independently, reducing reliance on government data. By making air quality data more accessible and localised, these sensors can help communities better understand pollutants, such as $PM_{2.5,}$ in their environment and take proactive health and environmental measures.

This is crucial, as $PM_{2.5}$ primarily generated from combustion sources, poses significant health risks, making it essential to assess its concentration and associated health effects. Several studies have evaluated the health potential of particulate matter using an air quality model known as the Human Health Risk Assessment Model and have found that high levels of $PM_{2.5}$, as well as its bound metal exposure, are linked with a range of health problems, including respiratory and cardiovascular problems. It is, therefore, imperative to continuously assess the effects of pollutant concentration on humans to provide effective mitigation actions and policy measures that safeguard public health (Morakinyo *et al.,* 2020; Alfeus *et al.,* 2022; Moradi *et al*., 2022; Saju *et al*., 2023).

Additionally, it is essential to predict ambient $PM_{2.5}$ concentrations, which would facilitate understanding trends in air pollution, thereby reducing its health risks and environmental



© University of Venda

impact (Morapedi & Obagbuwa, 2023; Hasnain *et al*., 2024). In developing areas like Thulamela Local Municipality, Limpopo Province, limited monitoring infrastructure presents challenges for effective air quality management. This limitation necessitates innovative approaches, such as machine learning (ML), which uses data-driven techniques to predict pollutant levels accurately. Furthermore, local monitoring, modelling, and prediction of air quality in Vhembe District will aid government decision-making and policy development. This study's aim, thus, was to estimate emissions of gaseous pollutants and particulate matter ($PM_{2.5}$ and $PM_{10}$), determine their dispersion, monitor $PM_{2.5}$ using low-cost sensors, assess potential health risks, and predict pollutant levels with high accuracy within Thulamela Municipality, Vhembe District, South Africa.

### 1.3 Statement of Problem

Air pollution has become a major problem in recent years due to an increase in emissions from different human activities (Chan & Yao, 2008), which continues to pose health threats to humans and the environment. Worldwide, it has been reported that about 98% of most cities do not meet WHO air quality guidelines, especially in low to medium-income countries (WHO, 2016). Countries like China, India, Pakistan, and Bangladesh, for example, have been reported as having very high pollution rates. Likewise, studies have revealed South Africa as one of the most polluted, in terms of carbon dioxide ($CO_2$) and nitrogen oxide ($NO_2$) (Akinshipe and Kornelius, 2015). In 2016, it was reported that over 95% of the world's population is breathing unhealthy air (WHO, 2016). As a result, over 6.1 million people have died from long-term exposure to contaminated air, with India and China jointly responsible for more than 50% of these deaths caused by $PM_{2.5}$ (Hadley *et al*., 2018). Air pollution has, hence, been declared the silent killer, causing 7 million deaths annually.

South Africa was reported as the second country with a high infant mortality rate among children between the ages of 0-5 years due to respiratory diseases caused by poor air quality (Doherty *et al*., 2016). In 2016, the WHO reported that towns like Diepkloof in Soweto, Gauteng; Middelburg, Witbank, and Ermelo in Mpumalanga; Durban in KwaZulu-Natal; and Waterberg in Limpopo were the most affected areas with bad air quality in South Africa. In Limpopo Province, a few potential "hot spot" zones due to poor air quality have been identified in the urban areas; these areas include Lephalale, Polokwane, Phalaborwa,



© University of Venda

Steelpoort and around the Vhembe District. Polokwane was identified as one of the four hot spot areas selected for the dispersion model in Limpopo Province because it has very high concentrations of sulphur dioxide and particulate matter (LEDET, 2013). The concentration levels of these pollutants were reported to exceed the WHO and National Ambient Air Quality Standards (NAAQS) standard limits, with little data for $SO_2$ and $PM_{10}$ (LEDET, 2013). Even though $PM_{2.5}$ has recently been monitored in other parts of South Africa (such as the National Air Quality Priority Areas, which contain the majority of the country's large industrial hubs and coal-fired power plants), the government has not yet been able to extend its monitoring program to some areas, particularly low-income areas of the country. Residential wood combustion (RWC) and vehicular emissions have been contributors to both ambient and indoor air pollution. These, however, remain a gap in understanding and quantifying emissions from these activities, as few studies have comprehensively addressed this issue. Due to the lack of comprehensive baseline data on air pollution and trends of air quality in semi-urban and rural areas, especially in Vhembe Districts, Limpopo Province, there is an urgent need to monitor the atmospheric air quality of the districts for a sustainable health-environment nexus. It is important to identify hotspot areas, monitor and determine the impact of air quality on humans and the environment.

In the Vhembe District, there is limited or no study that has been done to estimate the emission inventory and measure the ambient air quality associated with the combustion of domestic fuels (Madonsela, 2019). Vhembe District is mainly occupied by rural households and the informal population, and some of the households do not have basic services like electricity. Also, due to the high tariffs of electricity in South Africa, most of these households often resort to alternative sources of fuel for energy (Buthelezi *et al*., 2019). These alternative sources of energy involve the combustion of coal, wood, paraffin, and other materials, which leads to pollution of their surrounding atmosphere. Also, air pollution associated with the increase in transportation and industrialisation has posed some challenges to air quality; for example, the increase in vehicle use has contributed significantly to the increase in vehicle emissions like PM, $NO_2$, and ozone ($O_3$) (Tongwane *et al*., 2015). The overwhelming use of coal in industries has brought about several contributions to their surrounding air pollution, similar to the Vhembe District area of South Africa. With this problem, limited or no forms of monitoring and emission inventory



© University of Venda

measurements in the Vhembe District has been done currently in the area either by government agencies or academic scholars. This study, therefore, was conducted to evaluate emissions from domestic, and transport emissions in semi-urban, rural households within Thulamela Municipality, Vhembe District. This will help understand the various fuel types, combustion rates of these fuels and monitor the ambient air of the surroundings.

Currently, the day-to-day monitoring and management process of air quality across the globe, including South Africa, requires a lot of money. This problem has contributed to the inadequacies and increase in the insufficiency of air quality data availability across the world and South Africa; however, the use of low-cost monitors has been introduced by different scholars (Spinelle *et al*., 2015; Sousan *et al*., 2016; Han *et al*., 2017) and this has helped in filling the paucity of information on both ambient and indoor air quality data. It is urgent to know and acquire data on air quality in regions; thus, this study also seeks to validate and utilise low-cost sensor analyser, like the Dylos DC1700, to measure the ambient air quality in Vhembe District.

### 1.4 Motivation

The atmosphere is a complex and dynamic gaseous system that is essential for life on Earth (Steffen *et al*., 2005). Improving living standards for a rapidly growing world population and urbanisation tends to increase global pollutant emissions due to the different human activities that degrade air quality. Recently, research on atmospheric conditions have reported that surface pollutants can be influenced by meteorological factors like wind speed, wind direction, relative humidity as well as source conditions (Habeebullah *et al*., 2015), while pollutants could be transported to high altitudes (Munir *et al*., 2013; Seinfeld & Pandis, 2016). Many studies have reported on variations in urban air quality in the world (Joshi & Mahadev, 2011; Chen *et al*., 2015; Landrigan *et al*., 2019); however, there are a few reports on Africa, especially in South Africa (Sowden *et al*., 2008).

The human health risk assessment (HHRA) framework is a useful tool employed by the US Environmental Protection Agency (US EPA) in estimating potential human health risks due to exposure to pollutants, and this has been previously used in South Africa (Oosthuizen *et al*., 2015). From the epidemiological standpoint, studies have established the link between



© University of Venda

air pollutants, from different sources and human health implications, which could be short-term or long-term negative effects, depending on the exposure time (Mannucci & Franchini, 2017).

Currently, in South Africa, although there are a few hotspots for air quality monitoring in some of the urban areas, there is still limited data. Few monitoring stations that are in place at these hotspot areas are functioning, hence, there is much missing data in the semi-urban and rural areas (LEDET, 2013). This has made it impossible to determine the trends of air quality in the country, particularly in Limpopo Province, specifically, in the area selected for this study - Vhembe District (LEDET, 2013; LEDET, 2016); there is an urgent need for adequate air quality monitoring of the environmental quality to attain sustainable development in the area. This study, thus, aimed to monitor and compare pollution from different areas within the Thulamela Municipality of Vhembe District, where there are varied anthropogenic activities.

Geographical Information System (GIS) is a powerful tool that provides a flexible environment for entering digital data from various sources, thereby capturing, storing, querying, analysing, displaying, and producing geographic information (Dadhich & Hanaoka, 2010). Geospatial mapping provides a quick overview of pollutants in relatively critical areas that need more attention for improvement. Yet, there is no current existing map of the study area showing air pollutant geographical data, air quality plume, and no prediction practices have been suggested by earlier reports. This study addresses these gaps by analysing air quality dispersion within the selected study area and estimating the local impact of various emission sources on air quality management.

Advanced dispersion and machine learning (ML) modelling studies are essential for analysing the effects of various controlling factors on particulate matter (PM), particularly in understanding the association between meteorological data and PM concentrations. These analyses play a crucial role in enhancing air quality and management strategies. Data obtained from continuous monitoring are used to evaluate pollutant concentrations, assess their health impacts, model pollutant distribution, and predict future pollution levels in an area (Pawul & Sliwka, 2016). This research will form part of an ongoing project by Eskom,



© University of Venda

aimed at establishing baseline air quality data across different regions in South Africa, with a specific focus on respirable $PM_{2.5}$. The results will also fill a critical knowledge gap by employing an HHRA framework model to assess the potential impacts of $PM_{2.5}$ on human health in the Vhembe District. The study evaluated the compliance levels of the selected air pollutant using the South Africa National Ambient Air Quality Standards (NAAQS) and World Health Organization (WHO) guidelines for the period of monitoring. Information was also obtained on the spatial distribution of the deposited $PM_{2.5}$, contributing to air quality modelling and offering valuable insights into the management of air pollution-related health risks in the region. Furthermore, this was the first study to compare and apply advanced machine learning techniques in Thulamela Municipality for estimating $PM_{2.5}$ concentrations, providing stakeholders with an efficient tool for air quality management and mitigation strategies in a semi-urban setting.

## 1.5 Aim and Objectives of the Study

### 1.5.1 Aim of the study

The study aimed to quantify, monitor, model, and predict ambient air pollution, assess its potential health risks, and deploy predictive models for real-time air quality monitoring within Thulamela Municipality, Vhembe District, Limpopo Province, South Africa.

### 1.5.2 Specific Objectives

i. To quantify emissions of gaseous ($SO_2$, $NO_x$, and CO) and particulate matter ($PM_{2.5}$ and $PM_{10}$) pollutants and develop an emission inventory from various air pollution sources, within Thulamela Municipality, Vhembe District, Limpopo Province, South Africa.
ii. To evaluate the usefulness of a low-cost sensor air quality monitoring instrument for ambient $PM_{2.5}$ in a semi-urban and rural environment within the Thulamela Municipality, Vhembe District of Limpopo Province, South Africa.
iii. To estimate the particulate matter ($PM_{2.5}$) concentration and its distribution in the study area using the AERMOD model.
iv. To determine the potential health risks associated with exposure to $PM_{2.5}$ pollutants within the study area.



© University of Venda

v. To develop and evaluate machine learning models for predicting ambient $PM_{2.5}$ concentrations in the Thulamela Local Municipality.

### 1.6 Research Hypothesis

i. Emissions of gases and particulate matter ($PM_{2.5}$ and $PM_{10}$) from air pollution sources can be accurately quantified within the study area.
ii. Low-cost sensor instruments can be effectively used for monitoring ambient air quality in the study area.
iii. Particulate matter ($PM_{2.5}$) dispersion concentration can be estimated in the study area using the AERMOD model.
iv. There is a potential health risk associated with ambient $PM_{2.5}$ exposure in the study area.
v. Machine learning techniques can accurately predict ambient $PM_{2.5}$ concentrations, supporting air quality management plans and mitigation strategies.

### 1.7 Description of the Study Area

The study was conducted in Thulamela Local Municipality (LM), located in Vhembe District of Limpopo Province of South Africa (Figure 1.1). The Ambient air quality monitoring (AAQM) research study was carried out at several sites within the Municipality, including the University of Venda, Thohoyandou, Shayandima, Makumbane, Sibasa, Makwarela, and Tshilongoma town as well as Gondeni and Lunungwi villages within TLM.

TLM is situated in the eastern part of the Vhembe District, with coordinates 22° 45’ 24.24” S latitude and 30° 35’ 53.36” E longitude. The Municipality covers an area of approximately 5,835 km² and is home to about 47.7% of the district’s population. TLM has a total population of 618,462, with 54.9% female and 45.1% male, and a household population of 156,594. The Venda and Tsonga tribes account for the major population in the district (StatsSA, 2011). Most households in the area largely depend on government grants, pensions, and remittances from family members who migrate to urban centres to work. Household wealth is generally lower in this Municipality when compared to other municipalities in South Africa (Aron & Muellbauer, 2006). While most households use liquefied petroleum gas (LPG),



© University of Venda

electricity, and wood for cooking, Gondeni and Lunungwi villages primarily rely on fuelwood as their main source of energy for cooking.

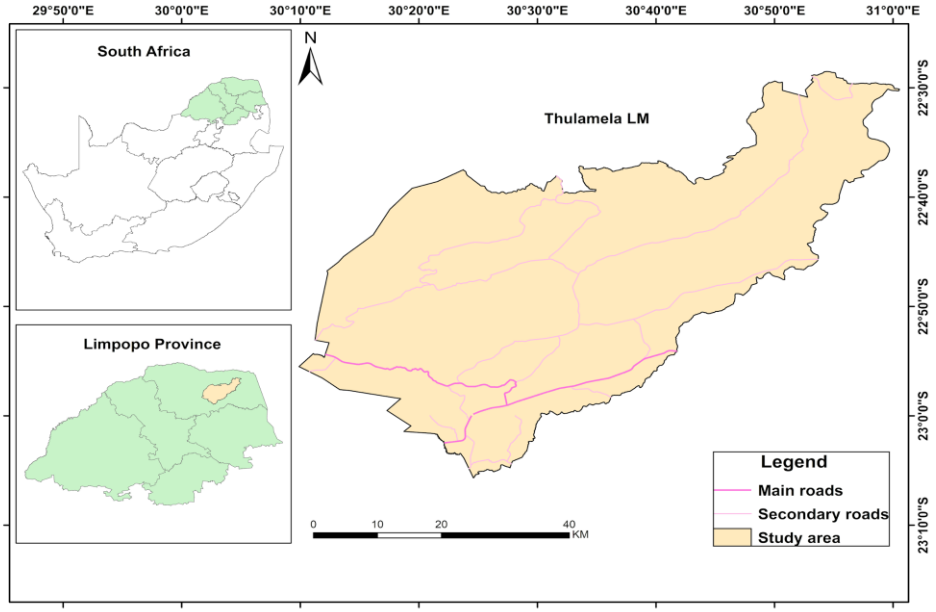



**Figure 1.1:** Study Area map of Thulamela Municipality. Data source: Limpopo Provincial Government in Surveyor General Department; map developed using ArcMap 10.8 (Esri software).

## 1.8 Thesis Outline and Research Methodology Framework

The thesis is divided into eight chapters, each serving a specific purpose, including an introduction (Chapter 1), literature reviews (Chapter 2), and Chapters 3–7 discuss the five specific objectives of the thesis and are presented in paper format; Chapter 8 covers the conclusions and recommendations. This section further presents the research methodology flow chart used in the study, during the questionnaire survey, fieldwork, modelling, data analysis, and interpretation (Figure 1.2).



© University of Venda

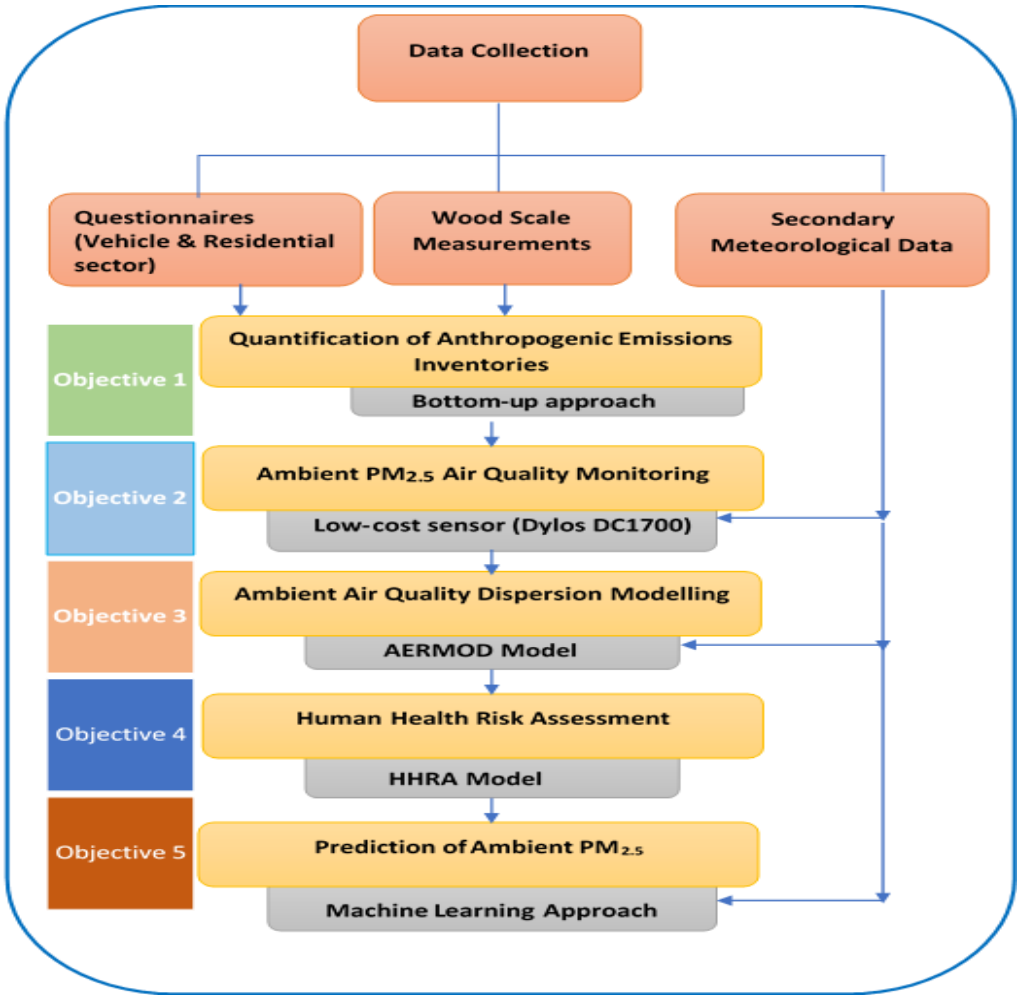



**Figure 1.2:** Flowchart showing the approach and tools for the research.

## References

Akinshipe, O. & Kornelius, G. 2015. Atmospheric emissions from clamp kilns in the South African clay brick industry. National Association for Clean Air (NACA) Annual Conference, Bloemfontein, South Africa.

Alfeus, A., Molnar, P., Boman, J., Shirinde, J., & Wichmann, J. (2022). Inhalation health risk assessment of ambient $PM_{2.5}$ and associated trace elements in Cape Town, South Africa. *Human and Ecological Risk Assessment: An International Journal*, *28*(8), 917–929.

Aniyikaiye, T. E., Piketh, S. J., & Edokpayi, J. N. (2023). Quantification of ambient $PM_{2.5}$ concentrations adjacent to informal brick kilns in the Vhembe District using low-cost



© University of Venda

sensors. *Scientific Reports*, *13*(1), 22453.

Aron, J., & Muellbauer, J. (2006). Estimates of household sector wealth for South Africa, 1970–2003. *Review of Income and Wealth*, *52*(2), 285–307.

Bird, T. & Von Gruenewaldt, R. 2016. Air Quality Impact Assessment for the Proposed Kriel Power Station FFP Retrofit.

Blignaut, J. N., Chitiga-Mabugu, M. & Mabugu, R. 2005. Constructing a greenhouse gas emissions inventory using energy balances: the case of South Africa for 1998.

Bonjour, S., Adair-Rohani, H., Wolf, J., Bruce, N. G., Mehta, S., Prüss-Ustün, A., Lahiff, M., Rehfuess, E. A., Mishra, V., & Smith, K. R. (2013). Solid fuel use for household cooking: country and regional estimates for 1980–2010. *Environmental Health Perspectives*, *121*(7), 784–790.

Buthelezi, S. A., Kapwata, T., Wernecke, B., Webster, C., Mathee, A., & Wright, C. Y. (2019). Household fuel use for heating and cooking and respiratory health in a low-income, South African coastal community. *International Journal of Environmental Research and Public Health*, *16*(4). https://doi.org/10.3390/ijerph16040550.

Chan, C. K., & Yao, X. (2008). Air pollution in mega cities in China. *Atmospheric Environment*, *42*(1), 1–42.

Chandrappa, R., Chandra Kulshrestha, U., Chandrappa, R., & Chandra Kulshrestha, U. 2016. Major issues of air pollution. *Sustainable Air Pollution Management: Theory and Practice,* 1-48.

Chattopadhyay, S. (2014). Spatial and temporal variations of ambient air quality in Burdwan town. *West Bengal, India*.Chen, W., Yan, L., & Zhao, H. (2015). Seasonal variations of atmospheric pollution and air quality in Beijing. *Atmosphere*, *6*(11), 1753–1770.

Chidhindi, P. (2022). *Using dispersion models as a regulatory tool in South Africa*. North-West University (South Africa). PhD diss., North-West University (South Africa), 2022.

Chopra, R. (2016). Environmental degradation in India: causes and consequences. *International Journal of Applied Environmental Sciences*, *11*(6), 1593–1601.

Coker, E., & Kizito, S. (2018). A narrative review on the human health effects of ambient air pollution in sub-Saharan Africa: An urgent need for health effects studies. *International Journal of Environmental Research and Public Health*, *15*(3). https://doi.org/10.3390/ijerph15030427

Dacunto, P. J., Klepeis, N. E., Cheng, K. C., Acevedo-Bolton, V., Jiang, R. T., Repace, J. L., Ott, W. R., & Hildemann, L. M. (2015). Determining $PM_{2.5}$ calibration curves for a low-cost particle monitor: Common indoor residential aerosols. *Environmental Sciences: Processes and Impacts*, *17*(11), 1959–1966. https://doi.org/10.1039/c5em00365b.

Dadhich, P. N., & Hanaoka, S. (2010). Remote sensing, GIS and Markov's method for land use change detection and prediction of Jaipur district. *J. Geomat*, *4*, 9–15.

Das, P., & Horton, R. (2018). Pollution, health, and the planet: time for decisive action. The



© University of Venda

Lancet, 391(10119), 407-408.

Doherty, T., Kroon, M., Rhoda, N., & Sanders, D. (2016). Ending preventable child deaths in South Africa: What role can ward-based outreach teams play? *South African Medical Journal*, *106*(7), 672–674.

Egondi, T., Muindi, K., Kyobutungi, C., Gatari, M., & Rocklöv, J. (2016). Measuring exposure levels of inhalable airborne particles ($PM_{2.5}$) in two socially deprived areas of Nairobi, Kenya. *Environmental Research*, *148*, 500–506.

European Environment Agency (EEA), 2019. Air Pollutant Emission Inventory Guidebook. "https://www.eea.europa.eu/publications/emep-eea-guidebook-2019".

European Environment Agency. 2013. Air quality in Europe - 2013 report. EEA Report No 9/2013. Copenhagen, Denmark: European Environment Agency.,

Franklin, B. A., Brook, R., & Pope III, C. A. (2015). Air pollution and cardiovascular disease. *Current Problems in Cardiology*, *40*(5), 207–238.

Freiman, M. T., & Piketh, S. J. (2003). Air transport into and out of the industrial Highveld region of South Africa. *Journal of Applied Meteorology and Climatology*, *42*(7), 994–1002.

Gordon, S. B., Bruce, N. G., Grigg, J., Hibberd, P. L., Kurmi, O. P., Lam, K. H., Mortimer, K., Asante, K. P., Balakrishnan, K., & Balmes, J. (2014). Respiratory risks from household air pollution in low and middle income countries. *The Lancet Respiratory Medicine*, *2*(10), 823–860.

Habeebullah, T. M., Munir, S., AA, A. H., Morsy, E. A., Seroji, A. R., & Mohammed, A. M. F. (2015). The interaction between air quality and meteorological factors in an arid environment of Makkah, Saudi Arabia. *International Journal of Environmental Science and Development*, *6*(8), 576.

Hadley, M. B., Baumgartner, J., & Vedanthan, R. (2018). Developing a clinical approach to air pollution and cardiovascular health. *Circulation*, *137*(7), 725–742.

Hamra, G. B., Guha, N., Cohen, A., Laden, F., Raaschou-Nielsen, O., Samet, J. M., Vineis, P., Forastiere, F., Saldiva, P., Yorifuji, T., & Loomis, D. (2014). Outdoor particulate matter exposure and lung cancer: A systematic review and meta-analysis. In *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.1408092.

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

Hasnain, A., Hashmi, M. Z., Khan, S., Bhatti, U. A., Min, X., Yue, Y., He, Y., & Wei, G. (2024). Predicting ambient $PM_{2.5}$ concentrations via time series models in Anhui Province, China. *Environmental Monitoring and Assessment*, *196*(5), 487. https://www.eea.europa.eu/publications/air-quality-in-europe-2013.

Joshi, P. C., & Mahadev, S. (2011). Distribution of air pollutants in ambient air of district Haridwar (Uttarakhand), India: A case study after establishment of State Industrial



© University of Venda

Development Corporation. *International Journal of Environmental Sciences*, *2*(1), 237–258.

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental pollution*, 151(2), 362-367.

Keuken, M. P., Jonkers, S., Zandveld, P., & Voogt, M. (2012). Elemental carbon as an indicator for evaluating the impact of traffic measures on air quality and health. *Atmospheric Environment*, *61*, 1–8.

Kiai, C., Kanali, C., Sang, J., & Gatari, M. (2021). Spatial Extent and Distribution of Ambient Airborne Particulate Matter ($PM_{2.5}$) in Selected Land Use Sites in Nairobi, Kenya. *Journal of Environmental and Public Health*, *2021*.

Koenig, J. Q. (2000). *Health effects of ambient air pollution: how safe is the air we breathe?* Springer Science & Business Media.

Landrigan, P. J., Fuller, R., Fisher, S., Suk, W. A., Sly, P., Chiles, T. C., & Bose-O'Reilly, S. (2019). Pollution and children's health. *Science of the Total Environment*, *650*, 2389–2394.

Laro, K. O., & Raheem, U. A. (2022). Ambient Air Quality and The Incidence of Selected Diseases: Some Urban Health Observations in Lagos, Nigeria. *International Journal of Environmental Studies and Sustainability*, *2*(1).

LEDET. (2016). Limpopo Environmental Outlook Report. Limpopo Economic Development Environment & Tourism (LEDET).

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, *525*(7569), 367–371. https://doi.org/10.1038/nature15371

Limpopo Department of Economic Development, Environment and Tourism (LEDET). (2013). Provincial Air Quality Management Plan. Limpopo Department of Economic Development, Environment and Tourism, Polokwane.

Madonsela, B. S. (2019). *Assessment of environmental exposure to air pollution within four neighbourhoods of the Western Cape, South Africa*. Cape Peninsula University of Technology.

Maes, J., Vliegen, J., Van de Vel, K., Janssen, S., Deutsch, F., De Ridder, K., & Mensink, C. (2009). Spatial surrogates for the disaggregation of CORINAIR emission inventories. *Atmospheric Environment,* 43(6), 1246-1254.

Mannucci, P. M., & Franchini, M. (2017). Health effects of ambient air pollution in developing countries. *International Journal of Environmental Research and Public Health*, *14*(9), 1–8. https://doi.org/10.3390/ijerph14091048.

Meth, O. (2018). New satellite data reveals the world's largest air pollution hotspot is Mpumalanga–South Africa. *Greenpeace Africa*.

Morakinyo, O. M., Mukhola, M. S., & Mokgobu, M. I. (2020). Concentration levels and carcinogenic and mutagenic risks of $PM_{2.5}$-bound polycyclic aromatic hydrocarbons in an urban–industrial area in South Africa. *Environmental Geochemistry and Health*,



© University of Venda

*42*(7), 2163–2178.

Morapedi, T. D., & Obagbuwa, I. C. (2023). Air pollution particulate matter ($PM_{2.5}$) prediction in South African cities using machine learning techniques. *Frontiers in Artificial Intelligence*, *6*, 1230087.

Munir, S., Habeebullah, T. M., Seroji, A. R., Morsy, E. A., Mohammed, A. M. F., Saud, W. A., Abdou, A. E. A., & Awad, A. H. (2013). Modeling particulate matter concentrations in Makkah, applying a statistical modeling approach. *Aerosol and Air Quality Research*, *13*(3), 901–910.

National Research Council (NRC), 2010. *Global sources of local pollution: An assessment of long-range transport of key air pollutants to and from the United States*. National Academies Press.

Njoku, P. O., Piketh, S., Makungo, R., & Edokpayi, J. N. (2023). Monitoring of subsurface emissions and the influence of meteorological factors on landfill gas emissions: A case study of a South African landfill. *Sustainability,* 15(7), 5989.

Nkosi, V., Wichmann, J., & Voyi, K. (2015). Chronic respiratory disease among the elderly in South Africa: any association with proximity to mine dumps? *Environmental Health*, *14*(1), 33. https://doi.org/10.1186/s12940-015-0018-7 Accessed 08.06.2023

Oosthuizen, M. A., Wright, C. Y., Matooane, M., & Phala, N. (2015). *Human health risk assessment of airborne metals to a potentially exposed community: a screening exercise*.

Paulin, L. M., Williams, D., Oberweiser, C., Diette, G. B., Breysse, P. N., McCormack, M. M., Matsui, E. C., Peng, R., Metts, T. A., & Hansel, N. N. (2013). Indoor air quality in central Appalachia homes impacted by wood and coal use. *Journal of Environmental Protection*, *4*(1), 67.

Pawul, M. & Śliwka, M. (2016). Application of artificial neural networks for prediction of air pollution levels in environmental monitoring. *Journal of Ecological Engineering,* 17**,** 190-196.

Phalen, R. F., & Phalen, R. N. (2013). Introduction to Air Pollution Science: A Public Health Perspective. In *Jones & Bartlett Learning, LLC*.

Pope Iii, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama*, *287*(9), 1132–1141.

Rai, P. K., & Panda, L. L. S. (2014). Dust capturing potential and air pollution tolerance index (APTI) of some road side tree vegetation in Aizawl, Mizoram, India: An Indo-Burma hot spot region. *Air Quality, Atmosphere and Health*. https://doi.org/10.1007/s11869-013-0217-8.

Rood, A. S. (2014). Performance evaluation of AERMOD, CALPUFF, and legacy air dispersion models using the Winter Validation Tracer Study dataset. *Atmospheric Environment*, *89*, 707–720.

Saju, J. A., Bari, Q. H., Mohiuddin, K. A. B. M., & Strezov, V. (2023). Measurement of



© University of Venda

ambient particulate matter ($PM_{1.0}$, $PM_{2.5}$ and $PM_{10}$) in Khulna City of Bangladesh and their implications for human health. *Environmental Systems Research*, *12*(1), 42.

Scorgie, Y, Burger, L. W., & Annegarn, H. J. (2003). Socio-economic Impact of Air Pollution Reduction Measures–Task 2: Establishment of Source Inventories, and Task 3: Identification and Prioritisation of Technology Options. *Report Compiled on Behalf of NEDLAC*, *25*.

Scorgie, Yvonne. (2012). *Urban Air Quality Management and Planning in South Africa*. Phd Thesis, University of Johannesburg (South Africa).

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics: From air pollution to climate change*. John Wiley & Sons.

Semple, S., Ibrahim, A. E., Apsley, A., Steiner, M., & Turner, S. (2015). Using a new, Low-Cost air quality sensor to quantify Second-Hand smoke (SHS) levels in homes. *Tobacco Control*, *24*(2), 153–158. https://doi.org/10.1136/tobaccocontrol-2013-051188.

Shabbir, R., & Ahmad, S. S. (2010). Monitoring urban transport air pollution and energy demand in Rawalpindi and Islamabad using leap model. *Energy*, *35*(5), 2323–2332.

Sharma, N., Agarwal, A. K., Eastwood, P., Gupta, T. & Singh, A. P. (2018). Introduction to air pollution and its control. *Air Pollution and Control***,** 3-7.

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

Sowden, M., Cairncross, E., Wilson, G., Zunckel, M., Kirillova, E., Reddy, V., & Hietkamp, S. (2008). Developing a spatially and temporally resolved emission inventory for photochemical modeling in the City of Cape Town and assessing its uncertainty. *Atmospheric Environment*, *42*(30), 7155–7164.

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2015). Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical*, *215*, 249–257.

StatsSA, Statistics South Africa (2011). Statistics South Africa. Formal Census, 2011. https://www.statssa.gov.za/

Steffen, W., Sanderson, R. A., Tyson, P. D., Jäger, J., Matson, P. A., Moore III, B., Oldfield, F., Richardson, K., Schellnhuber, H.-J., & Turner, B. L. (2005). *Global change and the earth system: a planet under pressure*. Springer Science & Business Media.

Sun, J., Wang, J., Wang, T., & Zhang, T. (2019). Urbanization, economic growth, and environmental pollution: Partial differential analysis based on the spatial Durbin model. *Management of Environmental Quality: An International Journal*, *30*(2), 483–494.

Tongwane, M., Piketh, S., Stevens, L., & Ramotubei, T. (2015). Greenhouse gas emissions from road transport in South Africa and Lesotho between 2000 and 2009.



© University of Venda

*Transportation Research Part D: Transport and Environment*, *37*, 1–13. https://doi.org/10.1016/j.trd.2015.02.017.

World Health Organization (WHO), 2016. Ambient air pollution: A global assessment of exposure and burden of disease. Geneva: World Health Organization. Available https://www.who.int/publications/i/item/9789241511353. Accessed 08.06.2023.

World Health Organization (WHO), 2020. Ambient (outdoor) air pollution: Key facts. Fact sheet. WHO, Geneva. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

World Health Organization. 2018. Ambient (outdoor) air quality and health: Fact sheet. Geneva. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health. Accessed 08.06.2023.



© University of Venda

# CHAPTER TWO

# LITERATURE REVIEW

## 2.1 Preamble

This chapter presents a review of relevant literature on air pollution, gaseous and particulate matter pollutants, and emission inventory for anthropogenic activities. Likewise, air quality monitoring, dispersion of pollutants, and the potential health risks of the air pollutants, as well as the use of machine learning techniques in air quality modelling, particularly their application and effectiveness.

## 2.2 Air Pollution

### 2.2.1 Overview and Health Impacts of Air Pollution

Air pollution is a problem that is affecting developing countries, and Rohde and Muller (2015) suggest that it has killed more people than AIDS, malaria, breast cancer, or tuberculosis. Air pollution is defined as the presence of one or more toxic suspended chemicals or contaminants in the atmosphere at levels that pose a health risk (Brunekreef and Holgate, 2002). Worldwide epidemiological studies show a consistent increase in cardiac and respiratory disease, lung cancer, and mortality from exposure to vehicle exhaust pollution (Hamra *et al*., 2014; Franklin *et al*., 2015). The phenomenon has been reported to cause climate change, and reduce photosynthetic pigment, stomata conductance, net photosynthetic rate, and grain protein contents (Rai and Panda, 2014).

### 2.2.2 Major Air Pollutants and Sources

There are many air pollutants that are suspended contaminants in the atmosphere from different sources (Figure 2.1). It comprises - dust, fumes, smoke, mists, gaseous pollutants, hydrocarbons, volatile organic compounds (VOCs), polycyclic aromatic hydrocarbons (PAHs), and halogen derivatives - which when inhaled in high concentrations by humans can cause several health effects (Rai and Panda 2014; Rai et al., 2013: Seinfeld and Pandis, 2016). WHO describes six major air pollutants or criteria air pollutants - particle matters (PM), ground-level ($O_3$), carbon monoxide (CO), sulfur dioxides, nitrogen dioxides, and lead (Pb) (Phalen and Phalen, 2012). The concentrations of these air pollutants vary widely



© University of Venda

depending on the sources of pollution, their distribution, meteorological conditions, and the topographical features in the vicinity (Dastider *et al*., 2023). According to the WHO, 4.2 million deaths are recorded every year because of exposure to ambient air pollution (Carr, 2020).

### 2.2.3 Global Air Pollution Trends and Monitoring

In 2016, ambient air pollution contributed to about 7.6% of premature deaths worldwide. Approximately 91% of these premature deaths occurred in low and middle-income countries, and the greatest numbers were from the South-Eastern Asia and Western Pacific regions (Figure 2.2). These premature deaths were as a result of human exposure to $PM_{2.5}$ or less in diameter, which could lead to ischaemic heart disease and strokes; 18% of these deaths were due to chronic obstructive pulmonary disease and acute lower respiratory infections and 6% from lung cancer (Babatola, 2018).

The WHO shows that 9 out of 10 people, globally, breathe air comprising a high level of pollutants; 91% of the world's population live in places where the WHO air quality guidelines are not met (WHO, 2018). More than 80% of the people living in urban areas are exposed to pollution levels exceeding WHO guideline limits (WHO, 2016a). Low and middle-income countries suffer from the highest levels of air pollution, hence, they are at higher risk for air pollution-related diseases such as cardiovascular disease, chronic obstructive pulmonary disease, acute lower respiratory infections, and lung cancer. This situation results from household combustion of polluting fuels for fires or stoves for cooking, heating, and lighting (WHO, 2016a).

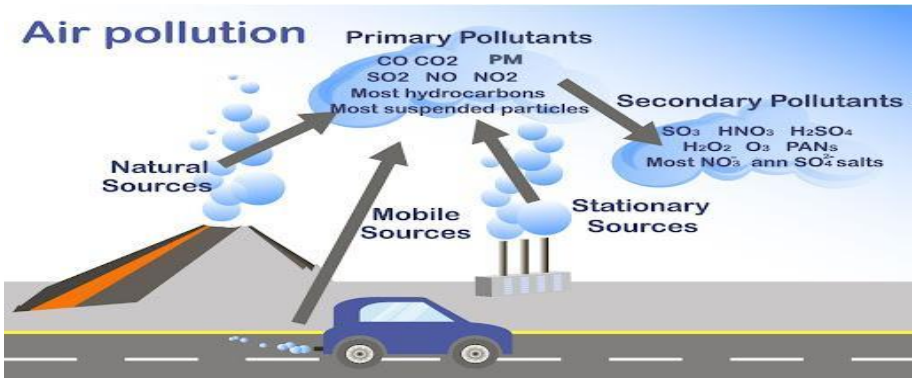



**Figure 2.1:** Major sources of anthropogenic air pollution (Source:www.clarity.io/blog).



© University of Venda

**United States**

According to the United States Environmental Protection Agency (US EPA), the national average concentration of air pollutants declined between 1999 and 2018, and this reduction was largely attributed to improvements in emissions controls on diesel engines and power plants, as well as the use of cleaner fuels for other activities. In the United States (U.S), the national average concentration of air pollutants declined between 1999 and 2018; examples of the air pollutant reduction include, Carbon monoxide (CO) (8-hours) by 74%; Pb (monthly average from 2010) by 82%, $NO_2$ (annual) by 57%; $NO_2$ (1 hour) by 50%; $O_3$ (8 hours) by 21%; $PM_{10}$ (24 hours) by 26%; $PM_{2.5}$ (Annual from 2000) by 39%; $PM_{2.5}$ (24 hours from 2010) by 34%; and $SO_2$ (1 hour) by 89% (US EPA, 2020) (Figure 2.3). This reduction is largely attributed to improvements in emissions controls on diesel engines and power plants, as well as the use of cleaner fuels for other activities. There are still areas of the country that face significant air pollution challenges, despite the overall improvements in air quality. Parts of California, for example, continue to have some of the highest levels of particulate matter and ozone in the country due to the State's large population, transportation infrastructure, as well as certain population groups, such as low-income communities and communities of colour, being disproportionately affected due to factors, such as proximity to major sources of pollution (US EPA, 2020).

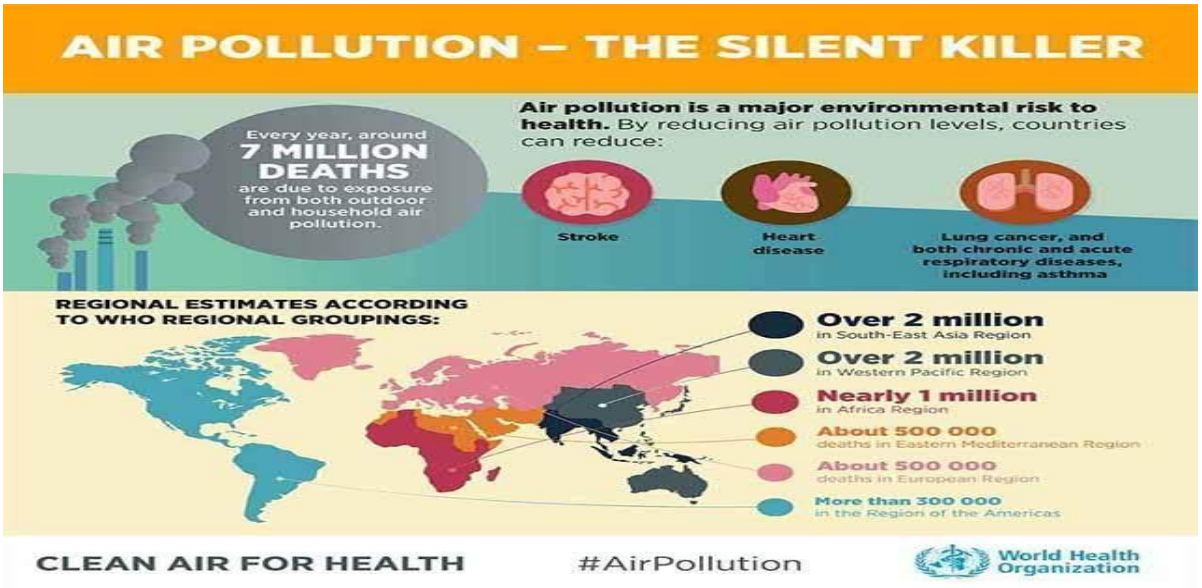



**Figure 2.2:** Illnesses and deaths across the world due to air pollution (Source: WHO, 2016a).



© University of Venda

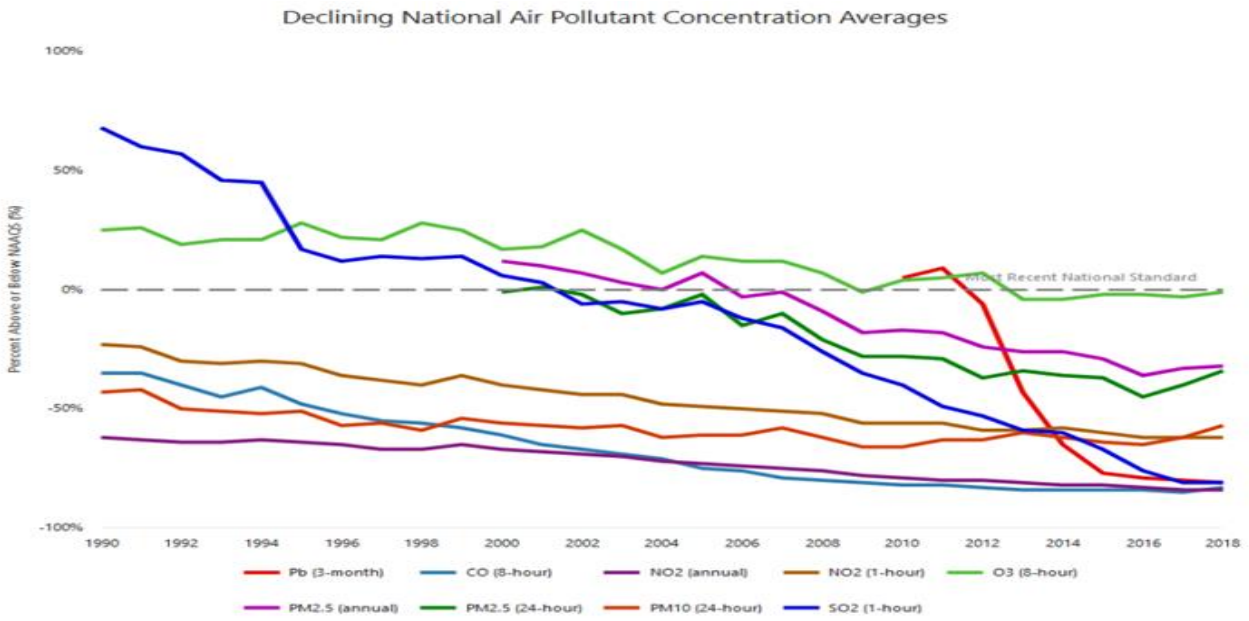



**Figure 2.3:** Shows the percentage of air pollutants in the U.S from the year 1990 to 2018 (Source:US EPA, Air quality system 2020).

A study by Carnegie Mellon University estimated that U.S. air quality worsened after 2016, leading to nearly 10,000 additional deaths in 2018 compared to what would have occurred had pollution levels remained at 2016 benchmarks. Approximately 43% of these excess deaths were recorded in California during the 2018 wildfire season (Burns *et al*., 2018; Carnegie Mellon University, 2019). Furthermore, $PM_{2.5}$ concentrations had increased by about 5.5% since 2016 - a trend attributed to rising natural gas consumption and increased motor vehicle use (Clay & Muller, 2019).

**China**

China is an emerging economy where urbanisation and industrialisation have significantly increased energy consumption across the country, bringing about a tremendous increase in ambient air pollution problems in recent times. The 70s and 80s brought about rolling black smoke emissions from stacks, resulting in some areas experiencing serious acid rain pollution (He, *et al*., 2002). Approximately 1.2 to 2 million premature deaths, as of 2012, were recorded in China from air pollution. Some health effects associated with air pollution in China include cardiopulmonary disease, respiratory dysfunction, and lung malfunction (Rohde and Muller, 2015). China then implemented an Ambient Air Quality Standard, which



© University of Venda

commenced national air monitoring and development; from 2015, 945 sites in 190 cities were in operation (Kim *et al*., 2019). As the years went by, China's ambient air quality continued to improve, however, it is still recorded as one of the most polluted countries in the world. China's air pollutants mainly consist of coal smoke, suspended PM, $NO_2$, Pb, and $SO_2$, a result of fast growth in vehicle usage, coal combustion emissions, and other sources.

In China, the $PM_{2.5}$ concentration levels of northern cities are higher compared to levels in southern China because of enormous emissions of PM and unfavourable meteorological conditions for pollution dispersion, although $SO_2$ and $NO_2$ levels between the northern and southern cities do not differ much. In 2004, major southern cities in China recorded an annual average PM concentration of 102 μg/$m^3$ and the northern parts, an annual average of 140 μg/$m^3$, and 121 μg/$m^3$. The annual average concentration of $SO_2$ and $NO_2$ nationwide were 66 μg/$m^3$, and 38 μg/$m^3$, respectively (Zhang and Cao, 2015). As of 2003, the total health cost associated with ambient air pollution, especially in urban areas, ranges from 157 to 520 billion Chinese yuan, which accounts for 1.2-3.3% of China's Gross Domestic Product (GDP) (Kan, *et al*., 2009). The emission of air pollutants is from different sectors, such as the industrial, transportation, agriculture, and domestic. Table 2.1 summarises the emission estimate for different sectors as of 2010 (Meng *et al.,* 2017).

**Table 2.1**: Emission estimates by sectors in China, 2010 (Gg/year) (Meng et al., 2017).

| Species | Power | Industry | Residential | Transportation | Solvent used | Agriculture | Total (Tg/year) |
|---|---|---|---|---|---|---|---|
| $SO_2$ | 7779.06 | 16372.99 | 4153.77 | 222.98 | | | 28.53 |
| $NO_2$ | 9265.93 | 9344.26 | 1726.82 | 6982.79 | | | 27.32 |
| CO | 3632.28 | 74739.76 | 71281.77 | 20323.67 | | | 169.98 |
| NMVOCs | 66.39 | 7878.63 | 5014.79 | 2351.98 | 7151.05 | | 22.46 |
| $NH_3$ | 0.00 | 272.57 | 427.65 | 25.42 | | 9701.62 | 10.43 |
| $PM_{10}$ | 978.19 | 6060.13 | 8068.53 | 497.78 | | | 15.60 |
| $PM_{2.5}$ | 847.27 | 5805.25 | 4452.15 | 493.31 | | | 11.60 |
| BC | 1.75 | 584.35 | 848.19 | 273.13 | | | 1.71 |
| OC | 0.02 | 578.64 | 2481.50 | 99.61 | | | 3.16 |



© University of Venda

**India**

India is recorded as one of the most populated countries in the world, and air pollutants have exceeded standards recommended by the WHO. Air pollution in India results from vehicular and industrial emissions and fuel use for domestic purposes such as cooking and other miscellaneous activities like burning of household wastes and emissions from small businesses and cremation grounds, although natural sources of air pollution are significantly contingent on seasons and location. The Global Burden of Disease (GBD) 2010 project on the risk associated with household air pollution in India suggests about 1.04 million premature deaths and 31.4 million disability-adjusted lives in the country. Household air pollution resulting from solid cooking fuels associated with $PM_{2.5}$ has contributed to approximately 627,000 premature deaths and 17.8 million disability-adjusted life. The GBD 2010 project in India projected the daily $PM_{2.5}$ exposure for children, women and men were 285 $\mu g/m^3$, 337 $\mu g/m^{3,}$ and 204 $\mu g/m^3$, respectively (Balakrishnan *et al.,* 2014). More recent estimates from the GBD 2021 study, along with recent assessments of ambient air quality in Delhi, further confirm that household and ambient air pollution continue to contribute significantly to premature mortality and disease burden in India (Prasetya *et al.,* 2023, GBD 2021-HAP Collaborators, 2025).

The global model for ambient air pollution exposure, in rural areas suggests a 2010 population weighted annual mean $PM_{2.5}$ of 27.2 $\mu g/m^3$; this is up to 6% rise from the year 1990. These emissions far in excess of the WHO Air quality levels (Balakrishnan, *et al*., 2014). The Indian national ambient air quality standards (NAAQS) for annual average of $PM_{10}$ and $PM_{2.5}$ are 60 and 40 $\mu g/m^3$, respectively. These are comparable to the WHO interim target-1 guideline values, but much higher than the recommended WHO air quality guideline values of 20 and 10 $\mu g/m^3$ or the U.S. Environmental Protection Agency (EPA) annual PM standard of 12 $\mu g/m^3$. The Indian NAAQS focuses mainly on the urban areas with extremely limited rural monitoring in place, although two-thirds of the population of India live in rural areas (Balakrishnan, *et al*., 2014; GHOSH, 2015). One-fourth of ambient PM in India is associated with solid household cooking fuels, hence, it will be very difficult to reach the current NAAQS levels without reducing household emissions, in addition to addressing



© University of Venda

vehicular, industrial, and other emissions from known urban sources (Balakrishnan, *et al*., 2014).

Former efforts in improving air pollution emissions from households using solid cooking fuels, have been with the developing of fuel-efficient biomass cook stoves (Smith & Pillarisetti, 2017). There has been, however, limited gainful education for the population on emissions or exposure or biomass cook stove technologies which tend to meet the theoretical-minimum-risk exposure distribution for household air pollution and ambient air pollution (Smith & Pillarisetti, 2017).

**Nigeria**

In Nigeria, numerous studies have shown extreme levels of ambient air pollution in most urban cities of the country. Typical air quality assessments in most cities showed that the levels of $NO_2$, $SO_2$, and total PM exceeded the existing federal government agency standards (Aliyu & Botai, 2018). The 2018 world air quality report ranks Nigeria as the 10$^{th}$ most polluted country in the world. The common sources of air pollution in these cities include bush burning, automobile emissions, generator emissions, pipeline explosions, industrial emissions, and gas flaring. In 2017, more than 114,000 people died in Nigeria from air pollution, which is recorded as the highest in Africa (Ana, 2011; Egbedi, 2017). Onitsha, which is one of the major business hubs in the South-East of Nigeria, was showing the most populated $PM_{10}$ emissions in the world as of the year 2016, at an annual mean concentration of 594 µg/m$^3$ (Egbedi, 2017); this is approximately 30 times above the WHO annual guideline of 20 µg/m$^3$. There has been, however, limited ambient air quality monitoring data in the country, leading to a limited track of air pollution in the country.

Nigeria has recorded unprecedented mortality rates of 307.4 for every 100,000 persons. In 2018 the average mean concentration in the country for $PM_{2.5}$ was 44.8 µg/m$^3$, which is approximately 4 times above the WHO guideline for ambient air quality (AirVisual, 2018). Kano state was recorded as the most polluted city in Africa (Offor *et al*., 2016). In 2017, over 64,000 deaths from household air pollution in Nigeria were reported by the Health Effects Institute and Institute for Health Metrics and Evaluation, which was a result of the burning of solid fuels, like charcoal and wood for cooking. There has been a decline of air pollutants



© University of Venda

into the environment from the year 2000, however, poor ambient air quality remains a major health risk in Nigeria, and it is related to thousands of premature deaths annually.

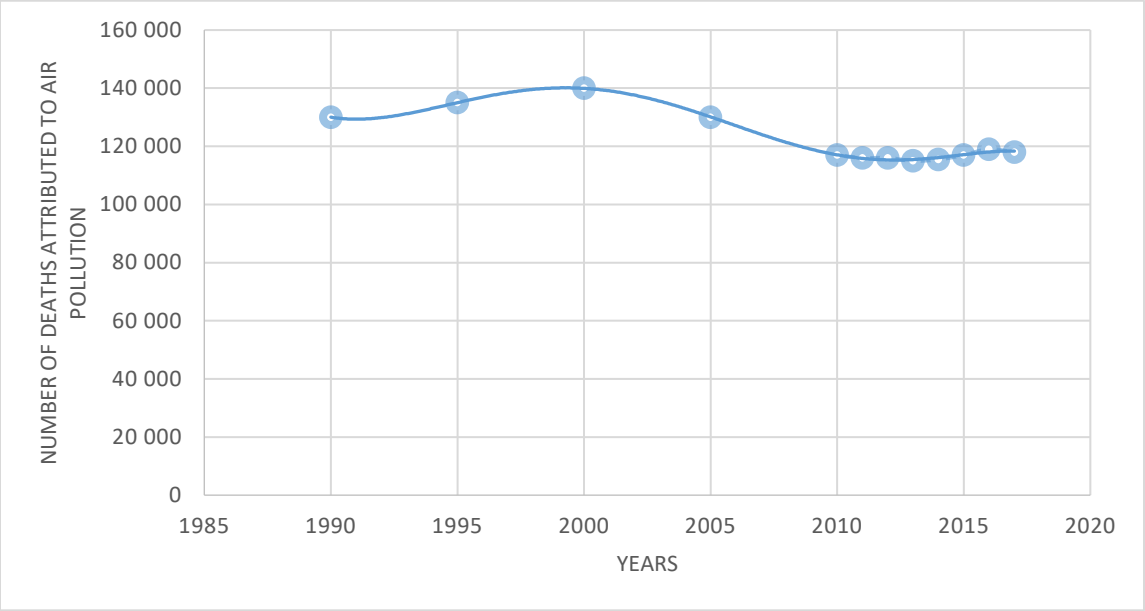



**Figure 2.4:** The number of deaths attributable to household air pollution in Nigeria from 1990 to 2017 (Chasant, 2019).

### South Africa

South Africa has national and international legal responsibilities to cut its air pollutant emissions. The Bill of Rights section of the South African Constitution includes an environmental clause, which states that - *everyone has the right to – (a) an environment that is not harmful to their health or well-being; and (b) have the environment protected, for the benefit of present and future generations, through reasonable legislative and other measures* (Republic of South Africa, 1996, Section 24). This is to be achieved through:

i. *Prevent pollution and ecological degradation;*

*ii. Promote conservation; and*

*iii. Secure ecologically sustainable development and use of natural resources while promoting justifiable economic and social development.*



© University of Venda

According to the Constitution of the Republic of South Africa (1996), all citizens have the right to an environment that is not harmful to their health or well-being (Section 24), and the right to access information held by the state, including environmental data (Section 32). This establishes a legal obligation on the government to ensure that South Africans breathe clean air that is not detrimental to human health (Van der Bank & Van der Bank, 2014). In line with these constitutional provisions, the government has established over 130 fully automated air quality monitoring stations, which form part of the National Ambient Air Quality Monitoring Network (NAAQMN), covering areas of high population density across the country (Setshedi, 2017). These stations provide vital data that comply with both national and international ambient air quality standards and offer essential information on the levels of air pollution to which South Africans are exposed. To ensure transparency and public access, real-time data from these stations is made available through the SAAQIS, accessible via its website and mobile application (Gwaze & Mashele, 2018). The importance of enforcing these rights was further underscored in the landmark *Deadly Air* case (*Minister of Environmental Affairs v Trustees for the Time Being of Ground Work Trust & Others*, 2025). In the case, the Supreme Court of Appeal confirmed that ongoing poor air quality in the Highveld Priority Area constituted a violation of Section 24(a), and ruled that the State has an immediate duty to implement and enforce air quality standards in accordance with its constitutional obligations (CER, 2025).

The understanding of ambient air pollutant measurements is very complex, especially issues concerning the measurement of air pollutant concentrations in association with the quality of air and its health effects. The Department of Environmental Affairs (DEA) established an official country-specific Air Quality Index (AQI) in line with best global practices, to make the data obtained simplified for the general public. The AQI was derived from the six criteria for pollutants ranging in a ten-point scale, from *good air quality* (scale 1) to *hazardous* (10) based on National Ambient Air Quality Standards (NAAQS). Also, there are five bands of air pollution ranging from ‘Low’, ‘Moderate’, ‘High’, ‘Very High’ to ‘Hazardous’ levels. The ‘Low’ band indicates that it is unlikely that there are any adverse effects on human health on short-term exposure, however, the ‘hazardous’ band indicates that the air pollution levels will activate emergency health conditions if exposed to humans, even for short periods (Gwaze and Mashele, 2018).



© University of Venda

South Africa's dependence on coal and increase in industrial and vehicular activities affects its emissions policies with the country, currently, emitting 10 times more nitrogen oxide than China and Japan. Satellite data across the six continents shows that a province in the east of South Africa was the world's largest $NO_2$ hotspot (Adeyemi, 2020). This region is where dozens of coal-fired power stations owned and operated by Eskom are situated. In South Africa, the worst air-polluted area is found in the suburb of Three Rivers in Vereeniging and the nearby townships of Sharpeville (at Vanderbijlpark) and Zamdela (at Sasolburg). In the Mpumalanga Highveld, the towns with the most heavily polluted air are Standerton, Delmas, eMalahleni (formerly Witbank), Middelburg, Hendrina, and Ermelo. In the Waterberg region, the towns of Lephalale (formerly Ellisras), Thabazimbi, and Mokopane (formerly Potgietersrus) were singled out for their polluted air (Hallowes and Munnik, 2006).

According to Joubert, (2019) air pollution levels generally in South Africa have remained largely stable over the last few years. In Cape Town, air pollution levels are low, however, high levels of air pollution were measured at the Wallacedene informal settlement near Kraaifontein, in Khayelitsha, and Cape Town's city center. In Johannesburg, air pollution is approximately six times higher than in Cape Town, with the townships of Alexandra, Diepsloot, and Ivory Park being the worst air-polluted areas. In Pretoria, air pollution levels are approximately three times higher than in Cape Town. Pretoria has its highest air-polluted areas in Olivenhoutbosch (near Centurion); Pretoria West in second place, and the suburb of Booysens third. In East Rand, air pollution levels are lower than in Johannesburg but higher than in Pretoria with the worst air-polluted area being in the Etwatwa Township near Springs, next to the Wattville area and Germiston. Likewise, in the city of Durban, the most polluted areas are the city centre, as well as Chatsworth, and Wentworth which are the industrial hub of Durban.

The elevated levels of inorganic and organic gaseous pollutant species are influenced by diverse natural and anthropogenic activities like power generation (coal-fired power stations), agriculture, mining operations, coal dumps, and transportation, in the Highveld of South Africa. The largest emissions from anthropogenic activities in the savanna region compared to other regions are found in South Africa, which is the most industrialized part of the continent (Piketh and Walton, 2004). Concentrations of $NO_2$, $O_3$, and $SO_2$ are at their



© University of Venda

peaks during winter, therefore, vary due to seasonal meteorological conditions (Lourens *et al*., 2011). Previous studies have shown that the concentrations of ozone are highest in the spring period (Feig *et al*., 2014; Lourens *et al*., 2011). According to Buthelezi and Davies (2015), exposure to emissions of CO, $O_3$, and $NO_2$ from industrial activities (mainly the transportation industry) in the Umlazi Township, South of Durban, KwaZulu-Natal Province is risky to humans and the environment.

### 2.2.4 Policy and Technological Interventions

There have been efforts made through South Africa's legislative and institutional frameworks, despite this, the persistent challenge of poor air quality, particularly in heavily-industrialised and densely-populated regions, highlights the need for more robust and innovative responses. To address these concerns effectively, both national and global trends have shifted toward integrated policy approaches and technological interventions. The next section examines various policy instruments and technological solutions that have been adopted in different contexts to mitigate air pollution, enhance monitoring capabilities, and promote cleaner production and consumption practices.

Policies and regulations which support cleaner transport, energy-efficient homes, power generation, industry, and better municipal waste management are essential in minimising key sources of ambient air pollution. Several successful interventions implemented globally across various sectors including transport, urban planning, and power generation, have contributed to notable reductions in air pollution (Gwilliam *et al*., 2004). Introducing clean technologies, for example, reducing industrial smokestack emissions and capturing methane ($CH_4$) from agricultural waste sites for use as biogas, are proven measures.

In addition, individuals can access affordable clean household energy solutions for cooking, heating, and lighting. The shift to clean modes of power generation, prioritizing rapid urban transit, walking and cycling networks in cities, as well as rail interurban freight and passenger travel, are all critical. Similarly, transitioning to cleaner heavy-duty diesel vehicles and adopting low-emission vehicles and fuels, including fuels with reduced sulfur content, can have significant impacts. Improving the energy efficiency of buildings and developing greener, more compact cities, can enhance overall sustainability. Renewable, combustion-



© University of Venda

free power sources, such as solar, wind, and hydropower, along with distributed energy generation systems like rooftop solar panels, have become increasingly viable. Strategies such as waste reduction, separation, recycling, and reuse or reprocessing further contribute to pollution control.

A practical example is the use of improved cooking stoves in rural Mexico, which has led to a reduction in respiratory symptoms (Thomas *et al.*, 2015). In addition to other charges, the city of London in 2018 introduced a daily special tax (Toxicity (T) Charge) of £10, for the most polluting vehicles entering the city. Cities such as Madrid and Paris have moved to ban old or diesel-powered cars, with similar measures adopted in various German cities. German cities have already started implementing limits on polluting cars. The US EPA introduced Tier 3 vehicle emission standards in 2017, which treat vehicle and fuel as an integrated system (Tier, 2014). These standards are expected to reduce both tailpipe and evaporative emissions from passenger vehicles (Tier, 2014).

### 2.3 Air Pollutants

Air pollutants include gases and PM that are found in the atmosphere at normal temperature and pressure. The six major criteria air pollutants include - particle matter (PM), ground-level ($O_3$), carbon monoxide (CO), sulfur dioxides ($SO_2$), nitrogen dioxides ($NO_2$), and lead (Pb) from different sources (Phalen and Phalen, 2012: Sharma *et al.*, 2018). Air pollutants vary widely depending on the sources of pollution, their distribution, meteorological conditions, and the topographical features in the vicinity (Dastider et al., 2023). The pollutant mass/volume concentration is usually measured as micrograms of pollutant per cubic meter of air ($\mu g/m^3$) or parts per million (ppm) (Sharma *et al.*, 2018).

#### 2.3.1 Particulate matter (PM)

PM is the mixture of particles and pollutants present in the air. These particles are a complex mixture of organic and inorganic particles which are a major part of air pollutants. PM is grouped not based on its geometric diameter, but its aerodynamic diameter and varied in - sizes (diameters), shapes, surface area, chemical composition, solubility, and redox reaction (Kero *et al.*, 2015). According to their sizes or diameter, PM is typically categorized



© University of Venda

into Total Suspended Particles (TSP) (which includes all particles up to 30 μm); Coarse particles (which includes particles with diameter from 2.5 – 10 μm ($PM_{2.5}$ – $PM_{10}$)); Fine particles (which includes particles with diameter less than or equal to 2.5 μm ($PM_{2.5}$); Very fine particles (includes particles with a diameter of less than or equal to 1 μm); and Ultrafine particles (which includes particle with a diameter of less than 0.1 μm (from 1 to 100 nanometres)). Particulate matter in air pollutants has both short-term and long-term effects. Recent publications suggest that finer particles ($PM_{2.5}$) are mostly associated with human health problems. The aerodynamic diameter of a PM shows a significant role in determining where within the respiratory tract the PM will finally settle in the body. Larger PM tends to be deposited in the upper section of the respiratory tracts, while smaller PM gets deposited in the lower part (alveolar area) of the lungs as illustrated in Figure 2.5 (Heyder, 2004). Particles that are smaller than 10 micrometres in diameter or about 1/7 the diameter of a single human hair or a fine beach sand, will get trapped in the lungs, while the smallest can pass through the lungs into the bloodstream (American Lung Association, 2013). Figure 2.6 shows an estimated 2020 annual average $PM_{2.5}$ concentrations ($\mu g/m^3$) around the world (Institute for Health Metrics and Evaluation, 2025).

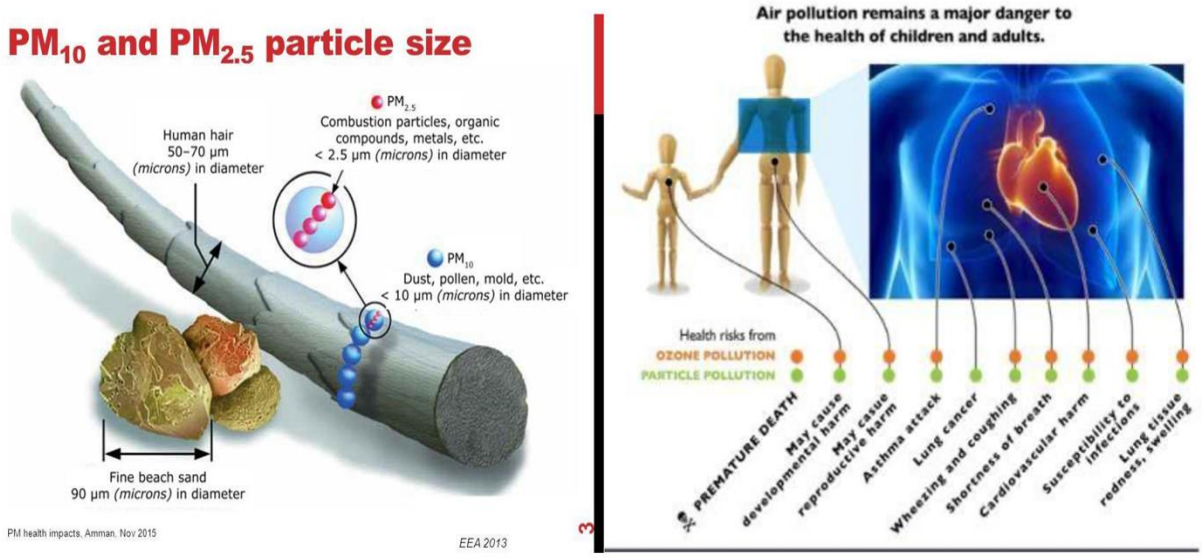



**Figure 2.5:** Overview of the relative sizes of $PM_{10}$ and $PM_{2.5}$ in relation to human hair (US EPA, 2008; American Lung Association, 2013).



© University of Venda

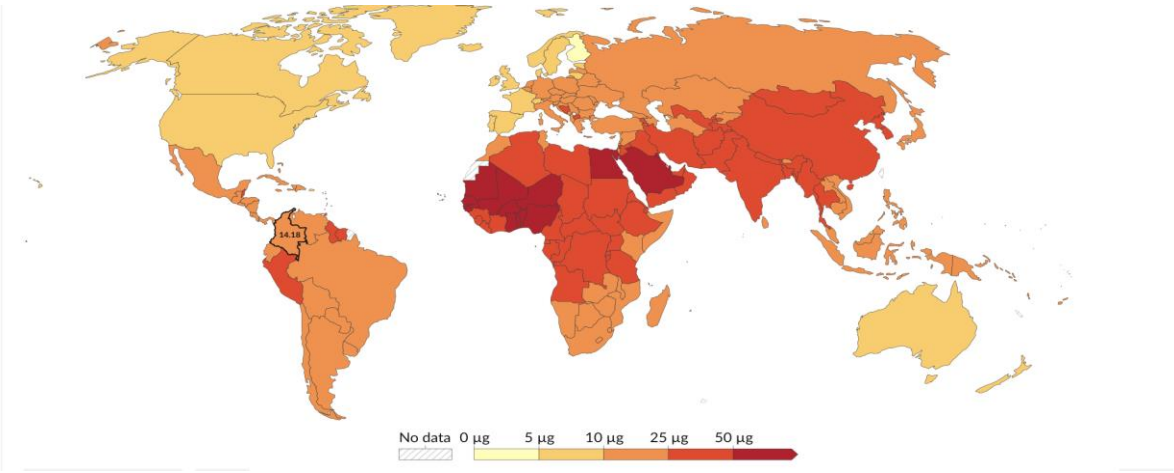



**Figure 2.6:** Estimated 2020 annual average $PM_{2.5}$ concentrations (µg/m$^3$) (Source: Institute for Health Metrics and Evaluation, 2025).

$PM_{2.5}$ predominantly originates from emissions during combustion like the burning of gasoline and diesel in motor vehicles, burning of coal-powered plants, and wood burning; they can be derived from industrial processes like smelters, paper mills, cement generation plants, and steel mills. Some constituents of $PM_{2.5}$ include transformation products like sulphate and nitrate particles – generated by conversion from primary Sulphur and nitrogen oxide emissions and secondary organic aerosol from Volatile Organic Compound (VOC) emissions (HEI, 2019). Ultrafine particles, however, are derived from initial combustions from vehicle exhaust and atmospheric photochemical reactions and have a very small life span - from minutes to hours - which grows as time goes on through coagulation and condensation to produce larger complex aggregates, adding up to the $PM_{2.5}$ – $PM_{10}$ ranges (HEI, 2019). Coarse particles are primarily resulting from the suspension or resuspension of dust, soil, or other crustal materials from roads, farming, mining, brakes, windstorms, and volcanoes; they also include sea salts, pollen, mold, spores, and other biological materials, depending on their origin (Morakinyo *et al.,* 2016).

The World Health Organization's air quality guidelines estimate that reducing the annual average of $PM_{2.5}$ concentrations from levels of 35 µg/m$^3$, commonly found in many developing cities, to the WHO guideline level of 10 µg/m$^3$, could reduce air pollution-related deaths by around 15% (WHO, 2018). In European countries, where PM concentrations in many cities do conform to WHO guideline levels, it is estimated that average life expectancy



© University of Venda

is 8.6 months lower than it would be, due to PM exposures. The effects of PM on human health are derived from the pathogenic effects of the particles and the area in which they are deposited in the lungs. Respirable particles that are smaller than 2.5 micrometres are among the most dangerous particulate matter because they can be drawn into the lungs, where they damage respiratory tissues (Choo and Jalaludin, 2015), triggering myocardial infection (IM) (Shah *et al*., 2013), cardiac arrhythmias and heart failure (Mill *et al*. 2007).

There are five methods or processes by which inhaled PM can be deposited in the respiratory system of humans, it includes impaction, settling, diffusion, interception, and electrostatic precipitation. Most PM with an aerodynamic diameter of more than 5 µm will deposit in the upper airways or larger lower airways from impaction. The impaction process is accountable for the largest amount of deposition in the respiratory system, based on the mass of the PM. The settling process consists of smaller PM (less than 5 µm), which passes further down the respiratory system into the distal airways. Diffusion occurs when mostly fine and ultrafine PM diffuse in the distal airways and consequently move into the bloodstream. The human body has mechanisms which it uses to fight PM particles in the body, however, adverse health effects from PM inhalation have been recorded by many scholars. Ultrafine particulates (< 0.1 µm) can pose a specific threat to human respiratory health, because of its very small particle size. These ultrafine particles can move further down into the lungs (alveolar region). Based on certain factors exhibited by ultrafine particles, for example, they can occur in abundance more than other PM size ranges (Kwon *et al.,* 2020). These particles have irregular geometry and a higher surface area-to-mass ratio, and due to their abundance in air and their irregular geometry, they have the potential to act as a carrier of hazardous gases (Kwon *et al.,* 2020). The ultrafine particles tend to go into deeper areas of the lungs than other larger particles.

Similar to the ultrafine particles, $PM_{2.5}$ is also of great importance due to the recorded adverse effects of long-term and short-term exposure to humans. The chemical composition of $PM_{2.5,}$ if analysed properly, can help understand the origin of the particle matter (Cao *et al*., 2012; Tan *et al*., 2017), and the understanding of their sources will help identify the main contributors of $PM_{2.5}$ in the environment. Scholars have associated $PM_{2.5}$ long-term or



© University of Venda

chronic exposures with premature death and short-term exposure to acute cardiorespiratory morbidity (Nkosi *et al*., 2016). The increase in oxidative stress, inflammation, and nervous system dysfunction, which could lead to bronchitis, asthma, myocardial infarction, and stroke, are a result of chronic exposure to $PM_{2.5}$ (Nkosi *et al*., 2016). The South African ambient air quality standard for the criteria pollutants is shown in Table 2.2. Developing countries like South Africa, have a heavy reliance on fossil fuels, resulting in productivity losses and mortality due to high concentrations of air pollution, namely, fine PM. A recent International Growth Centre (IGC) study indicates that 7.4% of all deaths in South Africa in 2012 were due to chronic exposure to fine PM, costing the country up to 6% of its GDP. The GBD study estimated that South Africa had 1800 deaths in 2012 attributable to fine PM and this number was based on global satellite and modelling views on the severity of air pollution in the country, however, measured fine PM values are 3-4 times higher than the GBD estimates (Altieri and Keen, 2016). At monitoring stations, in South Africa, where the fine PM data was available, only one station met the health guidelines recommended by the World Health Organization (Altieri and Keen, 2016).

**Table 2.2:** South African ambient air quality standard for $PM_{2.5}$, $PM_{10}$, $SO_2$, $NO_x$ and CO (DEAT 2009 and Garland et al., 2021)

| Pollutants | Averaging period | Concentration (µg/m$^3$) | Frequency of Exceedance | Compliance date |
|---|---|---|---|---|
| $PM_{2.5}$ | 24 hours | 40 | 4 | 1 January 2016 – 31 December 2029 |
| | 1 year | 20 | 0 | 1 January 2016 – 31 December 2029 |
| $PM_{10}$ | 24 hours | 75 | 4 | 1 January 2015 |
| | 1 year | 40 | 0 | 1 January 2015 |
| $SO_2$ | 10 minutes | 500 µg/m$^3$ (191 ppb) | 526 | Immediately |
| | 1 hour | 350 µg/m$^3$ (134 ppb) | 88 | Immediately |
| | 24 hours | 125 µg/m$^3$ (48 ppb) | 4 | Immediately |
| | 1 year | 50 µg/m$^3$ (19 ppb) | 0 | Immediately |
| $NO_2$ | 1 hours | 200 µg/m$^3$ (48 ppb) | 88 | Immediately |
| | 1 year | 40 µg/m$^3$ (19 ppb) | 0 | Immediately |
| CO | 1 hour | 30 mg/m$^3$ (26 ppm) | 88 | Immediately |
| | 8 hours (calculated on 1-hour averages) | 10 mg/m$^3$ (8.7 ppm) | 11 | Immediately |
| $O_3$ | 8 hour (running) | 120 µg/m$^3$ (61 ppb) | 11 | Immediately |



© University of Venda

### 2.3.2 Carbon monoxide (CO)

CO is a colourless and odourless gas, which is formed by incomplete combustion of fossil fuel, especially materials containing carbon, such as coal, wood, gasoline, kerosene, oil, and propane. The health risks associated with CO vary with its concentration and duration of exposure which can be deadly, especially through accidental smoke or exhaust inhalation in enclosed areas. CO emission sources are predominantly cigarette smoke, vehicle emissions, wood stoves, gas stoves, and ovens emissions (Seguel *et al*., 2017). The United States Environmental Protection Agency has estimated as much as 95% of CO comes from vehicles. A high level of CO is harmful to human health because CO has a great effect on oxygen delivery to the body's organs (for example, heart and brain) and tissues (for example, skin) (US EPA, 2006).

Effects range from subtle cardiovascular and neurobehavioral effects at low concentrations to unconsciousness and death after prolonged exposures or after acute exposures to high concentrations of CO (Raub *et al*., 2000). Normally, CO will cause headaches and even visual impairment, however, at comparatively high levels, CO can directly cause death, especially in people with heart diseases (US EPA, 2006; HEI 2010). CO effects on human health is to increase the presence of ischemic heart or peripheral vascular disease, decrease blood pressure, headaches, dizziness and vomiting. Ischemic heart disease has the capacity to impair exercise capability, heart attack, and ultimately premature death (Wenger *et al*., 2010). Carbon monoxide (CO) is attracted to hemoglobin with an affinity approximately 250 times greater than that of oxygen. At low carboxyhemoglobin (COHb) levels, around 2%, health effects are generally insignificant, however, exposures exceeding 40% can lead to severe health impacts. The severity of CO poisoning depends on both the concentration of CO and the duration of exposure. Long-term effects of CO exposure can cause anoxic brain injury, neurological problems (US EPA, 2006). CO is at a high level in South Africa, especially from motor vehicles Figure 2.7 (DEA, 2012).



© University of Venda

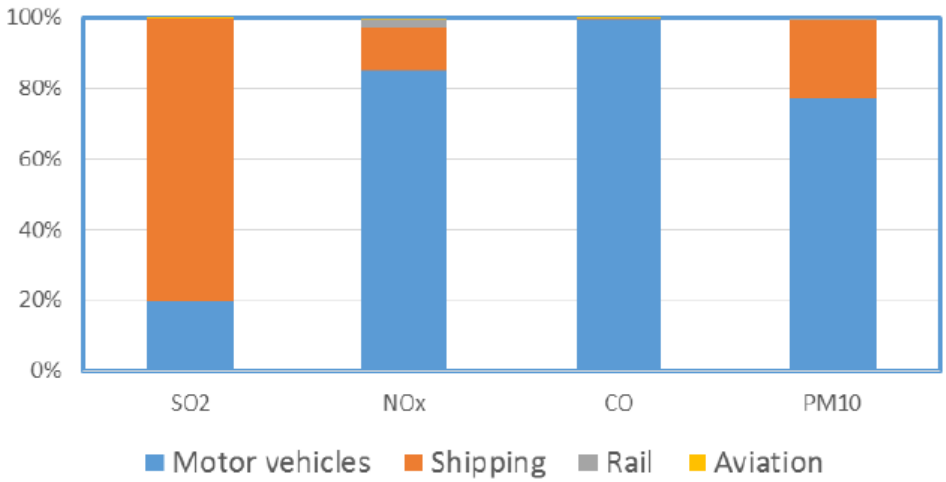



**Figure 2.7:** Transportation sector emissions showing high CO from motor vehicles. Source: (DEA, 2012).

### 2.3.3 Sulphur dioxide ($SO_2$)

$SO_2$ is a colourless, poisonous, and highly reactive gas that is mostly emitted from both natural and anthropogenic activities like the burning of fossil fuels, volcanic activities, and in various industrial processes (by roasting metal sulphide ores). Anthropogenic sources of $SO_2$ emissions are estimated for more than 70% of the emissions, of which half of it is from the combustion of fossil fuels. It is estimated that 65 million tonnes of $SO_2$ per year enter the atmosphere because of man's activities, primarily from the combustion of fossil fuels, to which the energy producing companies using coal are by far the greatest contributor (DEA, 2012). In the United States, it is estimated that almost 65% of $SO_2$ emissions are a result of coal-fired power stations (Massetti, et al., 2017). $SO_2$ can be detected by taste and odour in concentrations that ranged between 0.38 - 1.15 ppm and above 3 ppm, it has an irritating odour; at a concentration of 10,000 µg/m$^3$, it has a pungent smell. The breathing in of $SO_2$ through the mouth penetrates the lungs more than breathing through the nose. $SO_2$ has a relatively short lifespan in the atmosphere and when inhaled. When $SO_2$ enters the airways of the body, it is deposited along the airways and reaches the upper lungs and immediately passes through the mucous membrane into the blood where it becomes mixed with the alpha-globulin fraction of plasma (Mundial, 1998). In the moist pulmonary environment, $SO_2$ produces sulphurous acid which is a severe irritant. These in turn produce $H^+$, bisulphate, and sulphite which affect the muscles and nerves involved in bronchoconstriction. The effects of short-term exposure to $SO_2$ include difficulty in breathing, coughing, irritation of



© University of Venda

the lungs, throat, and shortness of breath. The long-term effects of $SO_2$ exposure include temporary loss of smell, headache, bronchitis, reduced fertility, dizziness, respiratory irritation, and aggravation of any existing cardiovascular disease. A dosage of $SO_2$ in the human body at a concentration of 5 ppm causes dryness of the nose and throat and at 400 ppm to 1000 ppm will likely cause immediate danger to life, respiratory depression, and premature death (Zhu and Maibach, 2020).

### 2.3.4 Nitrogen dioxide ($NO_2$)

Nitrogen oxides are primarily nitric oxide (NO), which are produced as a result of combustion at high temperatures, for example, emissions from motor engines and power plants. Once emitted, NO rapidly reacts with $O_3$ or $O_2$ to form $NO_2$ (secondary $NO_2$). Nitrogen dioxide is a suffocating, brownish gas, one of a family of highly reactive gases - the nitrogen oxides (NOx) and they are formed when fuel is burned at high temperatures. Nitrogen dioxide is also an irritant to humans and corrosive to metals. Primary $NO_2$ (that is, $NO_2$ directly emitted from a source rather than formed in the atmosphere) is usually emitted from diesel engines and a high amount of it can also be found in indoor air, especially from unventilated heaters and gas stoves. $NO_2$ is, therefore, a significant ambient gas that is commonly less toxic than $O_3$, but can increase the risk of respiratory disease if exposed in high concentration to humans (US EPA, 2006). The exposure of $NO_2$ of about 2 – 5 ppm interferes with the T-lymphocytes, particularly $CD8^+$ cells and natural killer cells that play an important role in host defences against viruses (Ghorani-Azam *et al*., 2016). Epidemiological studies show that $NO_2$ exposure causes respiratory infections in children, like coughing and wheezing. Other illnesses like eye, nose irritations, dyspnea, chest pain, headache, diaphoresis, fever, and pulmonary edema may also occur, although, exposure at concentrations between 0.2 – 0.6 ppm is harmless to humans (Latza, *et al.,* 2009; Samoli *et al.,* 2006).

### 2.3.5 Ground-level Ozone (GLO)

Ozone ($O_3$) is a colourless gas and a major component of the atmosphere. $O_3$ can be found at the ground level and the upper region of the atmosphere called the “troposphere”. GLO is formed as a result of photochemical reactions between nitrogen oxides (NOx) and Volatile Organic Compounds (VOCs) emitted from natural sources and/or due to anthropogenic



© University of Venda

activities in the presence of heat and sunlight (Hanna *et al*., 2011). The major anthropogenic sources are coal-fired plants and vehicle emissions, therefore, the emission of VOCs and NOx contributes to the seasonal differences in GLO concentrations in a particular area (Hanna *et al*., 2011). The formation of GLO is very sensitive to the amount of sunlight and temperature difference in an area; for example, if the ambient temperatures and sunlight levels remain high for several days and the air is relatively stagnant, GLO and its precursors can build up and produce more GLO than typically would occur on a single high-temperature day. Also, GLO can be conveyed into an area as a result of the emission from other sources found very far away, which results in highly elevated GLO levels even in areas of low emission of VOCs and NOx (NRC, 2010).

The GLO has plausible effects on health and the environment (US EPA, 2006). GLO is a powerful oxidant that accepts electrons from other molecules. There is a high level of polyunsaturated fatty acids in the surface fluid lining of the respiratory tract and cell membranes that underlie the lining fluid. The double bonds obtainable in these fatty acids are unstable, and GLO attacks unpaired electrons to form ozonides and progress through an unstable zwitterion or trioxolane (depending on the presence of water). These ultimately recombine into lipohydroperoxides, aldehydes, and hydrogen peroxide. These pathways are thought to initiate the propagation of lipid radicals and auto-oxidation of cell membranes and macromolecules. Children and older people are at high risk of GLO when sunny, so their respiratory illness and exacerbations of asthma are linked to GLO (US EPA, 2006). According to Sheffield *et al*. (2011), during the 1996 Olympic games in Atlanta, there was a 28% peak drop in GLO, which led to an 11% reduction in pediatric emergency department visits for asthma.

### 2.3.6 Lead (Pb)

Lead (Pb) is a highly toxic chemical element in the carbon group with symbol Pb, from the Latin word *Plumbum*, meaning “the liquid silver” with an atomic number of 82, it is a very poisonous and a global environmental toxicant. It is distributed in three main forms - metallic lead, lead salts, and organic lead-containing carbon - and its pollution results from both indoor and outdoor sources. The main sources of occupational Pb exposure are traditionally



© University of Venda

leaded gasoline, Pb-based paints, and batteries. Also, irrigation water wells and wastewaters, are other emission sources of the Pb into the environment. The exposure of Pb to humans has been generally known as a significant public health problem over the years and its absorption by humans is influenced by particle size and concentration. Approximately 90% of Pb distributed in the ambient air that is inhaled is small enough to be retained in the body. Pb at low doses in fetuses and children is dangerous (Ara and Usmani, 2015). An important milestone in air quality management was the phasing out of leaded petrol, which led to a substantial reduction in ambient lead concentrations worldwide and in South Africa (Pacyna & Pacyna, 2001). Following regulatory actions, including the elimination of leaded petrol in South Africa in 2006, significant declines in blood lead levels and atmospheric lead pollution have been recorded (Mathee *et al*., 2006). This reduction represents a key public health success, particularly for protecting children from the neurotoxic effects of lead exposure.

High levels of exposure to Pb by humans are now strictly controlled due to its adverse health effects. Pb potentially brings about oxidative stress and evidence shows a role of oxidative stress in the pathophysiology of lead toxicity. Pb is capable of inducing oxidative damage to the brain, heart, kidneys, and reproductive organs. The mechanisms for lead-induced oxidative stress include the effects of lead on membranes, DNA, and antioxidant defence systems of cells (Flora *et al.,* 2012). Exposure to lead is harmful and is associated with hypertension, kidney disease, neurodegenerative disease, cognitive impairment behavioral abnormalities, hearing deficits, neuromuscular weakness, and impaired cognitive functions in both humans and animals. A study on an assessment of blood of traffic police officers showed a higher levels of Pb, possibly from environmental pollution due to their jobs. Pb accumulates in the blood, bone, and soft tissue because it is not readily excreted, and Pb can affect the kidneys, liver, nervous system, and other organs (Flora *et al.,* 2012).

### 2.4 Drivers of Pollution and Emission by Sectors

The following sections provide an overview of the emission inventory by sector for the emission estimate.



© University of Venda

### 2.4.1 Residential Sector

About half of the world's population uses fuel wood, charcoal, and non-woody biofuels daily. Globally, the combustion of biofuel is the main cause of trace gases in the atmosphere, contributing to about 17% carbon dioxide, 13% carbon monoxide, and 6% nitrous oxide. WHO suggests that about 4.3 million premature deaths occur as a result of household air pollution every year, with about 3.5 million premature deaths occurring as a result of household air pollution directly and 0.5 million premature deaths occurring because of exposure to outdoor air pollution, which is associated with solid fuel combustion (WHO, 2014).

Burning of solid fuels is the primary source of domestic energy for those living in low-income settlements in developing countries, likewise in Africa, and this is practiced by approximately 3 billion people around the world (Bonjour *et al*., 2013). Most domestic fuel burning, however, is an inefficient and incomplete combustion process that takes place in the developing world due to economic reasons, where vital energy needs for cooking, heating, and lighting have to be met by biofuels. This act produces several harmful substances, such as $SO_2$, PM, CO, NOx, VOCs, and heavy metals that might have severe health consequences from an indoor air pollution source (Friedl *et al*., 2008). This residential fuel burning is a problem that has been gaining increasing concern over the past few years, especially in underdeveloped and developing countries, where domestic fuel burning is done regularly. A variety of fuels burnt emit different ranges of gases and associated particulate matter.

According to Karekezi (2002), Africa is observed to base its domestic fuel use on three different categories. In North Africa, the primary domestic fuel used is oil and gas. Southern Africa is dependent mainly on coal, and sub-Saharan Africa depends mainly on the combustion of biofuels. Also, most African countries, especially in low-income areas, use wood as their main fuel, primarily for cooking and heating. Domestic fuel-burning activities are an integral part of ambient and indoor air pollution both in urban and rural areas in South Africa. Studies have shown that the burning of domestic fuel, especially in low-income settlements and townships, contributes to the overall air quality problems experienced by



© University of Venda

most developing urban centres. Mostly, South Africa's population comprises low-income households that live within townships and these burn enormous amounts of coal, wood, and paraffin, so as to provide for their energy needs, which are the main sources of energy for cooking in households of South Africa (Figure 2.8).

Recently, South Africa has undertaken a wide range of electrification, through the Department of Energy's Integrated National Electrification Programme (DoE: INEP), which has brought about a tremendous upsurge in the number of electrified households in the country (DoE, 2013). Since 1993, the number of households using electricity for cooking increased by 26.4% from 47.5% to 73.9% in 2011. Also, the proportion of households using electricity for lighting increased from 57% in 1996 to 84% in 2011 (Naidoo *et al*., 2017). This has, however, significantly reduced the amount of domestic biofuel combustion, but it is still unclear if South African households will abandon the use of biofuel combustion as it is their most convenient and cheaper source of energy. The use of wood is common in Vhembe and Mopani Districts, paraffin in Capricorn and Waterberg, and coal in Sekhukhune District Municipality due to proximity to the collieries and this was identified as a potential air quality issue (StatsSA, 2011). Indoor air pollution was ranked as the number 15 risk factor for South Africa, higher than urban air pollution, as outlined by The Medical Research Council Burden of Disease Research (MRC, 2008). Due to their inability to access basic amenities, as a result of low monetary capacity, these households tend to fall back on the use of cheaper, convenient ways for energy supply in their homes - wood and coal. Figure 2.9 shows the fuel types used for cooking in households of Vhembe district and Thulamela municipality (StatsSA 2011).



© University of Venda

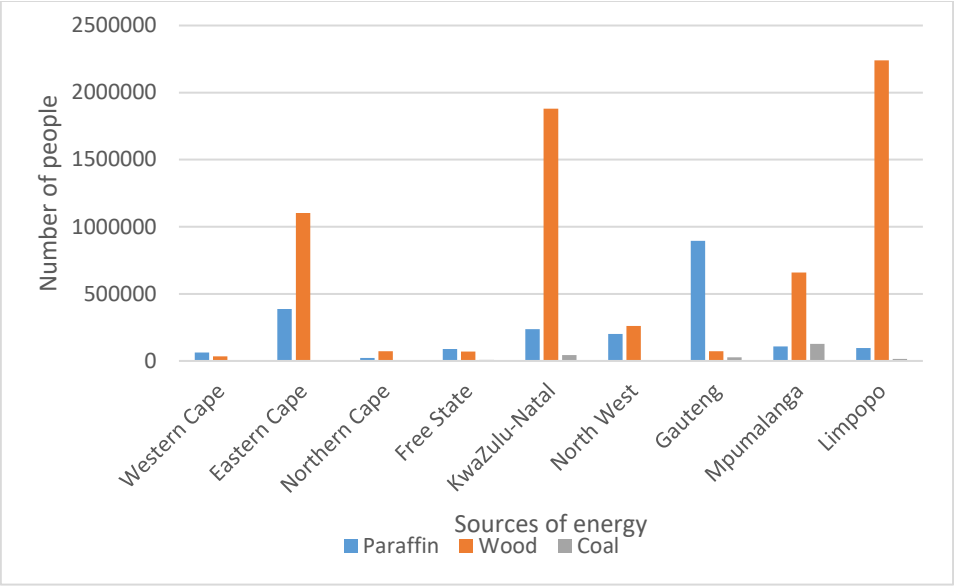



**Figure 2.8**: Main types of energy used for cooking in South Africa (StatsSA, 2011).

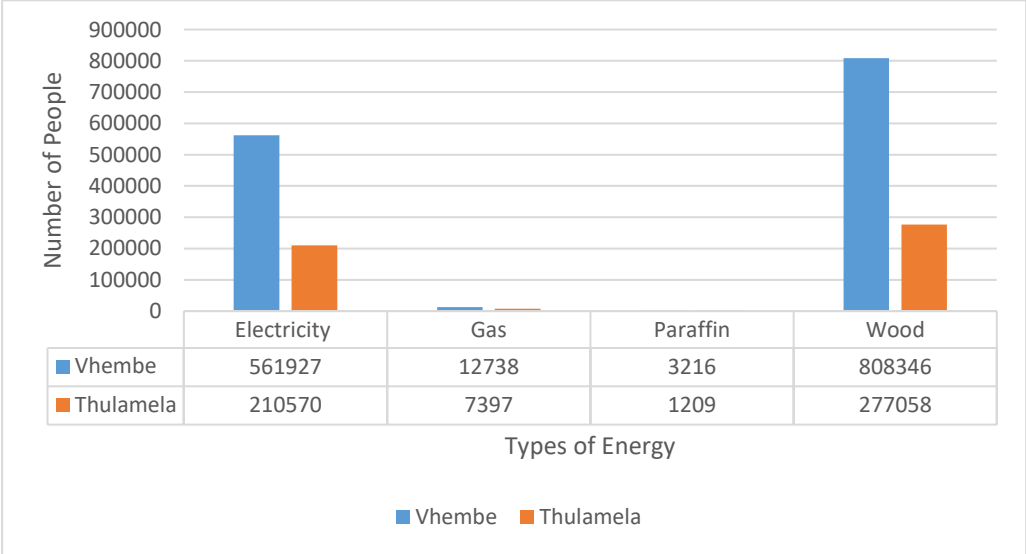



**Figure 2.9:** Main types of energy used for cooking in Vhembe District (StatsSA, 2011).

According to Ludwig *et al*. (2003), two major domestic fuel types used in South Africa during winter have been coal and wood, although, fuel sources like paraffin, liquid petroleum gas, animal dung, and electricity are used. During summer most common domestic fuels used are paraffin, candles, gas, and electricity. South African households usually utilize wood and coal because of their multi-functional nature and approximately 90% of these households burn biomass on an open fire. These multi-functional fuels are easily used for cooking and



© University of Venda

lighting, as well as for space heating purposes, however, the continuous use of wood and coal combustion has adverse effects on both the environment and human health.

The calculation of residential fuel combustion is a very challenging process because of the uncertainty of the domestic fuel consumed. Domestic fuel emissions vary across different countries depending on the fuel type and associated emission factors; for example, the combustion of agricultural residues tends to have higher emissions than dung (Battachayra *et al.,* 2000). Domestic fuel consumption, per capita like wood fuel depends on its easy accessibility and demand which ranges from 0.05 kg per capita per day in Lesotho to as high as 3.0 kg per capita per day in eastern Africa (Yevich and Logan, 2003). In West Africa, countries like Guinea have abundant forest cover, therefore they have enough fuelwood resources, hence, the highest fuelwood consumption of 3.2 kg per capita per day. Nigeria, Tanzania, Kenya, and Zaire use approximately 50% of their fuelwood consumption, which amounts to approximately 138 Tg yr of trace gas production. The coastal region of West Africa comprises lots of thick dense forest and high population density, which result in a range of 1.3 to 1.7 kg per capita per day consumption of fuelwood (Mulaudzi, 2006). In Southern Africa, Bertschi *et al.* (2003) showed that Zambia used more fuelwood than charcoal, with wood burning producing 9.8 Tg C $yr^{-1}$ of $CO_2$ and coal burning producing 3.3 Tg $yr^{-1}$. South Africa has the highest emissions from biofuel combustion, and Botswana and Namibia contribute low emissions from fuelwood combustion. This is mainly from the presence of the Kalahari Desert and the low biomass found in the Botswana and Namibia areas.

### 2.4.2 Vehicle Emission

Vehicular emission remains a threat to environmental health – a problem which is expected to increase as vehicle ownership increases in the world. Over 600 million people globally are exposed to hazardous levels of traffic–generated pollutants (Cacciola *et al*., 2002). Emissions from the transport sector depend mainly on the type of vehicle and fuel as well as the type of combustion engine, emission mitigation techniques, maintenance procedures, and the vehicle age. The major pollutants emitted from transport are carbon dioxide ($CO_2$), methane ($CH_4$), carbon monoxide (CO), nitrogen oxides ($NO_x$), nitrous oxide ($N_2O$), sulphur



© University of Venda

dioxide ($SO_2$), and non-methane volatile organic compounds (NMVOCs). South Africa's transport sector contributes approximately one-third of the country's total energy-related greenhouse gas emissions (Merven *et al.,* 2013). Transportation contributes 10% of the total emissions, which have increased by 32.2% - from 2000 to 2010 in South Africa. Road transport in South Africa produces 80-90% of the total transport emissions (DEA, 2014). This is estimated to generate 29.6 $MtCO_2$-eq, 29.8 $MtCO_2$-eq, 36.6 $MtCO_2$-eq and 43.4 $MtCO_2$-eq in 1990, 1994, 2000 and 2010, respectively (Tongwane *et al.,* 2015). The overall greenhouse gas emission from the transport station in South Africa from 2000 to 2010, however, shows that $CO_2$, $CH_4$, and $N_2O$ contribute to 444.21 $MtCO_2$-eq (97.86%), 3.19 $MtCO_2$-eq (0.70%), 6.52 $MtCO_2$-eq (1.44%), respectively. Motorcars and trucks produced more than 70% of the road transport emissions in South Africa (WWF-SA, 2016).

There are many different types of transportation sources, not all of which are significant sources of pollution. Unfortunately, the most utilised transportation sources at present are also major polluters. Motor vehicles are a very significant source of different pollutants such as CO, $CO_2$, Pb, PMs, hydrocarbons, $NO_2$, and even small amounts of oxides of sulphur may affect plant growth adversely (Krzyżanowski *et al*., 2005; Sharma *et al*., 2018). Pollutants from motor vehicles can react in the atmosphere to form different and even more reactive (and dangerous) pollutants such as photochemical smog. Vehicle volumes and emissions have reached critical levels in many cities within urban areas, and it is one of the largest sources of air pollution and causes of many environmental problems in the world. Vehicle emissions in Cape Town have been identified as one of the sources of brown haze (Chiloane, 2006). A recent study by Mabusela and Mamakoko (2017) also showed that the increase in high road traffic in the City of Johannesburg contributes significantly to the observed high levels of pollutants and greenhouse gas emissions in the city. Vehicular exhaust from the combustion of fuels, engine oils, tyre wear and tear, brake wear, vehicular exhaust catalysts (VEC), and road abrasion contain a wide range of pollutants and metals. Some of these metals are directly deposited on the exhaust pipe, while the remainder are dispersed into the surrounding environment, posing significant ecotoxicological and public health risks (Winther and Slento, 2010; see Figure 2.10).



© University of Venda

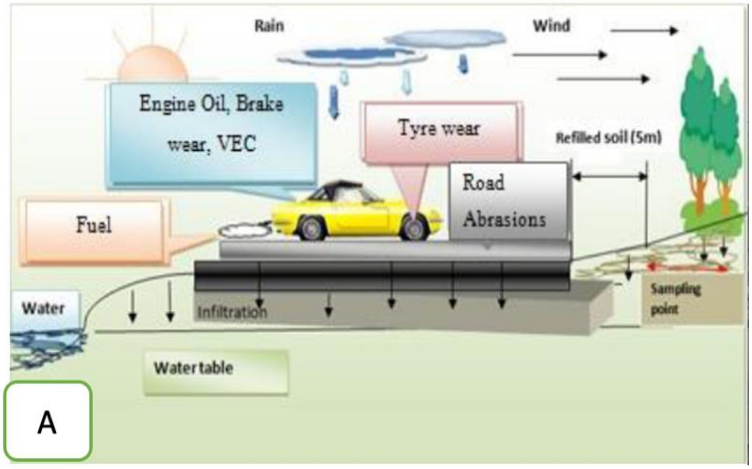



**Figure 2.10:** Schematic representation of the mechanisms of vehicular emissions and pollutant pathways into air, water, and soil environments (Winther and Slento, 2010).

In India, Singh *et al*. (1995) reported that pollutants from vehicles account for 60-70% of the pollutants found in the urban area and the two major pollutants are lead (Pb) and $SO_2$. Heavy metals originate from different sources and are transferred to the biosphere constituents of airborne particles (Odiyo *et al*., 2005; Bhuyan, *et al*., 2018). Recently, several studies have drawn attention to the gaseous pollutants and heavy metal accumulation in plants at high concentrations (Bonanno and Pavone, 2015; Zhang *et al*., 2016), causing serious health risk to people or animals eating these plants (Rai *et al*., 2013).

### 2.4.3 Industrial Emission

Energy is essential for sustaining life in all living organisms. Worldwide, it has been observed by several researchers that emissions from coal-fired power stations can be harmful to the environment and human health; the primary emitter with adverse environmental and/or human health impacts, coming from coal-fired power stations, includes PM, $SO_2$, $NO_x$, mercury (Hg) and GHG. In South Africa, coal is still the main source of material for the generation of electricity. Approximately 67% of South Africa's total primary energy supply comes from coal, which is used to generate about 90% of the country's electricity through coal-fired power stations (DoE, 2013). South African coal is consumed by different activities, like brickworks, smelters, mineral processing, broilers, and power generation; these



© University of Venda

activities release pollutants as well. Eskom as one of the largest energy utilities in the world is responsible for the generation of approximately 45% of Africa's electricity; 95% of South African electricity (LEDET, 2016; Manzini, 2016) and the remaining 5% of South African electricity is produced by coal-fired power plants owned by the private sector, such as, Kelvin Power Plant; municipalities (Rooiwal, Pretoria West and Bloemfontein power plants) and Sasol (Eskom, 2010; Morosele, 2020). The long-term effect of the constant generation of electricity and supply of energy from coal power stations often results in the emission of air pollutants and greenhouse gases (GHG). These have potential impacts on air quality and human health. The energy sector in South Africa is the biggest contributor to $SO_2$ and $NO_x$ emissions and the second highest contributor to PM likewise $NO_x$, Hg, and GHG emissions in the following order - 70%, 55%, 36%, 55%, 65%, and 78%, respectively; as compared to industrial, commercial and institutional fuel-burning, vehicle emissions, biomass burning, and domestic burning (Scorgie *et al.,* 2004; DEA, 2012). $SO_2$ and $NO_x$ are emitted from power stations industries in Limpopo.

Brick manufacturing is the fastest-growing industrial sector in many countries, which can either use biomass fuels such as fuelwood or coal for clamp kilns in rural areas for firing. Biomass fuels, especially fuelwood, are the main fuels used in clay brick firing in the traditional method (Alam, 2006). Biomass burning from brick-making industries is a significant source of trace and non-trace greenhouse gases, such as $CO_2$, $CH_4$, CO, $N_2O$, $NO_x$, and NO, which are regarded as an energy-inefficient process with a high potential for air pollution globally (World Bank, 1998; FAO,1999). The use of coal in brick-making accounts for most carbon emissions pollutants like $SO_2$, (Skinder *et al*., 2014), and nitrous oxides (DEA, 2012) (Figure 2.11). The clamp kiln is the major clay brick production technique in South Africa (Akinshipe and Kornelius, 2015). Most of the brick kiln plants result in the production of $SO_2$, $NO_x$, CO, and PM along with many other organic pollutants due to the burning of sub-standard waste material. With the rapid but unrestrained development, emissions from these sources are constantly increasing and pose negative environmental effects as well as adverse health, social, economic, and ecological impacts (DEA, 2012; Khan and Vyas, 2008). Sulphur dioxide ($SO_2$) and particulate matter have been identified as pollutants to be controlled by clamp kiln emitters and there are stipulated



© University of Venda

standards and methods by which ambient monitoring must be carried out (Akinshipe and Kornelius, 2015).

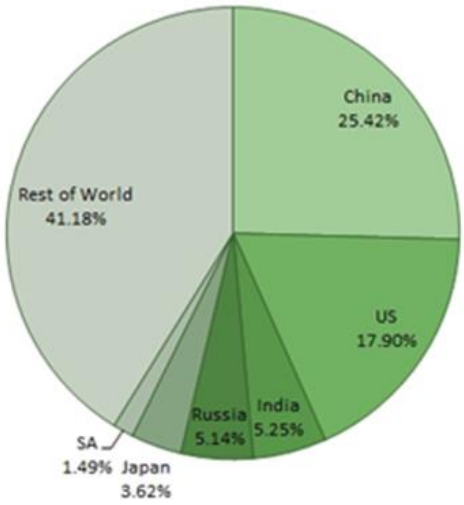



**Figure 2.11:** South Africa CO contribution to global emissions (Source: IEA, 2023).

The mining industries are a major economic activity in many developing countries. Mining has played a significant role in the development of the South African economy, with both positive and negative consequences. Particulate matter is the main air pollutant from coal mining and other mining activities when compared to other pollutants emitted because of vehicle and equipment combustion (that is, the burning of fuel by trucks, generators, and other machinery used in mining operations). South Africa has an industrialised economy, with significant mining and metallurgical activities and a large fraction of the South African mineral assets is concentrated in the Bushveld Igneous Complex (BIC) (Van *et al*., 2014). Pandey *et al.,* (2014) have also shown that the major sources contributing to air pollution are coal mining-related activities from Jharia Coal field (JCF) and active mine fires. Mine activities and dumpsites increase air pollution as many of them are not covered and long-term exposure of humans to air pollutants in the mining environment can cause several respiratory diseases, such as chronic bronchitis and asthma, cardiovascular, reproductive, and gastrointestinal problems (Wright *et al*., 2014). Studies conducted by Nkosi *et al*., (2015) in Gauteng and Northwest provinces reported a high level of chronic respiratory symptoms and diseases among elderly people in communities located about five kilometres (km)



© University of Venda

radius away from the mine dumpsites, causing certain regions in South Africa to be nationally designated due to their critical pollution levels.

#### 2.4.3.1 Highveld Priority Area (HPA)

Designated in 2007 under the National Environmental Management: Air Quality Act (2004), the Highveld Priority Area (HPA) spans parts of Gauteng and Mpumalanga provinces and includes municipalities such as eMalahleni, Govan Mbeki, and Steve Tshwete. These regions are home to 12 Eskom coal-fired power stations, Sasol's Secunda Petrochemical Complex, numerous metal smelters, and brickmaking operations, making it one of the most heavily industrialised and polluted regions in the country. According to the Department of Environmental Affairs (DEA, 2012) and follow-up reports, the HPA regularly exceeds national ambient air quality standards, particularly for particulate matter ($PM_{10}$ and $PM_{2.5}$) and sulphur dioxide ($SO_2$). These excesses are closely linked to industrial emissions, domestic coal burning, and vehicular sources. The 2017 Highveld Priority Area Air Quality Management Plan Review noted that poor enforcement, outdated technologies, and a lack of emission abatement in major facilities contributed to the regions' failure to meet air quality goals. A 2021 report by the Centre for Environmental Rights highlighted the continued public health crisis, including increased risks of asthma, lung cancer, cardiovascular diseases, and premature death, especially among vulnerable populations. The HPA serves as a case study of how industrial emissions, particularly from coal-fired power generation can cause long-term, widespread air quality degradation with substantial human health and environmental costs, if not effectively monitored and mitigated.

### 2.5 Air Quality Management in South Africa

Air quality management in South Africa is primarily governed by the National Environmental Management: Air Quality Act (NEMAQA, 2004), which provides a regulatory framework for controlling atmospheric emissions. This legislation, along with the National Ambient Air Quality Standards (NAAQS), sets threshold values for key pollutants such as - sulphur dioxide ($SO_2$), nitrogen dioxide ($NO_2$), carbon monoxide (CO), ozone ($O_3$), particulate matter ($PM_{10}$ and $PM_{2.5}$), lead, and benzene ($C_6H_6$) (DEAT, 2009). Under this framework, provincial



© University of Venda

and municipal authorities deploy monitoring stations to track pollutant levels, feed data into platforms such as SAAQIS, and inform Air Quality Management Plans (AQMPs), notably implemented in provinces such as Limpopo, Gauteng, and Mpumalanga (Thabethe *et al*., 2014; DEA, 2018). These standards guide local authorities in evaluating air quality and safeguarding public health.

Monitoring is overseen by the Department of Forestry, Fisheries and the Environment (DFFE, formerly DEA) and provincial departments like Limpopo's LEDET, which implemented its first Emissions Inventory in 2013. Recent analyses, however, reveal persistent challenges in spatial coverage and real-time data availability, especially in under-resourced and informal areas, thereby limiting effective exposure assessment (Wright *et al*., 2019). Since 2015, discourse on air quality governance has increasingly emphasised transparency, accountability, and data-driven policymaking. The 2015 State of Air Report and the Air Quality Governance Lekgotla introduced frameworks for cost–benefit evaluation of interventions, while portable monitoring networks have improved local-level enforcement., Monitoring, despite these advances, remains uneven, with critics citing limited deployment in rural areas and inadequate enforcement of emission inventories post-2013.

The Air Quality Index (AQI) is commonly used across South Africa to communicate pollution levels. It translates concentrations of $PM_{10}$, $O_3$, $NO_2$, and $SO_2$ into a five-band system (from *low* to *hazardous*) for public understanding (Reid & Rolfe, 2014). Studies like that of Joshi and Semwal (2011), however caution that AQI metrics may oversimplify localised pollution and fail to highlight temporal exceedances affecting vulnerable populations. Monitoring data are also often inconsistent and spatially sparse, particularly in informal settlements, limiting accurate exposure analysis. Wright *et al*. (2019) confirm that real-time reporting gaps remain a challenge for policy enforcement and public health planning.

Recently, machine learning approaches have enhanced $PM_{2.5}$ forecasting and show potential for predictive air quality management. A 2022 South African commentary noted that despite a robust legal framework, implementation remains weak, with the law failing to adequately address pollution in priority areas. Field studies and bibliometric reviews continue to show that $PM_{2.5}$ and $NO_2$ exposure disproportionately affect low-income and



© University of Venda

rural communities, highlighting the need for equity-focused, adaptive air governance. In summary, although South Africa has made regulatory progress since 2009, contemporary research (2015–2025) underscores ongoing monitoring limitations, enforcement issues, and exposure disparities. There is a clear need for integrated data systems, rigorous cost–benefit assessments, and community-oriented solutions to address systemic environmental injustices. These insights directly align with the thesis objective of connecting air pollution exposure to public health outcomes and enhancing responsive environmental governance.

### 2.6 Instruments for Monitoring Particulate Matter (PM) Pollutants

Effective air quality governance, as outlined in the previous section, relies heavily on accurate, continuous, and context-specific data on ambient pollutant levels, especially fine particulate matter ($PM_{2.5}$ and $PM_{10}$), which poses severe risks to human health. Monitoring these pollutants requires instruments that differ in cost, complexity, and methodological principles. This section reviews key instruments used to measure PM, comparing low-cost sensors with regulatory-grade equipment, and explores how these tools support both local exposure assessment and broader atmospheric modelling efforts.

There are various instruments for measuring and quantifying different characteristics of PM, and it is very important to choose the best instrument for the objective of the study. PM is usually measured based on its particle concentration and particle size. Particle size analyser identifies the behaviour of the particles in outdoor air and the instruments used for measurement of particle size distribution, measures based on the behaviour of the particles, which include, diffusion, aerodynamics, and optical and electrical mobility (Wang *et al*., 2016; Lei *et al*., 2024). It is essential to measure PM concentration, so as to standardise PM emission limits, for cleaner air, for breathing. The PM concentration can be in mass, number counts, and surface area. Also, the instruments for measuring PM concentration does so based on different methods, which could be - gravimetric, optical microbalance**,** and electric charge.

The gravimetric method for the measurement of PM concentration is based on the introduction of filters and cascade impactors. Filter-based gravimetric methods are usually robust and rather simple to operate, however, the process takes a substantial amount of



© University of Venda

time (McMurry, 2000). The removal of moisture from the filters could lead to the loss of potential health-related volatile compounds (Huang *et al.,* 2019). Some types of gravimetric instruments include, E-sampler (Moja *et al*., 2017), Harvard Impactors (HI), and Personal Environmental Monitors (PEM) (Wang *et al*., 2016).

Optical detection methods are lit by a light beam that radiates in all directions, which is also known as “scattering”. Some parts of the light beam are simultaneously transformed into other energy forms (absorption). Optical instruments used for measuring particle concentration, in real-time, are based on the principles of scattering, absorption, and light extinction. Some types of optical detection instruments using the beam scattering technique include Data Respirable Aerosol Monitor (DataRam 4), UCB-PATS (University of California Berkeley-Particle and Temperature Sensors), SidePak AM510, and DustTrak II 8530 Personal Aerosol Monitor (Amaral *et al.,* 2015). Some optical detection instruments associated with the light absorption technique include Spotmeters, Aethalometer, Photoacoustic Soot Sensor (PASS), and Laser Induced Incandescence (LII). Some types of optical detection methods associated with light extinction include the Cavity Ring Down (CRD) and Opacity Meter (Amaral *et al.,* 2015).

The microbalance method instruments evolve when the particles are collected from the surface oscillatory microbalance element, and those microbalances use the alteration of the resonance frequency to determine the PM (Giechaskiel *et al*., 2014). Tapered Element Oscillation Microbalance (TEOM) and Quartz Crystal Microbalance (QCM) are the two major measurement instruments that utilise the microbalance method. The measurements of concentration of PM can be measured using several instruments, considering the above review; some of the instruments include Beta-Attenuation Monitors (GRIMM EDM180; BAM 1020 (Met One Instrument); Dylos DC 1700; DustTrak II 8530; Sidepak AM 510 (TSI Inc.); Data Respirable Aerosol Monitor (DataRam 4) and UCB-PATS (University of California Berkeley-Particle and Temperature Sensors). In practice, instruments such as the GRIMM EDM180, BAM 1020, Dylos DC 1700, DataRAM-4, and others are widely used for PM monitoring across regulatory and research settings. Each has strengths and limitations in terms of accuracy, cost, calibration needs, and applicability to specific environments.



© University of Venda

### 2.6.1 GRIMM EDM180 Model

GRIMM EDM180 model (GRIMM Aerosol Technik GmbH & Co. KG) is a German Ambient Environmental Dust Monitoring instrument for continuous measurement of particles in the air (aerosols) and the equipment uses light scattering technology to count particles. A semiconductor laser serves as the light source for the measurement of these particles. Particle mass concentration, as $\mu g/m^3$ is reported for the various size distribution channels. The scattered signal from the particle passing through the laser beam is collected at approximately 90° by a mirror and transferred to a recipient-diode. The signal of the diode passes, after a corresponding reinforcement, a multi-channel size classifier (Kaonga, and Ebenso, 2011). A pulse height analyser then classifies the signal transmitted in each channel. These counts can be displayed and are also stored in the data storage card and then transferred for further analysis (Wernecke *et al*., 2021; Bozilov *et al*., 2022).

A volume-controlled internal pump draws in the ambient air to be studied at a rate of 1.2 litres per minute (Kaonga, and Ebenso, 2011). The sample moves through the laser diode detector, via the measurement cell, and is then gathered onto a filter. The pump also generates the necessary clean sheath air, which is filtered and passes through the sheath air regulator back into the optical chamber; this protects the laser-optic assembly from being contaminated by dust. During the auto-calibration, the reference-zero tests are also conducted using this particle free airflow -GRIMM EDM180 dust monitor.

### 2.6.2 BAM 1020 Monitoring System

BAM 1020 measures and records continuous airborne concentration levels of particulate matter ($PM_{2.5}$) levels using the principle of beta ray attenuation (Gobeli *et al*., 2008). The concentration unit is in $\mu g/m^3$ (Hafkenscheid and Vonk, 2015) and its flow rate is 16.7 litres/minute, which is adjustable (Motallebi *et al*., 2003). A sensitive scintillation detector is used to detect and count the beta particles. The automatic BAM consists of a size-selective inlet, a filter tape, a beta radiation source, and a beta radiation detector. Particles smaller than the cut-diameter of the size-selective inlet are collected at a single point on a length of filter tape, and the baseline attenuation signal is measured before each sampling event (Chung *et al*., 2001). This is used to monitor continuous $PM_{2.5}$ and $PM_{10}$ mass



© University of Venda

concentrations on the rooftop. An external pump pulls a measured amount of dust-laden air through a filter tape. After the filter tape is loaded with ambient dust, it is automatically placed between the source and the detector, thereby causing an attenuation of the beta particle signal. In addition, the volumetric and mass concentration of particulate matter on the filter tape is determined by the degree of attenuation of the beta particle signal (Chung *et al*., 2001). BAM 1020 is a more portable monitoring and has been used as a reference instrument in past studies (Holstius *et al*., 2014). Periodic leak and flow checks are performed to ensure the proper functionality of the BAM 1020 a few hours before taking the reading. The BAM 1020 is calibrated once a year.

### 2.6.3 DustTrak II 8530 Monitoring Instrument

The DustTrak II 8530 is a light scattering laser instrument that gives real-time aerosol mass readings and collects a gravimetric sample using a pump set at 3.0 L/min technique to quantify particulate matter concentration (Rivas *et al*., 2017). The DustTrak is a very practical and compact instrument, and it provides real-time $PM_{2.5}$ concentrations with a range suitable for different environments. In this study, it was configured to provide 5-minute averages expressing the output in $mg/m^3$. The instrument was zero checked and checked for proper function; the device is factory calibrated.

### 2.6.4 Sidepak AM 510 Monitoring Instrument

This is a laser light scattering instrument that draws air through an impactor using a pump set at 1.7 L/min (Jiang *et al*., 2011), which is a method to quantify the concentrations of particulate matter. The device logs every 5-minute recorded values, expressing the output in $mg/m^3$ with a sensitivity down to 0.001 $mg/m^3$. The device is factory calibrated for any type of particle size.

### 2.6.5 DataRAM 4 (Model DR-4000) Monitoring Instrument

The instrument is used to continuously monitor the real-time concentration and median particle size of airborne dust, smoke, mist, and fumes with a display of air temperature and humidity. The model can measure correlated values with $PM_{10}$, $PM_{2.5}$, $PM_{1.0,}$ and respiratory fractions. The two-wavelength particle detection system provides the volume median



© University of Venda

particle diameter of the sampled aerosol over a concentration range up to 400 mg/m$^3$. The model monitors the concentration of fine PM in ambient air by a combination of aerodynamic size preselection, low wavelength nephelometry, and concurrent sensing/correction for relative humidity. The model continuously measures $PM_{2.5}$ independent of particle size and moisture – without heating, diffusion drying or denuding the sample stream. Also, the instrument is intended for fast, simple field calibration, and the calibration process can take seconds to be achieved. The field calibration does not require additional equipment.

### 2.6.6 Dylos DC 1700 Sensor

Dylos DC 1700 sensor (Dylos Corporation, Riverside, CA) is a low-cost air quality monitor device that counts particle numbers per volume of air. The instrument uses a small fan to pull particles into a sensing zone. The instrument then measures particle number concentrations for two particle size of small size (lower than 0.5 µm) and large particle size (larger than 2.5 µm). A laser detector in the Dylos located approximately 90˚ to the airstream, measures scattered light intensity by the number of particles in the airstream and the changes of light intensity is reported as particle number counts. The data derived is automatically stored in internal memory at a 1-minute interval. The DC Dylos 1700 has an internal battery that runs up to 6 hours after being fully charged (Han *et al.,* 2017; Holstius *et al.*, 2014).

Castell *et al.* (2017) conducted a study on whether low-cost sensors can effectively monitor air quality and exposure estimate. Holstius *et al.* (2014), conducted a field study in Oakland California on the calibration of a small, portable, and reconfigurable platform around a low-cost, off shelf optical sensor, which was called the PANDA (Portable and Affordable Nephelometric Data Acquisition) system. The PANDA was then validated with government regulatory equipment, including the BAM 1020. Also, the GRIMM OPC (v. 1108), DustTrak II model 8530, and the Dylos DC 1700 were also compared using results derived from the PANDA system. To determine the validity and applicability of the PANDA system, various statistical analysis were incorporated into the analysis and the impact of ambient light, relative humidity, and temperature. The overall results showed that the comparison of the PANDA and other commercial instruments was remarkably good, and on a longer time scale, that is a healthy relevant scale on a 24 h average between the PANDA and the BAM



© University of Venda

1020. The results suggest that, at least in urban areas with similar aerosols and concentrations, additional deployments and calibrations may help to usefully enhance the resolution of $PM_{2.5}$ datasets. Also, the PANDA system can be used in more polluted regions, and it can be shown to resist saturation and wear.

Overall, low-cost sensors such as the Dylos DC 1700 and the PANDA system offer portability, affordability, and real-time monitoring capabilities, making them suitable for preliminary air quality assessments, especially in under-resourced settings. These devices, however, may suffer from limited accuracy, sensitivity to environmental conditions, for example, humidity, and require calibration against standard instruments. In contrast, reference-grade monitors such as the BAM 1020, GRIMM OPC, and DustTrak II provide higher precision and long-term data stability but are more expensive, stationary, and require technical expertise for operation and maintenance (Holstius *et al*., 2014; Castell *et al*., 2017; Ajayi *et al*., 2021). The Dylos DC 1700 and similar low-cost sensors provide useful, real-time data on particulate matter concentrations and are particularly beneficial in areas lacking extensive monitoring infrastructure. These instruments offer valuable insights into local air quality, however, broader assessments of pollutant dispersion over space and time require the application of atmospheric dispersion models. The following section introduces key modelling approaches used to estimate pollutant transport and distribution.

### 2.7 Dispersion Models for the estimation of pollutants

Dispersion modelling refers to the mathematical simulation of how air pollutants disperse in the atmosphere, typically within a pollution plume. These models require several key inputs, including accurate source emission data, meteorological conditions, and appropriate model selection based on source type and release scenario. A wide range of dispersion models exists, each varying in complexity, data requirements, and suitability. While steady-state models like the Gaussian Plume model are commonly used for regulatory assessments, non-steady-state models such as CALPUFF are better suited to variable meteorological and terrain conditions. Recent studies have compared the accuracy, computational demands, and limitations of these models in real-world applications (Rood, 2014; Gibson *et al*., 2013). The following subsections examine key dispersion modelling approaches, highlighting their strengths, limitations, and typical use cases.



© University of Venda

### 2.7.1 Gaussian Plume Models

Gaussian plume models are based on a Gaussian distribution of the plume in the vertical and horizontal directions under steady state conditions. The models are generally used in evaluating the effects of air pollution from local and urban sources predominantly for regulations. The Gaussian plume formula is shown in equation 2.1 (Arya, 1999).

$$C = \frac{Q}{2\pi u \sigma_y \sigma_z} exp(-\frac{y^2}{2{\sigma_y}^2})(exp\ exp\ \left(-\frac{(z-H)^2}{2{\sigma_z}^2}\right) + exp\ (-\frac{(z+H)^2}{2{\sigma_z}^2}) \quad (2.1)$$

Where, C is the downwind concentration at the receptor location, in μg/m$^3$;
(x,y,z) are the ground-level coordinates of the receptor relative to the source and wind direction, in meters (m);
Q is the mass flow of a given pollutant from a source located at the origin, in (μg/s);
U is the mean speed across the plume, in (m/s);
$\sigma_y$ and $\sigma_z$ are the Gaussian plume dispersion parameters.
H is the effective emission height, in (m).

Some assumptions made in the Gaussian plume model include (Ajayi *et al.*, (2021):

- Continuous emission from the source at a constant rate; no wind shear in the vertical;
- Steady-state flow and constant meteorological conditions;
- Constant mean transport wind in the horizontal (x-y) plane, and others.

Gaussian plume dispersion models are easy-to-use, this made them widely used for air pollution assessment (Arya, 1999). Some commonly used Gaussian plume models include AERMOD, AUSPLUME model, BLP, CALINE 3, OCD, and ISC version 2. These models are advantageous because they are computationally efficient, require relatively simple input data, and are suitable for screening-level and regulatory assessments. Their key limitations, however, lie in their assumptions of steady-state meteorological conditions, flat terrain, and constant emission rates, which reduce their accuracy in complex environments or under highly variable weather patterns.



© University of Venda

#### 2.7.1.1 Industrial source complex short term version 2 (ISC2)

The Industrial Source Complex Short Term Version 2 (ISC2) model is a steady-state Gaussian plume model mainly used to determine regulatory compliance with criteria pollutants that are released by industrial facilities. ISC2 was substituted by ISC Version 3, and the EPA preferred this model over version 2 until the endorsement of the AERMOD in December of 2005.

#### 2.7.1.2 AERMOD Model

AERMOD model went through a seven-step process until the model was proposed by the EPA to be adopted. The process involved in the development of the AERMOD includes initial model formulation; developmental evaluation; internal peer review and beta testing; revised model formulation; performance evaluation and sensitivity testing; external peer review; and submission to EPA's Office of Air Quality Planning and Standards (OAQPS) for consideration as a regulatory model. The AERMOD modelling system is comprised of two pre-preprocessors and the dispersion model - the AERMOD meteorological preprocessors (AERMET) and the AERMOD Terrain pre-processor (AERMAP). The AERMET provides the AERMOD with meteorological data and surface characteristics to calculate boundary layer parameters, for example, mixing height, friction and velocity; these are needed by the AERMOD. The AERMET's major aim is to calculate the boundary layer parameters for use by AERMOD.

The meteorological INTERFACE, internal to AERMOD, uses these parameters to generate profiles of the needed meteorological variables, AERMET then passes all meteorological observations to AERMOD. Surface characteristics in the form of albedo, surface roughness, and Bowen ratio, plus standard meteorological observations (wind speed, wind direction, temperature, and cloud cover), are input data into AERMET. AERMET then calculates the Problem Based Learning (PBL) parameters like the friction velocity, Monin-Obukhov length, convective velocity scale, temperature scale, mixing height, and surface heat flux. These parameters are then passed to the INTERFACE (which is within AERMOD) where similarity expressions (in conjunction with measurements) are used to calculate vertical profiles of



© University of Venda

wind speed, lateral and vertical turbulent fluctuations, potential temperature gradient, and potential temperature (USEPA, 2004).

The AERMAP uses gridded terrain data for the modelling area to calculate a representative terrain height scale associated with each receptor location. The terrain height scale, which is uniquely defined for each receptor location, is used to calculate the dividing streamline height. The gridded data applied to AERMAP is selected in the format of the Digital Elevation Model (DEM) data. AERMAP is also used to create receptor grids. The elevation for each specified receptor is automatically assigned through AERMAP. For each receptor, AERMAP passes the following information to AERMOD - the receptor's location, its height above mean sea level, and the receptor specific terrain height scale (Cimorelli *et al.,* 2003).

The AERMOD incorporates new or improved algorithms for;

- Dispersion in both the convective and stable boundary layers;
- Plume rise and buoyancy;
- Plume penetration into elevated inversions;
- Computation of vertical profiles of wind, turbulence, and temperature;
- The urban nighttime boundary layer;
- The treatment of receptors on all types of terrain from the surface to above the plume height;
- The treatment of building wake effects;
- An improved approach for characterising the fundamental boundary layer parameters;
- The treatment of plume meander.

Gibson *et al*., 2013 used the AERMOD to model the air dispersion of point and major line emissions of $PM_{2.5}$, $NO_x$, and $SO_2$ in several cities in Canada. The results of the study showed that the AERMOD model had similar results when compared with the observed $SO_2$ concentration, however, the author realised some discrepancies between the results modelled by AERMOD and the observed $PM_{2.5}$ and $NO_x$ concentrations in some cities. The authors still concluded that AERMOD is useful in understanding the surface impact of $PM_{2.5}$,



© University of Venda

$NO_x$, and $SO_2$ from point and major line sources at annual, monthly, and hourly averaging periods in the model domain. AERMOD offers several advantages, including improved accuracy over earlier Gaussian models due to its incorporation of boundary layer physics, terrain adjustment, and building downwash algorithms. It is widely accepted for regulatory use and performs well under a range of meteorological conditions when high-quality input data is available. Its limitations, however, include the assumption of steady-state conditions, reduced performance during low wind speeds or calm conditions, and a reliance on robust meteorological datasets processed through AERMET, which may not always be accessible. As such, while AERMOD remains a powerful tool for local-scale regulatory assessments, it may require complementary approaches in complex or highly variable environments (Cimorelli *et al*., 2003; U.S. EPA, 2019).

### 2.7.2 Gaussian Puff Models

Gaussian puff models are used to help estimate how air pollution disperses in the atmosphere. Some examples of Gaussian puff models include - INPUFF, RATCHET, CALPUFF, Artificial Neural Network (ANN) model. The general equation for the Gaussian puff model is shown in Equation 2.2 (Arya, 1999).

$$C = \frac{Q_{ip}}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\sigma_z} exp\ exp\left[-\frac{x^2}{2\sigma^2{}_x} - \frac{y^2}{2\sigma^2{}_y}\right]\{exp\ exp\left[-\frac{(z-H)^2}{2\sigma^2{}_x}\right] + exp\ exp\left[\frac{(z+H)^2}{2\sigma^2{}_z}\right] \qquad (2.2)$$

Where, C is the downward concentration at the receptor location (x,y,z);
$Q_{ip}$ is the instantaneous point source emission rate;
$\sigma_x$, $\sigma_y$, and $\sigma_z$ are the puff-diffusion parameters;
H is the effective emission height.

Gaussian puff models offer several advantages over steady-state models like the Gaussian plume, particularly in their ability to simulate time-varying and spatially dynamic emissions under non-steady meteorological conditions. They are well-suited for short-term accidental releases, complex terrains, or scenarios with rapidly changing wind fields. Models such as CALPUFF have been validated for long-range transport and regulatory applications, however, these models also have limitations. They require more detailed and time-resolved



© University of Venda

meteorological input data, are computationally more intensive, and their accuracy decreases if the input meteorology is sparse or of low quality. Additionally, while Gaussian puff models provide better realism, they still rely on simplifying assumptions that may not fully capture complex urban or topographically-influenced dispersion (Scire *et al*., 2000; Rood, 2014).

#### 2.7.2.1 INPUFF Model

INPUFF is a Gaussian Integrated PUFF model that simulates dispersion from a semi-instantaneous or continuous point source over a spatially or temporary variable wind field. The modelling scale is from tens of meters to tens of kilometres. The model simulates the accidental emission of substances over several minutes or models a more continuous plume from a stack. In the default mode, the model assumes a homogeneous wind field (Cheng *et al.,* 2021). The INPUFF model utilises dispersion algorithms and input features - optional stake downwash; optional buoyancy induced dispersion; wind speed extrapolation; temporally variable source characteristics; temporally and spatially variable wind field use; consideration of terrain effects; consideration of moving source; alternative subroutines for estimating dispersion; optional alternative sub-routine for estimating plume rise, as well as removal through gravitational settling and deposition. Some meteorological data required for computation include wind direction, wind speed, mixing height, stability class, the standard deviation of elevation angle, the standard deviation of azimuth angle, ambient air temperature, as well as anemometer height (Cheng *et al.,* 2021). One of its major strengths is that it bridges the gap between simple Gaussian plume models and more complex Eulerian models, however, limitations include the assumption of a homogeneous wind field in its default mode, which can reduce accuracy in complex terrains. Additionally, while the model offers several sub-routines, it may still oversimplify atmospheric processes compared to the more advanced Lagrangian or CFD models (Scire *et al*., 2000).

#### 2.7.2.2 CALPUFF Dispersion Model

CALPUFF dispersion model is based on the Lagrangian puff model and is a multi-layer, multi-species, non-steady-state puff dispersion model that is designed for regulatory use. CALPUFF model calculates the impact of temporal and spatial variability of



© University of Venda

micrometeorological conditions on pollutant transport, transformation, and removal (Guo *et al*., 2022). The US EPA has suggested the CALPUFF model can be used as a guideline model for long-range transport and on a case-by-case basis for near-field applications where the effects of non-steady-state may be significant (Rood, 2014). The modelling system of the CALPUFF comprises of primary modules including CALMET, CALPUFF, and CALPOST, which are all composed of the CALPUFF modelling system. The CALMET module is a meteorological model which makes a three-dimensional hourly wind field within a three-dimensional gridded modelling domain. Also, the CALPUFF model uses the CALMET-generated wind field and micrometeorological parameters to disperse a “puff”. The CALPOST concentration and deposition flux files produce time-average concentration and deposition output along with visibility impacts. The EPA has currently endorsed the CALPUFF model for long range dispersion (above 50 km) (Cimorelli, *et al*., 2003; Rood, 2014). CALPUFF offers several advantages over steady-state models. Its ability to simulate non-steady-state conditions, spatially and temporally varying wind fields, and chemical transformations makes it ideal for long-range transport, complex terrain, and coastal applications. It is particularly useful in regulatory contexts where plume behaviour changes over time and space, however, CALPUFF has high input data requirements, especially for accurate meteorological fields via CALMET, and greater computational demands than simpler models. Its performance is also sensitive to data quality, and it requires a steeper learning curve to configure and run correctly, as compared to models like AERMOD (Scire *et al*., 2000; Rood, 2014).

#### 2.7.2.3 The Regional atmospheric transport code for Hanford emission tracking (RATCHET)

The Regional Atmospheric Transport Code for Hanford Emission Tracking (RATCHET) is a Lagrangian puff model developed by Pacific Northwest Laboratories for the Hanford Dose Reconstruction Project. The model includes a surface wind field interpolator that allows the incorporation of multiple surface meteorological stations into a model simulation. Terrain complexities are not explicitly treated but are implicitly represented using multiple meteorological surface stations that reflect major topographical features. Surface roughness features are spatially variable across the model domain. Diffusion coefficients are estimated



© University of Venda

from atmospheric turbulence statistics that are inferred from estimates of atmospheric stability, surface roughness length, and the Monine Obukhov length (Rood, 2014).

Research has been done with several commercial air dispersion models for a lot of air monitoring, in both urban and rural areas, from different sources (Gibson *et al*., 2013; Rood, 2014). Gibson *et al.* (2013) simulated air dispersion from point and major line emissions of $PM_{2.5}$, $SO_{2,}$ and $NO_x$ in some major cities of Halifax, Pictou, Port Hawkesbury, and Sydney, Nova Scotia and Canada. For these locations, 2004 emission inventory was used to simulate the 50 km by 50 km domain over annual, monthly, and 1-hour averaging periods using statistical methods like $R^2$, fractional bias, normalized mean square error, and the fraction of prediction, to evaluate the results. A comparison of AERMOD evaluation and the Federal Government NAPS monitoring sites (discrete receptors) in the model domains was undertaken. Major sources of air pollution in these cities included power generation, construction activities, vehicle emissions, ship emissions, and domestic and industrial space heating via fossil and biomass fuels. The AERMOD model showed good agreement between modelled and observed $SO_2$ concentrations for the annual and monthly comparison, but showed poor results in estimating the hourly comparisons for $SO_2$ in Halifax and Sydney. Also, poor results were derived from the AERMOD model for simulating $SO_2$ in Port Hawkesbury over the same average time. The model's evaluation for $PM_{2.5}$ in Halifax, $PM_{2.5}$ in Pictou, and NO in Halifax showed poor comparison and modelling skills. Overall, the surface concentration from the point and major line sources in all domains from all metrics were below the National Air Quality Standards. The AERMOD model was shown to be a viable model for conducting dispersion modelling from point and line sources.

Rood (2014) conducted a performance evaluation of steady-state dispersion models (AERMOD and ISC2) and Lagrangian puff models (CALPUFF and RATCHET) using a winter validation tracer study dataset (performed previously in the year 1991 in Denver, Florida, USA). The major performance objectives of this study were to locate plume maximum, plume impact area, arc-integrated concentration, evaluate unpaired maximum one-and nine-hour average concentration, unpaired nine-hour average concentration, and paired ensemble means. The measures used to determine the performance included



© University of Venda

fractional bias, normalised mean square error, geometric mean variance, correlation coefficient, and a fraction of observation within a factor of two. The study showed that the Lagrangian puff models simulated the smallest variance, highest correlation, and highest number of predictions within a factor of two, compared to the steady-state models at distances of 8 km and 16 km. The steady-state models are less likely to underestimate the maximum one-and nine-hour average concentrations when compared to the Lagrangian puff models; the study, therefore, concluded that the performance of all the models are consistent and none outperformed each other based on the objectives of the study. Also, the recently adopted models (CALPUFF and AERMOD) did not show superiority over the legacy models (ISC2 and RATCHET).

Silverman *et al*. (2007) conducted a comparison of the ISC and AERMOD Dispersion Models with the incorporation of their improved version Plume Rise Model Enhancements (PRIME) algorithm; also, the potential human health risk receptor associated with the emissions was simulated. The switch from ISC to AERMOD and the incorporation of the PRIME algorithm simulates lower concentration estimates at the point of maximum ground-level concentration. For the meteorological dataset, 75% of the hours were suggested to be stable or neutral, and 25% of the hours unstable. The ISC simulated higher maximum ground level concentrations close to the sources for the 1 hour and the total averaging period, however, the AERMOD predicted lower maximum ground-level concentrations than the ISC, because of the ability of AERMOD to handle the steady periods. The switch from ISC to AERMOD and the integration of the PRIME algorithm simulated lower concentration estimates at the point of maximum ground-level concentration. Various studies on the AERMOD model have shown it simulates considerable reasonable results, especially when compared with other air dispersion models (Gibson *et al*., 2013; Rood, 2014).

### 2.7.3 Meteorological Data

Meteorological conditions play a fundamental role in air pollutant dispersion, influencing the transport, dilution, transformation, and removal of particulates and gaseous pollutants. Weather may directly contribute to the presence of fine particles in the atmosphere, and its effects depend largely on geographical location and local topography. Monitoring and



© University of Venda

forecasting of meteorological parameters are important to determine the pattern, flow of air pollutants and for accurate prediction of pollutants concentrations during the occurrence of air pollution (Wang and Song, 2018). Air quality models need sufficient hourly average meteorological data, both temporally and spatially at the surface. Key parameters include, hourly average data of wind speed (m/s), wind direction (degree), temperature (°C), relative humidity (%), atmospheric pressure (mbar), precipitation (cm), and global solar radiation ($Whm^{-2}$) for air quality monitoring (Habeebullah *et al*., 2015).

Models such as AERMOD and CALPUFF rely on dedicated pre-processors (AERMET and CALMET, respectively) to convert raw meteorological data into model-ready formats. These systems calculate boundary layer parameters, such as mixing height and friction velocity, which are critical for modelling plume behaviour, however, the accuracy of dispersion models is highly dependent on the quality, resolution, and representativeness of meteorological inputs. Inaccurate or sparse data can lead to substantial errors, particularly in complex terrains, coastal zones, or areas with strong diurnal variability. Assumptions of homogeneity in wind fields, for example, can misrepresent localised conditions, reducing model reliability (Rood, 2014; Pope *et al*., 2015); therefore, the placement of meteorological sensors should be carefully considered in relation to emission sources and prevailing wind patterns (Gruber, 1963; Forbes, 2015). In South Africa, long-range pollutant transport is enhanced by prevailing winds from August to October, raising concerns about the build-up of secondary pollutants, such as fine particulate matter ($PM_{2.5}$), which is especially harmful due to its ability to penetrate deep into the lungs (LEDET, 2013). As a result, the integration of high-resolution meteorological data, satellite observations, and improved forecasting systems is increasingly necessary to support accurate dispersion modelling and air quality management.

### 2.8 Air Pollution Health Risk Assessment (APHRA)

Air pollution is an important determinant of health. As stated in the previous chapter, several epidemiological studies have been conducted on the adverse effects of air pollution and human health (WHO, 2016). Figure 2.12 shows the probable severe effect on human health from air pollution. This has brought increasing concerns about air pollution, especially in developing countries, where air pollution has increased and the absence of strict air pollution



© University of Venda

policies. The quantitative estimates of the impacts of air pollution on health, therefore, become very important, as this allows policymakers and other stakeholders to develop and implement very effective policies.

An APHRA objective is to estimate the risk of past, current, and future exposure to air pollution and changes in exposure that may result from planned policies or other modifications to air quality (WHO, 2016). The APHRA could be quantitative or qualitative, which follows a systematic way in approaching its objective, this includes, firstly - estimating or quantifying the amount of air pollution concentration; then calculating the amount of probable contact (exposure) of the population in that specific area of concern, and finally, estimate the health risk involved in the exposure - that is, how harmful the air pollution concentration is to human health. The APHRA will help the process by responding to precise policy questions. APHRA is also required as part of the decision-making process for new air-related projects, regulations, and policies for improving air quality and the associated health of the people (Feuyit *et al.,* 2019).

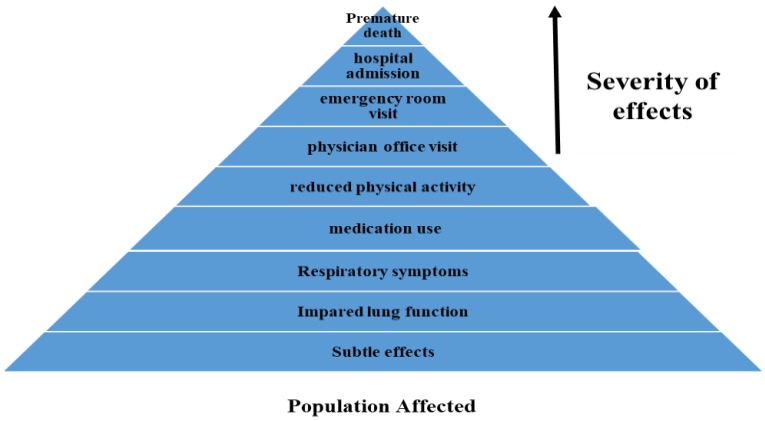



**Figure 2.12:** Health effects associated with air pollution (Source: WHO, 2016a).

According to WHO (2016b), in the estimation of APHRA, certain uncertainties are generally considered, which include;

- Air pollutants exist as a complex mixture – Regardless of the tremendous advancements experienced in the science underlying APHRA, it is however not still possible to completely know the certainty of the effects of air pollution on health.



© University of Venda

There are several epidemiological studies on the negative effects of air pollution on health, however, the observed effects ascribed to an individual air pollutant may be attributed to other pollutants in the air mixture.

- Baseline disease burden – The number of deaths accredited to air pollution may be uncertain, due to the uncertainties which arise from the projected population size and future death rate.
- Pollution exposure level – There is no complete certainty of geographical coverage of ground monitors, because most APHRA simply depend, to some extent, on modeling to estimate exposure. Modeling is also needed for estimates of future exposure based on predicted changes in air pollution as a result of new policies or technological improvements. Since air quality models are based on a set of assumptions, it is not possible to be certain that the estimated exposure will coincide with the actual ambient concentrations in a given location. Even if there could be full coverage by ground monitors, all APHRA assume that either measurement made at a specific location or model estimates of average exposure over a particular area, are representative of the exposure of the targeted population.
- The concentration-response function;
- The counterfactual level of air pollution, and
- The deliberate simplifications of the model.

Having highlighted some probable APHRA uncertainties, the analysis is typically presented as ranges, with the focus being on the confidence level around the mean. Also, usually 95% confidence levels are used to provide an estimate of the precise results, and this assumes that the levels are in the range within which there is a 95% probability in which the true value lies (Scott, 2008; Nunnari and Pozzi, 2022).

Several computer-based models are available that automate the process of APHRA (Table 2.3). These models offer a lot of advantages and have been made by the developers to be simpler, consistent, and able to compare among assessments and quality assurance. Most tools use similar approaches, relying on epidemiologically derived concentration-response functions and population-level exposure estimates to determine the proportion of cases of



© University of Venda

a particular health effect that may be attributable to a change in air quality. Automated tools are typically preloaded with health and demographic data and concentration-response functions, and some allow for user-specified inputs. The tools, however, vary in many aspects, and analysts should choose the one that most closely matches the context of the assessment (WHO, 2016b). This study utilized the Human Health Risk Assessment (HHRA) model of the United States Environmental Protection Agency's (US EPA) to analyse the carcinogenic and non-carcinogenic health risks associated with particulate matter ($PM_{2.5}$) exposure in Thulamela Municipality of Limpopo Province, South Africa. This was suggested based on WHO (2016b) analysis of the different APHRAs' suitability to get the best result.



© University of Venda

**Table 2.3:** Models for air pollution health impact assessment

| Models | Developing Institute | Geographical Scope | Health Endpoint Addressed | Reference |
|---|---|---|---|---|
| **AirCounts** | Abt Associates | Global (42 cities, additional 3000 under development) | Mortality | Abt Associates, (2015) |
| **AirQ+ (v2.2, 2023)** | WHO Regional Office for Europe | Global (user-defined populations) | Mortality and Morbidity | WHO, (2023) |
| **Aphekom** | French Institute of Public Health Surveillance | Global (current version focuses on Europe) | Mortality and morbidity | Pascal *et al*. (2013) |
| **Economic Valuation of Air Pollution (EVA)** | Aarhus University | Northern hemisphere, continental (e.g., Europe), national, city | Mortality and morbidity | Brandt *et al*., (2013) |
| **EcoSense** | University of Stuttgart | Europe | Mortality and morbidity | Friedrich & Bickel, (2001) |
| **BenMAP-CE** | US Environmental Protection Agency | Continental USA and China predefined; global user-defined | Mortality and morbidity | US EPA, (2021) |
| **Environmental Burden of Disease (EBD) Assessment Tool** | WHO | Global | Mortality and morbidity | WHO, (2016b) |
| **GMAPS2** | World Bank | Global | Mortality and morbidity | World Bank, (2017) |
| **Human Health Risk Assessment (HHRA)** | US EPA | Global | Risk characterization for Mortality and morbidity | US EPA, (2009) |
| **IOMLIFET** | Institute of Occupational Medicine | Anywhere with background mortality data and measured or predicted pollutant concentrations | Mortality and morbidity | IOM, (2014) |
| **Rapid Co-benefits Calculator** | US EPA & Stockholm Environment Institute (SEI) | Global (Under development for all countries) | Mortality | SEI, (2018) |
| **SIM-Air** | Urban Emissions | Asia, Africa, Latin America | Mortality | Guttikunda, (2008); Guttikunda & Jawahar, (2012) |
| **TM5-FASST** | European Commission Joint Research Centre | Global (56 source regions) | Mortality and morbidity | Van Dingenen *et al*. (2018) |



© University of Venda

### 2.8.1 Human Health Risk Assessment (HHRA) Model

The Human Health Risk Assessment (HHRA) model is a widely used framework for evaluating the potential health risks associated with exposure to environmental pollutants, including ambient $PM_{2.5}$. The United States Environmental Protection Agency (US EPA) defines a four-step process for HHRA: (1) hazard Identification, (2) exposure Assessment, (3) dose-response evaluation, and (4) risk characterisation (US EPA, 2009).

1. **Hazard Identification** involves ascertaining the types of health effects associated with $PM_{2.5}$, whether respiratory or cardiovascular outcomes, as well as potential carcinogenic effects.
2. **Exposure Assessment** estimates the magnitude, frequency, and duration of human exposure to $PM_{2.5}$, often using air quality monitoring data and demographic-specific exposure parameters.
3. **Dose-Response Assessment** determines the relationship between the $PM_{2.5}$ dose and the likelihood of adverse health effects.
4. **Risk Characterisation** integrates the above steps to estimate non-carcinogenic risk (Hazard Quotient, HQ) and carcinogenic risk (Excess Lifetime Cancer Risk, ELCR).

The application of the HHRA model in $PM_{2.5}$ exposure studies is typically operationalised using the Average Daily Dose (ADD) for non-carcinogenic effects and the Lifetime Average Daily Dose (LADD) for carcinogenic risk estimation (Das *et al*., 2020; US EPA, 2009). These exposure metrics are then used to determine the Hazard Quotient (HQ) for non-cancer risks and the Excess Lifetime Cancer Risk (ELCR) for potential carcinogenic effects. Researchers, such as Li *et al*. (2021), applied this approach in urban and industrial areas with high $PM_{2.5}$ concentrations, revealing HQ values >1 and ELCR values exceeding the acceptable thresholds ($10^{-6}$ – $10^{-4}$), especially among vulnerable populations, which indicates a heightened likelihood of cancer risks over a long period. The HHRA model serves as a vital tool in environmental health studies, offering evidence for regulatory actions and pollution reduction strategies designed to lessen the public health risks linked to exposure to air pollution.



© University of Venda

This model is especially important in studies related to air pollution, as it helps quantify the potential health risks associated with long-term exposure to fine particulate matter through inhalation. By combining pollutant concentration information with exposure parameters that are specific to different populations, the HHRA model facilitates the evaluation of the severity and reach of health risks among various demographic groups. In the context of this thesis, the HHRA framework was applied to assess the potential health risks of $PM_{2.5}$ exposure via inhalation across selected regions with varying pollution profiles. This aligns with the research objective of linking pollutant concentrations with potential public health impacts and supports the broader goal of informing evidence-based mitigation strategies.

### 2.9 Machine Learning

Machine learning is a subfield of artificial intelligence that uses algorithms to learn from data and make predictions or decisions without being explicitly programmed. It is a rapidly growing field with applications in various domains, such as healthcare, finance, and transportation (Rebala, *et al*., 2019). Machine learning, nestled within the expansive realm of artificial intelligence, constitutes a dynamic discipline wherein algorithms are adeptly harnessed to glean insights from data and autonomously formulate predictions or decisions, devoid of explicit programming. Machine learning represents a dynamic subset of artificial intelligence (AI) that has revolutionised the way computers learn and adapt to data patterns autonomously, rather than relying on explicit programming. This paradigm shift allows machines to discern intricate relationships and make informed predictions or decisions across a plethora of contexts, propelling advancements in numerous domains. As articulated by Mitchell (1997), machine learning enables systems to automatically improve their performance through experience, providing an efficient means to tackle complex problems that elude traditional rule-based programming.

Applications of machine learning permeate diverse sectors, showcasing its transformative potential, and these methods have transformed air quality forecasting by offering predictive capabilities that surpass traditional statistical approaches (Shen *et al*., 2020; Lu and Lee, 2025). Similarly, in transportation, machine learning algorithms power autonomous vehicles, traffic management systems, and predictive maintenance solutions, ushering in a new era



© University of Venda

of smart mobility (Filho *et al*., 2020). In essence, machine learning epitomizes a paradigm shift in computational intelligence, offering a powerful toolkit to extract meaningful insights from data and drive innovation across various domains. With its unprecedented capabilities, machine learning continues to redefine the boundaries of what is possible, ushering in a future where intelligent systems augment human decision-making and unlock new frontiers of knowledge and discovery.

### 2.9.1 Types of Machine Learning

Machine learning is a branch of artificial intelligence that enables computers to learn from data and improve their performance without explicit programming. Machine learning can be categorised based on different criteria and approaches. The categorisation can also be dependent on the nature of the data and the learning method. The various types of machine learning can include supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, deep learning, transfer learning and online learning. The basic types of machine learning have been classified into three, as seen in Figure 2.13.

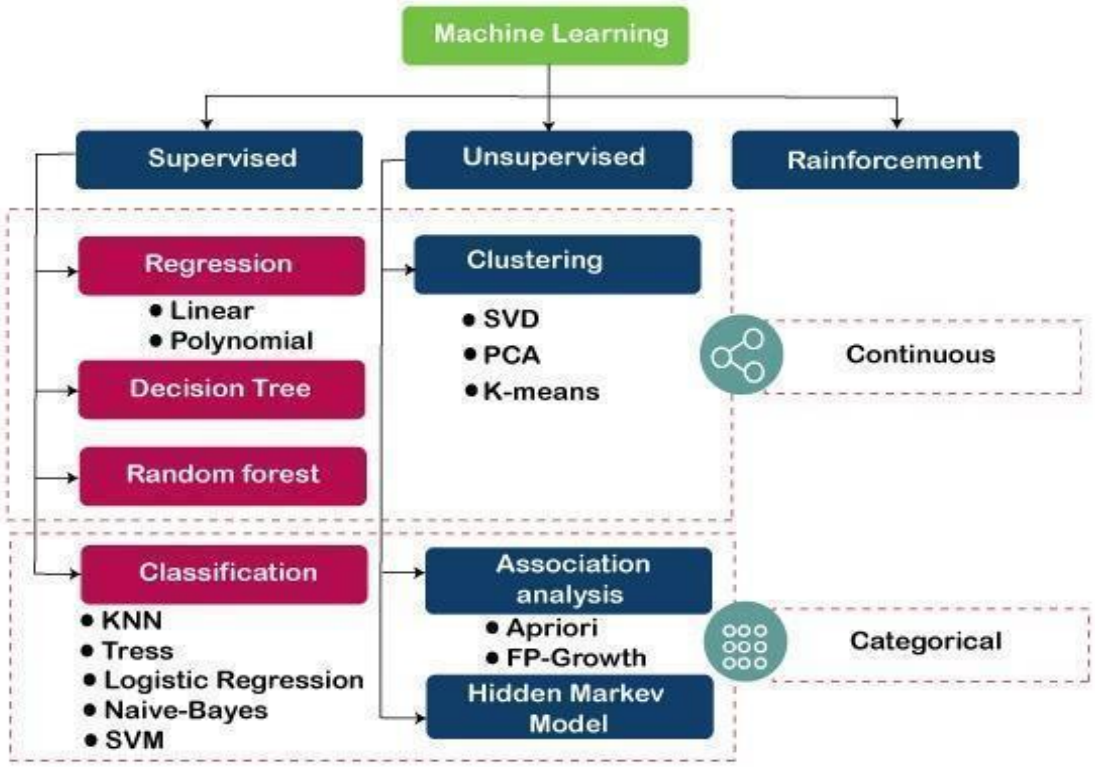



**Figure 2.13:**Types of machine learning.



© University of Venda

### 2.9.1.1 Supervised learning

This type of machine learning uses labelled data sets, which have both input and output variables. The goal of supervised learning is to learn a function that maps the input to the output, based on the training examples. Supervised learning can be further divided into two categories: classification and regression. Classification deals with predicting categorical outputs, such as spam or not spam, while regression deals with predicting continuous outputs, such as house prices or sales. Some examples of supervised learning algorithms are logistic regression, support vector machine, decision tree, random forest, trees, naive-bayes and k-nearest neighbours.

1. **Regression Analysis:** Regression is a cornerstone of supervised learning. It facilitates the prediction of continuous outcome variables based on input features. Linear regression and polynomial regression are two prevalent techniques employed in regression analysis, each offering distinct advantages and considerations in modelling relationships between variables.

a. **Linear Regression:** Linear regression assumes a linear relationship between the input features and the target variable (Figure 2.14). Mathematically, it can be represented as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

where, $\beta_0$ is the intercept, $\beta_i$ are the coefficients of $x_i$, and $\epsilon$ is the error term.

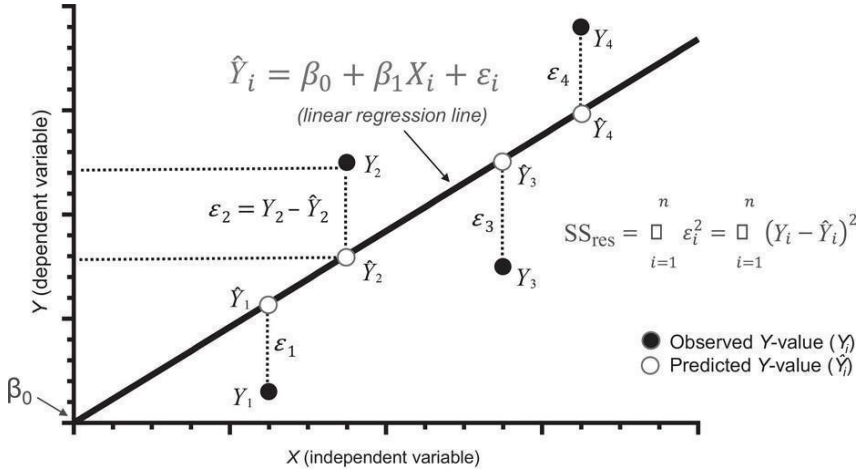



**Figure 2.14:** Illustration of the derivation of the residual sum of squares.



© University of Venda

Linear regression is a foundational method in statistical modelling. It assumes a linear relationship between input features and the target variable. As articulated by Hastie *et al*. (2009), a linear regression model assumes that the regression function $E(X)$ is linear in the inputs $X_1$,⋯, $X_p$. Linear regression models the target variable as a linear combination of the input features, with coefficients estimated through methods like ordinary least squares (OLS) regression. Its simplicity and interpretability make it a popular choice across various domains; for instance, in a recent study by Hoxha (2025), exploring the housing market, generalised linear regression proved effective in predicting property prices based on factors such as location, size, and amenities. Linear regression's assumption of linearity, however, may limit its applicability in capturing complex, non-linear relationships inherent in real-world datasets.

b. **Polynomial Regression:** Polynomial regression extends linear regression by accommodating non-linear relationships between input features and the target variable. As outlined by Ostertagova, (2012), polynomial regression fits a polynomial function to the data, allowing for more flexible modelling of curvilinear patterns. Recent research by Shah *et al*. (2021) demonstrated the utility of Multivariate Polynomial Regression Ecological Models in ecological studies, where it effectively captured non-linear associations between environmental factors and species diversity. Nevertheless, polynomial regression's susceptibility to overfitting, particularly with higher-degree polynomial terms, necessitates careful regularisation and model validation techniques to prevent spurious relationships and ensure robust generalisation (Figure 2.15). In essence, linear and polynomial regression represent versatile tools for modelling relationships in supervised learning contexts. While linear regression offers simplicity and interpretability, polynomial regression enables the capture of non-linear associations. By leveraging insights from recent literature and employing appropriate regularisation and interpretation methods, practitioners can harness the strengths of both approaches to effectively model complex data patterns and make informed predictions.



© University of Venda

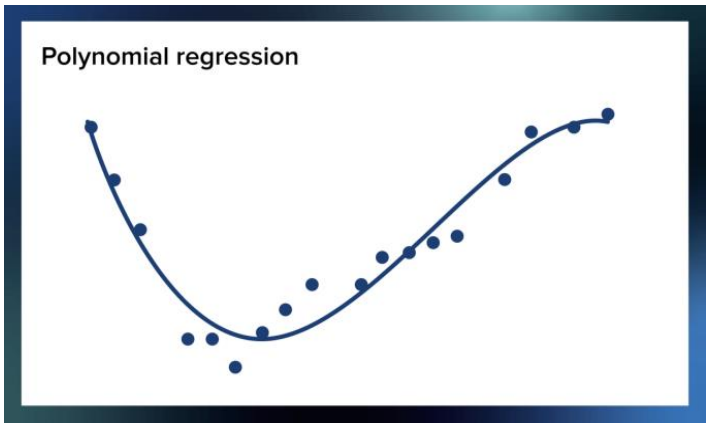



**Figure 2.15:** Polynomial Regression.

**2. Random Forest:** Random Forest stands as one of the most powerful and versatile algorithms in supervised learning, renowned for its robustness, scalability, and ability to handle complex data structures. This review delves into the principles, methodologies, applications, and recent advancements in Random Forest algorithms, shedding light on its significance in predictive modelling and data-driven decision-making. Random Forest is an ensemble learning method that constructs multiple decision trees during training and aggregates their predictions to produce a final output. Each decision tree in the Random Forest is trained on a bootstrapped subset of the original dataset, and at each node, a random subset of features is considered for splitting (Breiman, 2001). By combining the predictions of diverse trees through voting or averaging, Random Forest mitigates overfitting and enhances generalisation performance, making it well-suited for a wide range of classification and regression tasks (Figure 2.16).

Recent advancements in Random Forest algorithms have focused on improving scalability, interpretability, and predictive accuracy while addressing inherent limitations such as bias and variance. Variants such as Extremely Randomized Trees (Geurts *et al*., 2006) and Rotation Forest (Rodriguez *et al*., 2006) introduce additional randomness in the tree-building process, further diversifying the ensemble and reducing the risk of model overfitting. Moreover, techniques like feature importance ranking and partial dependence plots enable stakeholders to gain insights into the relative contributions of input features to model predictions, enhancing transparency and interpretability.



© University of Venda

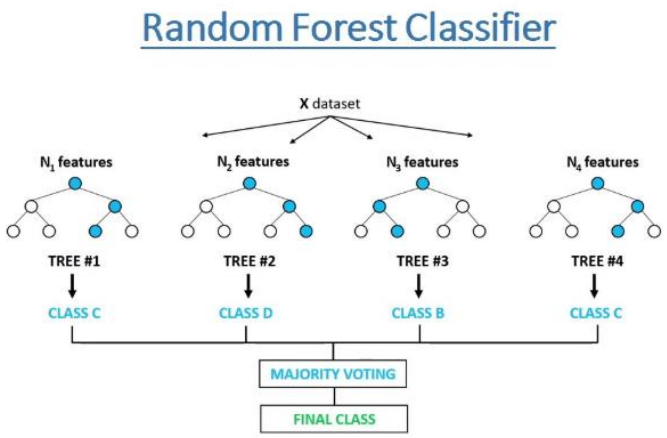



**Figure 2.16**: Random Forest

## 2.10 Applications and Real-World Impact of Machine Learning in Predictive Modelling

Predictive modelling serves as a powerful tool for leveraging data-driven insights to anticipate trends and inform strategic decision-making. As articulated by Provost and Fawcett (2013), predictive modelling empowers organisations to anticipate and proactively respond to dynamic environments, thereby unlocking opportunities and mitigating risks. By utilising machine learning algorithms, organisations can estimate outcomes across various domains, enabling proactive responses to emerging challenges. In healthcare, predictive modelling facilitates early disease detection, supports treatment planning, and enhances resource allocation, ultimately improving patient outcomes and operational efficiency (Obermeyer & Emanuel, 2016). Additionally, in engineering and manufacturing, predictive models are used for equipment failure prediction, process optimisation, and improving system reliability (Kusiak, 2018).

Predictive model selection is a critical aspect of predictive analytics that requires strategic evaluation of models to determine their suitability for specific tasks. Various techniques have been employed to optimise model selection and are broadly categorised into empirical, theoretical, and hybrid approaches. The Empirical techniques rely on data-driven experimentation to assess model performance. Methods such as cross-validation, bootstrap resampling, and hold-out validation enable practitioners to evaluate models by testing them on different subsets of data. These techniques help mitigate overfitting and ensure that the



© University of Venda

models generalise well to unseen data. By providing direct performance comparisons, empirical methods offer practical insights into model effectiveness across diverse applications. Theoretical approaches emphasise mathematical principles and model complexity. Techniques such as information criteria (for example, Akaike Information Criterion and Bayesian Information Criterion) and complexity measures help estimate a model's generalisation error (Hastie *et al*., 2009). These methods guide model selection by balancing the goodness-of-fit with model simplicity, thereby preventing overfitting and enhancing interpretability. Hybrid approaches integrate empirical validation with theoretical foundations to achieve a balanced model selection strategy. Bayesian model selection, structural risk minimisation, and the minimum description length principle combine statistical evaluation with theoretical rigor to optimise predictive performance (Ghahramani, 2015). These methods offer a robust framework for selecting models that perform well in both practical applications and theoretical assessments.

In predictive modelling, the choice of technique depends on the nature of the data, complexity of the problem, and available computational resources. By leveraging a combination of empirical, theoretical, and hybrid methodologies, researchers can enhance the accuracy, reliability, and interpretability of predictive models, across various domains.

*A. Predictive Model Processes*

I. Data collection and preprocessing: This step involves gathering relevant data and transforming it into a suitable format through cleaning, filtering, and encoding.
II. Data splitting and sampling: This step involves dividing the data into training, validation, and testing sets to prevent overfitting and ensure model reliability.
III. Model selection and training: This involves choosing an appropriate predictive model based on the type and characteristics of the data and the problem and training it, using learning algorithms.
IV. Model evaluation and tuning: This step involves assessing the model performance using evaluation metrics and adjusting the parameters for optimisation.



© University of Venda

V. Model deployment and monitoring: This involves applying the model to new data, to make predictions or recommendations, track performance and reliability of the model over time, and update it as needed.

### 2.10.1 Prediction of Future Air Quality Using Machine Learning Models

As machine learning continues to evolve, predictive modelling remains a driving force in innovation and informed decision-making. Its adaptability across industries ensures that organisations and researchers can leverage predictive insights to enhance efficiency, reduce uncertainties, and unlock new opportunities. The ability to harness vast amounts of data for accurate prediction not only optimises operations but also fosters sustainable advancements in science, business, and public health. The application of predictive modelling extends to the environmental sciences, where it aids in forecasting climate patterns, monitoring ecosystem changes, and optimising resource management (Mills *et al*., 2018). One significant application is in air pollution prediction, where machine learning models analyse historical pollutant levels, meteorological conditions, and urban activities to estimate future air quality trends (Masih, 2019). These predictions support policymakers in implementing timely interventions to mitigate pollution-related health risks, such as respiratory diseases and cardiovascular complications associated with $PM_{2.5}$ exposure (Liu *et al*., 2022). By integrating large-scale environmental datasets, predictive models can help identify pollution hotspots, assess long-term exposure effects, and guide sustainable urban planning strategies (Verma, *et al*., 2024).

Predicting future air quality using machine learning models has gained significant attention due to their ability to accurately estimate pollutant concentrations, such as particulate matter ($PM_{2.5}$ and $PM_{10}$), nitrogen oxide ($NO_x$), sulfur dioxide ($SO_2$), and ozone ($O_3$), with particular emphasis on fine particulate matter ($PM_{2.5}$). In this study, MLR, PRR and Random Forest Regression were used to model and predict the ambient $PM_{2.5}$ in the study area of South Africa due to their below-mentioned advantages. Multilinear Regression (MLR) serves as a fundamental approach, utilising multiple independent variables to establish linear relationships between air quality parameters and contributing factors, such as meteorological conditions, fuelwood combustion emissions, traffic emissions, and industrial



© University of Venda

activities. By leveraging historical air pollution data, MLR can identify significant predictors of $PM_{2.5}$ concentrations and generate predictive equations that assist in forecasting air quality trends; however, due to the complex and nonlinear nature of air pollution dynamics, polynomial regression extends or builds on MLR by incorporating higher-degree terms, allowing for a more flexible model that better captures intricate variations in $PM_{2.5}$ levels over time (Wang *et al*., 2022). Several studies have used these regression models to predict air quality, demonstrating their utility in understanding the impact of environmental variables on pollutant concentrations (Zhang *et al*. 2017; Ozer, 2024).

In contrast, machine learning techniques, such as Random Forest (RF), offer a more sophisticated and robust approach to air quality prediction. RF, an ensemble learning method, constructs multiple decision trees using random subsets of data to enhance predictive accuracy and mitigate overfitting (Breiman, 2001). This technique is particularly efficacious in capturing complex, nonlinear relationships between air pollution determinants and pollutants, such as $PM_{2.5}$ concentrations, rendering it a preferred choice in numerous air quality studies (Zhang *et al*., 2023). Through the analysis of extensive datasets comprising meteorological variables, land-use patterns, and emission sources, RF has demonstrated effectiveness in predicting future air quality levels across various seasons and locations, delivering both short-term and long-term predictions of $PM_{2.5}$ levels and other pollutants (Ji *et al.,* 2023; Babu and Thomas, 2023; Zhang *et al*., 2023). The robustness of RF facilitates more reliable predictions compared to traditional regression models, assisting policymakers in implementing proactive air pollution mitigation strategies. Its capacity to integrate diverse environmental variables also renders RF an indispensable tool for addressing the growing challenges in air quality management and public health.

In synthesising the literature on air quality monitoring, dispersion modelling, machine learning, and human health risk assessment, it is evident that studies commonly employ a combination of fixed regulatory monitors and low-cost sensors. Gaussian-based dispersion models such as AERMOD and CALPUFF simulate the transport and distribution of air pollutants, and health risk frameworks like the USEPA's HHRA model are applied to evaluate the potential health effects of pollutant exposure. In recent years, machine learning approaches have enhanced prediction accuracy, captured complex non-linear relationships



© University of Venda

in air quality data, and supported real-time forecasting systems. Regulatory-grade monitors provide high precision but often lack the spatial coverage offered by low-cost sensors. Dispersion models offer powerful tools for estimating pollutant distribution but are highly dependent on accurate meteorological data and rigorous model validation. Similarly, health risk assessments enable quantifiable exposure-risk analysis, yet they often rely on average dose assumptions and offer limited differentiation across demographic groups. This study integrates these methodological components to provide a site-specific analysis of $PM_{2.5}$ exposure, drawing on both observational data and modelling techniques. This approach lays the groundwork for the following chapters on air quality monitoring, modelling, risk assessment, and data-informed recommendations aligned with the study's objectives.

## References

Abt Associates. (2015). *AirCounts model documentation*.

Adeyemi, A. A. (2020). *$PM_{2.5}$ Chemical Composition, Source Apportionment and Geographical Origin of Air Masses in Pretoria, South Africa*. Doctoral dissertation, University of Pretoria (South Africa).

Ajayi, O. O., Ngene, B. U., & Ogbiye, S. A. (2021, March). Plume model: a simple approach to air quality control. In *IOP Conference Series: Materials Science and Engineering*, 1036(1): 012017). IOP Publishing. Accessed 02/04/23.

Akinshipe, O., & Kornelius, G. (2015). Atmospheric emissions from clamp kilns in the South African clay brick industry. In National Association for Clean Air (NACA) Annual Conference, Bloemfontein, South Africa.

Alam, S.A. (2006). Use of biomass fuels in the brick-making industries of Sudan: Implications for deforestation and greenhouse gas emission.

Aliyu, Y.A., Botai, J.O. (2018). Appraising city-scale pollution monitoring capabilities of multi-satellite datasets using portable pollutant monitors. *Atmospheric environment,* 179:239-249.

Altieri, K., & Keen, S. (2016). The cost of air pollution in South Africa. International Growth Centre Blog. Accessed 15/04/23.

Amaral, S. S., De Carvalho, J. A., Costa, M. A. M., & Pinheiro, C. (2015). An overview of particulate matter measurement instruments. *Atmosphere,* 6(9), 1327-1345.

American Lung Association (2013). American Lung Association State of the Air 2013, Health Effects of Ozone and Particle Pollution. Retrieved from http://www.stateoftheair.org/2013/health-risks. Accessed 02.04.22.



© University of Venda

Ana, G. R. (2011). Air Pollution in the Niger Delta Area: Scope, Challenges and Remedies. *The Impact of Air Pollution on Health, Economy, Environment and Agricultural Sources*, 181.

Ara, A., & Usmani, J. A. (2015). Lead toxicity: a review. *Interdisciplinary toxicology,* 8(2), 55-64.

Aron, J., & Muellbauer, J. (2006). Estimation of household sector wealth for South Africa, 1970-2003. *Review of income and wealth,* 52(2), 285-307.

Arya, S. P. (1999). Air pollution meteorology and dispersion. Oxford University Press.

Babatola, S. S. (2018). Global burden of diseases attributable to air pollution. *Journal of public health in Africa,* 9(3).

Babu, S., & Thomas, B. (2023). A survey on air pollutant $PM_{2.5}$ prediction using random forest model. *Environmental Health Engineering and Management Journal*, 10(2), 157-163.

Balakrishnan, K., Cohen, A., & Smith, K. R. (2014). Addressing the burden of disease attributable to air pollution in India: the need to integrate across household and ambient air pollution exposures.

Bertschi, I. T., Yokelson, R. J., Ward, D. E., Christian, T. J., & Hao, W. M. (2003). Trace gas emissions from the production and use of domestic biofuels in Zambia measured by open-path Fourier transform infrared spectroscopy. *Journal of Geophysical Research: Atmospheres,* 108(D13).

Bhattacharya, S. C., Salam, P. A., & Sharma, M. (2000). Emissions from biomass energy use in some selected Asian countries. *Energy,* 25(2), 169-188.

Bhuyan, P., Deka, P., Prakash, A., Balachandran, S., Hoque, R.R. (2018). Chemical characterization and source apportionment of aerosol over mid Brahmaputra Valley, India. *Environmental Pollution,* 234:997-1010.

Bonanno, G., Pavone, P. (2015). Leaves of Phragmites australis as potential atmospheric biomonitors of Platinum Group Elements. *Ecotoxicology and Environmental Safety* 114:31–37.

Bonjour, S., Adair-Rohani, H., Wolf, J., Bruce, N.G., Mehta, S., Prüss-Ustün, A., Lahiff, M., Rehfuess, E.A., Mishra, V., Smith, K.R. (2013). Solid fuel use for household cooking: country and regional estimates for 1980–2010. *Environmental health perspectives,* 121(7):784.

Bozilov, A., Tasic, V., Zivkovic, N., Lazovic, I., Blagojevic, M., Misic, N. and Topalovic, D., (2022). Performance assessment of NOVA SDS011 low-cost PM sensor in various microenvironments. *Environmental Monitoring and Assessment,* 194(9), pp.1-15.

Brandt, J., Silver, J. D., Christensen, J. H., Andersen, M. S., Bønløkke, J. H., Sigsgaard, T., ... & Frohn, L. M. (2013). Assessment of health-cost externalities of air pollution at the national level using the EVA model system. *Atmospheric Chemistry and Physics, 13*(15), 7725–7746. https://doi.org/10.5194/acp-13-7725-2013.

Breiman L (2001) Random forests. *Mach Learn* 45:5–32



© University of Venda

Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. *The Lancet,* 360(9341), 1233-1242.

Burns, D., *et al*. (2019). The impact of air pollution on mortality in the United States: Evidence from changes in wind direction (NBER Working Paper No. 26381). National Bureau of Economic Research. https://doi.org/10.3386/w26381. Accessed 02.04.19.

Buthelezi, S.A. and Davies, T.C., 2015. Carbon monoxide (CO), ozone ($O_3$) and nitrogen dioxide (NO2) exposure from vehicular transportation and other industrial activities in the vicinity of Umlazi Township, South of Durban, KwaZulu-Natal Province, South Africa. *Transactions of the Royal Society of South Africa*, *70*(3), pp.277-283.

Cacciola, R. R., Sarva, M., & Polosa, R. (2002). Adverse respiratory effects and allergic susceptibility in relation to particulate air pollution: flirting with disaster. *Allergy,* 57(4), 281-286.

Cao, J., Wang, Q., Chow, J.C., Watson, J.G., Tie, X., Shen, Z., Wang, P., An, Z. (2012). Impacts of aerosol compositions on visibility impairment in Xi'an China. *Atmos. Environ*. 59:559–566.

Carr, T. (2020). Book Review: Urban Health Issues: Exploring the Impacts of Big-City Living. *Reference & User Services Quarterly,* 59(2), 143-143.

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., …& Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?" *Environment International,* 99, 293-302.

Centre for Environmental Rights (CER). (2025). *Deadly Air case in Supreme Court of Appeals*. https://cer.org.za/news/deadly-air-case-in-supreme-court-of-appeals.

Chasant, M. (2019). Nigeria Leads Africa in Air Pollution-Related Deaths: Updated Estimates. Retrieved from https://www.atcmask.com/blogs/blog/air-pollution-Nigeria.

Cheng, B., Kumar, A. P. S., & Wang-Li, L. (2021). Inverse AERMOD and SCIPUFF dispersion modeling for farm-level PM10 emission rate assessment. *Transactions of the ASABE*, *64*(3), 801-817.

Chiloane, K. E. (2006). Volatile organic compounds (VOC's) analysis from Cape Town haze ll study (Master's thesis, University of the Witwatersrand, Johannesburg (South Africa)).

Choo, C.P., Jalaludin, J. (2015). An overview of indoor air quality and its impact on respiratory health among Malaysian school-aged children. *Reviews on environmental health,* 30(1):9-18.

Chung, A., Chang, D. P., Kleeman, M. J., Perry, K. D., Cahill, T. A., Ducher, D.,…& Stroud, K. (2001). Comparison of real-time instruments used to monitor airborne particulate matter. *Journal of the Air & Waste Management Association*, 51(1), 109-120.

Cimorelli, A.J., Perry, S.G., Venkatram, A., Weil, J.C., Paine, R.J., Wilson, R.B., Lee, R.F. and Peters, W.D. (2003). AERMOD description of model formulation. US



© University of Venda

Environmental Protection Agency Rep. *EPA* 454/R-03-002d. https://gaftp.epa.gov/aqmg/SCRAM/models/preferred/aermod/aermod_mfd.pdf

Clay, K., & Muller, N. Z. (2019). Recent increases in air pollution: evidence and implications for mortality (No. w26381). National Bureau of Economic Research.

Das, A., Singh, G., Habib, G., & Kumar, A. (2020). Non-carcinogenic and carcinogenic risk assessment of trace elements of $PM_{2.5}$ during winter and pre-monsoon seasons in Delhi: a case study. *Exposure and Health,* 12, 63-77.

Dastider, S. R., Chattopadhyay, S., & Chattopadhyay, G. (2023). Viewing the interrelationship between some pollutants under the purview of binary and composite fuzzy relation in a post-monsoon and winter scenario over a Metro city of India. *International Journal of Environmental Research*, *17*(6), 59.

DEA: Department of Environmental Affairs (2012). Fuel Burning Emissions: Contribution to total particulate, sulphur dioxide and nitrogen oxide emissions. http://soer.deat.gov.za/1322.html. Accessed on 18/06/ 2018.

DEA: Department of Environmental Affairs (2014). GHG Inventory for South Africa 2000–2010. Department of Environmental Affairs, Pretoria.

DEAT, (2009). National Environmental Management: Air quality Act 2004 (Act No 39 of 2004)- National Ambient Air Quality, Government gazette No. 32816. Department of Environmental Affairs and Tourism, Pretoria.

Directorate: Atmospheric Quality Information. (n.d.). *Briefing on the SAAQIS launch*. Retrieved from SAAQIS documentation (PDF).

DoE: Department of Energy (2013). Draft 2012 Integrated Energy Planning Report. Department of Energy, Pretoria.

Doherty, T., Kroon, M., Rhoda, N., & Sanders, D. (2016). Ending preventable child deaths in South Africa: What role can ward-based outreach teams play? *South African Medical Journal*, 106(7), 672-674.

Egbedi, H. (2017). Welcome to Onitsha: the city with the world's worst air. Retrieved from https://www.theguardian.com/cities/2017/feb/13/polluted-onitsha-nigeria-perpetual-dust-city-world-worst-air

Environmental Protection Agency. https://www.epa.gov/scram/air-quality-dispersion-modeling-preferred-and-recommended-models#aermod.

Eskom (2010). Eskom Annual Report. Eskom Holdings Limited. Available from: http://www.eskom.co.za/OurCompany/Investors/IntegratedReports/Pages/Annual_ State ments.aspx.

FAO: Food and Agriculture Organisation (1999). Cultivating Our Futures. FAO-Netherlands conference on the multifunctional character of agriculture and land. The Netherlands, 189.

Feuyit, G., Nzali, S., Lambi, J. N., & Laminsi, S. (2019). Air quality and human health risk assessment in the residential areas at the proximity of the Nkolfoulou landfill in Yaoundé metropolis, Cameroon. *Journal of Chemistry,* 2019.



© University of Venda

Filho, Geraldo & Meneguette, Rodolfo & Torres Neto, José & Valejo, Alan Demétrius & Weigang, Li & Ueyama, Jó & Pessin, Gustavo & Villas, Leandro. (2020). Enhancing intelligence in traffic management systems to aid in vehicle traffic congestion problems in smart cities. *Ad Hoc Networks.* 107. 102265. 10.1016/j.adhoc.2020.102265.

Flora, G., Gupta, D., & Tiwari, A. (2012). Toxicity of lead: a review with recent updates. *Interdisciplinary toxicology*, 5(2), 47-58.

Food and Agriculture Organisation (FAO), (1999). Cultivating Our Futures. FAO-Netherlands conference on the multifunctional character of agriculture and land. The Netherlands, 189.

Forbes, P. (2015). *Monitoring of air pollutants: sampling, sample preparation and analytical techniques* (Vol. 70). Elsevier.

Franklin, B.A., Brook, R., Pope, C.A. (2015). Air pollution and cardiovascular disease. *Current problems in cardiology*, 40(5):207-238.

Friedl, A., Holm, D., John, J., Kornelius, G., Pauw, C.J., Oosthuizen, R., van Niekerk, A.S. (2008). Air pollution in dense, low-income settlements in South Africa. In proceedings: National Association for Clean Air (NACA), 2008, Nelspruit, Mpumalanga.

Friedrich, R., & Bickel, P. (Eds.). (2001). *Environmental external costs of transport*. Springer.

Garland, Rebecca M., Bianca Wernecke, Gregor Feig, and Kristy Langerman. "The new WHO global air quality guidelines: what do they mean for South Africa? *Clean Air Journal* 31, no. 2 (2021): 1-2.

GBD 2021 HAP Collaborators. (2025). Global, regional, and national burden of household air pollution, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet,* 405(10485), 1167–1181. https://doi.org/10.1016/S0140-6736(24)02840-X.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature,* 521(7553), 452-459.

Ghorani-Azam, A., Riahi-Zanjani, B., & Balali-Mood, M. (2016). Effects of air pollution on human health and practical measures for prevention in Iran. *Journal of research in medical sciences,* 21(1), 65.

Ghosh, S. (2015). Development and application of Semiparametric Regression Models for estimation of short term health effects related to Air Pollution in Chennai.

Gibson, M. D., Kundu, S., & Satish, M. (2013). Dispersion model evaluation of $PM_{2.5}$, $NO_x$ and $SO_2$ from point and major line sources in Nova Scotia, Canada, using AERMOD Gaussian plume air dispersion model. *Atmospheric Pollution Research,* 4(2), 157-167.

Giechaskiel, B., Maricq, M., Ntziachristos, L., Dardiotis, C., Wang, X., Axmann, H., Bergmann, A. and Schindler, W., (2014). Review of motor vehicle particulate



© University of Venda

emissions sampling and measurement: From smoke and filter mass to particle number. *Journal of Aerosol Science,* 67, .48-86.

Gobeli, D., Schloesser, H. &PottbergMet T. (2008). Met One Instruments BAM-1020 Beta Attenuation Mass Monitor US-EPA $PM_{2.5}$ Federal Equivalent Method Field Test Results, Paper # 2008-A-485-AWMA.

Gruber, C. W., & Schumann, C. E. (1963). Cincinnati's Radioactive Fallout Monitoring Program. *Journal of the Air Pollution Control Association*, 13(5), 195-204.

Guarino, B., Stone, Z., & Morrar, S. (2018, November 14). California's deadliest wildfire is also a massive air-quality problem. The Washington Post. Retrieved from https://www.washingtonpost.com/nation/2018/11/15/californias-deadliest-wildfire-is-also-massive-air-quality-problem/ Accessed date: 25/07/2024.

Guo, X., Li, S., Zhang, Y., Wu, B., & Guo, W. (2022). Applications of dynamic simulation for source analysis of soil pollutants based on atmospheric diffusion and deposition model. *Science of The Total Environment*, *839*, 156057.

Guttikunda, D., & Jawahar, P. (2008). Simple Interactive Models for Better Air Quality. *India: Vehicular Air Pollution Information System*, 1-20.

Guttikunda, S. K., & Jawahar, P. (2012). Application of SIM-air modeling tools to assess air quality in Indian cities. *Atmospheric Environment, 62*, 551–561. https://doi.org/10.1016/j.atmosenv.2012.08.072.

Gwaze, P., & Mashele, S. H. (2018). South African Air Quality Information System (SAAQIS) mobile application tool: bringing real time state of air quality to South Africans. *Clean Air Journal,* 28(1), 3-3.

Gwilliam, K. M., Kojima, M., & Johnson, T. (2004). *Reducing air pollution from urban transport.* Washington, DC: World Bank.

Habeebullah, T.M., Munir, S., AA, A.H., Morsy, E.A., Seroji, A.R., Mohammed, A.M. (2015). The interaction between air quality and meteorological factors in an arid environment of Makkah, Saudi Arabia. *International Journal of Environmental Science and Development,* 6(8):576.

Hafkenscheid, T., & Vonk, J. (2015). Evaluation of equivalence of the MetOne BAM-1020 for the measurement of $PM_{2.5}$ in ambient air.

Hallowes, D., & Munnik, V. (2006). The Groundwork Report 2006: Poisoned Spaces: Manufacturing Wealth, Producing Poverty. Groundwork.

Hamra, G.B., Guha, N., Cohen, A., Laden, F., Raaschou-Nielsen, O., Samet, J.M., Vineis, P., Forastiere, F., Saldiva, P., Yorifuji, T., Loomis, D. (2014). Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environmental health perspectives,* 122(9):906.

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



© University of Venda

Hanna, A.F., Yeatts, K.B., Xiu, A., Zhu, Z., Smith, R.L., Davis, N.N., Talgo, K.D., Arora, G., Robinson, P.J., Meng, Q. and Pinto, J.P. (2011). Associations between ozone and morbidity using the Spatial Synoptic Classification system. *Environmental Health,* 10, 1-15.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. https://doi.org/10.1016/j.buildenv.2020.107025.

He, K., Huo, H., & Zhang, Q. (2002). Urban air pollution in China: current status, characteristics, and progress. *Annual review of energy and the environment,* 27(1), 397-431.

HEI: Health Effects Institute (2019). State of Global Air 2019. Data source: Global Burden of Disease Study 2017. IHME, 2018.

Heyder, J. (2004). Deposition of inhaled particles in the human respiratory tract and consequences for regional targeting in respiratory drug delivery. *Proceedings of the American Thoracic Society,* 1(4), 315-320.

Holstius, D. M., Pillarisetti, A., Smith, K. R., & Seto, E. (2014). Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California. *Atmospheric Measurement Techniques,* 7(4), 1121-1131.

Hoxha, V. (2025). Comparative analysis of machine learning models in predicting housing prices: a case study of Prishtina's real estate market. *International Journal of Housing Markets and Analysis*, 18(3), 694-711.

Huang, Y., Su, T., Wang, L., Wang, N., Xue, Y., Dai, W., *et al.* (2019). Evaluation and characterization of volatile air toxics indoors in a heavy polluted city of northwestern China in wintertime. *Science of the Total Environment*, *662*, 470-480.

IEA: International Energy Agency. (2023). Greenhouse Gas Emissions from Energy. https://www.iea.org/data-and-statistics/data-product/greenhouse-gas-emissions-from-energy. Accessed ate: 25/07/2024.

Institute for Health Metrics and Evaluation (2025). Processed by Our World in Data. “Average annual exposure to $PM_{2.5}$ air pollution” [dataset]. Institute for Health Metrics and Evaluation, “World Development Indicators” [original data]. Retrieved June 12, 2025. From https://ourworldindata.org/grapher/average-exposure-pm25-pollution.

IOM. (2014). *IOMLIFET Technical Documentation*. Institute of Occupational Medicine.

Ji, Y., Zhi, X., Wu, Y., Zhang, Y., Yang, Y., Peng, T., & Ji, L. (2023). Regression analysis of air pollution and pediatric respiratory diseases based on interpretable machine learning. *Frontiers in Earth Science*, 11, 1105140.

Jiang, R. T., Acevedo-Bolton, V., Cheng, K. C., Klepesis, N. E., Ott, W. R., & Hildemann, L. M. (2011). Determination of response of real-time SidePak AM510 monitor to secondhand smoke, other common indoor aerosols, and outdoor aerosols. *Journal of Environmental monitoring*, 13(6), 1695-1702.



© University of Venda

Joshi, P.C. and Mahadev, S. (2011). Distribution of air pollutants in ambient air of district Haridwar (Uttarakhand), India: A case study after establishment of State Industrial Development Corporation. *International Journal of Environmental Sciences*, 2(1):237.

Joubert, J.J. (2019). These areas have the worst air pollution in South Africa. Retrieved from https://www.thesouthafrican.com/news/worst-air-pollution-in-south-africa-cities-and-suburbs/ Accessed ate: 06/07/2024.

Kan, H., Chen, B., & Hong, C. (2009). Health impact of outdoor air pollution in China: current knowledge and future research needs.

Kaonga, B., & Ebenso, E. (2011). An Evaluation of Atmospheric Aerosols in Kanana, Klerksdorp Gold Mining Town, North-West Province of South Africa. In Air Quality Monitoring, Assessment and Management. INTECH Open Access Publisher Rijeka, Croatia.

Karekezi, S. (2002). Renewables in Africa—meeting the energy needs of the poor. *Energy policy*, 30(11-12), 1059-1069.

Kero, I., Naess, M. K., & Tranell, G. (2015). Particle size distributions of particulate emissions from the ferroalloy industry evaluated by electrical low pressure impactor (ELPI). *Journal of occupational and environmental hygiene,* 12(1), 37-44.

Khan, R., Vyas, H. (2008). A study of impact of brick industries on environment and human health in Ujjain city (India). *Journal of Environmental Research and Development,* 2(3).

Kim, J. M., Lee, N., & Xiao, X. (2019). Directional dependence between major cities in China based on copula regression on air pollution measurements. *PloS one*, 14(3).

Krzyzanowski, M., Kuna-Dibbert, B., & Schneider, J. (Eds.). (2005). Health effects of transport-related air pollution. WHO Regional Office Europe.

Kusiak, A. (2018). *Engineering Design: A Practical Guide.* CRC Press.

Kwon, H. S., Ryu, M. H., & Carlsten, C. (2020). Ultrafine particles: unique physicochemical properties relevant to health and disease. Experimental & Molecular Medicine, 1-11.

Landrigan, P.J. (2017). Air pollution and health. *The Lancet Public Health,* 2(1):4-5.

Latza, U., Gerdes, S., & Baur, X. (2009). Effects of nitrogen dioxide on human health: systematic review of experimental and epidemiological studies conducted between 2002 and 2006. *International journal of hygiene and environmental health,* 212(3), 271-287.

LEDET. (2016). Limpopo Environmental Outlook Report. Limpopo Economic Development Environment & Tourism (LEDET).

Lei, T., Xiang, W., Zhao, B., Hou, C., Ge, M., & Wang, W. (2024). Advances in analysis of atmospheric ultrafine particles and application in air quality, climate, and health research. *Science of The Total Environment*, 175045.

Li, F., Yan, J., Wei, Y., Zeng, J., Wang, X., Chen, X., Zhang, C., Li, W., Chen, M. and Lü, G., (2021). PM2. 5-bound heavy metals from the major cities in China:



© University of Venda

Spatiotemporal distribution, fuzzy exposure assessment and health risk management. *Journal of Cleaner Production, 286*, 124967.

Limpopo Department of Economic Development, Environment and Tourism (LEDET). (2013). Provincial Air Quality Management Plan. Limpopo Department of Economic Development, Environment and Tourism, Polokwane.

Liu, L., Wu, X., Li, S., Li, Y., Tan, S., & Bai, Y. (2022). Solving the class imbalance problem using ensemble algorithm: application of screening for aortic dissection. *BMC Medical Informatics and Decision Making*, 22(1), 82.

Lourens, A.S., Fourie, G.D., Burger, J.W., Pienaar, J.J., Read, C.E., Jordaan, J.H., Van Zyl, P.G. and Beukes, J.P., 2011. Spatial and temporal assessment of gaseous pollutants in the Highveld of South Africa. *South African Journal of Science*, *107*(1), pp.1-8.

Lu, Q. O., & Lee, C. C. (2025). Innovative Geo-AI model: An enhance of outdoor PM estimations based on land use and outdoor environmental factors in a highly polluted area. *Chemosphere*, *373*, 144178.

Ludwig, J., Marufu, L. T., Huber, B., Andreae, M. O., & Helas, G. (2003). Domestic combustion of biomass fuels in developing countries: A major source of atmospheric pollutants. *Journal of atmospheric chemistry*, 44(1), 23-37.

Mabusela, X. and Mamakoko, T. (2017). An overview of Joburg's vehicular emissions measurements and control methods.

Maji, S., Ahmed, S., Siddiqui, W. A., & Ghosh, S. (2017). Short term effects of criteria air pollutants on daily mortality in Delhi, India. *Atmospheric environment,* 150, 210-219.

Manzini, L. (2016). *Greenhouse Gas Emissions Assessment for Electricity Generation from Coal: An Eskom Power Station*. University of Johannesburg (South Africa).

Masih, A. (2019). Machine learning algorithms in air quality modeling. *Global Journal of Environmental Science & Management* (GJESM), 5(4).

Massetti, E., Brown, M. A., Lapsa, M., Sharma, I., Bradbury, J., Cunliff, C., & Li, Y. (2017). Environmental quality and the US power sector: air quality, water quality, land use and environmental justice. *ORNL*, *772*, 1-169.

Mathee, A., Rollin, H. B., von Schirnding, Y. E., Levin, J., & Naik, I. (2006). Reductions in blood lead levels among school children following the introduction of unleaded petrol in South Africa. *Environmental Research, 100*(3), 319–322.

McMurry, P.H. (2000). A review of atmospheric aerosol measurements. *Atmospheric environment,* 34(12-14), 1959-1999.

Meng, W., Zhong, Q., Yun, X., Zhu, X., Huang, T., Shen, H., ... & Wang, X. (2017). Improvement of a global high-resolution ammonia emission inventory for combustion and industrial sources with new data from the residential and transportation sectors. *Environmental science & technology,* 51(5), 2821-2829.

Mills, K., Ahalt, S., & York, W. (2018). *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. John Wiley & Sons



© University of Venda

Mills, N.L., Tornqvist, H., Gonzalez, M.C., Vink, E., Robinson, S.D., Soderberg, S., Boon, N.A., Donaldson, K., Sandstrom, T., Blomberg, A., Newby, D.E. (2007). Ischemic and thrombotic effects of dilute diesel-exhaust inhalation in men with coronary heart disease. *New England Journal of Medicine*, 357(11):1075-1082.

Minister of Environmental Affairs v Trustees for the Time Being of GroundWork Trust & Others, 2025 (549/2023) ZASCA 43.

Mitchell, T. M. (1997). *Machine Learning.* McGraw-Hill.

Moja, S.J. and Thivhafuni, P.H.U.M.U.D.Z.O., 2017. And PM10 Levels within the baphalaborwa municipality of South Africa. *Air pollution xxv*, *211*, p.97.

Morakinyo, O. M., Mokgobu, M. I., Mukhola, M. S., & Hunter, R. P. (2016). Health outcomes of exposure to biological and chemical components of inhalable and respirable particulate matter. *International journal of environmental research and public health,* 13(6), 592.

Morosele, I. P. (2020). *An Assessment of the Impacts of Commissioning Coal-Fired Power Stations on Ambient Air Quality*. University of Johannesburg (South Africa).

Motallebi, Nehzat, Hien Tran, Bart E. Croes, and Lawrence C. Larsen. "Day-of-week patterns of particulate matter and its chemical components at selected sites in California." *Journal of the Air & Waste Management Association* 53, no. 7 (2003): 876-888.

MRC: Medical Research Council (2008). 'South Africa Comparative Risk Assessment' Summary Report. http://www.mrc.ac.za/sites/default/files/files/2017-07-03/crasumrpt.pdf. Accessed 02.04.19.

Mulaudzi, S. K. (2006). *An estimation of emissions from domestic biofuel combustion over southern Africa*. University of the Witwatersrand, Johannesburg (South Africa).

Mundial, B. (1998). Pollution prevention and abatement handbook. Pulp and Paper Mills.

Naidoo, S., Piketh, S., & Curtis, C. (2017). Domestic Fuel Combustion in Un-electrified Low-income Settlements in South Africa.

NEMAQA (National Environmental Management: Air Quality Act). (2004). *Act No. 39 of 2004*. Government Gazette, Republic of South Africa.

Nkosi, V., Wichmann, J. and Voyi, K., 2016. Comorbidity of respiratory and cardiovascular diseases among the elderly residing close to mine dumps in South Africa: A cross-sectional study. *South African Medical Journal,* 106(3), pp.290-297.

Nkosi, V., Wichmann, J., Voyi, K. (2015). Chronic respiratory disease among the elderly in South Africa: any association with proximity to mine dumps? *Environmental Health,* 14(1):33.

NRC: National Research Council (2010). International impacts on local and regional air pollution: In Global sources of local pollution: An assessment of long-range transport of key air pollutants to and from the United States. *National Academies Press,* 11-34.



© University of Venda

Nunnari, S., & Pozzi, M. (2022). *Meta-analysis of inequality aversion estimates* (No. 9851). CESifo Working Paper.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *The New England Journal of Medicine,* 375(13), 1216-1219. https://doi.org/10.1056/NEJMp1606181

Odiyo, J.O., Bapela, H.M., Mugwedi, R., Chimuka, L. (2005). Metals in environmental media: A study of trace and platinum group metals in Thohoyandou, South Africa. *Water SA*, 31(4):581-588.

Offor, I. F., Adie, G. U., & Ana, G. R. (2016). Review of particulate matter and elemental composition of aerosols at selected locations in Nigeria from 1985–2015. *Journal of Health and Pollution,* 6(10), 1-18.

Ozer, M. M. (2024). Predictive Modeling of Urban Air Pollution Using Machine Learning and Unmanned Aerial Vehicle Platforms. In *Innovative Applications of Artificial Neural Networks to Data Analytics and Signal Processing* (pp. 79-115). Springer, Cham.

Pacyna, E. G., & Pacyna, J. M. (2001). An assessment of global and regional emissions of trace metals to the atmosphere from anthropogenic sources worldwide. *Environmental Reviews, 9*(4), 269–298.

Pandey, B., Agrawal, M., Singh, S. (2014). Assessment of air pollution around coal mining area: emphasizing on spatial distributions, seasonal variations and heavy metals, using cluster and principal component analysis. *Atmospheric pollution research,* 5(1):79-86.

Pascal, M., Corso, M., Chanel, O., Declercq, C., Badaloni, C., Cesaroni, G., ... & Medina, S. (2013). Assessing the public health impacts of urban air pollution in 25 European cities: Results of the Aphekom project. *Science of the Total Environment, 449*, 390–400. https://doi.org/10.1016/j.scitotenv.2013.01.077.

Phalen, R.F., & Phalen, R.N. (2012). Introduction to air pollution science. Jones & Bartlett Publishers.

Piketh, S.J. and Walton, N.M., 2004. Characteristics of atmospheric transport of air pollution for Africa. In *Air Pollution* (pp. 173-195). Springer, Berlin, Heidelberg.

Pope, R. J., Savage, N. H., Chipperfield, M. P., Ordóñez, C., & Neal, L. S. (2015). The influence of synoptic weather regimes on UK air quality: regional model studies of tropospheric column NO 2. *Atmospheric Chemistry and Physics,* 15(19), 11201-11215.

Prasetya, T. A. E., Taufik, M. R., Wardani, R. W. K., Septiarini, T. W., & Rosanti, E. (2023). Characteristics of ambient air pollutions in Delhi, India. Asian Journal of Water, Environment and Pollution, 20(3), 1-9.Geurts, Pierre & Ernst, Damien & Wehenkel, Louis. (2006). Extremely Randomized Trees. *Machine Learning,* 63. 3-42. 10.1007/s10994-006-6226-1.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data,* 1(1), 51-59.



© University of Venda

Rai, P.K., & Panda, L.L. (2014). Dust capturing potential and air pollution tolerance index (APTI) of some road side tree vegetation in Aizawl, Mizoram, India: an Indo-Burma hot spot region. *Air Quality, Atmosphere & Health,* 7 (1):93-101.

Rai, P.K., Panda, L.L., Chutia, B.M., Singh, M.M. (2013). Comparative assessment of air pollution tolerance index (APTI) in the industrial (Rourkela) and non-industrial area (Aizawl) of India: An Eco management approach. *African Journal of Environmental Science and Technology*, 7(10):944-948.

Raub, J.A., Mathieu-Nolf, M., Hampson, N.B., Thom, S.R. (2000). Carbon monoxide poisoning—a public health perspective. *Toxicology,* 145(1):1-14.

Rebala, Gopinath & Ravi, Ajay & Churiwala, Sanjay. (2019). An Introduction to Machine Learning. 10.1007/978-3-030-15729-6.

Reid, S., & Rolfe, P. (2014). *Air quality governance in South Africa: A review of current trends and challenges*. Council for Scientific and Industrial Research (CSIR).

Republic of South Africa (2012). The 2012 National Framework for Air Quality Management in the Republic of South Africa. Available at: https://www.dffe.gov.za/sites/default/files/gazetted_notices/nema_national_framework_aqm_gn115_0.pdf Accessed 02/10/22.

Republic of South Africa. (1996). *Constitution of the Republic of South Africa, 1996 (Act No. 108 of 1996)*. https://www.justice.gov.za/legislation/constitution/.

Rivas, I., Mazaheri, M., Viana, M., Moreno, T., Clifford, S., He, C., ... & Alvarez-Pedrerol, M. (2017). Identification of technical problems affecting performance of DustTrak DRX aerosol monitors. *Science of the Total Environment,* 584, 849-855.

Rodríguez, Juan & Kuncheva, Ludmila & Alonso, Carlos. (2006). Rotation Forest: A New Classifier Ensemble Method. *IEEE transactions on pattern analysis and machine intelligence*. 28. 1619-30. 10.1109/TPAMI.2006.211

Rohde, R. A., & Muller, R. A. (2015). Air pollution in China: mapping of concentrations and sources. *PloS one,* 10(8).

Rood, A.S. (2014). Performance evaluation of AERMOD, CALPUFF, and legacy air dispersion models using the Winter Validation Tracer Study dataset. *Atmospheric Environment,* 89, 707 – 720.

Samoli, E., Aga, E., Touloumi, G., Nisiotis, K., Forsberg, B., Lefranc, A., … & Brunstein, R. (2006). Short-term effects of nitrogen dioxide on mortality: an analysis within the APHEA project. *European Respiratory Journal,* 27(6), 1129-1138.

Scire, J. S., Strimaitis, D. G., & Yamartino, R. J. (2000). A user's guide for the CALPUFF dispersion model. *Earth Tech, Inc,* 521, 1-521.

Scorgie, Y., Burger, L.W., Annegarn, H.J. (2004). Socio-Economic Impact of Air Pollution Reduction Measures – Task 2: Establishment of Source Inventories, and Task 3: Identification and Prioritisation of Technology Options. Report compiled on behalf of NEDLAC, 25 June 2003.

Scott I (2008) Interpreting risks and ratios in therapy trials. Australian Prescriber, 31: 12–16.



© University of Venda

Seguel, J. M., Merrill, R., Seguel, D., & Campagna, A. C. (2017). Indoor air quality. *American journal of lifestyle medicine,* 11(4), 284-295.

SEI. (2018). *Rapid Co-benefits Calculator: A screening tool for evaluating climate and air pollution policies*. Stockholm Environment Institute.

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics: From air pollution to climate change* (3rd ed.). Wiley.

Setshedi, T. (2017). The Operation, Maintenance & Overall Management of the NAQI Stations in South Africa. Retreived from https://www.environment.gov.za/sites/default/files/briefingsession.pdf. Accessed date: 12/10/23.

Shah, A.S., Langrish, J.P., Nair, H., McAllister, D.A., Hunter, A.L., Donaldson, K., Newby, D.E., Mills, N.L. (2013). Global association of air pollution and heart failure: a systematic review and meta-analysis. *The Lancet,* 382(9897):1039-1048.

Shah, V., Jagupilla, S. C. K., Vaccari, D. A., & Gebler, D. (2021). Non-linear visualization and importance ratio analysis of multivariate polynomial regression ecological models based on river hydromorphology and water quality. *Water,* 13(19), 2708.

Sharma, N., Agarwal, A.K., Eastwood, P., Gupta, T., Singh, A.P. (2018). Introduction to Air Pollution and Its Control. *Air Pollution and Control*. Springer, 3-7.

Sheffield, P. E., Knowlton, K., Carr, J. L., & Kinney, P. L. (2011). Modeling of regional climate change effects on ground-level ozone and childhood asthma. *American journal of preventive medicine*, 41(3), 251-257.

Shen, J., Valagolam, D., & McCalla, S. (2020). Prophet forecasting model: A machine learning approach to predict the concentration of air pollutants ($PM_{2.5}$, $PM_{10}$, $O_3$, $NO_2$, $SO_2$, CO) in Seoul. South Korea. PeerJ, 8, e9961. https://doi. org/ 10. 7717/ peerj. 9961 Accessed date: 12/10/23.

Silverman, K.C., Tell, J. G., Sargent, E. V., &Qiu, Z. (2007). Comparison of the industrial source complex and AERMOD dispersion models: case study for human health risk assessment. *Journal of the Air & Waste Management Association,* 57(12), 1439-1446.

Singh, N., Yunus, M., Srivastava, K., Singh, S.N., Pandey, V., Misra, J.A., Ahmad, K.J. (1995). Monitoring of auto exhaust pollution by roadside plants. *Environmental Monitoring and Assessment*, 34(1):13-25.

Skinder, B.M., Sheikh, A.Q., Pandit, A.K., Ganai, B.A. (2014). Brick kiln emissions and its environmental impact: A Review. *Journal of Ecology and The Natural Environment,* 6(1):1-11.

Smith, K. R., & Pillarisetti, A. (2017). Household air pollution from solid cookfuels and its effects on health. *Injury Prevention and Environmental Health,* 133.

Snyder, M. G., Dacre, H. F., & Gray, S. L. (2022). Evaluation of assumptions in Gaussian plume modelling for regulatory applications. *Atmospheric Environment,* 277, 119089. https://doi.org/10.1016/j.atmosenv.2022.119089



© University of Venda

StatsSA: Statistics South Africa (2011). 2011 Census products, Statistics South Africa, Pretoria. http://www.statssa.gov.za/?page_id=3955.

Tan, J., Zhang, L., Zhou, X., Duan, J., Li, Y., Hu, J., & He, K. (2017). Chemical characteristics and source apportionment of PM2. 5 in Lanzhou, China. *Science of the Total Environment*, 601, 1743-1752.

Thabethe, N. D., Engelbrecht, F. A., & McGregor, J. L. (2014). Modelling the impact of air pollution on climate change in Gauteng, South Africa. *Environmental Research Letters*, *9*(8), 085001.

Thomas, E., Wickramasinghe, K., Mendis, S., Roberts, N., & Foster, C. (2015). Improved stove interventions to reduce household air pollution in low and middle income countries: a descriptive systematic review. *BMC public health*, 15(1), 650.

Tier, E. S. (3) (2014). Motor Vehicle Emission and Fuel Standards. Off. Transp. *Air Qual.* EPA-420-F-14, 9, 5-9.

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.

US EPA, (2009). Risk Assessment Guidance for Superfund Volume 1: Human Health Evaluation Manual (Part F, Supplemental Guidance for Inhalation Risk Assessment). U.S. Environmental Protection Agency Washington, DC. EPA-540-R-070-002. Accessed on 14th July 2024.

US EPA, (2019). Compilation of air pollutant emissions factors (AP-42). https://www.epa.gov/air-emissions-factors-and-quantification/ap-42-compilation-air-emissions-factors. Accessed date: 18/06/21.

US EPA, (2020). Our Nation's Air. Retrieved from https://gispub.epa.gov/air/trendsreport/2019/#welcome.

US EPA: United State Environmental Protection Agency (2004). User's Guide for the AMS/EPA Regulatory Model–AERMOD.https://www.epa.gov/scram/air-quality-dispersion-modeling-preferred-and-recommended-models#aermod.

US EPA: United State Environmental Protection Agency (2006). Six Common Air Pollutants. http://www.epa.gov/air/urbanair/index.html. Accessed 10/05/2018.

US EPA. (2021). *BenMAP-CE: Environmental Benefits Mapping and Analysis Program – Community Edition*. https://www.epa.gov/benmap Accessed date: 20/07/22.

Van der Bank, C. M., & Van der Bank, M. (2014). Environmental protection in South Africa: Human rights approach. *Mediterranean Journal of Social Sciences*, 5(23), 2059-2065.

Van Dingenen, R., Dentener, F., Crippa, M., Leitao, J., Marmer, E., Rao, S., ... & Valentini, L. (2018). TM5-FASST: A global atmospheric source–receptor model for rapid impact analysis of emission changes on air quality and short-lived climate pollutants. *Atmospheric Chemistry and Physics, 18*(21), 16173–16211. https://doi.org/10.5194/acp-18-16173-2018.



© University of Venda

Van Zyl, P. G., Beukes, J. P., Du Toit, G., Mabaso, D., Hendriks, J., Vakkari, V., Tiitta, P., Pienaar, J. J., Kulmala, M., Laakso, L. (2014). Assessment of atmospheric trace metals in the western Bushveld Igneous Complex, South Africa. *South African Journal of Science,* 110:1-10.

Verma, P., Verma, R., Mallet, M., Sisodiya, S., Zare, A., Dwivedi, G., & Ristovski, Z. (2024). Assessment of human and meteorological influences on PM10 concentrations: Insights from machine learning algorithms. *Atmospheric Pollution Research*, 15(6), 102123.

Wang, J., & Song, G. (2018). A deep spatial-temporal ensemble model for air quality prediction. *Neurocomputing*, *314*, 198-206.

Wang, Z., Calderón, L., Patton, A. P., Sorensen Allacci, M., Senick, J., Wener, R., ... & Mainelis, G. (2016). Comparison of real-time instruments and gravimetric method when measuring particulate matter in a residential building. *Journal of the Air & Waste Management Association*, *66*(11), 1109-1120.

Wang, Z., Li, H., Chen, H., Ding, Z. and Zhu, J., 2022. Linear and nonlinear framework for interval-valued $PM_{2.5}$ concentration forecasting based on multi-factor interval division strategy and bivariate empirical mode decomposition. *Expert Systems with Applications*, *205*, 117707.

Wenger, N. K., Boden, W. E., Carabello, B. A., Carney, R. M., Cerqueira, M. D., & Criqul, M. H. (2010). *Cardiovascular Disability: Updating the Social Security Listings.* National Academy of Sciences.

Wernecke, B., Wright, C.Y., Hey, J.V., Piketh, S., Burger, R., Kunene, Z., Panchal, R., Millar, D., Oosthuizen, D.N., Batini, C. & John, C., 2021. Opportunities for the application of low-cost sensors in epidemiological studies to advance evidence of air pollution impacts on human health. *Clean Air Journal*, 31(1), pp.1-4.

Winther, M., & Slento, E. (2010). Heavy metal emissions for Danish road transport.

World Bank (2017). *The Global Platform for Sustainable Cities (GPSC): Summary Report*. World Bank.

World Bank (WB), (1998). Greenhouse gas assessment handbook: A practical guidance document for the assessment of project-level greenhouse gas emissions. Global Environment Division, the World Bank, USA. 168p.

World Health Organization (WHO), (2014). 7 million premature deaths annually linked to air pollution. 2014. http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/. Accessed 10.05.18.

World Health Organization (WHO), (2016b). *Ambient air pollution: A global assessment of exposure and burden of disease*. World Health Organization.

World Health Organization (WHO), (2018). Ambient (outdoor) air pollution. Key facts. World Health Organization, Geneva. https://www. who. int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.



© University of Venda

World Health Organization (WHO), (2023). *AirQ+ version 2.2 – Quantifying the health impacts of air pollution*. https://www.euro.who.int/en/health-topics/environment-and-health/air-quality/tools/airq. Accessed Date: 18/06/24.

World Health Organization (WHO). (2016a). WHO global urban ambient air pollution database (update 2016). Geneva, Switzerland: WHO. Accessed Date: 18/06/24.

World Wide Fund for Nature – South Africa (WWF-SA). (2018). Transport emissions in South Africa. How mrcs the transport sector give rise to greenhouse gas emissions? https://dtnac4dfluyw8.cloudfront.net/downloads/wwf_2016_transport_emissions_in_south_africa.pdf. Accessed 14.06.2020.

Wright, C. Y., Nkosi, V., Wichmann, J., & Mathee, A. (2019). Indoor air pollution, public housing and respiratory health in children. *South African Medical Journal*, *109*(9), 654–658.

Wright, C.Y., Matooane, M., Oosthuizen, M.A., Phala, N. (2014). "Risk Perceptions of Dust and Its Impacts among Communities Living in a Mining Area of the Witwatersrand, South Africa." *Clean Air Journal*, 24(1):22-27.

Yevich, R., & Logan, J. A. (2003). An assessment of biofuel use and burning of agricultural waste in the developing world. *Global biogeochemical cycles*, 17(4).

Zhang, H., Zhang, Y., Wang, Z., Ding, M., Jiang, Y., Xie, Z. (2016). Traffic-related metal(loid) status and uptake by dominant plants growing naturally in roadside soils in the Tibetan plateau, China. *Science of the Total Environment* 573:915–923.

Zhang, J., Ding, W. (2017). Prediction of air pollutants concentration based on an extreme learning machine: the case of Hong Kong. *International journal of environmental research and public health*, 14(2):114.

Zhang, Y. L., & Cao, F. (2015). Fine particulate matter ($PM_{2.5}$) in China at a city level. *Scientific reports,* 5, 14884.

Zhang, Y., Sun, Q., Liu, J., & Petrosian, O. (2023). Long-term forecasting of air pollution particulate matter ($PM_{2.5}$) and analysis of influencing factors. *Sustainability,* 16(1), 19.

Zhu, H., & Maibach, H. I. (2020). *Skin decontamination.* Springer International Publishing.



© University of Venda

# CHAPTER THREE

# QUANTIFICATION AND INVENTORY OF EMISSIONS FROM ANTHROPOGENIC SOURCES: GASEOUS ($SO_2$, $NO_x$, CO) AND PARTICULATE MATTER ($PM_{2.5}$, $PM_{10}$) IN THULAMELA MUNICIPALITY OF VHEMBE DISTRICT, SOUTH AFRICA

## Abstract

*A comprehensive inventory of anthropogenic emissions was required to understand the sources of air pollution and design effective mitigation strategies. This study created an emission inventory for two major anthropogenic sources of criteria air pollutants ($SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$) in the Thulamela Municipality of South Africa for 2021 using the "bottom-up" approach methodology. Activity information corresponding to various sectors, related emission factors, and other secondary data were acquired from the field, literature, and regulatory bodies. More reliable survey-based data and local emission factors had been used to reduce the uncertainties in activity and emission rates. Results indicated that the emissions of $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ for the study domain in the Vhembe district for the year 2021 were 66.3, 1821.7, 21002.6, 2348.7, and 2649.3 tonnes per annum (tpa), respectively. Residential wood combustion (RWC) was the primary source of the criteria pollutants, accounting for 96% of $PM_{10}$ and $PM_{2.5}$, and 89% of CO, while the vehicle sector was the leading source of $NO_x$ (89%) and $SO_2$ (60 %). The gridded spatial distribution of the emissions revealed that they were mostly concentrated in the southwest and northwest of the area, where population density is a factor.*

**Keywords:** Emission inventory, Anthropogenic sources, Criteria air pollutants, Ambient air pollution, Bottom-up approach, Vhembe District

## 3.1 Introduction

Air pollution is a growing concern for public health in many parts of the world, with emissions from various sources contributing to the problem; for example, a fast-developing economy in India has caused people to move from rural to urban areas, resulting in a significant rise in population and urbanisation, leading to high levels of air pollution (Mentese *et al*., 2020; Singh, 2020). Likewise, increased demand for energy, industrial, and agricultural products has resulted in high levels of air pollution in South Africa, which has detrimental effects on human health (HEI, 2018; Muyemeki *et al*., 2022). Emission inventories play a crucial role in addressing these air pollution challenges.



© University of Venda

An emission inventory is a comprehensive and systematic record of pollutants released into the atmosphere by various sources (anthropogenic and natural) within a given area, such as a city or country. They are essential for understanding air pollution sources, magnitudes, and ways to improve air quality and protect public health as emission rates, local topography, and prevailing weather conditions affect an area's air quality. Additionally, it has been noted that long-distance air pollution transport from affected areas can impact regional and global areas with high pollutant concentrations (Piketh *et al*., 2002; Freiman & Piketh, 2003; Wang, 2007). For policymakers, regulators, and researchers, accurate inventories enable them to identify and target the sources of pollution that are most harmful to public health and to establish effective control measures.

Studies by researchers such as Koenig (2000) and Lelieveld *et al*. (2015, 2020) have highlighted the severe impacts of air pollution on public health, which include respiratory problems, heart disease, and premature deaths. South Africa has made significant efforts to reduce emissions and improve air quality, although pollution from various sources across different regions remains a significant issue. Burning fuels for residential activities (cooking and heating), industrial activities, and transportation, for example, all contribute to the high levels of pollutants in the air. These pollutants include particulate matter, sulphur dioxide, nitrogen oxides, and carbon monoxide, known as criteria pollutants. As a result, scholars have developed emission inventories to monitor and manage air pollution in the country; however, significant improvement in emission estimates is still ongoing (Scorgie *et al*., 2003; Blignaut *et al*., 2005; Scorgie, 2012; Seymore *et al*., 2014; Naidoo *et al*., 2019).

Spatial disaggregation of emissions has become an essential aspect of air quality management. It enables the creation of detailed emission inventories for developing air quality models, identifying pollution sources and transport routes, mapping pollutant distribution in the atmosphere, and devising effective control measures (Sindhwani *et al*., 2015; Jagarnath and Thambiran, 2018; Xu *et al*., 2020; Lestari *et al*., 2022). Geographical Information Systems (GIS) and remote sensing help identify pollution hotspots and evaluate control measures (Al-Ramahi, 2020; Sharma *et al*., 2021; Joshi, 2023). It is imperative in areas with high pollution sources, such as urban and industrial sites, that accurate



© University of Venda

inventories are required for assessing the impacts on public health and the environment (Lelieveld *et al*., 2015; Landrigan *et al*., 2018). In this study, spatial disaggregation of emissions was utilized to develop a high-accuracy inventory for Thulamela Municipality, providing valuable information for further research on the impact of anthropogenic emissions on air quality and climate change at different scales.

In the case of the Thulamela Municipality in South Africa, categorised as a low-income area, commercial practices, vehicle activities, and domestic land use are common and are usually significant sources of air pollution. Scholars have consistently studied how these activities contribute to ambient air pollution, including the Vaal Triangle Airshed Priority Area (VTAPA) (Freiman & Piketh, 2003; Scorgie, 2012; DEA, 2018; Naidoo *et al*., 2019; Chidhindi, 2022). These studies, however, are not relevant, specifically, to the Thulamela Municipality due to the heterogeneity of air composition; therefore, there is an urgent need for compiling emission inventories for vehicle and residential wood combustion (RWC) emissions within these regions. Understanding emissions in this area is essential for atmospheric studies, particularly anthropogenic emissions, which may have adverse health consequences and reduce visibility. Estimating emissions, thus, is necessary within Thulamela Municipality to obtain a picture of the nature of the emissions from various activities, thereby regulating the air quality in the environment.

This study aims to develop a 2021 emission inventory for the study domain (objective one), focusing on vehicle and residential emissions, source contributions from the two sources, and spatial disaggregation of the emissions. On-site data collection was conducted for both sources, and an inventory was created for the study area using a 1 km grid spatial resolution.

### 3.1.1 Study Domain

The study was conducted in Thulamela Municipality, particularly Thohoyandou and its neighbouring cities, in the Vhembe District of Limpopo Province, South Africa (Figure 3.1). The Municipality is located in the eastern part of the Vhembe District, with an approximate area of about 5835 $km^2$ and has 156,594 households (StatsSA, 2011). Rural and urban land use is mixed in this area, with most residents living in rural areas and relying on agriculture for income. Thohoyandou is mostly a residential town, one of the most densely-populated



© University of Venda

cities in Thulamela and has many commercial, warehouses and industrial establishments. Likewise, Lunungwi and Gondeni is a rural and low-income settlements located in Thulamela Municipality, where the activity data for residential emissions was collected.

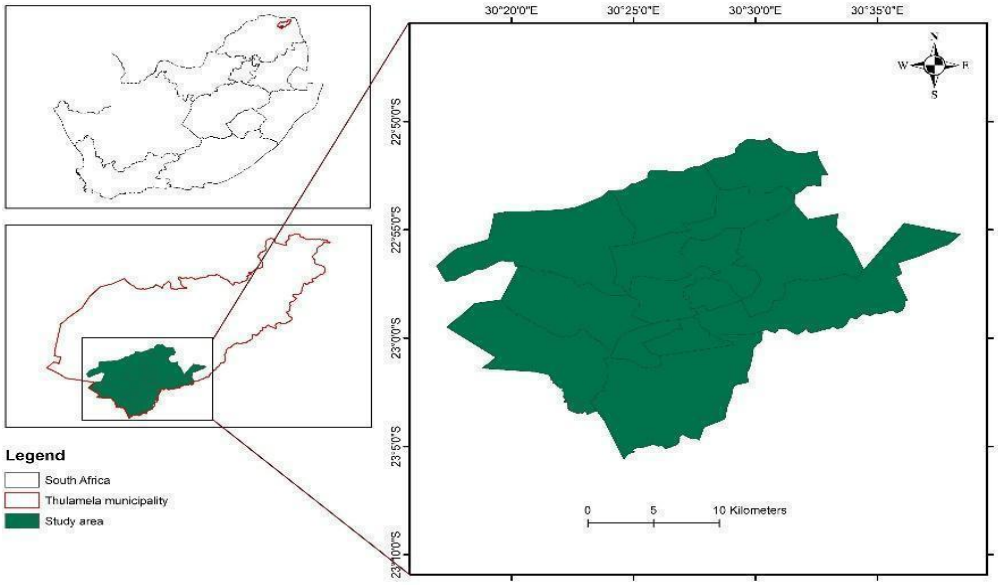



**Figure 3.1:** Study domain map of Thulamela Municipality.

## 3.2 Materials and Methods

The emission of criteria pollutants, namely $SO_2$, $NO_x$, CO, $PM_{2.5,}$ and $PM_{10}$, due to residential and vehicular emissions within the study domain, was estimated using a bottom-up approach, and the detailed methodology for the two sectors is described in this section. The emission inventory was spatially allocated using a 1 km grid resolution. The vehicle emissions were allocated based on the road network in the study area, while the residential emissions were allocated based on the location and number of households.

### 3.2.1 Sampling Sites and Data Collection

#### 3.2.1.1 Sampling sites

The sampled locations were selected based on emissions from two anthropogenic activities within the study area, which include residential wood burning and vehicle emissions, along with meteorological data of the area, which were obtained from the South African Weather



© University of Venda

Service (SAWS); other secondary data were also obtained from Statistics South Africa (StatsSA). Five different locations (Appendix 3.1) based on the availability of drivers and vehicle activities were selected for the Vehicle Emission Inventory (VEI) to record odometer readings for 7 days. For the residential wood combustion (RWC) sampling, 30 households within the villages (15 households per village) were selected using a purposive and snowball sampling method to measure the amounts of fuelwood consumed for both summer and winter periods in the residential sector (Figure 3.2) and the latitude and longitude questionnaires as shown in Appendix 3.2. The emissions of the gaseous ($SO_2$, $NO_x$, CO) and particulate pollutants ($PM_{2.5}$ and $PM_{10}$) were estimated based on the activity data of each emission source category and suitable emission factors selected from the literature.

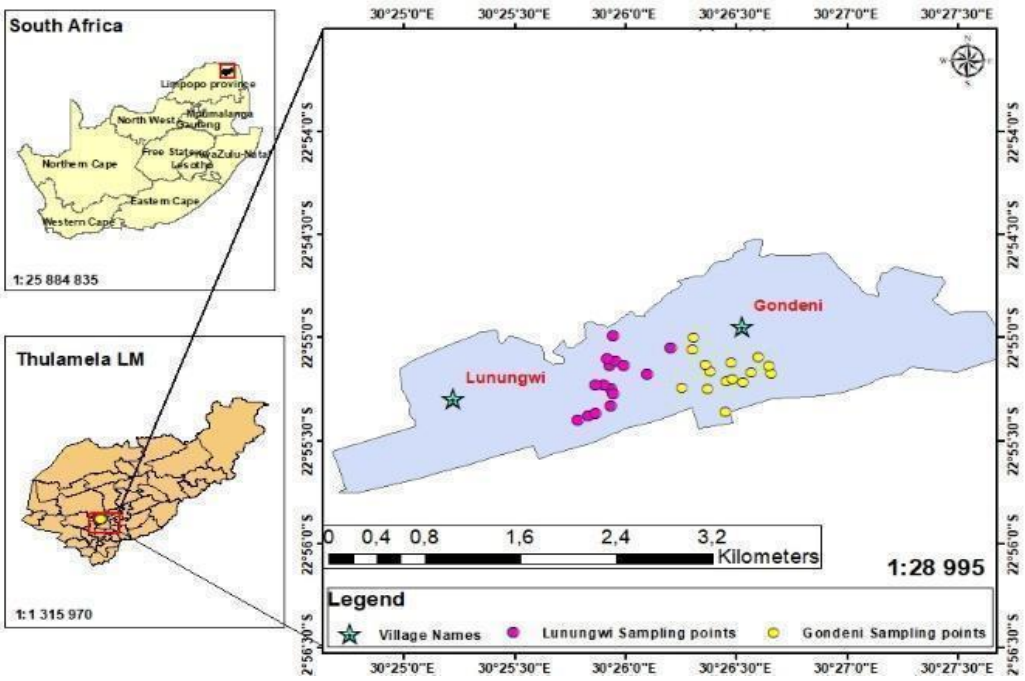



**Figure 3.2**: Sampling point where fuelwood usage was measured.

### 3.2.1.2 Data collection

Data were collected through a questionnaire survey, scale measurement, field visits, and literature, which includes both primary and secondary data. The primary data were achieved through questionnaires and wood scale measurement, while secondary data were obtained from literature reviews and analysis of documents like Statistics South Africa (StatsSA) and



© University of Venda

eNatis. For vehicle emissions, a questionnaire-based quantitative vehicle survey was developed to collect data randomly. In total, 100 questionnaires were administered in the sampling locations, including a taxi rank (Figure 3.3) based on the availability of drivers and the vehicle activities around the study area. Logbooks were given to drivers to record their odometer readings for 7 days (Appendix 3.3 contains the questionnaire and logbook). Out of the questionnaires distributed, seventy-five were retrieved in this proportion: 20, 31, 20, and 4 for passenger cars (PC), heavy passenger vehicles (HPV), light-duty vehicles (LDV), and heavy-duty vehicles (HDV), in accordance with vehicle categories (Enitan *et al*., 2021). The variables that had been requested from the participants included odometer readings before and after each day of service, kilometres travelled per day, the date of record, registration date, car model, and service date.



**Figure 3.3:** Thulamela taxi rank as one of the locations where the questionnaire was administered.

The questionnaire was also planned to enable the collection of data on fuelwood consumption rate (activity rate) as well as factors affecting them for residential emission estimates. A total of 180 questionnaires with closed and open-ended questions were administered to the participants of the two villages to obtain additional information regarding fuelwood use, employing purposive and snowball methods (see Appendix 3.4). The purpose of the questionnaires was to collect and analyse data regarding fuel wood consumption, demographics, fuel type and availability, as well as consumption rates. Emphasis was



© University of Venda

placed on household activities such as cooking, heating, and lighting. Based on the results from the above information, we generated the details in Table 3.1. The amount of wood consumed by each household in the study area was measured daily using a crane scale and recorded for both summer and winter.

**Table 3.1:** Residential fuel use within Thulamela municipality (Data was derived during field study)

| | **Summer** | **Winter** |
|---|---|---|
| Cooking | Wood and Electricity | Wood and Electricity |
| Heating | Electricity | Wood and Electricity |
| Lightning | Electricity | Electricity |

### 3.2.2 Development and Estimation of Sector-Based Emission Inventory

The emission inventory was developed using a bottom-up approach that combines activity and emission factor data for different sources. Several researchers have used this to develop anthropogenic emission inventories (Sindhwani *et al*., 2015). The bottom-up strategy involves gathering information on processes, activity rates, and their levels along with their estimated emission factors (EFs) from each source. This approach has more accurate estimates than the top-down approach (Nicholls *et al*., 2015). Emission inventories of the pollutants for the area were developed for residential and vehicular sources. All calculations that were used to estimate pollutants' emissions were based on the general formula used by the US EPA in the AP 42 Compilation of Air Pollutant Emission Inventory (USEPA, 1997; 2006).

#### 3.2.2.1 Vehicle sector

Emissions from various vehicle types were estimated using a bottom-up approach based on the Vehicle-Based Method Vehicle Kilometres Travelled (VBM-VKT), by multiplying the VKT with the emission factors for each vehicle type and the number of days, as shown in equation (3.1) (Gurjar *et al*., 2004; Ntziachristos & Samaras, 2018; EEA, 2019).

$$E_{i;j} = VKT_j \times EF_{i,j} \times 365 \qquad (3.1)$$

Where,



© University of Venda

$E_i$ is the total annual vehicle emission of pollutants(i), in ton/year; for all vehicle (j), $i$ is an index of the pollutant type which includes $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$.
$VKT_j$ is the daily kilometres travelled per vehicle type, and
$EF_{i,j}$ is the emission factors of pollutants(i), per vehicle type (j).

**A. Activity Data for vehicle emission**

***Vehicle population (VP)***

The vehicle population (VP) data for the study area from 2012 and 2018 was obtained from the National Traffic Information System (eNaTis). Due to the unavailability of data for 2019 to 2021, the number of vehicles for this period was extrapolated using a trend analysis method, specifically applying the trend function to the seven years of available datasets from 2012 to 2018. This method accurately predicts future values based on observed patterns and changes. Vehicles were classified into four types: passenger cars (PC), heavy passenger Vehicles (HPV), light-duty vehicles (LDV), and heavy-duty vehicles (HDV), while motorcycles were disregarded due to their insignificance.

***Vehicle kilometre travel (VKTs)***

Vehicle kilometre travel (VKT) is a significant tool in the estimation of vehicle emission (Huo *et al*., 2012; Limanond *et al*., 2010). The calculation of VKT for vehicles, for this study, was based on registered vehicles and odometer readings (Jung *et al*., 2017). The odometer data were collected and recorded before and after each day of activities, while, the odometer readings of vehicles and kilometres travelled per day were collected through a questionnaire survey (Appendix 3.3)

The daily average kilometre travelled (AKT) was calculated using equation 3.2 while the AKT value obtained was converted to daily VKT for all vehicle types using equation 3.3 (de Azevedo & Cardoso, 2009; Shabadin *et al*., 2014).

$$AKTj(km) = \frac{Dj}{Total\ samples} \quad (3.2)$$

$$VKT(km) = AKT\ \times Vehj \quad (3.3)$$



© University of Venda

Where,

$AKT_j$: Average distance travelled per vehicle type (j), VKTj: Daily kilometre travelled per vehicle type (j), Vehj: Number of registered vehicles per type (j), and Dj: Total kilometre travelled per different vehicle types (j).

**B. Emission factor (EF) for the vehicle sector**

An emission factor (EF) is a variable that specifies the mass of pollutants emitted from an activity normalised by the intensity of the activity. It is typically expressed in units of mass per unit of activity. The EF of $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ for different vehicle types reported by Ramirez *et al*., (2019) and Zhao *et al*., (2018), as shown in Table 3.2, were used in this study. These emission factors were chosen because they are more current, comprehensive, relevant, and applicable to our study's requirements, ensuring accuracy and reliability in our emission inventory.

**Table 3.2:** Emission factors (g/km) for $SO_2$, $NO_X$, CO, $PM_{2.5}$, and $PM_{10}$ for calculation of vehicular emissions.

| **References** | **Pollutants (g/km)** | | | | |
|---|---|---|---|---|---|
| **Vehicle type** | **$SO_{2(a)}$** | **$NO_{X(a)}$** | **$CO_{(a)}$** | **$PM_{2.5}$** | **$PM_{10(a)}$** |
| Passenger cars (PCs) | 0.062 | 0.149 | 1.265 | 0.004(b) | 0.005 |
| Heavy passenger vehicles (HPCs) | 0.084 | 0.249 | 1.336 | 0.038b | 0.006 |
| Light-duty vehicles (LDVs) | 0.019 | 1.747 | 2.152 | 0.100a | 0.109 |
| Heavy-duty vehicles (HDVs) | 0.018 | 1.461 | 2.194a | 0.085a | 0.092 |

Sources: (a: Ramirez *et al*., 2019; b: Zhao *et al*., 2018)

### 3.2.2.2 Emission for the residential sector

Emission rate ($ER_{rwc}$) (g/h) in the residential sector per village and area was estimated using Eq.3.4 (Gurjar et al., 2004; Nkosi et al., 2018), where the activity rate (AR) (the fuelwood consumption) was multiplied by the emission factor of each pollutant. Emission estimates were multiplied by the number of households (HH) using wood for residential activities in



© University of Venda

the study area to determine the fuelwood emissions per area (Thulamela Municipality) (Nkosi *et al*., 2018). The number of HH and villages using fuelwood for residential activities in the study area was obtained from StatsSA and the community survey 2016 (StatsSA, 2011).

$$E_{rwc}\ (g/hr) = AR_{rwc}\ (kg/hr) \times EF\ (g/kg) \times \boldsymbol{n}\ (HH) \quad (3.4)$$

Where, $E_{rwc}$ (g/hr) is the RWC emission rate, $AR_{rwc}$ is the fuelwood burned per HH per hour (wood consumption rate) (kg per hour), EF is the emission factor (g/kg of a fuel burned), and n(HH) is the number of HH that burn fuelwood in the area.

**A. Activity data for residential emissions**

***Scale measurement of fuelwood***

The amount of wood used by each HH was measured daily using a crane scale and recorded for both summer and winter (Figure 3.4). The crane scale, as shown in Figure 3.5, measures up to 300 kg (ACS Crane Scales). This amount of wood consumed was used to calculate the activity rate or fuelwood consumption rate, per hour, for a day, by each household and the area, utilising equation 3.5. The activity rate ($AR_{rwc}$) (wood consumption rate) is the amount of wood consumed (kg) per time.

$$AR_{rwc}\ (Kg/hr) = Q(kg)/\text{number of hours} \quad (3.5)$$

Where, $AR_{rwc}$ is the activity rate and the amount or quantity of wood consumed (kg) per time, and Q is the quantity of fuelwood consumed.

The activity rate (kg/hr) of each household was calculated for morning and evening cooking sessions, and their daily average was used for the hourly and daily value of each season by all 15 households in each village. Wood consumption data were collected for 4 months of the year, 2 months (December and January) for summer, and 2 months (July and August) for winter, in 2020, at Lunungwi and Gondeni villages. The total burning period of 5 hours in winter and 4 hours in summer was used to calculate the activity and emission rate,



© University of Venda

considering both morning and evening burning times. It was assumed that all the houses using wood used three-legged stands and fireplaces in the villages of the municipality, as indicated in the questionnaire survey.




**Figure 3.4:** Collecting the amount of wood used at a) Gondeni village, and b) Lunungwi village.



**Figure 3.5:** Crane scale used for wood measurement.

### B. Emission factors (EF)

Emission factors (EF) are the rate at which pollutants are emitted from wood burning and are crucial for estimating emissions in the inventories and for air pollution. Fuel mass-based emission factors ($EF_m$) of gaseous and particulate pollutants are reported as grams of pollutants per kilogram of fuel consumed (g/kg). Reference was made to several sources, including the US-EPA. Ap-42 1993, 1996, and 2006 and Scorgie, 2012, to get the appropriate EF to use for the condition of RWC in Vhembe District, South Africa. Residential combustion processes and activities can vary between countries, therefore, it is important to consider local emission factors, if available, or those previously used in local and regional



© University of Venda

studies. This is to ensure accurate emission estimates for residential combustion. Based on this, the emission factors from Scorgie (2012) and US EPA (Scorgie, 2012; and US-EPA Ap-42 1993, 1996, and 2006) were selected for the emissions estimate in this study (Table 3.3).

**Table 3.3:** Emission factors for residential fuelwood use

| Pollutants | Wood (g/kg) |
|---|---|
| $PM_{2.5}$ | 15.3[a] |
| $PM_{10}$ | 17.3[abc] |
| $SO_2$ | 0.2[a] |
| CO | 126.3[b] |
| $NO_x$ | 1.3[bd] |

(Source: a: USEPA 1993, b: USEPA 1996, c: USEPA 2006, and d: Scorgie, 2012)

### 3.2.3 Ethical Considerations

This research involved field data collection and sampling, necessitating permission to access the study area. The study was approved by the University of Venda's ethical committee (certificate number: SES/19/ERM/20/0511) and adhered to the University Code of Ethics guidelines (Appendix 3.5). Permission was also obtained from relevant authorities and local traditional leadership. Participants were assured of confidentiality and anonymity during data collection, and their participation was voluntary

### 3.2.4 Quality Assurance and Control (QA/QC)

Activity data, such as fuel consumption and VKT, are often based on estimates or measurements that require careful documentation and validation. Similarly, emission factors used to calculate emissions from these activities were derived from reliable and representative data sources. Regular checks of data inputs were carried out, including the review and validation of data sources, as well as the application of consistent methodologies for data handling and calculations. We examined the accuracy and completeness of the inventory data by comparing it with others. The calculations and assumptions used to estimate emissions were also rechecked to ensure they were reasonable and appropriate. This verification process helped verify consistency and strengthened the reliability of emission inventories.



© University of Venda

## 3.3 Results and Discussion

### 3.3.1 Vehicle Sector

#### 3.3.1.1 Vehicle population, age, and kilometre travelled

***A. Vehicle Population (VP)***

The vehicle population trends for the four-vehicle classifications (PC, HPV, HDV, and HDV) and their proportion from 2012 to 2021 in Thulamela Municipality were examined. Figure 3.6 illustrates these trends, highlighting passenger cars as the dominant vehicle type with the highest population compared to others. This trend may be driven by the increased vehicle ownership, the affordability of vehicles as well as population and economic growth. This is consistent with earlier research conducted by Progiou and Ziomas (2011), Song and Hao (2019), and Lv *et al*. (2019), which also reported a significant population of passenger cars in similar areas. This consistency in the findings supports the notion that passenger cars remain the most common type of vehicle in Thulamela Municipality. The total population for all vehicle types (PC, HPV, HDV, HDV) from 2012 to 2021 were 27699, 30315, 32520, 35803, 36421, 31970, 36421, 37790, 38981, and 40173 for PCs, HPVs, LDVs, and HDVs respectively (eNaTIS, 2020) with an estimated Compound Annual Growth Rate (CAGR) of 4.2%. According to Kumar & Tripathi (2014), an increase in the growth rate of the vehicular population could result in an increase in vehicular emissions leading to a spike in ambient air pollution.

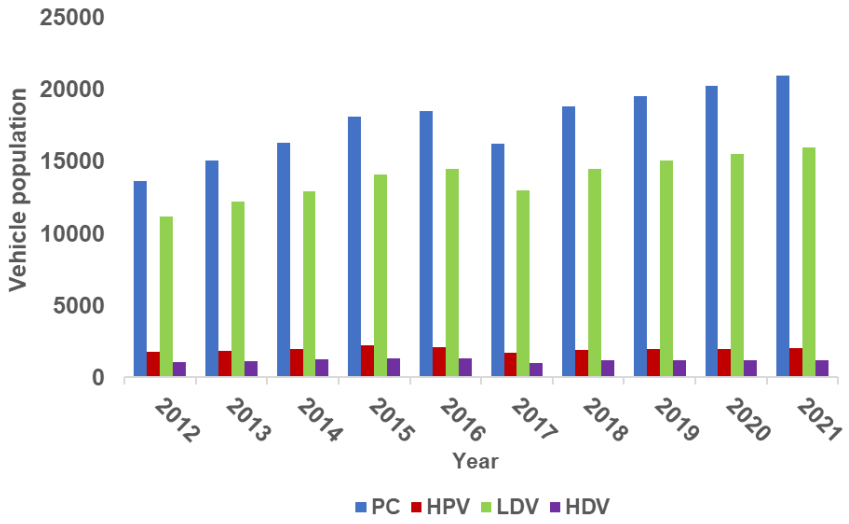



**Figure 3.6:** Population of Vehicles by Category in Thulamela, from 2012 to 2021 (Source: eNaTIS, 2020).



© University of Venda

### *B. Vehicle age*

Vehicle age significantly impacts fuel efficiency, VKT, and emissions. Figure 3 shows the variation in average vehicle km/l of fuel consumed with vehicle age. PCs within the Thulamela Municipality show a clear decline in VKT per litre as they age from the 2005 to 2016 model, except for a slight increase seen between 2001 and 2004. Interestingly, this PC trend aligns with Huo *et al*. (2012) finding that vehicle km/l decreases with vehicle age. For HPV, LDV, and HDV vehicles, the analysis of the typical vehicle kilometres travelled per litre values revealed a clear declining trend with age. This pattern is consistent with earlier research by Limanond *et al*. (2009). The analysis for the sampled HDV vehicles concentrated on the years 2009-2012 and 2017-2020, excluding vehicles from the years 2001-2008 and 2013-2016. In 2017–2020, 3.57 km/l were observed compared to 2009–2012, which was reduced to 3.08 km/l due to the impact of vehicle age on fuel efficiency (Figure 3.7).

Research shows that as vehicles age, their fuel efficiency decreases, leading to fewer kilometres travelled per litre (km/l) and may result in higher emissions, thus, a positive correlation was found between vehicle age and distance travelled, with newer models having higher VKT than older ones (Kenworthy & Laube, 1996; Stead, 1999; Limanond *et al*., 2009; Munyon *et al*., 2018). Factors such as wear and tear, lack of maintenance, and outdated technology could contribute to this decline. To improve fuel efficiency and reduce carbon footprint, Thulamela Municipality vehicle owners should consider regular servicing and upgrading to newer models. Linear regressions between the average distance travelled and the vehicle age for all vehicle types from 2001-2020 was conducted (Figure 3.8). These show a positive correlation, which suggests a strong association between the average distance travelled per litre and vehicle age, as the coefficient was greater than 0.87, hence, this suggests that vehicle kilometres travelled per litre decreases with age.



© University of Venda

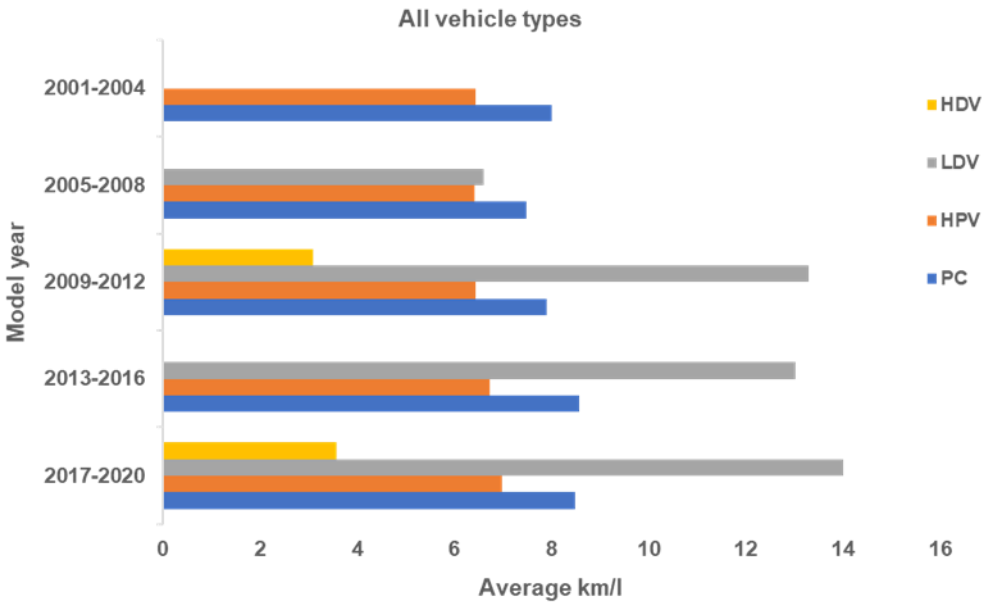



**Figure 3.7:** Average kilometre travelled by all the vehicles.

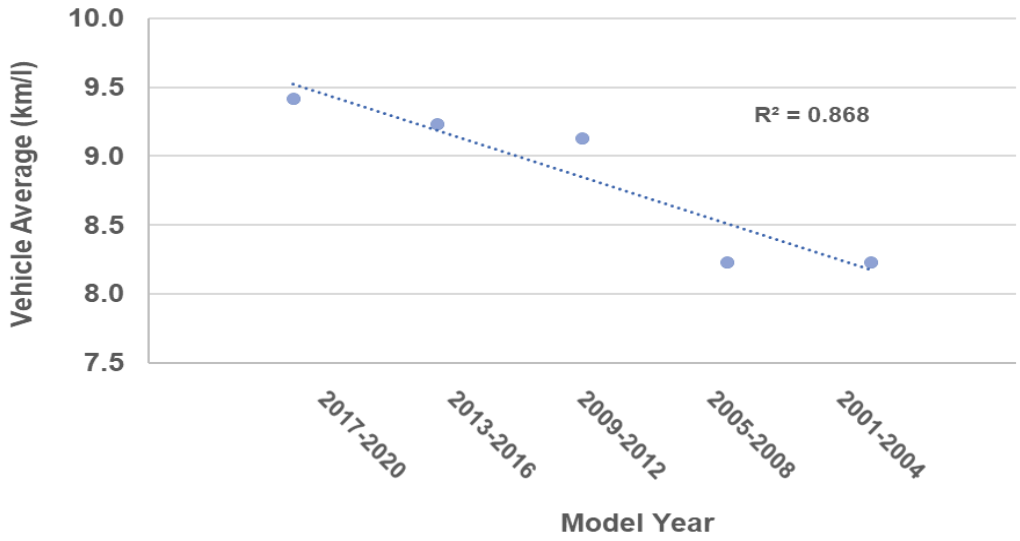



**Figure 3.8**: Linear regressions between average distance travelled and vehicle age.

### *C. Vehicle Kilometre Travelled (VKT)*

The mean daily VKT for PC, HPV, LDV, and HDV ranged from $7.21 \times 10^5$ km, $2.93 \times 10^5$ km, $2.54 \times 10^6$ km, and $2.30 \times 10^5$ km in 2012 to $1.11 \times 10^6$ km, $3.30 \times 10^5$ km, $3.64 \times 10^6$ km, and $2.57 \times 10^5$ km in 2021. Vehicle population follows this order; PC> LDV > HPV > HDV, while the calculated trend for VKT ranged from LDV > PC >HPV > HDV. In this result,



© University of Venda

PC has a higher number of vehicles, however, LDV has the highest calculated VKT among the vehicle types (Table 3.4).

The LDV results align with the previous study conducted by Singh *et al*. (2017). Furthermore, the increased presence of PC coincides with the findings of Cai and Xie (2007) and Progiou and Ziomas (2011). According to Cai and Xie (2007) and Lu *et al*. (2017), the high number of PCs can be ascribed to an increase in GDP as personal cars become more affordable in the study area (Vhembe District Municipality, 2019); the high estimated VKT in LDVs could be linked to an increase in agricultural activities and the establishment of small and medium enterprises (SMEs) within the area. The distance travelled for PC and LDV shows an upward trend throughout the years. A possible reason for this could be the increase in the number of vehicle owners, the demand for transport, the affordability of vehicles, and economic growth. Contrary to this, HPV and HDV show that the distance travelled increases with the years, followed by a drastic decline and then a steady growth. The fluctuation in VKT for these vehicle types could be attributed to factors such as the weight of the load carried by these cars, road types, the rules assigned to the road, materials used in road construction, and a decline in vehicle population, potentially influenced by post-2016 election municipality boundary changes.

**Table 3.4**: Mean VKT by different vehicle types in the study area

| | Average daily VKT ($10^5$ km) | | | |
|---|---|---|---|---|
| **Year** | **PC** | **HPV** | **LDV** | **HDV** |
| 2012 | 7.21 | 2.93 | 25.43 | 2.30 |
| 2013 | 7.96 | 3.07 | 27.78 | 2.46 |
| 2014 | 8.62 | 3.29 | 29.46 | 2.67 |
| 2015 | 9.55 | 3.68 | 32.12 | 2.88 |
| 2016 | 9.76 | 3.51 | 32.99 | 2.79 |
| 2017 | 8.57 | 2.84 | 29.56 | 2.15 |
| 2018 | 9.95 | 3.15 | 32.92 | 2.52 |
| 2019 | 10.31 | 3.27 | 34.26 | 2.56 |
| 2020 | 10.69 | 3.28 | 35.32 | 2.57 |
| 2021 | 11.07 | 3.30 | 36.38 | 2.57 |



© University of Venda

### 3.3.1.2 Vehicle emissions inventories at Thulamela Municipality from 2012 to 2021

***A. Annual emission trend by all vehicles for different pollutants from 2012 to 2021***

The annual total emission trend of $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ pollutants for all the vehicle types was calculated from 2012 to 2021. The estimation was based on vehicle population, VKT, and the emission factor of each vehicular pollutant. This is the first vehicular emission inventory of $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ pollutants developed for this Municipality, covering a time frame of 10 years. The total emissions of the criteria pollutants for the 10-year study period from all vehicle types were found to be 547.16, 22326.02, and 32781.12 for $SO_2$, $NO_x$, and CO (Figure 3.9a). Additionally, the emissions were 1291.76 tons for $PM_{2.5}$ and 1367.82 tons for $PM_{10}$ pollutants, respectively (Figure 3.9b). Notably, CO was the chief pollutant constituent in all the vehicle types, with LDV as the highest producer. The vehicular emission for each pollutant generally increased with the study period but declined in 2017 due to a reduction in vehicle population, as observed by Enitan *et al*. (2021). This reduction followed the 2016 municipal boundary changes that transferred the Malamulele area from Thulamela Local Municipality to Collins Chabane Local Municipality. The boundary adjustment led to a decrease in population and changes in governance structures within Thulamela, influenced by administrative restructuring and economic shifts (Thiba, 2019; Nemaxwi, 2019; Netswera, 2023). The boundary changes reduced traffic volumes and vehicle emissions in the affected areas, contributing to improved air quality and lower exposure risks for residents.

The trend in CO emission rate was higher when compared with other pollutants for all vehicle types, and this was similar to previous studies (Cai and Xie, 2007; Singh *et al*., 2017). It was observed that the trend of the pollutants for all vehicle types increased from 2012 to 2021 for the estimated pollutants (Figure 3.9). In 2021, CO has the highest share of 56%, followed by $NO_x$ (38%), $PM_{10}$ (3%), $PM_{2.5}$ (2%), and $SO_2$ (1%) with a growth rate of approximately 39.66%, 40.73%, 40.57%, 39.87%, and 40.84% for $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$, respectively. In general, LDVs, PCs, and HDVs vehicles were the major vehicle types contributing to the total emissions within the study period, with little emission from HPVs**.** LDVs and PCs were found to be major polluters of $SO_2$ and CO due to their large population**.** LDVs and HDVs contributed the most towards $NO_X$ gaseous pollutants as well as $PM_{2.5}$ and $PM_{10}$ particulate pollutants due to their high emission factor. This is similar to a previous



© University of Venda

study conducted, where HDV also had higher $NO_x$ emissions than PC (Cai and Xie, 2007, 2013). The contributions of emissions into the atmosphere fluctuate due to changes in the number of vehicles in the area and emission factors.

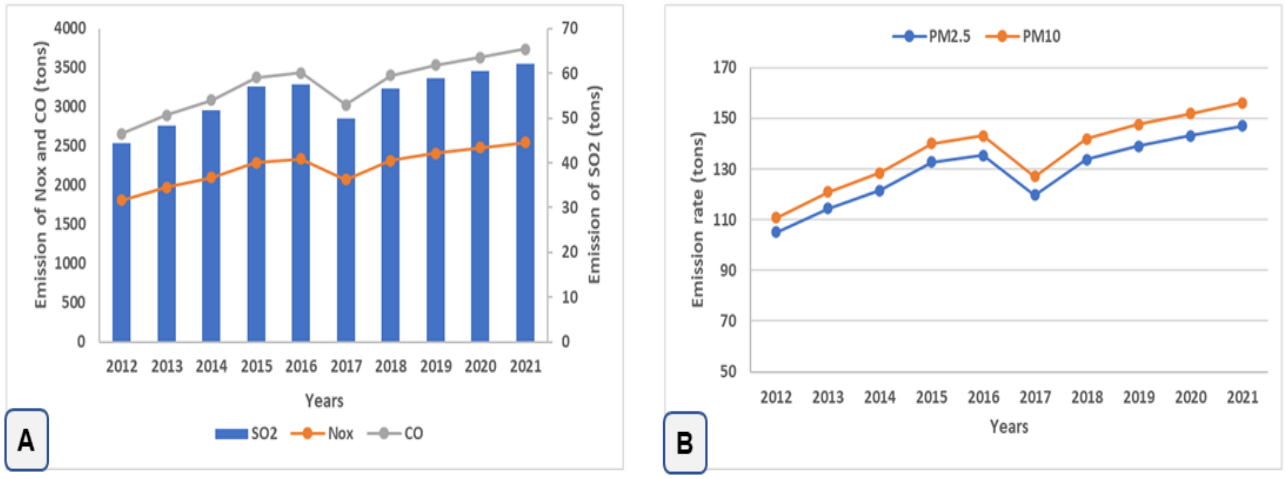



**Figure 3.9:** The total a) gaseous emissions and b) particulate emissions estimated for all vehicles during the study period (2012–2021).

### *B. Trends analysis of total emission by each vehicle category*

The contribution to the total emission of $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ from each vehicle type for each year was calculated for the study period as shown in Figure 3.10, and the emissions varied due to (i) different vehicle types (ii) vehicle population, and (iii) kilometres travelled. It is evident that LDVs are the most significant contributors to the total emissions of $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ as compared to the other vehicle types. Their contribution varied from 3830.13 tons in 2012 to 5479.74 tons in 2021, followed by PC. The PC had an emission rate of 390.93 tons in 2012 from the five pollutants, which increased in 2021 to 600.02 tons. For HPV and HDV, the $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ emissions varied from 183.28 and 323.20 tons to 206.24 and 361.65 tons per year, respectively. The highest contribution from HPV and HDV was observed in 2015, but declined in 2017 and later rose in 2018. The fluctuation in emission rate might be due to trends of the VKT discussed earlier and the impact of post-2016 election municipality boundary and location changes.



© University of Venda

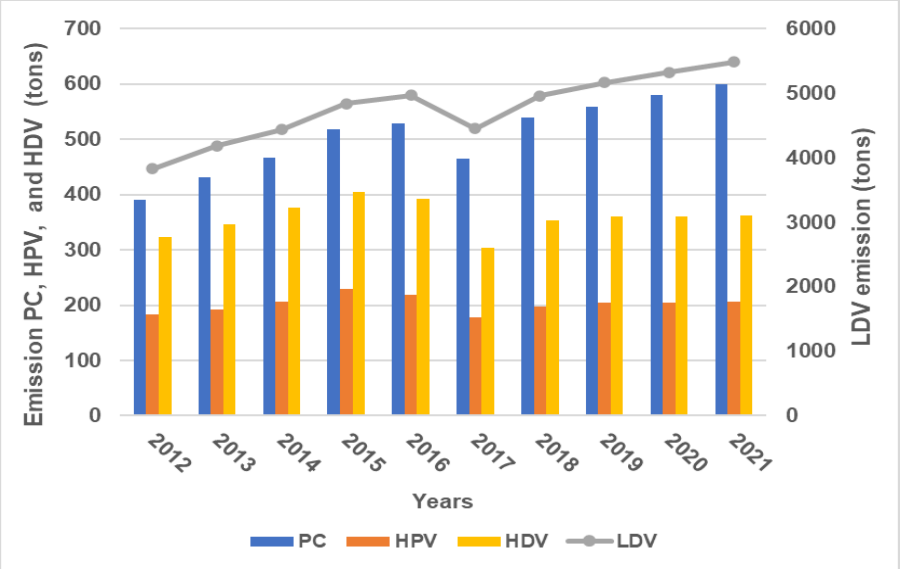



**Figure 3.10:** Total emission trend by each vehicle type for the study period.

### *C. Contribution of each vehicle category to the emission of $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ vehicles from 2012 to 2021*

Figures 3.11 and 3.12 illustrate the vehicle contribution, in proportion to emissions of each pollutant, by each vehicle type. Vehicle contribution to the emissions of each pollutant varied due to different vehicle populations, VKT, and different levels of emission factors of the pollutants. The emission of $SO_2$ is one major fuel-related pollutant that can be enormous with a significant increase in vehicle population, especially heavy-duty diesel vans. According to Figure 3.11, LDVs and PCs are the main contributors to $SO_2$, responsible for about 17.63 and 16.32 in 2012 to 25.23 and 25.05 tons of emissions respectively, in 2021. This is mainly due to the continuous growth of their population, which is contrary to the findings of Lv *et al*. (2019), since HPV and HDV accounted for about 8.99 to 10.11 and 1.51 to 1.69 tons, respectively, over the 10 years.

The principal contributors to NOx emissions during the study period were LDVs and HDVs that ranged from 1621.33 to 2319.63 and 122.65 to 137.24 tons, respectively and not $SO_2$ (Figure 3.11). The HDV values align with the findings of Lang *et al*. (2014) and Lv *et al*. (2019), suggesting consistency and agreement between these studies and the current one. The average emissions of $NO_X$ for the study period are 2016.42, 135.86, 50.96, and 29.37 tons for LDV, HDV, PC, and HPV, respectively. Studies show that prolonged exposure to



© University of Venda

nitrogen oxides ($NO_x$) in the atmosphere can reduce lung function and increase susceptibility to respiratory infections, including common colds to severe conditions such as pneumonia. Additionally, individuals with pre-existing respiratory conditions like asthma and COPD are particularly vulnerable to worsened symptoms. These harmful effects extend to animals and contribute to declining air quality in ecosystems infections (Darcın, 2017; Almetwally *et al*., 2020; Hajirasouliha & Zabiegaj, 2020; Ko & Kyung, 2022).

According to the findings, LDVs have been identified as the largest contributors to CO vehicle emissions, ranging from 1997.20 to 2857.38 tons, with an average emission of 2483.87 tons for the study period. Following LDVs, PCs were found to be the second contributor to CO emissions. These findings support previous research (Lang *et al*., 2014; Lv *et al*., 2019). In addition, as the economy developed more people started using PCs, and its contribution to the total vehicular CO emissions in the study area signifies an upward trend. Furthermore, PC, HDV, and HPV recorded average emissions of 432.65, 204.02, and 157.57 tons, respectively, during the study period. In general, all the vehicle types emit an average of 3278.11 tons of CO into the atmosphere during the study period.

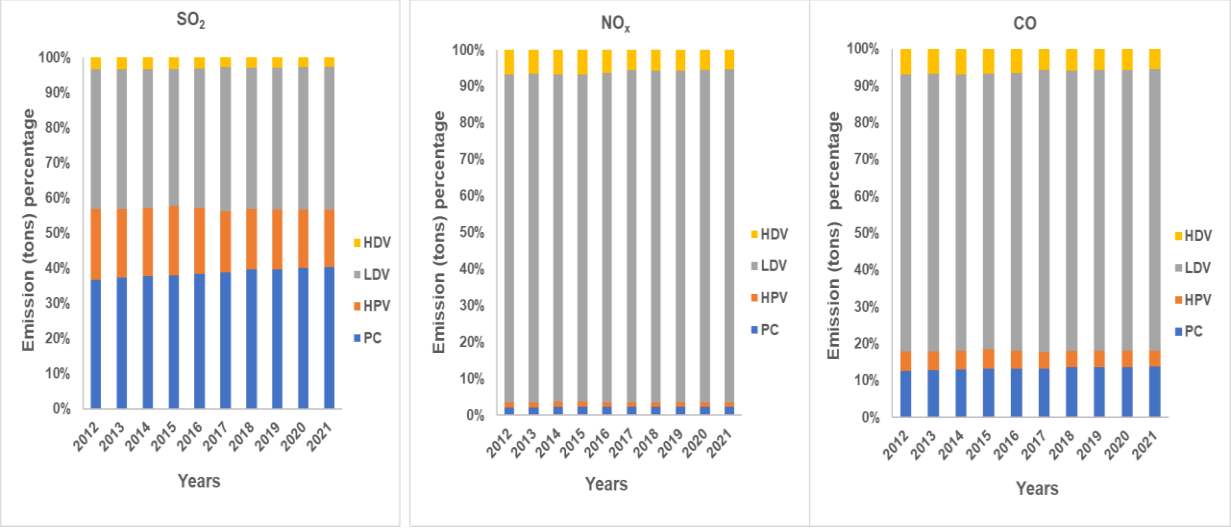



**Figure 3.11:** Contribution of vehicle type to emissions of $SO_2$, $NO_x$, and CO from 2012 to 2021.

During the research period in Thulamela municipality, LDVs and HDVs were identified as the predominant sources of $PM_{2.5}$ and $PM_{10}$ emissions. Figure 3.12 illustrates the substantial contribution of LDVs (92.81 to 132.78 tons) and HDVs (7.14 to 7.98 tons) to $PM_{2.5}$ vehicle



© University of Venda

emissions. Similarly, $PM_{10}$ emissions indicated LDVs as the main source, with levels ranging from 101.16 to 144.73 tons, followed by HDVs at 7.72 to 8.64 tons. These findings are consistent with prior research, such as Sahu *et al*. (2011), who found comparable patterns in $PM_{2.5}$ and $PM_{10}$ emissions from vehicles. This highlighted the impact of vehicle movement on the availability of particulate matter, particularly in the urban environment.

The research identified a number of influencing variables impacting LDV and HDV emissions. These variables include emission factors and the vehicle population in the study area, as they demonstrate temporal change and affect emission levels. This observation concurs with Lv *et al*. (2019), research results, where LDVs were identified as the primary source of emissions while HDV emissions were reduced. In addition to these findings, Liu *et al*. (2022) conducted a study that indicated that both LDVs and HDVs contributed to $PM_{10}$ emissions, with HDVs accounting for a higher percentage (34.33%) than LDVs (27.77%). These results highlight how both vehicle types contribute to $PM_{10}$ pollution. Overall, these findings add to our understanding of the sources and factors influencing $PM_{2.5}$ and $PM_{10}$ emissions, confirming the need for specific measures to reduce vehicle-related pollution in Thulamela and similar municipalities.

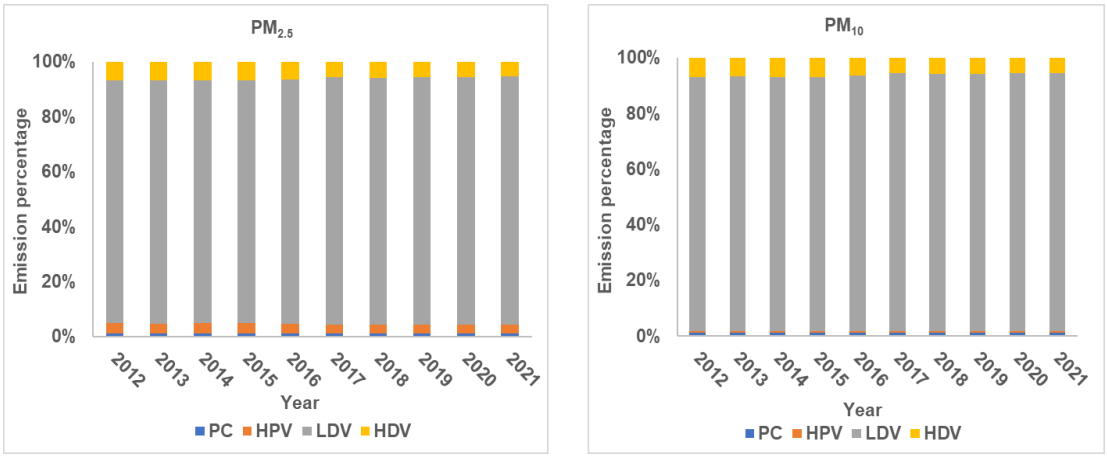



**Figure 3.12:** Contribution of each vehicle type to $PM_{2.5}$ and $PM_{10}$ emissions from 2012 to 2021.



© University of Venda

***D. The Correlation between $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ vehicle emissions and GDP from 2012 to 2021***

The Compound Annual Growth Rate (CAGR) of 6.69% during the period of 2012–2021 in Thulamela Municipality had driven the increase of vehicle population, which simultaneously increased the amount of emissions. The ratios of $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ emissions from vehicles to GDP decreased steadily during the study period, as shown in Figure 3.13. This trend indicates that the pollutant emissions from the vehicles have increased, and this is an indication of developments in the study area due to urbanisation and industrialisation. The Compound Annual Growth Rate (CAGR) of 3.78%, 3.87%, 3.86%, 3.8%, and 3.88% was revealed for $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ emissions, respectively, during the period of study. Likewise, Table 3.5 shows that there was a positive correlation between the $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ of vehicle emissions and the GDP with passing years, as most of the related coefficients were greater than 0.85. This suggests that an increase in GDP and the number of vehicles, in turn, results in increased vehicle emissions. Economic prosperity enriched the people; however, it resulted in the emission of a large quantity of pollutants, which pose health risks.

**Table 3.5:** Pearson correlation matrix between GDP and the pollutants within the study area

| *Variables* | *GDP (R Billions)* | $SO_2$ | $NO_x$ | *CO* | $PM_{2.5}$ | $PM_{10}$ |
|---|---|---|---|---|---|---|
| **GDP (R Billions)** | 1 | | | | | |
| **$SO_2$** | 0.85 | 1 | | | | |
| **$NO_x$** | 0.89 | 0.99 | 1 | | | |
| **CO** | 0.88 | 0.99 | 1.00 | 1 | | |
| **$PM_{2.5}$** | 0.88 | 0.99 | 1.00 | 1.00 | 1 | |
| **$PM_{10}$** | 0.89 | 0.99 | 1.00 | 1.00 | 1.00 | 1 |

### 3.3.1.3 2021 Vehicle emissions of the study domain in Thulamela LM

Figure 3.14 shows the road network within the study domain, and the 2021 vehicle emissions of each pollutant contribution were estimated for Thulamela Municipality. The total number of vehicles per vehicle category is presented in Table 3.6, and this study allocated 64% of the 2021 vehicle population for the study domain, considering the study area population densities. The total vehicle emissions of CO, $NO_x$, $PM_{10}$, $PM_{2.5}$, and $SO_2$ for 2021 from different vehicle types were found to be approximately 2390.7, 1630.1, 99.9, 94.1,



© University of Venda

and 39.7 tons, in that order as shown in Table 3.7. In addition, the Table reveals the contribution of each vehicle category, and shows that the vehicles are mainly emitting CO, $NO_x$, $PM_{10}$, and $PM_{2.5}$. According to the study's findings, $PM_{10}$ and $NO_x$ are mainly associated with LDVs, HDVs, and PCs. LDVs accounted for a high percentage of both $PM_{10}$ (92.71%) and $NO_x$ (91.07%) emissions, while HDVs contributed a notable proportion of $PM_{10}$ (5.54%) and $NO_x$ (5.39%) emissions, and the emissions from PCs were 1.29% and 2.36%, respectively. These findings are consistent with a prior study by Progiou and Ziomas (2011), who discovered that HDVs contribute significantly to $PM_{10}$ and $NO_x$ emissions. In contrast, LDVs and PCs are the main sources of CO emissions, while LDVs and HDVs are the main sources of $PM_{2.5}$ emissions. In 2021, LDV contributed a maximum share to the five estimated criteria pollutants in the area, followed by PC, HDV and HPV, respectively, despite having a lower population compared to PC; this contribution is directly related to the estimated VKT in the area**.** This study's baseline vehicle emission inventory was also compared with other South African inventories to assess its suitability for further study, revealing general alignment with prior studies but with varying magnitudes (Thambiran and Diab, 2011; Scorgie, 2012; DEA, 2018)**.**



© University of Venda

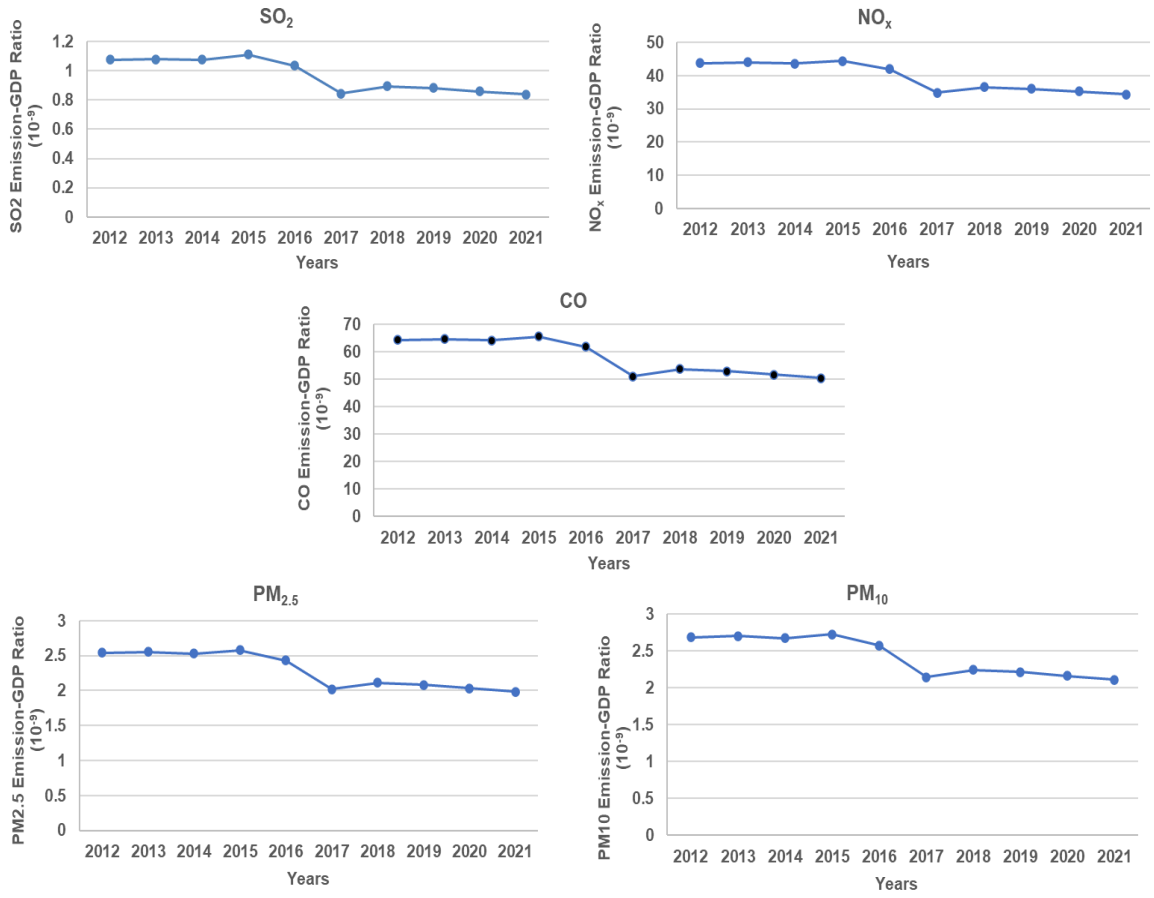



**Figure 3.13:** Trend ratio of $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ emission to GDP during the study period (2012-2021).

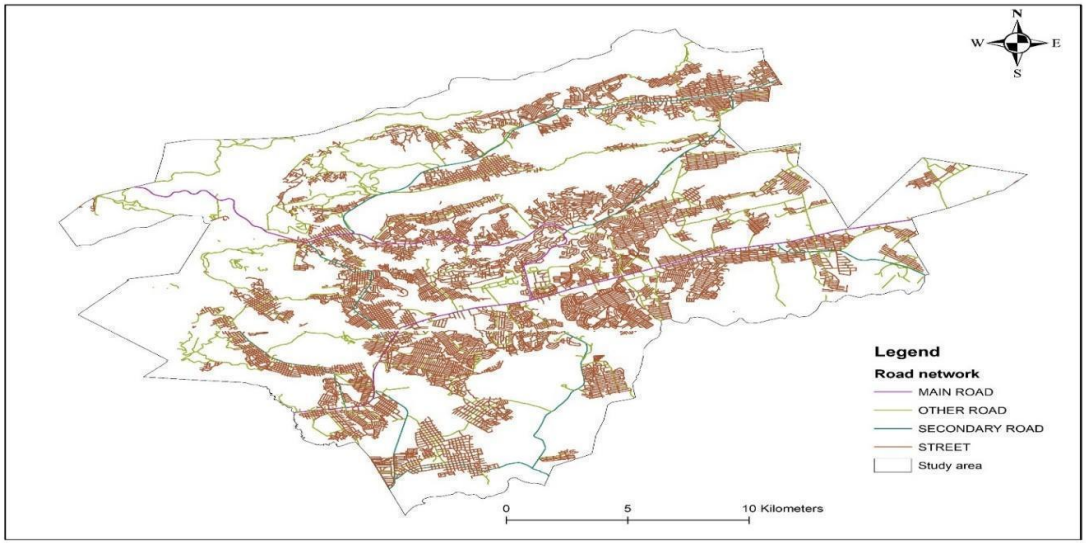



**Figure 3.14:** Road network within the study domain.



© University of Venda

**Table 3.6:** VKT for the year 2021 in the study area.

| Types of Vehicles | Total km/day | Number of samples | AKT (km) | RV 2021 | DVKT 2021 | 64% of RV for the study domain 64% of RV | 64% of RV for the study domain DVKT |
|---|---|---|---|---|---|---|---|
| PC | 1055.38 | 20 | 52.769 | 20978 | 1106988.1 | 13426 | 708472.4 |
| HPV | 5105 | 31 | 164.6774 | 2003 | 329848.9 | 1282 | 211103.3 |
| LDV | 4550.6 | 20 | 227.53 | 15988 | 3637749.6 | 10232 | 2328160 |
| HDV | 855 | 4 | 213.75 | 1204 | 257355 | 771 | 164707.2 |
| Total | 11565.98 | 75 | 658.7264 | 40173 | 5331941.6 | 25711 | 3412443 |

*Registered vehicles (RV), Daily vehicle kilometres travelled (DVKT), average kilometres travelled (AKT)

**Table 3.7:** Estimated emissions (tons/year) of each vehicle type for 2021 in the study domain

| Vehicle Type | $SO_2$ | $NO_x$ | CO | $PM_{2.5}$ | $PM_{10}$ |
|---|---|---|---|---|---|
| PC | 16.0 | 38.5 | 327.1 | 1.0 | 1.3 |
| HPV | 6.4 | 19.2 | 102.9 | 2.9 | 0.5 |
| LDV | 16.2 | 1484.6 | 1828.7 | 85.0 | 92.6 |
| HDV | 1.1 | 87.8 | 131.9 | 5.1 | 5.5 |
| Total | 39.7 | 1630.1 | 2390.6 | 94.0 | 99.9 |

### D. Spatial distribution pattern for $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ emission

The 2021 total annual pollutants emission for the area (which is the first vehicular emission estimation so far) was distributed or allotted to all grids within the domain, considering both domain population densities and vehicle population. Spatially, the highest vehicle emissions per grid (with 1-km grid spacing) are 0.0114, 0.5887, 0.7895, 0.0337, and 0.0365 ton/year for $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$, respectively, for the study domain. Based on this study, areas with high emissions levels were identified (Figure 3.15), where the hotspots with high emissions are primarily situated in areas with heavy human activity, such as major cities, roads, and commercial areas. These areas have a dense population and more vehicles, leading to elevated emission levels compared to nearby regions. This finding aligns with a previous study, which also attributed the increased emission loads to the rapid growth of the population and subsequent increase in vehicles within the area (Sindhwani and Goyal, 2014). After



© University of Venda

examining the vehicle sector emissions in this section, we will explore residential emissions, which represent the second source of emissions in the next section.

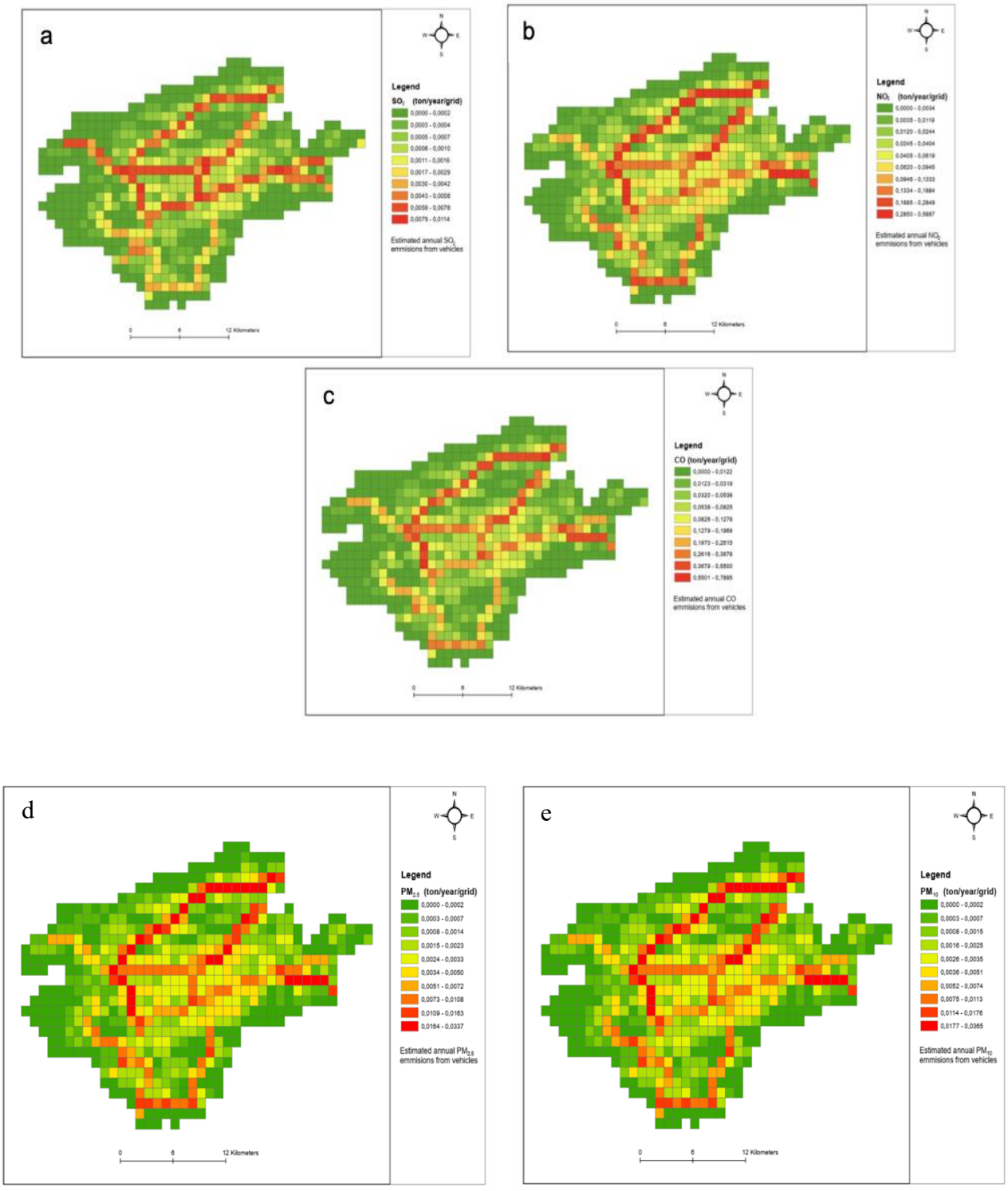



**Figure 3.15:** 2021 estimated annual emissions of a) $SO_2$, b) $NO_x$, c) CO, d) $PM_{2.5}$, and e) $PM_{10}$ from vehicles (all classes) for the domain in the 1 km per grid model.



© University of Venda

### 3.3.2 Residential Sector

#### 3.3.2.1 Socio-economic and demographic factors influencing household energy use

The socio-economic attributes of the respondents in the study area are shown in Table 3.8. The result reveals that 68.75% and 75% of the household heads were Female while 31.25% and 25% were male in Gondeni and Lunungwi villages, respectively. It could, therefore, be inferred that there were more women than men as household heads. This is probably due to their roles and responsibilities in the family, especially in the provision of domestic energy for indoor cooking, heating, and lighting. Similarly, our findings are consistent with those of other researchers, which indicate that women are usually entrusted with household duties and are responsible for making fuel consumption decisions (Nkosi *et al*., 2017; Francioli and Pharoah, 2018). It was also observed that many of the respondents (30%) were between 31 and 40 years of age, followed by 21 and 30 years of age (22.5%) from Gondeni, while, most of the respondents (27.5%) from Lunungwi were above 50 years, followed by 21 and 30 years of age at 23.75%.

For family size, households with 5 or more members were found to have the highest percentage in Gondeni (37.5%), followed by households with 3 members (27.5%), while in Lunungwi, households with 4 members accounted for 43.8%, followed by households with 5 or more members (38.8%), as shown in Table 3.8. This means that more than half of the household heads had 4 or more members who depend on them, which could probably increase the demand for domestic energy. Most households indicated a preference for using fuelwood for cooking due to its cost-effectiveness, energy-saving benefits, and ability to prepare sufficient food for the entire family. Danlami (2019) and Masekela and Semenya, (2021) reported similar findings where household size was an important factor affecting solid fuel consumption in households.

Findings in Table 3.8 show that most of the respondents (41.25%) had higher education, and 22.5% had no formal education in Gondeni whereas most of the respondents (33.75%) had tertiary education and 30% had no formal education in Lunungwi village. It could be presumed from the results of this study that more than two-thirds of respondents



© University of Venda

have higher education or lower, which may influence the type of domestic energy they use for cooking and heating. The results show that most of the respondents in Gondeni are unemployed (36%), with 25% employed full-time compared to other employment statuses. Similarly, a significant portion of respondents in Lunungwi are unemployed (46%), while 30% are self-employed. In Gondeni, it appears that most families are occupied by unemployed adults of working age, many of whom are still in school, living with their parents, and depending on them for a living, however, more unemployed people with no formal education, as well as those who were older than 50 lived in Lunungwi than in Gondeni.

The study indicates that the educational background of the household head plays a vital role in selecting energy sources apart from wood. This result is consistent with previous research which also revealed the influence of education on fuel use (Uhunamure *et al*., 2017; Netshipise and Semenya, 2022). Notably, educated and employed individuals still prefer wood to save money for other priorities like education and health. In addition to the educational background, the decision on fuel choice for domestic use is substantially influenced by employment status and higher levels of unemployment, further highlighting the complex relationship between education, economics, and energy preferences.



© University of Venda

**Table 3.8:** Chi-square Cross-tabulation of Respondents' Socio-demographic Attributes by Village

| | Gondeni | | Lunungwi | | Pearson | Cramers V |
|---|---|---|---|---|---|---|
| Per village | Frequency | Percentage | Frequency | Percentage | | |
| | | | **Gender** | | | |
| Female | 55 | 68.75 | 60 | 75 | 0.379 | 0.07 |
| Male | 25 | 31.25 | 20 | 25 | | |
| Total | 80 | 100 | 80 | 100 | | |
| | | | **Age (Years)** | | | |
| 11-20 | 12 | 15 | 7 | 8.75 | 0.197 | 0.194 |
| 21-30 | 18 | 22.5 | 19 | 23.75 | | |
| 31-40 | 24 | 30 | 15 | 18.75 | | |
| 41-50 | 11 | 13.75 | 17 | 21.25 | | |
| Above 50 | 15 | 18.75 | 22 | 27.5 | | |
| | 80 | 100 | 80 | 100 | | |
| | | | **Education** | | | |
| No formal education | 18 | 22.5 | 24 | 30 | 0.222 | 0.166 |
| Primary education | 5 | 6.25 | 8 | 10 | | |
| High education | 33 | 41.25 | 21 | 26.25 | | |
| Tertiary education | 24 | 30 | 27 | 33.75 | | |
| | 80 | 100 | 80 | 100 | | |
| | | | **Employment** | | | |
| Unemployed | 29 | 36.3 | 37 | 46.3 | 0.131 | 0.188 |
| Self-employed | 18 | 22.5 | 24 | 30.0 | | |
| Part-time employed | 13 | 16.3 | 8 | 10.0 | | |
| Full time employed | 20 | 25.0 | 11 | 13.8 | | |
| | 80 | 100.0 | 80 | 100.0 | | |
| | | | **Family size** | | | |
| 1-2 people | 11 | 13.8 | 2 | 2.5 | .001 | .310 |
| 3 people | 22 | 27.5 | 12 | 15 | | |
| 4 people | 17 | 21.3 | 35 | 43.8 | | |
| 5 or more | 30 | 37.5 | 31 | 38.8 | | |
| | 80 | 100.0 | 80 | 100.0 | | |

### 3.3.2.2. Energy used and the consumption pattern in the study area

Table 3.9 summarises the different energy sources that Gondeni and Lunungwi respondents use to meet their households' needs. This result shows that most people used more than one source of fuel as their energy of choice in the questionnaires distributed to 160 people, therefore, the total number of respondents for each energy use is greater than the total number of respondents. Wood and electricity were the predominant fuel sources for household activities in the study area (the two villages).



© University of Venda

Wood was used by 37% of households in the summer and 43% in the winter. Meanwhile, 28% of households relied on electricity in the summer and 29% in the winter. Additionally, about 32% of respondents mainly used electricity and others during both the summer and winter periods. Wood and electricity are not the only energy sources utilised, based on the questionnaires, as paraffin and candles are also used. Wood was commonly used by all interviewed households compared to other energy sources throughout the study period, with a peak in wood usage in the winter season.

**Table 3.9:** Respondent frequency of different energy types mainly used for household activities with seasons as highlighted in the field study questionnaires.

| Energy type | Summer | | Winter | |
|---|---|---|---|---|
| | Frequency | Percentage (%) | Frequency | Percentage (%) |
| Candle | 1 | 0.4 | 1 | 0.3 |
| Electricity | 74 | 28.2 | 97 | 29.5 |
| Electricity & others | 84 | 32.1 | 84 | 25.5 |
| Paraffin | 5 | 1.9 | 5 | 1.5 |
| Wood | 98 | 37.4 | 142 | 43.2 |
| Total | 262 | 100.0 | 329 | 100.0 |

Generally, wood emerged as the most widely used energy source for heating and cooking throughout the year, with a noticeable peak during winter season, as illustrated in Figure 3.16, which is attributed to wood's natural availability and lower cost. In the winter months, residents usually choose multiple fuels among the options presented in the questionnaire, while wood remains the preferred choice for cooking both in summer and winter. These findings validate an earlier research discovery made by the Department of Energy (DoE 2013) and Semenya and Machete (2019). According to these surveys, most South African communities, including those with access to electricity, still prefer solid fuels, such as fuelwood for cooking and heating purposes. This pattern indicates the simultaneous use of various energy sources, either individually or in combination.

Likewise, the utilisation of multiple energy sources (referred to as energy stacking) is prevalent in this area, consistent with findings from previous research (Lloyd, 2014; Naidoo *et al*., 2014a; Makonese *et al*., 2016). These studies highlight that low-income



© University of Venda

households commonly rely on a combination of fuels to meet their basic energy needs. Additionally, the questionnaire survey revealed that most households within the study area predominantly use three-legged stoves (as depicted in Figure 3.17) and fireplaces for cooking. As demonstrated in Figure 3.18, about 32% of the respondents continue to use wood for cooking in winter and approximately twenty-two percent (22%) during summer. For the comparison of wood usage among the 4 seasons, a Kruskal-Wallis Test (Nonparametric ANOVA) carried out, showed an extremely significant difference in the usage of wood during the seasons (P = 0.0007).

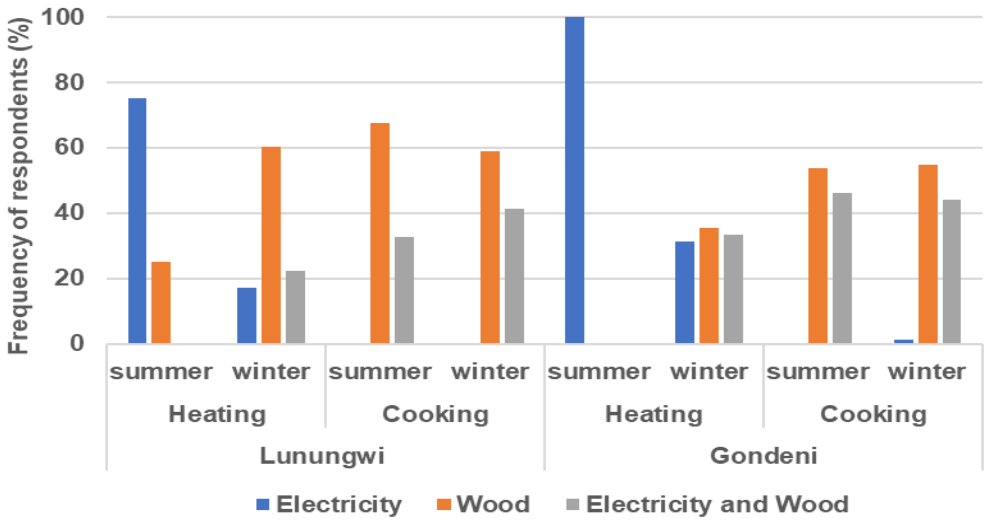



**Figure 3.16**: Percentages of the energy sources consumed by households for cooking and heating.



**Figure 3.17:** Three-legged stove used by households for cooking and other domestic activities.



© University of Venda

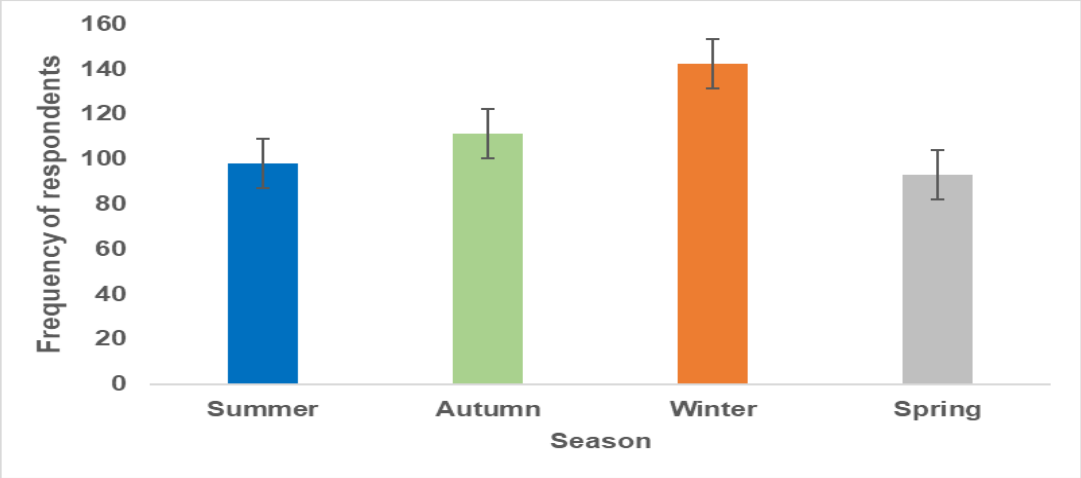



**Figure 3.18:** Wood usage in different seasons recorded in this study.

The findings from the survey (Figure 3.19) showed that most of the households use fuelwood for longer hours in a day, during winter compared to summer. Notably, both Lunungwi and Gondeni villages rely more on wood for fuel consumption, with Lunungwi residents using wood for longer hours in the summer and winter seasons than the residents of Gondeni. Lunungwi's higher wood consumption might be attributed to factors like larger family sizes, higher unemployment rates, and lack of formal education. Furthermore, both Lunungwi and Gondeni villages exhibit higher fuel consumption during winter than in summer, indicating a consistent seasonal trend. Based on these findings, it is reasonable to suggest that most communities in the study area use more wood during winter, highlighting a possible seasonal dependency on wood as the primary fuel source. The influence of seasons on household fuel consumption becomes evident through the variations observed between winter and summer (Makonese *et al*., 2016).



© University of Venda

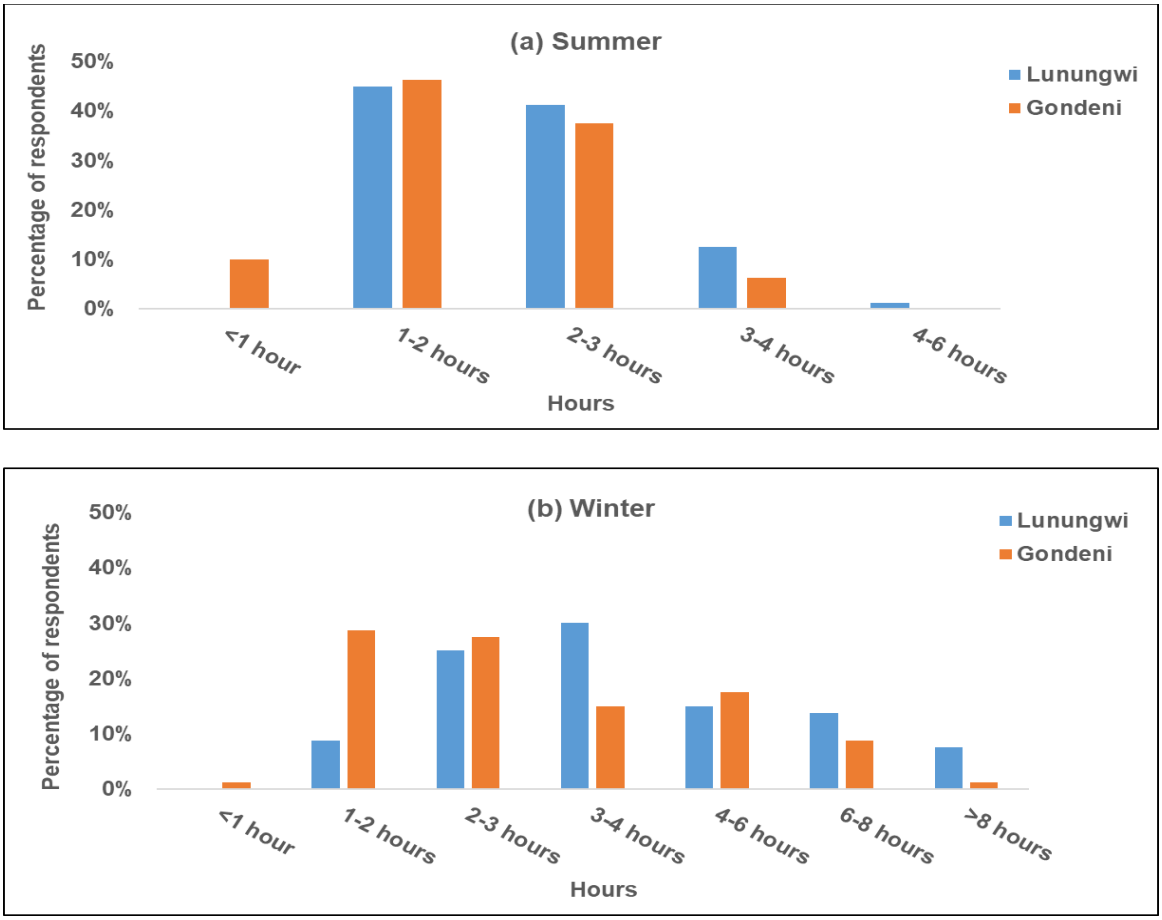



**Figure 3.19:** Percentage of responses on average daily hours of wood usage in the villages for summer and winter.

### 3.3.2.3 Preferred tree species for fuelwood and their sources in the study area

Household fuelwood preferences depend on local woods' availability, accessibility, and affordability (Kituyi *et al*., 2001; Lavhelani, 2021). Affordability can be measured in terms of the time and money spent collecting fuelwood. This economic factor significantly influences local fuel-use patterns, as most households prefer fuelwood over other fuels and prioritise cost-effective options for cooking and heating; furthermore, transportation costs also impact the fuelwood supply (Karimu *et al*., 2016; Semenya & Machete, 2019; Lavhelani, 2021). According to Boahene (1998), about 90% of the energy needs of tropical Africa are met by fuel wood, which is still significantly less expensive than most alternative fuels, hence, even if the price of fuel wood was to rise, the demand would not significantly decline because there are no suitable substitutes.



© University of Venda

Thulamela Municipality in the Vhembe District has diverse plant species, trees, grasses, and shrubs for fuelwood in the study area. Table 3.10 shows the scientific names, common names, and local names of the nine preferred tree species by households. The most used trees for fuel wood in the two villages are the mango tree, followed by *muluwa, mupane, mutulume,* and avocado. Other trees used in the villages but not so popular are *munzere, mugwiti, murenzhe,* and *muvhula* (Figure 3.20a). In general, the primary source of most fuelwood in both villages is through harvesting from the field, while some are purchased from local vendors, except for *mopane* wood, which is predominantly bought (Figure 3.20b and 3.21). Furthermore, the study's findings indicated that, except for *mopane* (100%) and *muluwa* (64%), which are mainly targeted for their energy content, as they produce flames that last longer when it burns, all the other trees are targeted due to their availability (Figure 3.20c). Tree species that are not primarily food sources but that burn well, produce gradual flames and smoke less are preferred for fuelwood (Mucina and Rutherford, 2006). In addition, the respondents indicated that except for *mutulume*, which is mainly sourced by the riverside, and *muluwa,* which is sourced by the mountains, all the other trees are primarily harvested from plains, except *mopane*, for which all respondents stated that they buy (Figure 3.20d).

**Table 3.10:** Preferred Tree species for fuelwood

| S/No | Tree species for fuelwood | | |
|---|---|---|---|
| | Scientific name | Common name | Local name |
| 1 | *Persea americana* | Avocado | Muafukhada |
| 2 | *Colophospermum mopane* | Mopane | Mugwiti |
| 3 | *Breonadia salicina* | Matumi | Muluwa |
| 4 | *Senegalia ataxacantha* | Flame thorn | Mupani |
| 5 | *Bridelia micrantha* | Mitzeerie | Murenzhe |
| 6 | *Combretum molle* | Velvet bushwillow | Mutulume |
| 7 | *Mangifera indica L* | Mango | Munngo |
| 8 | *Dichrostachys cinerea (L.)* | Sickle bush | Munzere |
| 9 | *Parinari curatellifolia* | Mobola-plum | Muvhula |



© University of Venda

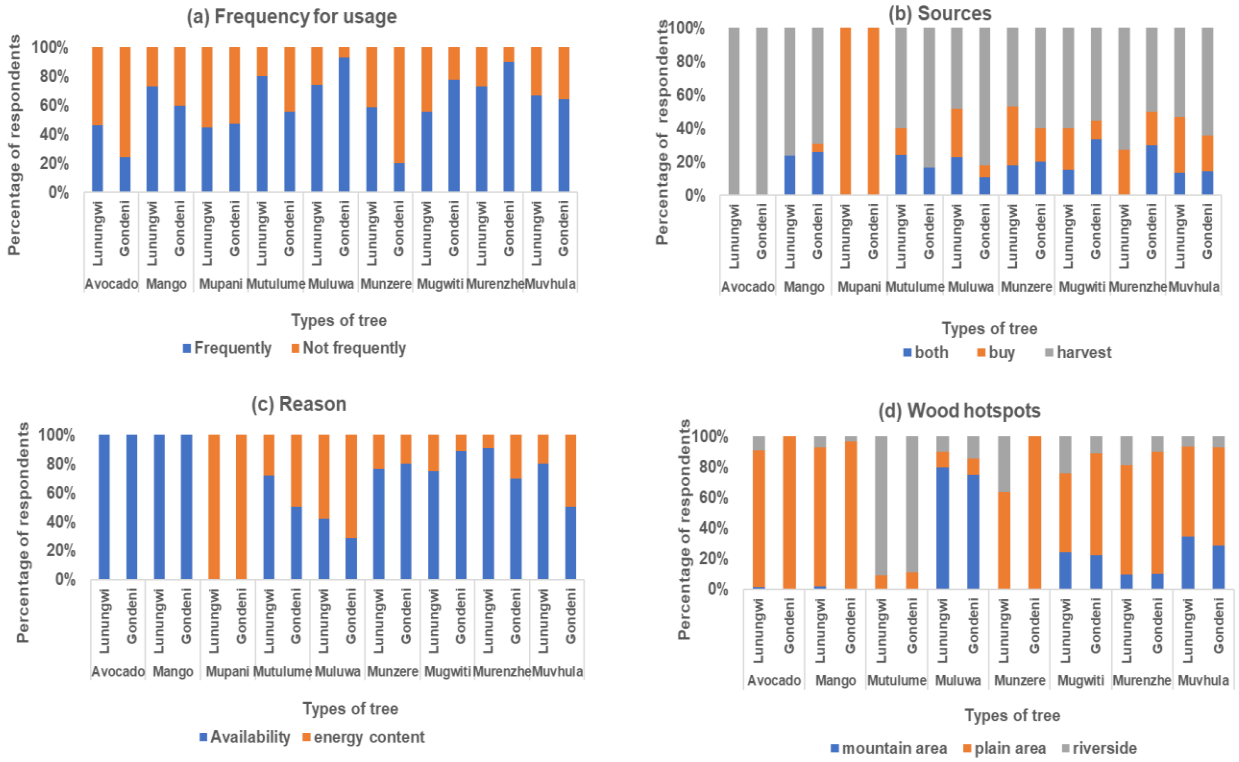



**Figure 3.20:** Percentage of respondents' frequency for: a) uses of different wood types in the villages, b) sources, c) reasons for using specific wood types in the villages, and d) hotspots where wood is sourced in the village.




**Figure 3. 21:** Various sources of wood through (a and b) harvest in the field, c) buying of different types, and d) buying of Mopani wood (Source: Fieldwork).



© University of Venda

### 3.3.2.4 Empirical analysis using Chi-square analysis

A Chi-square test was done to determine the relationship between the villages and respondents' perspectives on their choice of cooking energy sources, considering seasonal variations, reasons for choice, level of satisfaction, awareness, and condition of use. Table 3.11 shows that in Gondeni, 43 (54%) respondents depend on only wood for cooking, and 37 (46%) use both wood and electricity to form their energy mix during summertime. Lunungwi has only 54 (68%) of its respondents using only wood for cooking purposes, and 26 (32%) mix their energy. In winter, the highest number of respondents using only wood was observed, with 42 (55%) in Gondeni and 47 (59%) in Lunungwi, compared to using only electricity and multiple fuels.

Based on the Chi-square test results, there was no significant association between seasons and energy sources for cooking ($p > 0.05$). This observation implies that the choice of energy source for cooking is not influenced by the season, emphasising that factors other than seasonal variations play a more significant role in determining the energy sources used for cooking.

**Table 3. 11:** Cross-tabulation between the choice of energy source used for cooking during the seasons using chi-squared test

| Season | Area | Electricity | Wood | Electricity and wood | Total | Value | df | Asymptotic Significance (2-sided) | Cramer's V |
|---|---|---|---|---|---|---|---|---|---|
| Summer | Gondeni | 0 | 43 | 37 | 80 | 3.168[a] | 1 | 0.075 | 0.141 |
| | Lunungwi | 0 | 54 | 26 | 80 | | | | |
| | Total | 0 | 97 | 63 | 160 | | | | |
| Winter | Gondeni | 1 | 42 | 34 | 77 | 1.239[a] | 2 | 0.538 | 0.089 |
| | Lunungwi | 0 | 47 | 33 | 80 | | | | |
| | Total | 1 | 89 | 67 | 157 | | | | |

Furthermore, the Chi-square ($\chi^2$) analyses also revealed a significant association between the average number of daily hours of wood use seasonally ($p<0.05$) as more fuelwood is used for long hours per day during winter compared to summer (Table 3.12). Cramer's V, a widely used measure of nominal association, assesses the strength of a relationship independently of the sample size and serves solely as an index of association strength, however, a limitation of Cramer's V is that it cannot be utilized to compare the strength of one relationship to another correlation. For Cramer's V values, 0.0 to 0.30



© University of Venda

indicates no relationship to a weak relationship strength; 0.31 to 0.70 indicates a moderate relationship; and 0.71 to 1.0 shows that the strength of the relationship is strong (McBride, 2005; McHugh, 2013; Akoglu, 2018).

Likewise, Chi-square ($\chi 2$) analyses show a significant association between age and satisfaction with the fuel wood burning system and awareness of fuel wood burning contributing to air pollution ($p<0.05$) (Table 3.10). A significantly higher proportion of the younger generation (18-40 yrs) were not satisfied with the fuel wood burning system compared to the older groups (40-50 yrs and 50+ yrs). Similarly, a significantly higher proportion of the younger generation (18-40 yrs) indicated that they are aware of fuel wood burning contribution to air pollution compared to the older age groups (40-50 yrs and 50+ yrs). Based on these empirical results, age influences household fuel choices, with older people being more resistant to new technology and preferring traditional fuels. Factors like cultural norms and familiarity with traditional practices contribute to this preference. Designing interventions to promote cleaner, more sustainable energy sources among older populations is, hence, crucial. Moreover, the impact of age on fuel selection can also be attributed to factors like cultural practices and familiarity with established traditions (Buba *et al*., 2017; Semenya and Machete, 2019).



© University of Venda

**Table 3.12:** Crosstab for village comparison versus respondent's perspective of reason, satisfaction, awareness, and condition

| | Association | Chi-square | df* | p-value | Cramer's V |
|---|---|---|---|---|---|
| Factors associated with the village | The main reason to use fuelwood burning | 17.275[a] | 6 | 0.008† | 0.329 |
| | Satisfaction with the fuel wood burning system used | 10.103[a] | 2 | 0.006† | 0.251 |
| Factors associated with seasonality | The average number of daily hours of wood used | 84.391 | 6 | 0.000† | 0.514 |
| | daily hours of wood use in summer | 10.823 | 4 | 0.029† | 0.260 |
| | daily hours of wood use in winter | 18.243 | 6 | 0.006† | 0.338 |
| Factors associated with age | The main reason to use a fuelwood burning | 9.923[a] | 12 | 0.623 | 0.249 |
| | Satisfaction with the fuel wood burning system used | 16.035[a] | 4 | 0.003† | 0.317 |
| | Awareness of fuel wood burning contribution to air pollution | 30.839[a] | 2 | 0.000† | 0.439 |
| | condition of the atmosphere in your area | 5.874[a] | 4 | 0.209 | 0.192 |

*df=degree of freedom, †statistically significant

### 3.3.2.5 Fuelwood consumption by households based on scale measurement

In this field study, the average weights of wood consumed by each of the fifteen participants in Gondeni and Lunungwi were measured, then totalled, and the results from their scale measurements are presented in Table 3.13. Due to COVID-19 and limited time, spring and autumn average daily consumption were unavailable. Therefore, the spring and autumn wood consumption value per HH was extrapolated from Naidoo *et al*. (2014b) and Morales *et al*. (2018) using the equivalence fraction of the composite approach in equations 3.6 and 3.7.

$$x1(Autum) = \frac{36.65\ x\ winter\ value}{46.55} \quad (3.6)$$

$$x2(Spring) = \frac{14.13\ x\ Summer\ value}{13.49} \quad (3.7)$$



© University of Venda

**Table 3.12:** Mean, total and maximum amounts of wood burnt in the mornings and evenings by all 15 households in the villages for summer and winter 2020/2021.

| | | Gondeni | | | Lunungwi | | |
|---|---|---|---|---|---|---|---|
| **Seasons** | | Morning | Evening | Daily | Morning | Evening | Daily |
| **Winter** | Mean ± Stdev | 6.49 ± 2.00 | 8.29 ± 1.44 | 10.73 ± 1.10 | 7.09 ± 2.63 | 11.27± 2.44 | 15.10 ± 2.17 |
| | Maximum | 12.45 | 16.2 | 17.56 | 16.36 | 20.8 | 24.75 |
| | Minimum | 0.9 | 1.2 | 6.2 | 1.1 | 3.39 | 4.45 |
| | Total | 927.32 | 2939.5 | 3863.82 | 1524.65 | 3995.84 | 5436.49 |
| **Summer** | Mean ± Stdev | 4.97 ± 2.87 | 6.70 ± 1.32 | 6.50 ± 1.29 | 7.38 ± 2.36 | 6.29 ± 1.99 | 8.03 ± 2.21 |
| | Maximum | 10.11 | 11.5 | 11.5 | 15.9 | 13.47 | 16.87 |
| | Minimum | 3.1 | 1.95 | 1.95 | 1.45 | 0.6 | 0.6 |
| | Total | 626.85 | 1641.88 | 2268.73 | 1779.54 | 1052.43 | 2833.39 |

The amount of wood consumed by each household varies greatly depending on the time of the day and seasons as illustrated in Figure 3.22. Table 3.13 illustrates the mean (average daily consumption), the standard deviation, and total amounts of wood consumed by all 15 households in each village during the morning, evening, and daily burning periods of each season. The summer and spring consumption was very low compared to winter, with averages of 6.50 ± 1.29 kg and 6.81 ± 3.45 kg per HH per day for Gondeni, while Lunungwi had 8.03 ± 2.21 kg and 8.41 ± 3.45 kg per HH per day respectively. It is no surprise that the maximum wood consumption of 10.73 ± 1.10 kg and 15.10 ± 2.17 kg per household were observed during the winter season, and 8.68 ± 2.91 kg and 12.21 ± 2.91 kg consumption in the autumn season for Gondeni and Lunungwi, respectively.

- *Average Daily Wood Consumption per Household by Season in Thulamela LM, South Africa*

Based on actual scale measurement, Figure 3.23 shows a rise in wood consumption during the colder months. This rise corresponds to lower temperatures and is influenced by both convenience and temperature, resulting in more frequent evening use (Naidoo *et al*., 2014a; Nkosi *et al*., 2018). This pattern aligns with the questionnaire survey and is consistent with the findings of Tshikalanke (2007) and Naidoo *et al*. (2014a), where higher fuel usage during winter and evenings was reported.



© University of Venda

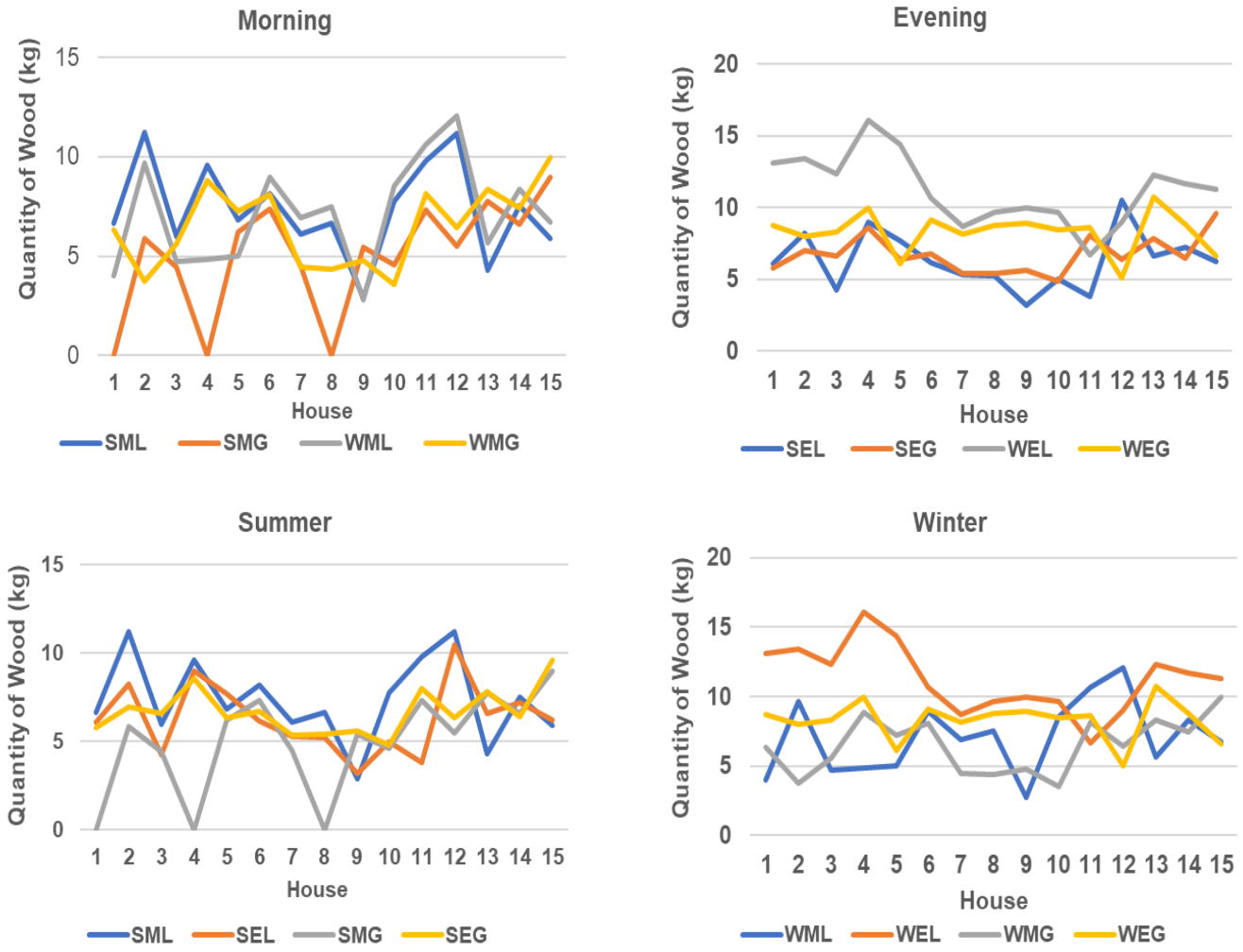



**Figure 3.22:** Average amount of wood consumed in a) morning and evening, and b) summer and winter per household (HH) in the 2 villages.

No significant differences in daily average fuelwood consumption were found among the various seasons of the year based on data from the two villages (Kruskal-Wallis test, $p = 0.129$). Fuel usage is presented per household to account for the area's higher consumption due to a larger number of households; therefore, the estimation of wood usage for the study domains was derived from the average household fuel consumption of 3491 kg per year. This value aligns with a prior study conducted in the Vaal Triangle Airshed Priority Area (VTAPA) of South Africa, where the firewood consumption rate was recorded as 3000 kg/year per household (DEA, 2018).



© University of Venda

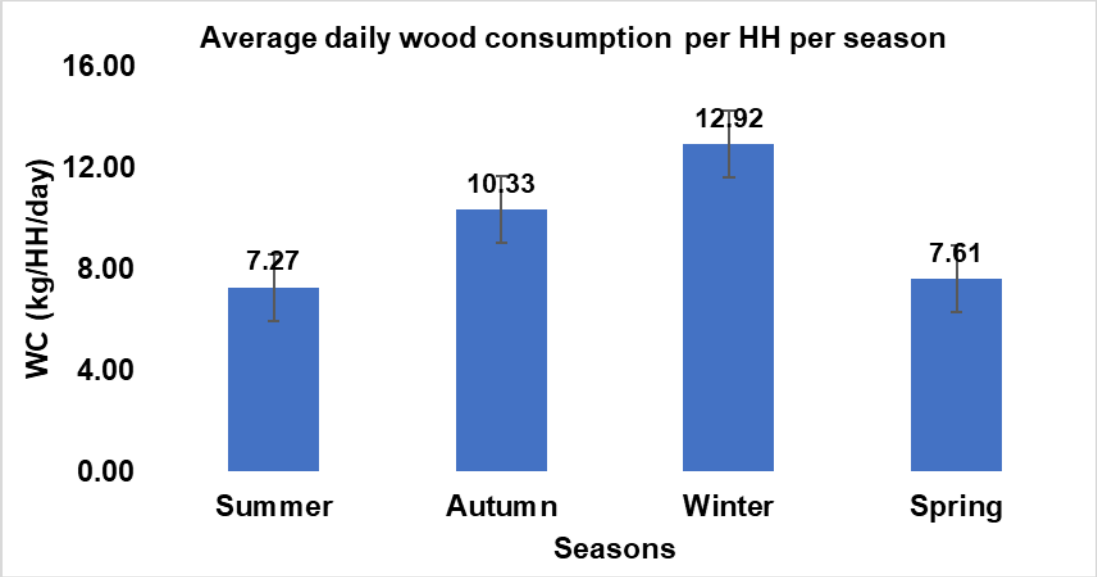



**Figure 3.23:** Average daily wood consumption per Household (HH) by season recorded in this study.

### 3.3.2.6 Emissions rate for residential fuelwood combustion in the study domain

Residential burning of low-grade fuels such as wood and biomass is a major source of air pollution in many parts of the world. Household fuel produces high levels of gaseous and particulate emissions due to the combustion of low-quality fuel and poor stove operation (Bonjour *et al*., 2013). Factors such as the quantity, type, size, chemistry, and moisture content of the fuel used, the age and condition of stoves, as well as local climate conditions, can have a significant impact on emissions from residential burning. These factors can vary over time and space, leading to variations in emissions (Shen *et al*., 2013; Lelieveld *et al*., 2015; Win *et al*., 2018; Cincinelli *et al*., 2019). Fuelwood is the predominant type of fuel used by all the HH in the two villages and within the Municipality, thus contributes significantly to the emission of the five criteria pollutants which include, $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$.

Figures 3.24 and 3.25 show the areas (43 area sources) using wood for domestic activities in 2021 and the corresponding number of households (42,222) employing wood in each area, based on 2001, 2011, and 2016 census results. According to Statistics South Africa (via the official Census and Community Survey), there is an increase in population as well as HH number in the Municipality (StatsSA, 2011). The consumption



© University of Venda

rates were only measured in the two villages to get the average value. The annual estimates are based on the assumption that all HH consumed the same amount of fuelwood per day, and it was also assumed that consumption rates would remain as they are in this study and only population density changed per area or village.

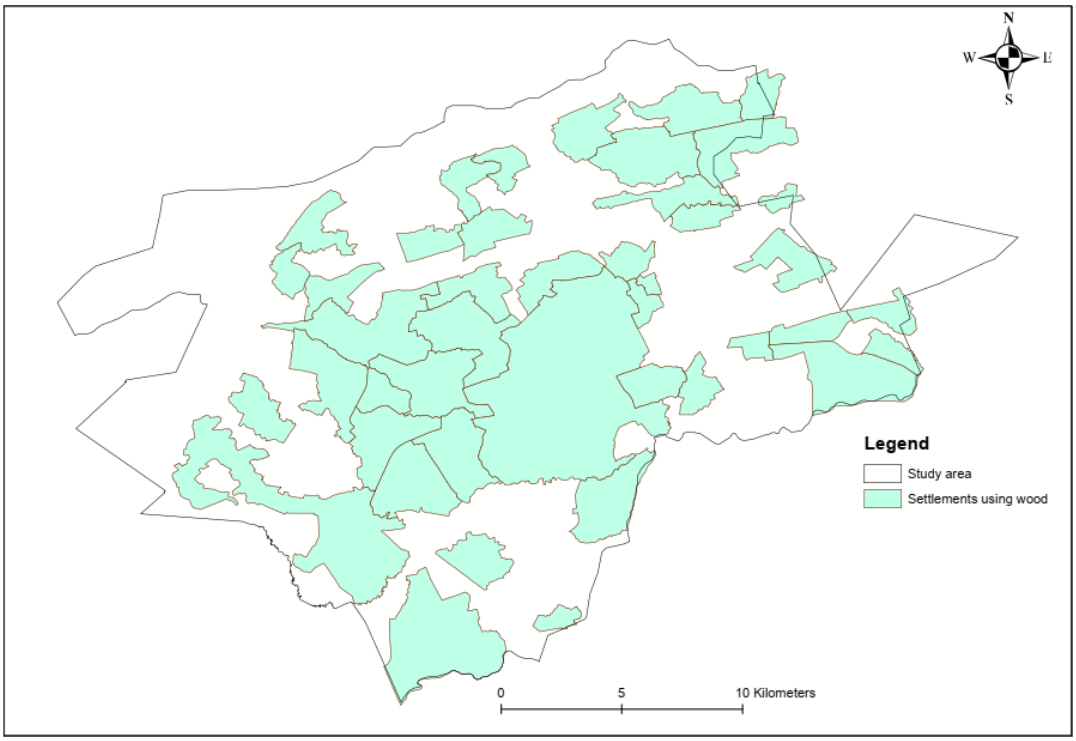



**Figure 3.24:** Areas using wood for domestic activities within the study domain.

Figure 3.26 illustrates emissions rates (g/s) for ($SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$) from RWC, revealing higher emissions during winter compared to summer. These findings support previous studies demonstrating increased $PM_{10}$ (0.0092 g/s) and CO (0.067 g/s) emission rates during winter fuelwood combustion (Naidoo *et al*., 2014b). This pattern is attributed to intensified burning, more burning events, longer durations, as well as the practice of boiling water and space heating, consistent with earlier research (Tshikalanke, 2007; Naidoo *et al*., 2014a).

Additionally, Table 3.14 provides the annual emission inventory for the area, showing the residential sector significantly contributing to CO, $PM_{2.5}$, and $PM_{10}$ emissions due to the large number of households that use wood for domestic activities and high EF. A higher



© University of Venda

prevalence of fuelwood use was observed in Limpopo Province (northern section) of South Africa than in other areas, highlighting regional variations in fuel consumption patterns. The fuelwood usage was disaggregated to a gridded spatial resolution suitable for use on the model grids using the StatsSA Dwelling Frame database.

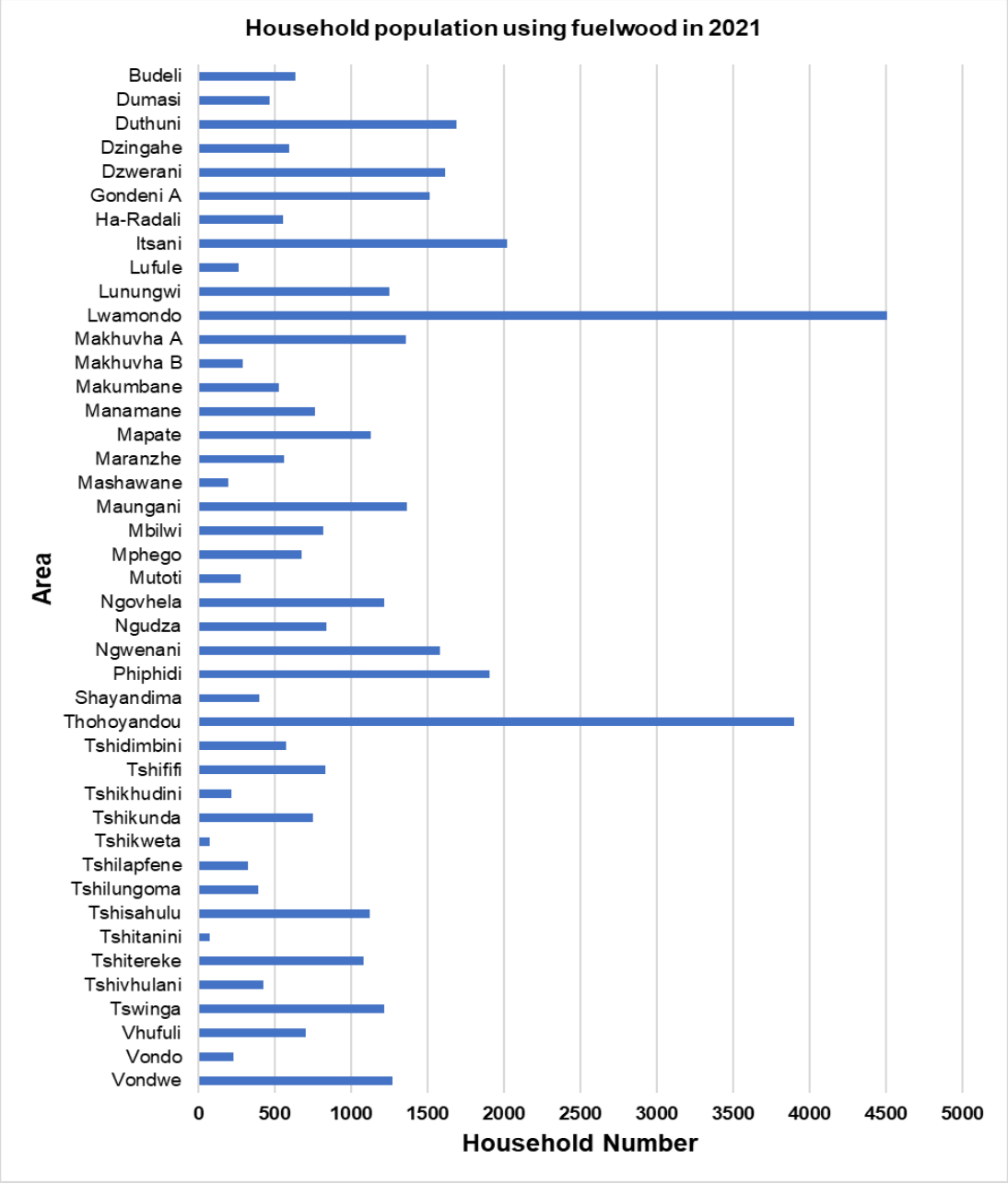



**Figure 3.25:** Number of households using wood in 2021 within the study area.



© University of Venda

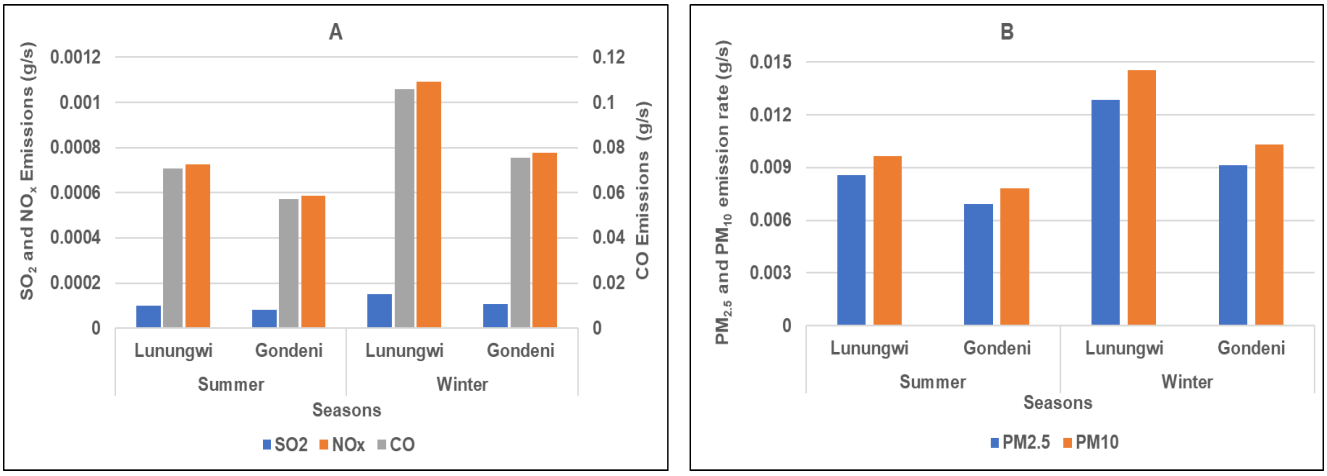



**Figure 3.26:** A) $SO_2$, $NO_x$, and CO gaseous emission, and B) $PM_{2.5}$ and $PM_{10}$ particulate emission by village in the study domain.

**Table 3.13:** Estimated RWC emission for 2021 ($10^{-4}$ tpa)

| Pollutants | $SO_2$ | $NO_x$ | CO | $PM_{2.5}$ | $PM_{10}$ |
|---|---|---|---|---|---|
| **Gondeni** | 5.38 | 38.82 | 3771.96 | 456.94 | 516.67 |
| **Lunungwi** | 7.19 | 51.92 | 5044.22 | 611.06 | 690.93 |
| **Average per HH** | 6.28 | 45.37 | 4408.09 | 534.0 | 603.80 |
| **Study Domain*** | **26.53** | **191.57** | **18611.92** | **2254.65** | **2549.38** |

* is tpa

Figures 3.27 and 3.28 display RWC emissions distributed within the study domain, which are primarily concentrated on the SW (which accounts for the majority) and NW sides, however, there are also limited emissions in the SE and NE areas. The spatial distribution of these emissions aligned with higher population density patterns, particularly in areas where wood is commonly used for household activities. This relationship indicates that densely-populated communities relying on wood combustion contribute more significantly to overall emissions, hence aligning with findings from previous studies on spatial patterns of residential fuel use and emissions (DEA, 2018; Kanemoto *et al*., 2020; Paunu *et al*., 2021).



© University of Venda

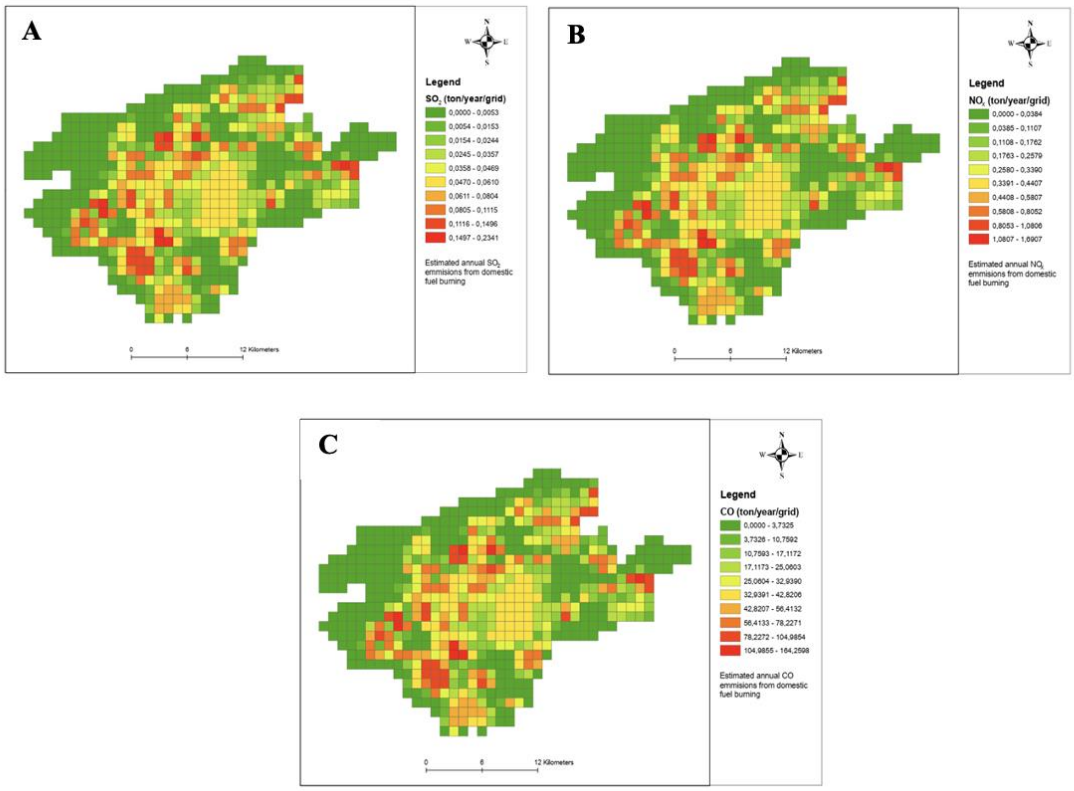



**Figure 3.27**: *Spatial distribution of annual A) $SO_2$, B) $NO_x$, and C) CO emissions (tpa) from Residential Wood Combustion (RWC) within the 1 km model domain.*

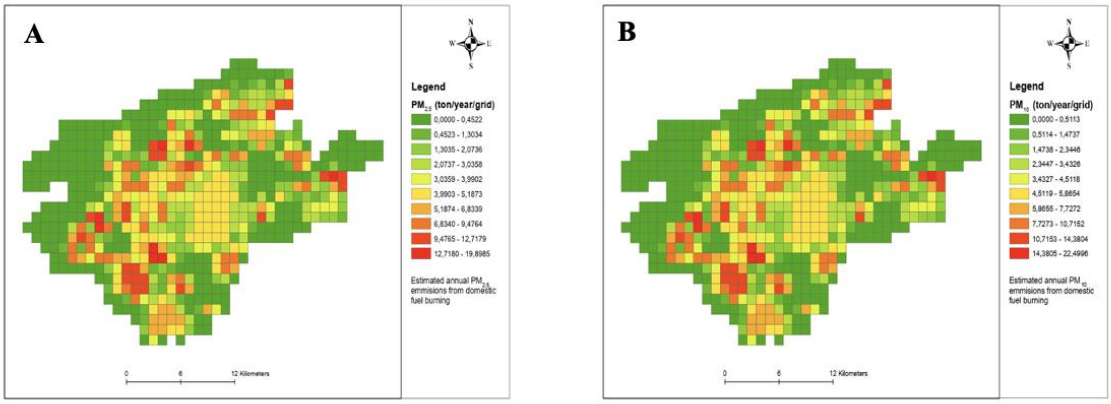



**Figure 3.28:** Spatial distribution of annual A) $PM_{2.5}$, and B) $PM_{10}$ emissions (tpa) from RWC in the 1 km model domain.



© University of Venda

### 3.4 Source contributions to total annual emissions

The source contributions to the total annual emissions of various pollutants ($SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$) are presented in Figure 3.29, with corresponding values in Table 3.15. Residential wood combustion (RWC) has emerged as the primary source of most pollutants, accounting for 96% of $PM_{10}$ and $PM_{2.5}$, and 89% of CO. In contrast, vehicles were the dominant contributors to $NO_x$ (89%) and $SO_2$ (60 %) emissions. These findings align with previous research conducted in the Vaal Triangle Airshed Priority Area (VTAPA) by the Department of Environmental Affairs (DEA, 2018), Naidoo *et al*. (2019), and Muyemeki *et al*. (2022), which consistently identified vehicles as the primary source of $NO_x$ emissions. Additionally, Mentese *et al*. (2020) also found that the traffic sector contributed significantly to $NO_x$ emissions. Among the two anthropogenic sources examined in the study domain, RWC emerged as the dominant producer of criteria pollutants, except for $NO_x$ and $SO_2$. This observation is consistent with earlier research by Naidoo *et al*. (2019) and studies conducted in VTAPA, South Africa, except for $SO_2$ emissions (DEA, 2018).

**Table 3.14:** 2021 emission inventories (ton per annum) per sector, for the study domain in Thulamela Municipality

| Sources | $SO_2$ | $NO_x$ | CO | $PM_{2.5}$ | $PM_{10}$ |
|---|---|---|---|---|---|
| **Vehicle emission** | 39.7 | 1630.1 | 2390.7 | 94.1 | 99.9 |
| **Residential wood combustion** | 26.5 | 191.6 | 18611.9 | 2254.7 | 2549.4 |
| **Total** | 66.3 | 1821.7 | 21002.6 | 2348.7 | 2649.3 |

### 3.5 Limitations of the study

The data used in the analysis relied on multiple sources, which may require further completion or updating. The accuracy and precision of emissions data for a particular area depend on the quality and certainty of the population (vehicle and household), VKT, fuel consumption, and other emissions data. For the residential sector, the seasonal consumption rate for the area was determined using measured summer and winter data and data from earlier studies to extrapolate the other two seasons' data. Only wood was



© University of Venda

quantified as the primary fuel used in the area. We used thirty households in two (2) villages to quantify the total fuelwood usage, and the average results of the two villages are representative of the domain emission per HH; therefore, there are inherent uncertainties related to having a small sample size.

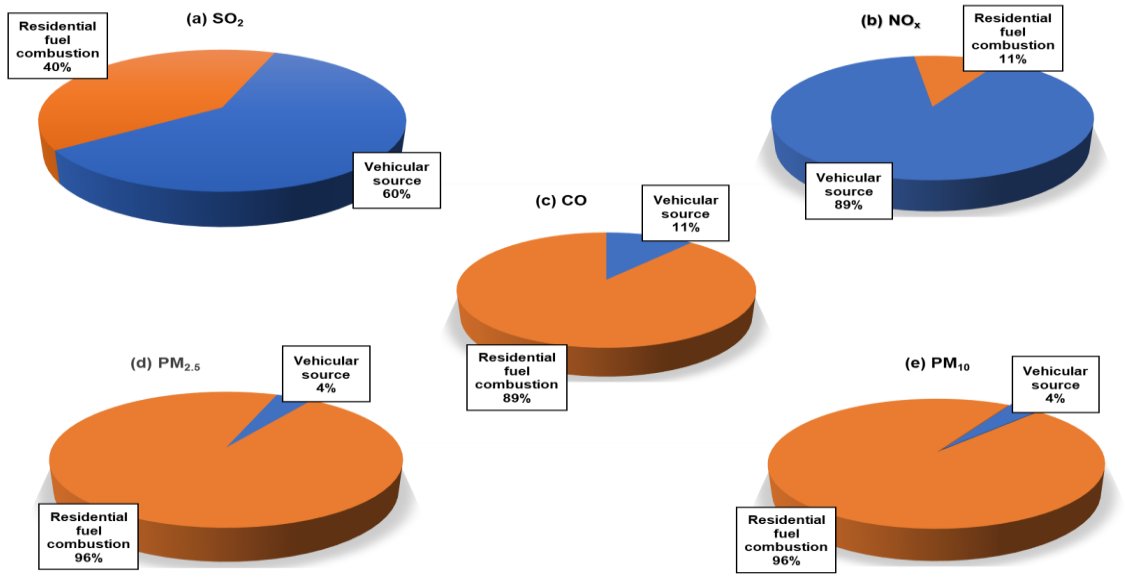



**Figure 3.29:** Source contributions to the total annual emissions by pollutants for the study.

### 3.6 Conclusion

This study employed an extensive bottom-up approach to develop an emission inventory for Thulamela Municipality in Limpopo Province for the year 2021, focusing on both the vehicle and residential sectors. By integrating qualitative and quantitative techniques, including eNaTis data, questionnaires, hanging scale measurements, and secondary data sources, the inventory estimated $SO_2$, $NO_x$, CO, $PM_{2.5}$, and $PM_{10}$ emissions. The results underscore the substantial impact of vehicle and domestic combustion on pollution levels within the Thulamela LM area, with LDVs, PCs, and HDVs emerging as primary contributors to $NO_x$, CO, and $SO_2$ emissions. In addition, the increasing mean daily vehicle kilometres travelled (VKT) of the area and the prominence of wood consumption (higher in the winter) further contribute to emissions, particularly CO, $PM_{2.5}$, and $PM_{10}$. This comprehensive inventory fills a critical gap in the literature by showing that



© University of Venda

Thulamela Municipality significantly contributes to pollution levels. Validating the current inventory using the AERMOD model enabled an accurate assessment of pollutant concentrations in densely populated areas and their determination of compliance with acceptable limits. The model facilitates evaluation of pollutant levels against standard thresholds and supports the development of suitable mitigation measures to reduce pollution, improve air quality, and promote clean and sustainable living environments for everyone. Further research is also recommended to understand the influence of other activities like biomass burning and brickmaking on emissions, to enable more targeted pollution reduction strategies.

## References

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

Al-Ramahi, F. K. M. (2020). Spatial analysis of Radon Gas concentration distributed at Baghdad City using remote sensing and geographic information system Techniques d. *Iraqi Journal of Agricultural Sciences*, *51*(Special).

Almetwally, A. A., Bin-Jumah, M., & Allam, A. A. (2020). Ambient air pollution and its influence on human health and welfare: an overview. *Environmental Science and Pollution Research*, *27*, 24815–24830.

Blignaut, J. N., Chitiga-Mabugu, M., & Mabugu, R. (2005). *Constructing a greenhouse gas emissions inventory using energy balances: the case of South Africa for 1998*.

Boahene, K. (1998). The challenge of deforestation in tropical Africa: reflections on its principal causes, consequences and solutions. *Land Degradation & Development*, *9*(3), 247–258.

Bonjour, S., Adair-Rohani, H., Wolf, J., Bruce, N. G., Mehta, S., Prüss-Ustün, A., Lahiff, M., Rehfuess, E. A., Mishra, V., & Smith, K. R. (2013). Solid fuel use for household cooking: Country and regional estimates for 1980-2010. *Environmental Health Perspectives*, *121*(7), 784–790. https://doi.org/10.1289/ehp.1205987.

Buba, A., Abdu, M., Adamu, I., Jibir, A., & Usman, Y. I. (2017). Socio-economic determinants of households' fuel consumption in Nigeria. *International Journal of Research-Granthaalayah*, *5*(10), 348–360.

Cai, H., & Xie, S. (2013). Temporal and spatial variation in recent vehicular emission inventories in China based on dynamic emission factors. *Journal of the Air and Waste Management Association*, *63*(3), 310–326. https://doi.org/10.1080/10962247.2012.755138.

Cai, H., & Xie, S. D. (2007). Estimation of vehicular emission inventories in China from 1980 to 2005. *Atmospheric Environment*, *41*(39), 8963–8979.



© University of Venda

https://doi.org/10.1016/j.atmosenv.2007.08.019.

Chidhindi, P. (2022). Using dispersion models as a regulatory tool in South Africa. North-West University (South Africa).

Cincinelli, A., Guerranti, C., Martellini, T., & Scodellini, R. (2019). Residential wood combustion and its impact on urban air quality in Europe. *Current Opinion in Environmental Science and Health*, *8*, 10–14. https://doi.org/10.1016/j.coesh.2018.12.007.

Danlami, A. H. (2019). Assessment of factors influencing firewood consumption in Bauchi state, Nigeria. *Journal of Sustainability Science and Management*, *14*(1), 99–109.

Darçın, M. (2017). How air pollution affects subjective well-being. *Well-Being and Quality of Life: Medical Perspective*, *211*.

de Azevedo, C. L., & Cardoso, J. L. (2009). Estimation of annual traffic volumes a model for portugal. *ECTRI–FEHRL–FERSI Young Researchers Seminar 2009*, 1–15.

DEA Department of Environmental Affairs. (2018). The Second-Generation Vaal Triangle Airshed Priority Area Air Quality Management Plan: Draft Baseline Assessment Report. https://saaqis.environment.gov.za/pagesfiles/vtapa%20second%20generation%20aqmp_draft%20baseline%20assessment%20report_public%20comment.pdf

DEAT. (2008). Executive Summary of the Vaal Triangle Airshed Priority Area Air Quality Management Plan. Government Gazette, 31615, 4–39.

Department of Energy (DoE). (2013). Draft 2012 Integrated Energy Planning Report. Department of Energy, Pretoria.

EEA. (2019). EMEP/EEA air pollutant emission inventory guidebook. Publications Office of the European Union Luxembourg.

Enitan I.T., Odiyo J.O, Piketh S.J, Edokpayi J.N. (2021). Vehicle Emission Inventory for Thulamela Municipality, Limpopo province, South Africa. 2021 Annual Conference Proceedings of the National Association for Clean Air (NACA), South Africa, pp 31-36.

Francioli, A. P., & Pharoah, D. R. (2018). *investigating energy usage among low income households and implications for fire risk*. https://scholar.sun.ac.za

Freiman, M. T., & Piketh, S. J. (2003). Air transport into and out of the industrial Highveld region of South Africa. *Journal of Applied Meteorology*, *42*(7), 994–1002. https://doi.org/10.1175/1520-0450(2003)042<0994:ATIAOO>2.0.CO;2

Gurjar, B. R., Van Aardenne, J. A., Lelieveld, J., & Mohan, M. (2004). Emission estimates and trends (1990-2000) for megacity Delhi and implications. *Atmospheric Environment*, *38*(33), 5663–5681. https://doi.org/10.1016/j.atmosenv.2004.05.057.

Hajirasouliha, F., & Zabiegaj, D. (2020). Effects of Environmental Emissions on the Respiratory System: Secrets and Consequences. In *Environmental Emissions*. intechopen.

HEI, Health Effects Institute (2018). *State of Global Air 2018: A special report on global exposure to air pollution and its disease burden.* Boston, MA: Health Effects



© University of Venda

Institute. https://www.stateofglobalair.org/sites/default/files/soga-2018-report.pdf

Huo, H., Zhang, Q., He, K., Yao, Z., & Wang, M. (2012). Vehicle-use intensity in China: Status and future trend. *Energy Policy*, *43*, 6–16. https://doi.org/10.1016/j.enpol.2011.09.019

Jagarnath, M., & Thambiran, T. (2018). Greenhouse gas emissions profiles of neighbourhoods in Durban, South Africa–an initial investigation. *Environment and Urbanization*, *30*(1), 191–214.

Joshi, J. (2023). A systematic study on integration of gis and remote sensing for urban environmental analysis. *A Journal for New Zealand Herpetology*, *12*(3), 2992–3000.

Jung, S., Kim, J., Kim, J., Hong, D., & Park, D. (2017). An estimation of vehicle kilometer traveled and on-road emissions using the traffic volume and travel speed on road links in Incheon City. *Journal of Environmental Sciences (China)*, *54*, 90–100. https://doi.org/10.1016/j.jes.2015.12.040

Kanemoto, K., Shigetomi, Y., Hoang, N. T., Okuoka, K., & Moran, D. (2020). Spatial variation in household consumption-based carbon emission inventories for 1200 Japanese cities. Environmental Research Letters, 15(11), 114053.

Karimu, A., Mensah, J. T., & Adu, G. (2016). Who adopts LPG as the main cooking fuel and why? Empirical evidence on Ghana based on national survey. *World Development*, *85*, 43–57.

Kenworthy, J. R., & Laube, F. B. (1996). Automobile dependence in cities: an international comparison of urban transport and land use patterns with implications for sustainability. *Environmental Impact Assessment Review*, *16*(4–6), 279–308.

Kituyi, E., Marufu, L., Huber, B., Wandiga, S. O., Jumba, I. O., Andreae, M. O., & Helas, G. (2001). Biofuel consumption rates and patterns in Kenya. *Biomass and Bioenergy*, *20*(2), 83–99.

Ko, U. W., & Kyung, S. Y. (2022). Adverse Effects of Air Pollution on Pulmonary Diseases. *Tuberculosis and Respiratory Diseases*, *85*(4), 313–319.

Koenig, J. Q. (2000). *Health effects of ambient air pollution: how safe is the air we breathe?* Springer Science & Business Media.

Kumar, A., & Tripathi, S. (2014). Study of vehicular pollution and its mitigation measures. *KIIT International Symposium on Advances in Automotive Technology*.

Landrigan, P. J., Fuller, R., Acosta, N. J. R., Adeyi, O., Arnold, R., Baldé, A. B., Bertollini, R., Bose-O'Reilly, S., Boufford, J. I., & Breysse, P. N. (2018). The Lancet Commission on pollution and health. *The Lancet*, *391*(10119), 462–512.

Lang, J., Cheng, S., Zhou, Y., Zhang, Y., & Wang, G. (2014). Air pollutant emissions from on-road vehicles in China, 1999-2011. *Science of the Total Environment*, *496*(X), 1–10. https://doi.org/10.1016/j.scitotenv.2014.07.021

Lavhelani, F. G. (2021). *Quantifying the use of firewood as a source of energy in Northwest Province: a case study for Matlwang*. North-West University (South Africa).

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The



© University of Venda

contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, *525*(7569), 367–371. https://doi.org/10.1038/nature15371.

Lelieveld, Jos, Pozzer, A., Pöschl, U., Fnais, M., Haines, A., & Münzel, T. (2020). Loss of life expectancy from air pollution compared to other risk factors: a worldwide perspective. *Cardiovascular Research*, *116*(11), 1910–1917.

Lestari, P., Arrohman, M. K., Damayanti, S., & Klimont, Z. (2022). Emissions and spatial distribution of air pollutants from anthropogenic sources in Jakarta. *Atmospheric Pollution Research*, *13*(9), 101521.

Limanond, T., Pongthanaisawan, J., Watthanaklang, D., & Sangphong, O. (2009). An analysis of vehicle kilometers of travel of major cities in Thailand. *ATRANS Final Report 2009*.

Limanond, T., Pongthanaisawan, J., Watthanaklang, D., & Sangphong, O. (2010). *An Analysis of Vehicle Kilometers of Travel of Major Cities in Thailand*.

Liu, G., Sun, S., Zou, C., Wang, B., Wu, L., & Mao, H. (2022). Air pollutant emissions from on-road vehicles and their control in Inner Mongolia, China. *Energy*, *238*, 121724.

Lloyd, P. (2014). The energy profile of a low-income urban community. *Twenty-Second Domestic Use of Energy*, 1–6.

Lu, H., Ma, H., Sun, Z., & Wang, J. (2017). *Analysis and Prediction on Vehicle Ownership Based on an Improved Stochastic Gompertz Diffusion Process*. *2017*.

Lv, W., Hu, Y., Li, E., Liu, H., Pan, H., Ji, S., Hayat, T., Alsaedi, A., & Ahmad, B. (2019). Evaluation of vehicle emission in Yunnan province from 2003 to 2015. *Journal of Cleaner Production*, *207*, 814–825. https://doi.org/10.1016/j.jclepro.2018.09.227.

Makonese, T., Masekameni, D. M., & Annegarn, H. J. (2016). Energy use scenarios in an informal urban settlement in Johannesburg, South Africa. *Proceedings of the 24th Conference on the Domestic Use of Energy, DUE 2016*. https://doi.org/10.1109/DUE.2016.7466703.

Masekela, M. E., & Semenya, K. (2021). Factors influencing the use of firewood post-electrification in rural South Africa: The case of Ga-Malahlela village. *Journal of Energy in Southern Africa*, *32*(3), 24–40.

McBride, G. B. (2005). A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. *NIWA Client Report: HAM2005-062*, *45*, 307–310.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica*, *23*(2), 143–149.

Mentese, S., Mirici, N. A., Elbir, T., Tuygun, G. T., Bakar, C., Otkun, M. T., & Oymak, S. (2020). A comprehensive assessment of ambient air quality in Çanakkale city: Emission inventory, air quality monitoring, source apportionment, and respiratory health indicators. *Atmospheric Pollution Research*, *11*(12), 2282–2296. https://doi.org/10.1016/j.apr.2020.07.005.

Morales, D., Molares, S., & Ladio, A. H. (2018). Seasonal Variation in the consumption of biomass fuel in a rural community of Patagonia, Argentina. *Ethnobiology and Conservation*, *7*.



© University of Venda

Mucina, L., & Rutherford, M. C. (2006). *The vegetation of South Africa, Lesotho and Swaziland.* South African National Biodiversity Institute.

Munyon, V. V., Bowen, W. M., & Holcombe, J. (2018). Vehicle fuel economy and vehicle miles traveled: An empirical investigation of Jevon's Paradox. *Energy Research and Social Science*, *38*(November 2016), 19–27. https://doi.org/10.1016/j.erss.2018.01.007.

Muyemeki, L., Petrus, R., Stuart, B., Piketh, J., Rafaj, P., & Kiesewetter, G. (2022). Integrated assessment of strategies to reduce air pollution in the Vaal Triangle Priority Area, South Africa. *Air Quality, Atmosphere & Health*, *3*, 2157–2169. https://doi.org/10.1007/s11869-022-01242-8.

Naidoo, M., Naidoo, S., and Garland, R.M, 2019. Development of a model-ready emissions inventory for a coastal city in South Africa, 2019. https://www.geiacenter.org/sites/default/files/site/community/geia-conferences/2019/geia_posters_2019/Naido%20poster.pdf.

Naidoo, S., Piketh, S. J., & Curtis, C. (2014a). Quantification of emissions generated from domestic burning activities from townships in Johannesburg. *Clean Air Journal= Tydskrif Vir Skoon Lug*, *24*(1), 34–41.

Naidoo, S. (2014b). Quantification of emissions generated from domestic burning activities from townships in Johannesburg. University of the Witwatersrand, South Africa.

National Traffic Information System (NaTIS) 2020. https://www.natis.gov.za/

Nemaxwi, T. P. (2019). *The impact of change management on service delivery at Thulamela Local Municipality*. University of Johannesburg (South Africa).

Netshipise, L. F., & Semenya, K. (2022). Evaluating factors influencing firewood consumption in households at the Thulamela Local Municipality, South Africa. *Journal of Energy in Southern Africa*, *33*(2), 48–62. https://doi.org/10.17159/2413-3051/2022/v33i2a9741.

Netswera, M. M. (2023). Implications of Municipal Boundary Determination on Social Integration of Diverse Communities in South Africa. *African Journal of Governance and Development*, *12*(1), 43–61.

Nicholls, D., Barnes, F., Acrea, F., Chen, C. L., Buluç, L. Y., & Parker, M. M. (2015). Top-down and bottom-up approaches to greenhouse gas inventory methods-a comparison between national-and forest-scale reporting methods. *General Technical Report-Pacific Northwest Research Station, USDA Forest Service*, *PNW-GTR-906*.

Nkosi N.C. 2018. Quantifying particulate emissions from domestic burning in kwaDela Township, Mpumalanga. North-West University. (Dissertation – MSc).

Nkosi, C. N., Piketh, J. S., Burger, P. R., & Harrold, A. J. (2017). Variability of domestic burning habits in the South African Highveld: A case study in the KwaDela Township (April 2017). *2017 International Conference on the Domestic Use of Energy (DUE)*, 23–29.



© University of Venda

Ntziachristos, L., & Samaras, Z. (2018). EMEP/EEA—Air Pollutant Emission Inventory Guidebook 2016—Update Jul. 2018. *European Environment Agency (EEA), Copenhagen, Denmark, Available at: Https://Www. Eea. Europa. Eu/Publications/Emep-Eea-Guidebook-2016/Part-b-Sectoral-Guidance-Chapters/1-Energy/1-a-Combustion/1-a-3-Bi (Last Access: 2 November 2021)*.

Paunu, V.V., Karvosenoja, N., Segersson, D., López-Aparicio, S., Nielsen, O.K., Plejdrup, M.S., Thorsteinsson, T., Niemi, J.V., Vo, D.T., van der Gon, H.A.D. and Brandt, J., 2021. Spatial distribution of residential wood combustion emissions in the Nordic countries: How well national inventories represent local emissions?. Atmospheric Environment, 264, p.118712.

Piketh, S. J., Swap, R. J., Maenhaut, W., Annegarn, H. J., & Formenti, P. (2002). Chemical evidence of long-range atmospheric transport over southern Africa. *Journal of Geophysical Research: Atmospheres*, *107*(D24), ACH-7.

Progiou, A. G., & Ziomas, I. C. (2011). Science of the Total Environment Road traffic emissions impact on air quality of the Greater Athens Area based on a 20-year emissions inventory. *Science of the Total Environment, The*, *410–411*, 1–7. https://doi.org/10.1016/j.scitotenv.2011.09.050.

Ramirez, J., Pachon, J. E., Casas, O. M., & González, S. F. (2019). A new database of on-road vehicle emission factors for Colombia: A case study of bogota. *CTyF - Ciencia, Tecnologia y Futuro*, *9*(1), 73–82. https://doi.org/10.29047/01225383.154

Sahu, S. K., Beig, G., & Parkhi, N. S. (2011). Emissions inventory of anthropogenic $PM_{2.5}$ and PM10 in Delhi during Commonwealth Games 2010. *Atmospheric Environment*, *45*(34), 6180–6190.

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

Scorgie, Y., Kneen, M. A., Annegarn, H. J., & Burger, L. W. (2003). Air pollution in the Vaal Triangle-quantifying source contributions and identifying cost-effective solutions. *Clean Air Journal= Tydskrif Vir Skoon Lug*, *13*(2), 5–18.

Semenya, K., & Machete, F. (2019). Factors that influence firewood use among electrified Bapedi households of Senwabarwana Villages, South Africa. *African Journal of Science, Technology, Innovation and Development*, *11*(6), 719–729. https://doi.org/10.1080/20421338.2019.1572336.

Seymore, R., Inglesi-Lotz, R., & Blignaut, J. (2014). A greenhouse gas emissions inventory for South Africa: A comparative analysis. In *Renewable and Sustainable Energy Reviews* (Vol. 34, pp. 371–379). Elsevier Ltd. https://doi.org/10.1016/j.rser.2014.03.028.

Shabadin, A., Johari, N. M., & Jamil, H. M. (2014). Car Annual Vehicle Kilometer Travelled Estimated from Car Manufacturer Data-An Improved Method. In *World Research & Innovation Convention on Engineering & Technology*. https://www.researchgate.net/publication/281020242.

Sharma, V., Sharma, M., Pandita, S., Kour, J., & Sharma, N. (2021). Application of



© University of Venda

geographic information system and remote sensing in heavy metal assessment. In *Heavy Metals in the Environment* (pp. 191–204). Elsevier.

Shen, G., Xue, M., Wei, S., Chen, Y., Zhao, Q., Li, B., Wu, H., & Tao, S. (2013). Influence of fuel moisture, charge size, feeding rate and air ventilation conditions on the emissions of PM, OC, EC, parent PAHs, and their derivatives from residential wood combustion. *Journal of Environmental Sciences (China)*, *25*(9), 1808–1816. https://doi.org/10.1016/S1001-0742(12)60258-7.

Sindhwani, R., & Goyal, P. (2014). Assessment of traffic-generated gaseous and particulate matter emissions and trends over Delhi (2000–2010). *Atmospheric Pollution Research*, *5*(3), 438–446.

Sindhwani, R., Goyal, P., Kumar, S., & Kumar, A. (2015). Anthropogenic emission inventory of criteria air pollutants of an urban agglomeration - National capital region (NCR), Delhi. *Aerosol and Air Quality Research*, *15*(4), 1681–1697. https://doi.org/10.4209/aaqr.2014.11.0271.

Singh, N. (2020). Air pollution exposure studies related to human health. *Air Pollution and Environmental Health*, 141–177.

Singh, R., Sharma, C., & Agrawal, M. (2017). Emission inventory of trace gases from road transport in India. *Transportation Research Part D: Transport and Environment*, *52*, 64–72.

Song, X., & Hao, Y. (2019). Vehicular emission inventory and reduction scenario analysis in the Yangtze River Delta, China. *International Journal of Environmental Research and Public Health*, *16*(23), 4790.

South Africa Weather Service (SAWS). https://www.weathersa.co.za/

StatsSA, Statistics South Africa (2011). Statistics South Africa. Formal Census, 2011. https://www.statssa.gov.za/

Stead, D. (1999). Relationships between transport emissions and travel patterns in Britain. *Transport Policy*, *6*(4), 247–258.

Thambiran, T., & Diab, R. D. (2011). Air pollution and climate change co-benefit opportunities in the road transportation sector in Durban, South Africa. Atmospheric Environment, 45(16), 2683–2689.

Thiba, M. C. (2019). *Evaluating the criteria for allocation of development projects in the context of spatial development frameworks in Thulamela Local Municipality*.

Tshikalanke, R. P. (2007). *Spatial and Temporal Variation in Domestic Biofuel Consumption rates in southern Africa*.

Uhunamure, S. E., Nethengwe, N. S., & Musyoki, A. (2017). Driving forces for fuelwood use in households in the Thulamela municipality, South Africa. *Journal of Energy in Southern Africa*, *28*(1), 25–34.

US EPA (1993). Emission Factor Documentation for Ap-42 Section 1.9, Residential Fireplaces, Ap 42, Fifth Edition, Volume I Chapter 1: External Combustion Sources, Air Emissions Factors and Quantification. Prepared by E.H. Pechan & Associates, Inc. Rancho Cordova, CA 95742. https://www.epa.gov/sites/default/files/2020-



© University of Venda

09/documents/background_document_ap42_section_1.9_residential_fireplaces.pdf.

US EPA (1996). AP 42, Fifth Edition, Volume I Chapter 1: External Combustion Sources. https://www3.epa.gov/ttn/chief/ap42/ch01/index.html.

U.S. EPA (1997). Exposure Factors Handbook (1997, Final Report). U.S. Environmental Protection Agency, Washington, DC, EPA/600/P-95/002F a-c, 1997. https://cfpub.epa.gov/ncea/efp/recordisplay.cfm?deid=12464.

US EPA (2006). Guidance for Quantifying and Using Emission Reductions from Voluntary Woodstove Changeout Programs in State Implementation Plans. https://nepis.epa.gov/Exe/ZyPDF.cgi/P1000N3M.PDF?Dockey=P1000N3M.PDF

Vhembe district municipality 2019/20 IDP REVIEW http://www.vhembe.gov.za/media/content/documents/2019/6/o_1dgf54agfs86u3u14ig1bvt1e1u1a.pdf?filename=VHEMBE%20DISTRICT%20MUNICIPALITY%20IDP%202019%20I%202020%20REVIEW.pdf. Accessed on 13/10/2020.

Wang, K. (2007). Long-range transport of the April 2001 dust clouds over the subtropical East Asia and the North Pacific and its impacts on ground-level air pollution: A Lagrangian simulation. *Journal of Geophysical Research: Atmospheres*, *112*(D9).

Win, Z. C., Mizoue, N., Ota, T., Kajisa, T., Yoshida, S., Oo, T. N., & Ma, H. (2018). Differences in consumption rates and patterns between firewood and charcoal: A case study in a rural area of Yedashe Township, Myanmar. *Biomass and Bioenergy*, *109*, 39–46.

Xu, B., You, X., Zhou, Y., Dai, C., Liu, Z., Huang, S., Luo, D., & Peng, H. (2020). The study of emission inventory on anthropogenic air pollutants and source apportionment of $PM_{2.5}$ in the changzhutan urban agglomeration, China. *Atmosphere*, *11*(7), 739. https://doi.org/10.3390/atmos11070739.

Zhao, D., Chen, H., Shao, H., & Sun, X. (2018). Vehicle Emission Factors for Particulate and Gaseous Pollutants in an Urban Tunnel in Xi'an, China. Journal of Chemistry, 2018. https://doi.org/10.1155/2018/896485.



© University of Venda

# CHAPTER FOUR

# EVALUATION OF DYLOS DC1700 LOW-COST SENSORS (LCS) FOR AMBIENT $PM_{2.5}$ AIR QUALITY MONITORING IN THE THULAMELA MUNICIPALITY, VHEMBE DISTRICT OF SOUTH AFRICA

## Abstract

*Global concerns over the health risks associated with ambient $PM_{2.5}$ pollution highlight the critical need for accurate monitoring for effective air quality management. Resource constraints, however, often hinder the availability of comprehensive monitoring systems, especially in peri-urban and rural regions, like the Vhembe District, where conventional methods are costly and spatially limited. There is, therefore, an urgent need to evaluate the performance of low-cost sensors like the Dylos DC1700 to provide air quality data. This study assessed $PM_{2.5}$ levels using Dylos DC1700 LCSs at fourteen monitoring sites for one year, covering urban, suburban, and rural areas to capture spatial variability across the district. The study found a strong correlation between Dylos DC1700 and GRIMM Environmental Dust Monitor EDM180 for $PM_{2.5}$ measurements and a strong correlation coefficient of determination ($R^2$) value of 0.75, confirming the reliability and adequacy of Dylos for $PM_{2.5}$ measurements. The study also revealed significant spatial and temporal variations in $PM_{2.5}$ concentrations, with higher levels observed near anthropogenic sources such as vehicle emissions, residential wood combustion, and unpaved roads. Annual $PM_{2.5}$ concentrations exceeded WHO guidelines and South Africa's NAAQS, except for one site that fell below the NAAQS standard. Monthly data exhibited a seasonal pattern, with rural areas primarily affected by domestic fuel burning and urban areas influenced by road traffic and residential fuel burning. Diurnal variations indicated peak concentrations during morning and evening hours from 6:00 to 8:00 and 17:00 to 20:00, suggesting potential exposure risks during those periods. Using a bivariate polar plot (BVP), the study identified hotspots of elevated $PM_{2.5}$ concentrations in densely populated areas near major intersections between residential and high-population density areas. Source identification highlighted significant contributions from vehicle emissions, traffic, and residential wood burning to poor air quality.*

**Keywords**: Ambient Air pollution (AAP), Particulate matter ($PM_{2.5}$), Dylos DC1700, Low-cost sensors (LCSs), Vhembe District, Bivariate polar plot (BVP)

## 4.1 Introduction

Ambient air pollution (AAP) is a significant environmental and health concern worldwide, affecting both developed and developing nations (WHO, 2016; Landrigan *et al*., 2018;



© University of Venda

Almetwally *et al*., 2020). According to the World Health Organization (WHO), AAP was responsible for 4.2 million premature deaths in 2016, with 91% occurring in low- and middle-income countries, particularly in Southeast Asia and the Western Pacific regions (WHO, 2016; Cohen *et al*., 2017; Landrigan *et al*., 2017). Urbanisation and industrialisation have led to the emission of air pollutants from various sources (Li *et al*., 2016; Khan *et al*., 2020). Particulate matter (PM) consists of solid or liquid particles suspended in the air, with distinct characteristics such as size, shape, and composition. Ultrafine particles (<0.1 μm), fine particulate matter ($PM_{2.5}$), and $PM_{10}$ encompass different size ranges, each with varying properties and impacts on human health (Kim *et al*., 2015; Seinfeld & Pandis, 2016; Almetwally *et al*., 2020; Williams *et al.,* 2021).

Notably, fine particulate matter ($PM_{2.5}$) is a major component of AAP. It is emitted from both natural and anthropogenic sources, contributing to global air quality decline, with local meteorological conditions influencing concentration levels. These fine particles pose substantial risks to human health since they penetrate deep into the lungs and bloodstream, leading to adverse health effects, such as - respiratory issues, cardiovascular disease, asthma exacerbations, reduced lung function, and increased mortality rates (Miller & Xu, 2018; Rajak & Chattopadhyay, 2019; Wang *et al*., 2019). Long-term exposure to $PM_{2.5}$ has also been linked to lung cancer and neurological disorders (HEI, 2018; Landrigan *et al*., 2018; Vanker *et al*., 2023).

The current monitoring infrastructure in South Africa is limited, posing challenges for comprehensive air quality assessment and management, despite the well-known health risks linked to $PM_{2.5}$ pollution. Conventional monitoring systems that use fixed monitoring stations, although effective, are limited due to high costs and complex setups, lack of technical personal, resulting in spatial and temporal coverage gaps (Castell *et al*., 2017). Most equipment often faces challenges in deployment due to the high variability in air pollution distribution (Steinle, *et al*., 2015). To combat air pollution effectively, accurate measurement and monitoring of $PM_{2.5}$ concentrations are imperative, hence, there is a pressing need for alternative monitoring approaches to address these limitations and enhance understanding of $PM_{2.5}$ pollution in this region. Introducing low-cost sensors



© University of Venda

(LCSs), portable direct-reading instruments based on light scattering principles can enhance epidemiological studies and provide better insights into exposure assessment across various pollution levels (Hegde *et al*., 2020).

Numerous studies have investigated the efficacy of low-cost sensors, such as the Dylos DC1700, in assessing ambient pollutant concentrations (Vercellino *et al*., 2018; Hegde *et al*., 2020; Li *et al*., 2020; Mei *et al*., 2020). These sensors offer complementary capabilities to traditional monitoring methods, particularly in areas with limited infrastructure (Steinle, *et al*., 2015; Sousan *et al*., 2016; Han *et al*., 2017; Vercellino *et al*., 2018; Mei *et al*., 2020). The Dylos DC1700 is a portable, affordable, and reliable device for measuring particle number concentrations (PNC) and estimating particle mass concentrations (PMC) in controlled and general environments. It detects particles in two bins, small particles below 0.5 µm and large particles above 2.5 µm. This device has shown correlations with $PM_{2.5}$ mass concentrations determined by various Federal Equivalent Method (FEM) and real-time continuous optical instruments, such as Beta-attenuation monitor (BAM 1020), GRIMM, DustTrak, and SidePak (Klepeis *et al*., 2013; Williams *et al*., 2014; Dacunto *et al*., 2015; Gao *et al*., 2015; Semple *et al*., 2015; Sousan *et al*., 2016) but, relying solely on these instruments can be challenging, as all particulate matter sensors operate on similar principles (Liu *et al*., 2017).

Challenges such as calibration and data reliability are some of the limitations of this equipment that need to be addressed in order to ensure accurate measurements and reliable air quality data to improve environmental health. Calibration of Dylos DC1700 LCS using advanced monitoring equipment such as a GRIMM instrument, is crucial and highly needed. In addition to monitoring the air quality, it is essential to identify the potential air pollution sources for an effective management and improvement of environmental health through techniques like bivariate polar plots (BVP) that use statistical measurements to pinpoint or identify pollutants' origins and sources from multiple locations (Belelie *et al*., 2019; Kanchanasuta *et al*., 2020; Sooktawee *et al*., 2020a). Understanding and monitoring of these sources are essential for comprehensive air quality management strategies.



© University of Venda

South Africa, with its diverse polluting sources, including vehicular emissions, domestic wood combustion, and industrial activities, faces spatial and temporal variations of air pollutants such as $PM_{2.5}$, underscoring the importance of effective monitoring and management strategies (Wright *et al*., 2014; Arowosegbe *et al*., 2021). Previous research in South Africa has predominantly focused on assessing air quality in areas with high industrial activity and power plants, but has overlooked comprehensive monitoring of $PM_{2.5}$ pollutant from other sources, such as vehicular traffic and residential wood combustion, especially within areas like the Vhembe District in Limpopo Province of South Africa.

This study, hence evaluates the effectiveness of low-cost sensors in monitoring ambient $PM_{2.5}$ levels across diverse environmental settings, including high-traffic road intersections and residential areas of Thulamela Local Municipality (LM), Vhembe District in Limpopo Province. Furthermore, it assesses the usefulness of Dylos DC1700 in monitoring ambient $PM_{2.5}$ air quality in the Thulamela Municipality environment. Specifically, it seeks to compare and validate the efficacy of this LCS instrument (Dylos DC1700) with standard instruments, the Environmental Dust Monitor (GRIMM EDM180), to measure ambient $PM_{2.5}$ concentration levels in the environment. Addressing this gap is crucial for providing valuable insights into cost-effective monitoring solutions that could enhance regional air quality management efforts, protect public health, and inform policy decisions aimed at reducing air pollution.

## 4.2 Materials and Methods

### 4.2.1 Sampling Site

Fourteen sampling locations at residential and vehicular sectors within Thulamela Municipality, Vhembe District were selected using a purposive and snowball method (Figure 4.1 and as indicated in Appendix 4.1). The monitoring sites were strategically distributed at various spatial levels, including inner city, sub-city, and rural areas, covering two main types of anthropogenic activities associated with dense populations in the area. Specifically, nine locations were selected in vehicular/traffic and commercial areas in the



© University of Venda

study domain of Thulamela Municipality core–city: Thohoyandou, Shayandima, Makumbane, Mvhuzoluzo Mall, University of Venda, Phangami Mall, Makwarela, Thavhani Mall, and Sibasa. These sites are characterised as having much vehicular, commercial, and residential fuel-burning activities. In addition, five residential sites in Gondeni (RG) and Lunungwi village (RL), which are approximately 17 km outside the city, were selected due to their substantial wood consumption for residential activities (RWC). Based on the study's common activities, sources, and geographical areas, the fourteen sites were categorised into five groups (RG, RL, VA, VB, and VC) to facilitate the analysis, - (1) RG: RG1, RG2, and RG3, (2) RL: RL4, and RL5, (3) VA: VS1, VTV2, VP4, and VTS9, (4) VB: VMK3, VUG5, and VUI6 and (5) VC: VV7 and VMV8 (Appendix 4.1).

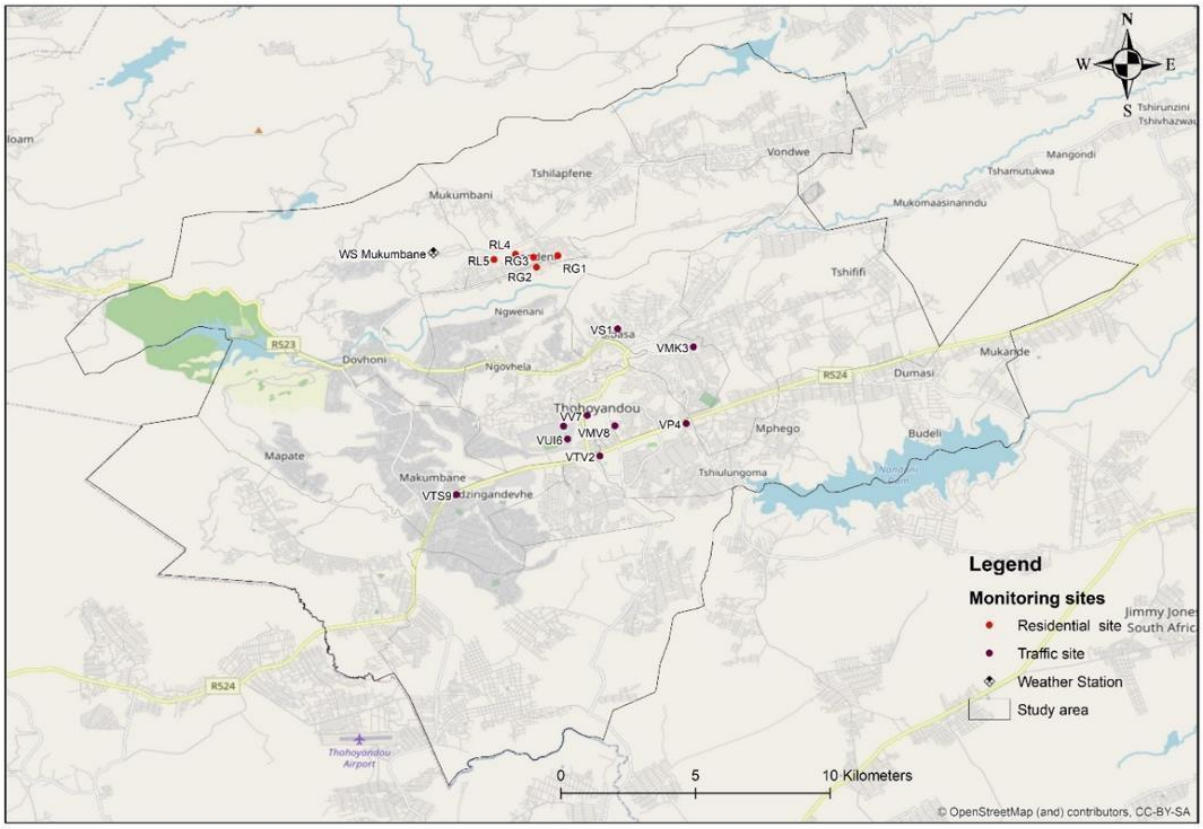



**Figure 4.1:** Map of the study area showing the monitoring sites.



© University of Venda

### 4.2.2 Experimental and Reference Instruments

#### 4.2.2.1 Dylos DC1700 particle counters

Dylos DC1700 (Dylos Corporation, Riverside, CA) particle counters for $PM_{2.5}$ Air Quality Monitor is a relatively low-cost laser particle counter device that counts particle numbers per volume of air with a 1.0 L/min flow rate, but does not measure mass concentrations of particles. The device records particle counts in two size bins. The upper range of the device is estimated to be close to 20 µm (Broday & Collaborators, 2017), but this has not been reliably quantified (Personal communication, Dylos Corporation, 2015). The Dylos DC1700 with a small fan cannot maintain constant airflow at the sampling inlet but sends an airstream to a laser detector located at approximately 90° to the airstream. The PNC for $PM_{2.5}$ is obtained from the Dylos DC1700 as recorded in the literature (Dacunto *et al*., 2015; Franken *et al*., 2019; Semple *et al*., 2013, 2015).

#### 4.2.2.2 Grimm 180 as Reference Instruments

Details on the GRIMM EDM180 model was elaborately discussed in subsection 2.5.1 GRIMM EDM180 model.

### 4.2.3 Modification of Dylos DC1700 to Measure Ambient Outdoor

The Dylos DC1700 specification did not meet all the requirements needed for an ambient outdoor real-time particle mass monitor, so modification was incorporated to accommodate a rechargeable battery at the Potchefstroom Campus (Potchefstroom), North-West University. Briefly, the equipment was modified by placing it in an aluminium case with the inlet protruding to the outside to allow air to be pulled into the device. To prevent airflow resistance, space was left around the bottom vent. Due to 6 hours battery life span which cannot be used for 24 hours, a 12V 12AH valve-regulated lead-Acid battery was connected to a solar panel through a five-star solar charger controller as shown in Figure 4.2.



© University of Venda

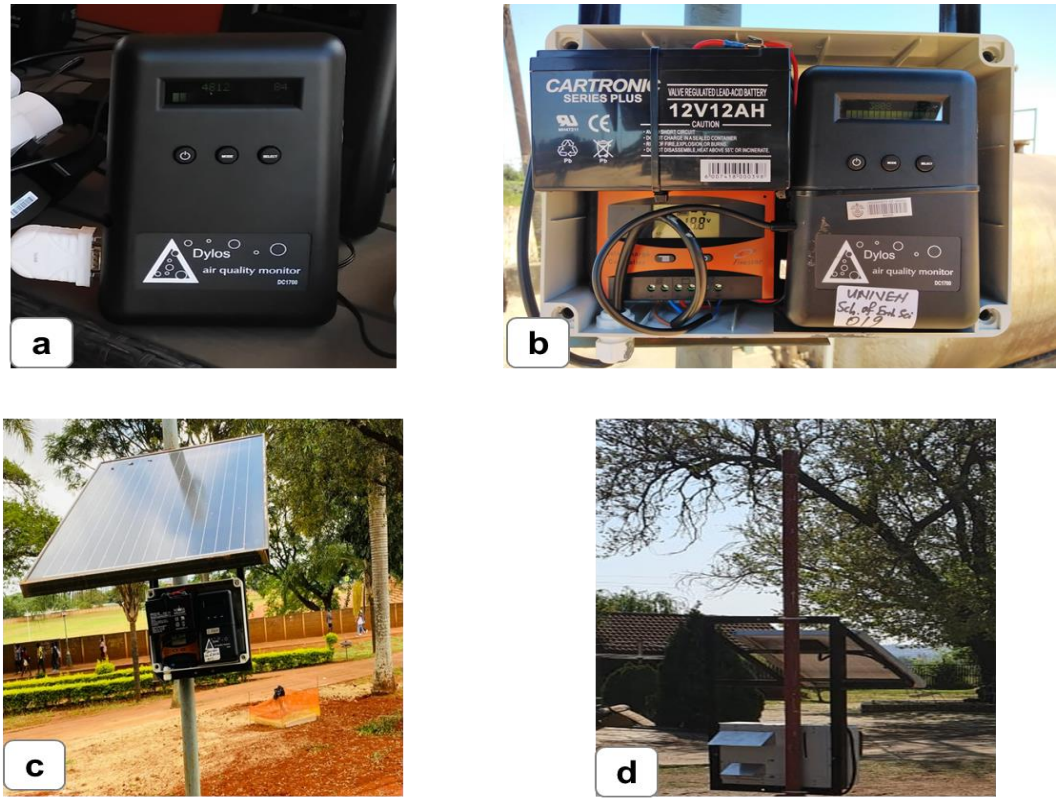



**Figure 4.2:** Modified Dylos DC1700 with solar panel.

### 4.2.4 Calibration procedure

In this study, Dylos DC1700 was calibrated using reference instruments known as GRIMM Environmental Dust Monitor 180 (EDM180) (German Ambient Dust Monitoring equipment). The monitoring station for the calibration was located on the Vanderbijlpark (Vaal) Campus of the North-West University, which is situated within the Vaal-Triangle Airshed Priority Area (VTAPA) - an ambient air pollution hotspot in central South Africa (Figure 4.3). The reference instrument (Figure 4.4a) and Dylos DC1700 were collocated and compared to determine their relationship. The reference instrument (GRIMM EDM180) continuously measures the concentrations of $PM_{10}$, $PM_{2.5}$, and $PM_1$, but only $PM_{2.5}$ measurements were considered in this research as Dylos DC1700 exclusively measures $PM_{2.5}$. Furthermore, data were collected from Dylos DC1700 and the GRIMM EDM180 for six months, from April 26 to October 29, 2021 (Wernecke *et al.*, 2021). A steel frame was constructed and designed so that the LCS inlets would sample at the same height as the reference instruments during the LCS-SA campaign, as shown in Figure 4.4b. The LCS was mounted on a pole about 4 m above the ground level.



© University of Venda

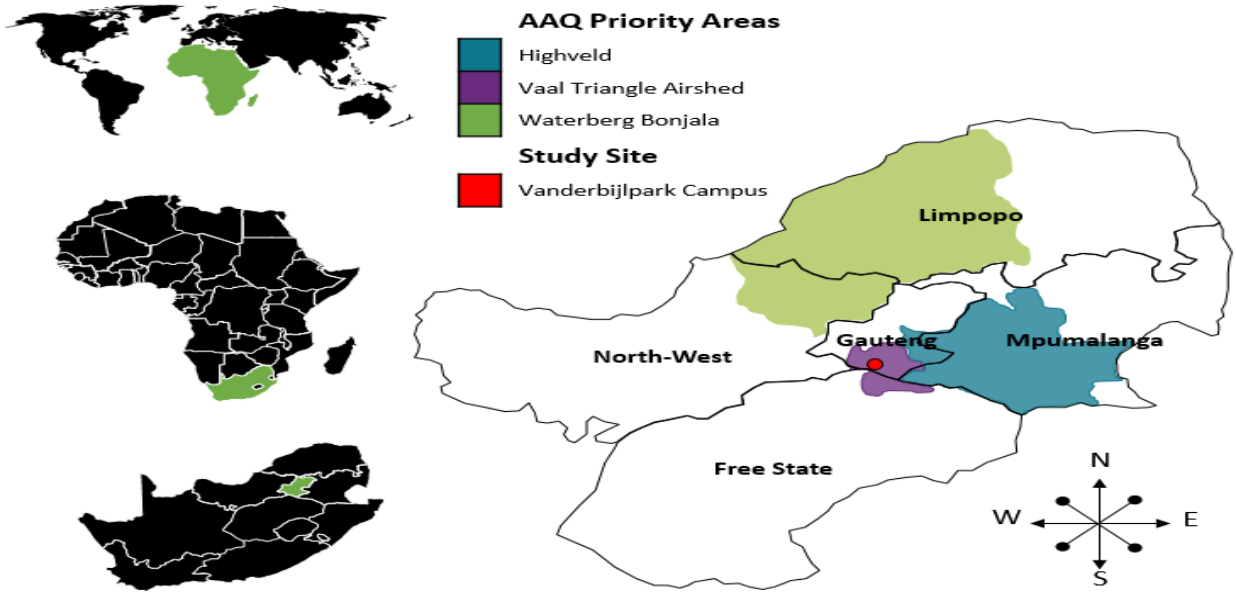



**Figure 4.3:** Monitoring site for calibration (Source: Wernecke et al., 2021).

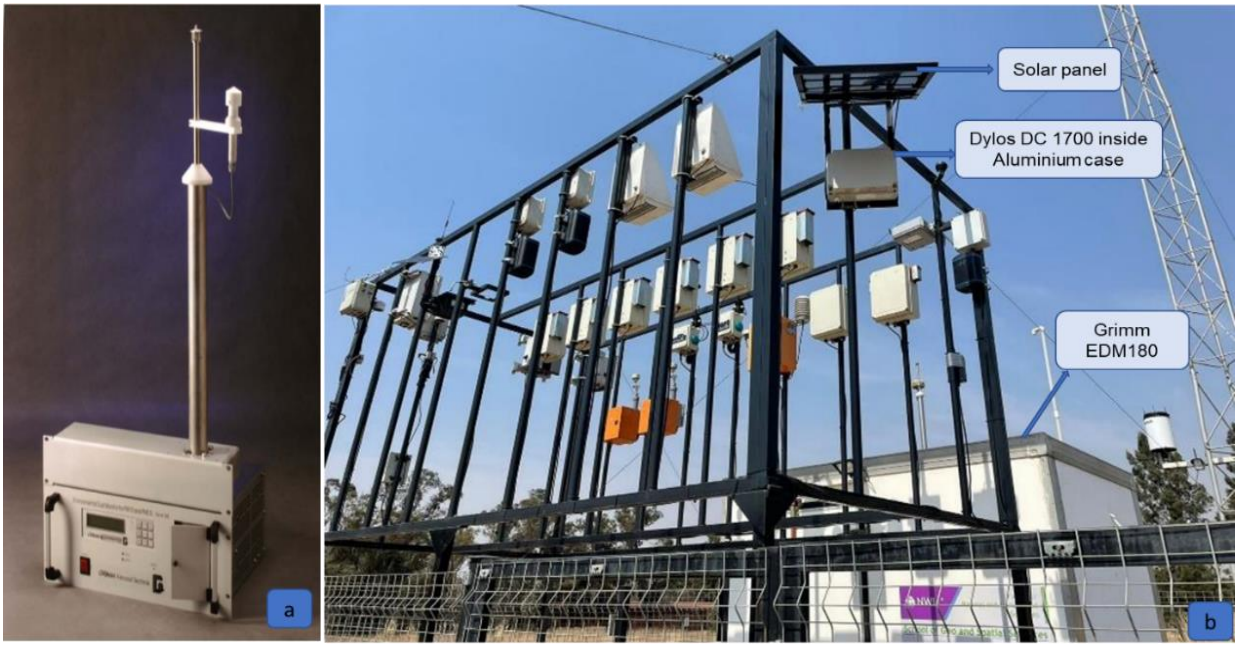



**Figure 4.4:** Diagram showing (a) GRIMM EDM180 (Source: GRIMM EDM180 user manual) and (b) Dylos DC1700 and the reference instrument in the same location.



© University of Venda

### 4.2.5 Data Collection

The modified Dylos DC1700 systems were installed to continuously monitor and measure the ambient $PM_{2.5}$ levels in the selected fourteen sampling locations (Figure 4.1) using a purposive and snowball method (Jackson-Morris *et al*., 2016). The Ambient Air Quality (AAQ) monitoring of $PM_{2.5}$ was carried out for a period of 12 months between March 2021 and February 2022 to accommodate all four seasons (summer, autumn, winter, and spring) in the study areas. Additionally, meteorological data of the study area were obtained from the South African Weather Service (SAWS) for analysis with the $PM_{2.5}$ data.

#### 4.2.5.1 Data logging

Dylos Logger software was used to download recorded data on Dylos DC1700. Before the commencement of data collection, all the instruments were synchronised to their internal clocks and the logging interval for data collection was set at 1-minute intervals for the study period. The measured data were automatically stored in the instrument's internal memory. $PM_{2.5}$ air quality data were downloaded from all the instruments at different locations at least once a week using male and female serial port cable, dylos logger software and laptop as shown in Figure 4.5, and the instruments were zero calibrated once a month. The downloaded text file from Dylos DC1700 contains three variables: particle number counts per 0.01 $ft^3$ for small particles (particle size >0.5 µm), large particles (particle size >2.5 µm) and date/time. The Particle Number Concentration (PNC) for $PM_{2.5}$ was determined by subtracting the number count of the large-size bin from that of the small-size bin as stated in earlier research and shown in equation 4.1 (Semple *et al*., 2013, 2015; Dacunto *et al*., 2015; Franken *et al*., 2019). The PNC was later converted to Particle Mass Concentration (PMC) in $PM_{2.5}$ (µg/$m^3$).

$$PNC_{PM_{2.5}} = PNC_{small} - PNC_{large} \quad (4.1)$$



© University of Venda




**Figure 4.5**: Data collection process; a) opening of the equipment using screwdriver, and b) downloading the data using Male and female cable, Dylos logger software and laptop.

### 4.2.6 Quality Control of Data

A quality control check was performed to identify any missing data. The 1-hour readings from Dylos DC1700 units were averaged from 1-minute measurements observed at each site without replacing the missing data. Typically, a dataset must achieve a minimum data capture rate of 80% for quality control and data processing (SANAS, 2000), but this was adjusted and lowered to account for the capabilities of the low-cost instruments. Additionally, we kept hourly readings data with a completeness rate of at least 40% and utilised a 70% data capture rate to calculate monthly average data, as shown in Table 4.1 for each location. Data gaps were attributed to factors such as the equipment's limited battery life, insufficient power supply from the solar panel, and unfavourable weather conditions, including low temperatures and rain. The time series plot for all the sites are displayed in Appendix 4.2.



© University of Venda

**Table 4.1:** Data completeness (%) per month at each $PM_{2.5}$ monitoring site from March 2021–February 2022. Sites with 70% and higher are bolded

| Monitoring sites | 2021 | | | | | | | | | | 2022 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | March | April | May | June | July | Aug | Sept | Oct | Nov | Dec | Jan | Feb |
| RG1 | **83.2** | **96.1** | **89.5** | **84.4** | **88.4** | 25.7 | 31 | 16.4 | 27.5 | 27.8 | 10.6 | 24.6 |
| RG2 | **83.2** | **100** | 41.3 | 41.8 | 12.4 | 9.8 | 5.0 | 2.2 | 0 | **92.7** | **96.2** | **95.4** |
| RG3 | **100** | **100** | **89.2** | **84.3** | **93.4** | 66.7 | 68.5 | 37.4 | 54.0 | 51.9 | **96.2** | **96.0** |
| RL4 | **83.1** | **100** | **89.0** | 56.8 | 60.1 | 28.9 | 10.1 | 7.7 | 0 | **92.9** | **96.5** | **95.5** |
| RL5 | **82.9** | **99.6** | **88.8** | **91.5** | **91.8** | 2.6 | **71.9** | 48.7 | 63.8 | 61.7 | **96.2** | **95.7** |
| VS1 | **87.4** | **100** | **84** | **93.3** | 62.2 | 30.2 | 53.9 | 58.2 | **81** | 32.1 | 25.8 | **86.9** |
| VTV2 | **91.8** | **100** | **100** | **100** | **94.2** | **91.4** | **87.5** | 21.2 | 0 | 0 | 0 | 0 |
| VMK3 | **100** | **100** | **100** | **100** | **88.6** | **76.6** | **85.7** | 58.9 | 69.6 | 66.7 | **93.4** | **86.9** |
| VP4 | **89.9** | **100** | **100** | **100** | 67.7 | 54.8 | 46.4 | 23.4 | 18.6 | **98.3** | **100** | 52.8 |
| VUG5 | 86.3 | 100 | 77.0 | 59.3 | 36.2 | 11.3 | 9.4 | 10.2 | 12.6 | 98.8 | 100 | 97.6 |
| VUI6 | **88.7** | **100** | **71.4** | 19.6 | 9.5 | 3.6 | 0 | 20.8 | **80** | 38.7 | 19.8 | 19.6 |
| VV7 | **91.1** | **100** | 19.8 | 29.0 | 29.3 | 27.3 | 56.5 | 61.8 | **79.7** | 41.0 | **100** | **100** |
| VMV8 | **86.6** | **100** | **83.3** | 68.5 | 41.0 | 66.0 | 9.0 | 5.6 | 29.4 | **92.2** | **82.7** | **100** |
| VTS9 | **91.3** | **100** | **100** | **100** | 67.9 | **75.9** | **83.5** | 52.4 | 53.9 | **99.5** | **100** | **100** |

### 4.2.7 Conversion of PNC to PMC and Statistical Analysis

For the conversion of particle number count (PNC) to particle mass concentration (PMC), a descriptive statistical analysis was performed first to study the mean PNC with standard deviation (SD) for Dylos DC1700 data and the mean concentration with standard deviation (SD) for GRIMM EDM180. The obtained PNC was later converted to the corresponding PMC ($\mu g/m^3$) using the correction factor derived from a simple linear regression model during calibration. This model assessed the relationship between Dylos DC1700 and GRIMM EDM180 values; although the data from the reference instrument should be plotted on the independent variable (X-axis) and Dylos DC1700 data on the dependent variable (Y-axis), the measured $PM_{2.5}$ mass concentrations from the GRIMM EDM180 were used as the dependent variables (Y-axis), while PNC from the Dylos DC1700 served as the independent variable (X-axis) to simplify comparisons of our results with previous studies (Klepeis *et al*., 2013; Semple *et al*., 2013, 2015; Dacunto *et al*., 2015). The average $PM_{2.5}$ data of hourly and daily interval readings obtained from the Dylos DC1700 device were analysed using statistical software such as Excel 2016,



© University of Venda

Xlstat, and the Openair package in R Studio 4.04 (Carslaw & Ropkins, 2012; Carslaw, 2019). The findings were compared to the National Ambient Air Quality Standards (NAAQS) and World Health Organisation (WHO) air quality guidelines to assess the study area's air quality status and their compliance with the regulation.

## 4.3 Results and Discussion

### 4.3.1 Assessment of Dylos DC1700 LCS for Measuring Ambient $PM_{2.5}$ level

Calibration of Dylos DC1700 with Reference Standards was done to assess the relationship between measurements obtained in Dylos PNC and GRIMM EDM180 using a linear regression analysis of the 10-minute averages data from the LCS and reference instrument, which were collocated for this purpose. The calibration process established a robust calibration equation or correction factor (CF) as given in equation 4.2, thus, minimising the measurement uncertainty and enabling accurate conversion of Dylos PNC measurements into corresponding $PM_{2.5}$ measurements, as recorded by the GRIMM. This calibration investigated the high correlation between the PNC bin and GRIMM measurements of $PM_{2.5}$. The equation gave the relationship between the two equipment, which played a pivotal role in converting Dylos PNC into PMC, thereby providing a means to obtain accurate and reliable measurements of particulate matter.

$$PMC_{PM_{2.5}(\mu g/m^3)} = 0.0059(PNC_{PM_{2.5}}) + 12.977 \qquad (4.2)$$

Based on the correlation analysis, the obtained coefficient of determination ($R^2$) value of 0.75 indicated a strong correlation between Dylos PNC bins and GRIMM measurements (Figure 4.6). This correlation analysis served as an essential preliminary step, further affirming the reliability of Dylos PM measurements and the adequacy of the Dylos sensor calibration.



© University of Venda

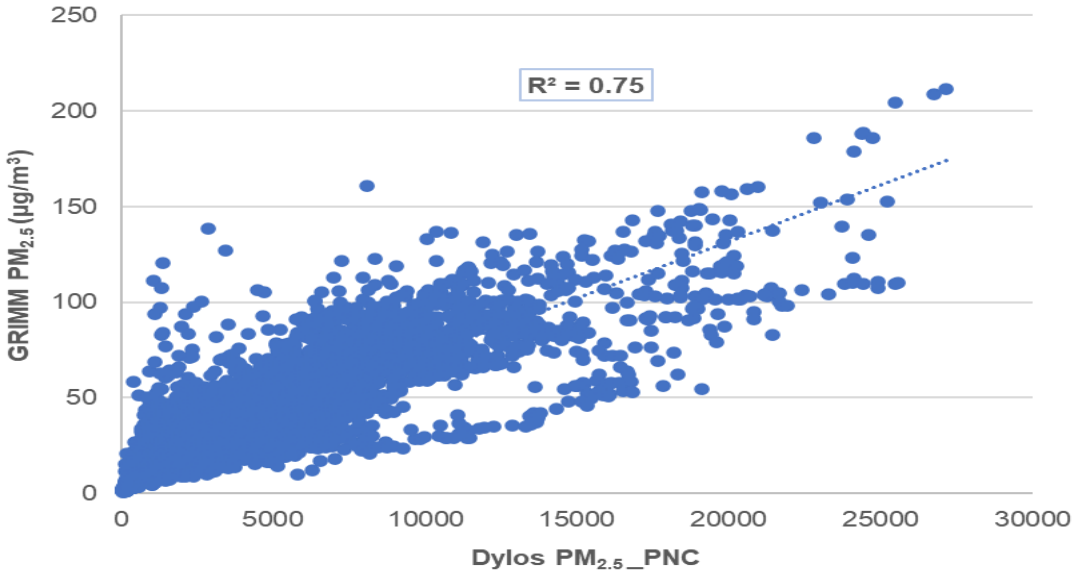



**Figure 4.6:** Scatterplots showing the relationship between GRIMM EDM180 and Dylos DC1700 particle number count (PNC) with the corresponding $R^2$ value.

#### 4.3.1.1 Performance evaluation of Dylos DC1700 against the reference instrument (GRIMM EDM180)

The $PM_{2.5}$ calculated for Dylos DC1700 using the calibration equation is referred to as the "estimated" or "corrected" $PM_{2.5}$ (PMC) concentrations (µg/m$^3$). The $PM_{2.5}$ concentration values obtained from both GRIMM reference instruments and Dylos DC1700 at ambient environments were compared using the 10-minute readings. The $PM_{2.5}$ readings showed that the mass concentration ranged from 13.321-173.148 µg/m$^3$ with a mean and standard deviation of 32.647 ± 22.087 µg/m$^3$. A subsequent linear regression analysis showed a positive correlation between the average $PM_{2.5}$ of Dylos and GRIMM, with a coefficient of determination ($R^2$) of 0.751 after converting from PNC to PMC (see Figure 4.7). This value aligns with the $R^2$ values reported in previous studies, as shown in Table 4.2, where Dylos DC1700 was calibrated with the GRIMM reference instrument. Additionally, Figure 4.8 illustrates that Dylos performed better with a similar trend to GRIMM values. The strong correlation and close graphical fit between Dylos and GRIMM also suggest that Dylos can be a reliable tool for estimating ambient $PM_{2.5}$ levels. Furthermore, the consistent demonstration of similar trends and high $R^2$ across multiple studies indicates that Dylos offers a dependable and cost-effective alternative for estimating ambient $PM_{2.5}$ levels (AQ-SPEC; Williams *et al.*, 2014; Jovašević-Stojanović



© University of Venda

*et al*., 2015; Manikonda *et al*., 2016; Han *et al*., 2017; Vercellino *et al*., 2018; Hegde *et al*., 2020), benefiting researchers and policymakers through a calibration process that maintains accuracy while reducing costs. It was noticed, however, that Dylos tends to underestimate $PM_{2.5}$ mass concentrations during periods of elevated pollution levels, potentially affecting its precision in specific areas.

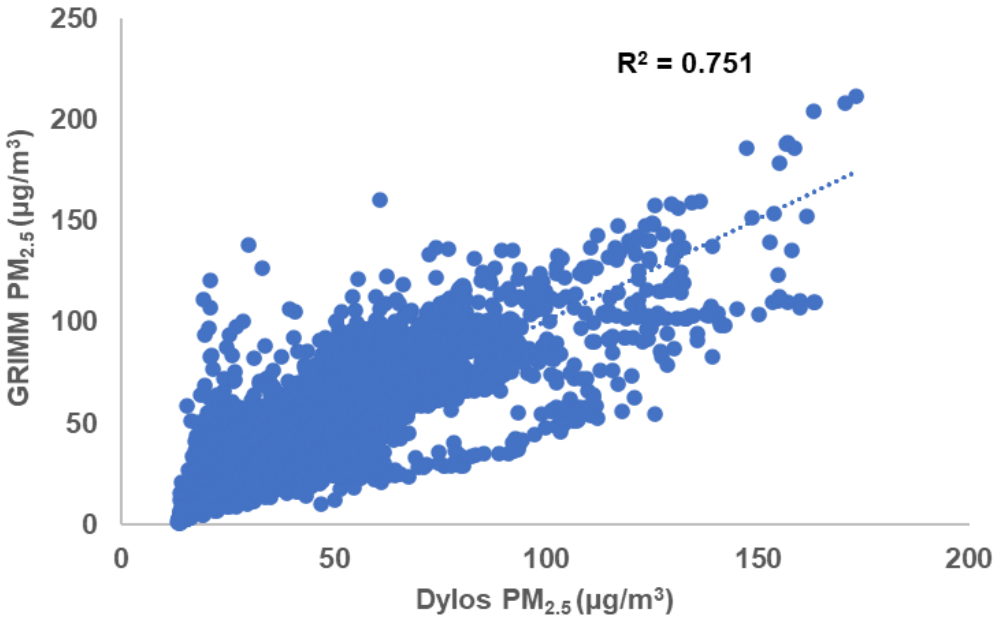



**Figure 4.7:** Scatterplots of concentration for GRIMM EDM180 and Dylos DC1700, with the corresponding $R^2$ value using the corrected $PM_{2.5}$ (PMC) concentrations (µg/m$^3$).



© University of Venda

**Table 4.2:** Calibration of Dylos DC1700 with reference instruments in previous studies

| Study | Instruments | Location | $R^2$ |
|---|---|---|---|
| Vercellino *et al*. (2018) | Dylos DC1100 against Grimm 1.109 | United States (University of Utah, Utah) | 0.571 - 0.678 (Mass)<br>0.858 - 0.922 (Count) |
| Williams et al. (2014) | Dylos DC1100 against Grimm EDM180 | United States (Durham, North Carolina) | 0.548 |
| Manikonda *et al*. (2016) | Dylos DC1700 against aerodynamic particle sizer (APS) 3321 | United States (Clarkson University, New York) | 0.87 |
| Sousan *et al*. (2016) | Dylos DC1700 against aerodynamic particle sizer (APS) 3321 | United States (Lowa) | 0.98 |
| Han *et al*. (2017) | Dylos DC1700 against Grimm 11-R | United States (Houston Texas) | 0.778 |
| West *et al*. (2020) | Dylos DC1700 against Grimm EDM180 | Kenya (Nairobi) | 0.26 |
| Jovašević-Stojanović *et al*. (2015) | Dylos 1700 against Grimm monitor | China (Xi'an) | 0.917 |
| AQ-SPEC (Air Quality Sensor Performance Evaluation Center) | Dylos DC1700 against Grimm EDM180 | United States (Los Angeles County, California) | 0.67 - 0.78 |
| This study | Dylos DC1700 against Grimm EDM180 | Thulamela LM, Limpopo Province, South Africa | 0.75 (Count)<br>0.751 (Mass) |



© University of Venda

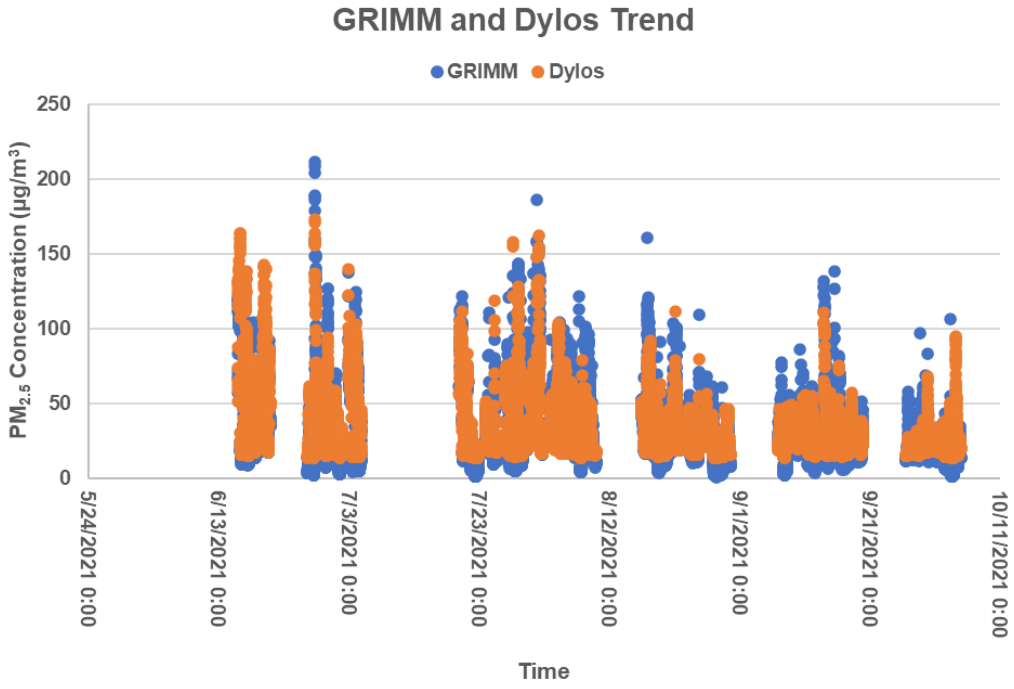



**Figure 4.8:** The time series of $PM_{2.5}$ concentrations (µg/m$^3$) obtained from Grimm EDM180 and Dylos DC1700.

### 4.3.2 Spatial Distribution of Annual Ambient $PM_{2.5}$ Concentrations in Thulamela Local Municipality (LM)

Table 4.3 shows the hourly statistical summary and annual concentrations of $PM_{2.5}$ at the fourteen sites, including the traffic and residential sites. These sites were categorised based on land use activities and distributed across various parts of the Municipality, (as outlined in Appendix 4.1). The highest annual concentration was observed at site VTV2 in Thavhani Mall (32.3 µg/m³), while the lowest was recorded at site RL4 in Lunungwi Village (19.1 µg/m³) (Appendix 4.1). The average $PM_{2.5}$ concentrations at all sites exceeded both the World Health Organization's (WHO) guideline for annual air quality (5 µg/m³) (WHO, 2021) and South Africa's National Ambient Air Quality Standard (NAAQS) of 20 µg/m³ (DEAT, 2009; DEA, 2012), except for site RL4 (19.1 µg/m³), that is below the NAAQS guideline (Table 4.3 and Figure 4.9). This means that residents in the study area, except those near site RL4, are at a higher risk of experiencing health issues because of exposure to PM, particularly as the levels exceeded NAAQS and WHO standards. Prolonged exposure to high levels of PM has been associated with severe health risks,



© University of Venda

including respiratory problems, cardiovascular diseases, and even premature death, as fine PM can deeply penetrate the lungs and enter the bloodstream (Gupta *et al*., 2016; Yunesian *et al*., 2019; Nawaz & Henze, 2020).

As seen in the results, the spatial distribution of $PM_{2.5}$ concentrations across the fourteen monitoring sites displayed significant variations. The urban areas exhibited elevated levels which could be attributed to vehicular emissions, traffic, and residential wood-burning. Importantly, hotspots were identified near major traffic intersections, residential areas, and zones. These places are known as high population density areas within the Thulamela LM, especially in central locations like Thohoyandou, Sibasa, Tshilungoma, along the Makumbane-Tshilidzini Hospital Road, and the Punda Maria main road (see Figure 4.10).

**Table 4.3:** Summary statistics of Ambient $PM_{2.5}$ concentrations in Thulamela LM

| Site | N(hour) | Min | Max | 1st Quartile | Median | 3rd Quartile | Mean | SD |
|---|---|---|---|---|---|---|---|---|
| **RG1** | 3242 | 13.00 | 86.09 | 17.67 | 20.37 | 25.54 | 23.01 | 8.09 |
| **RG2** | 3386 | 12.92 | 151.99 | 16.87 | 18.74 | 22.11 | 21.65 | 12.31 |
| **RG3** | 4791 | 13.40 | 91.18 | 16.83 | 19.18 | 24.24 | 22.21 | 8.41 |
| **RL4** | 4051 | 13.37 | 131.27 | 16.00 | 18.34 | 22.21 | 19.91 | 5.99 |
| **RL5** | 5214 | 13.48 | 104.15 | 17.12 | 19.35 | 23.53 | 21.35 | 6.49 |
| **VS1** | 3834 | 13.67 | 117.91 | 18.52 | 21.40 | 27.16 | 24.26 | 8.98 |
| **VTV2** | 4878 | 14.05 | 138.53 | 21.11 | 27.74 | 40.67 | 32.30 | 14.30 |
| **VMK3** | 6053 | 13.77 | 145.10 | 17.71 | 20.43 | 25.31 | 23.00 | 9.11 |
| **VP4** | 4328 | 13.40 | 122.70 | 17.28 | 20.61 | 30.61 | 25.81 | 12.47 |
| **VUG5** | 4070 | 12.98 | 91.52 | 17.16 | 19.34 | 23.39 | 21.42 | 7.10 |
| **VUI6** | 2487 | 13.99 | 78.22 | 17.98 | 21.06 | 26.70 | 23.87 | 8.52 |
| **VV7** | 3388 | 14.10 | 77.09 | 17.60 | 19.77 | 23.77 | 21.54 | 5.89 |
| **VMV8** | 3955 | 13.51 | 144.48 | 17.78 | 20.44 | 26.21 | 23.80 | 10.26 |
| **VTS9** | 6185 | 13.81 | 143.70 | 18.66 | 22.86 | 31.86 | 27.26 | 12.43 |



© University of Venda

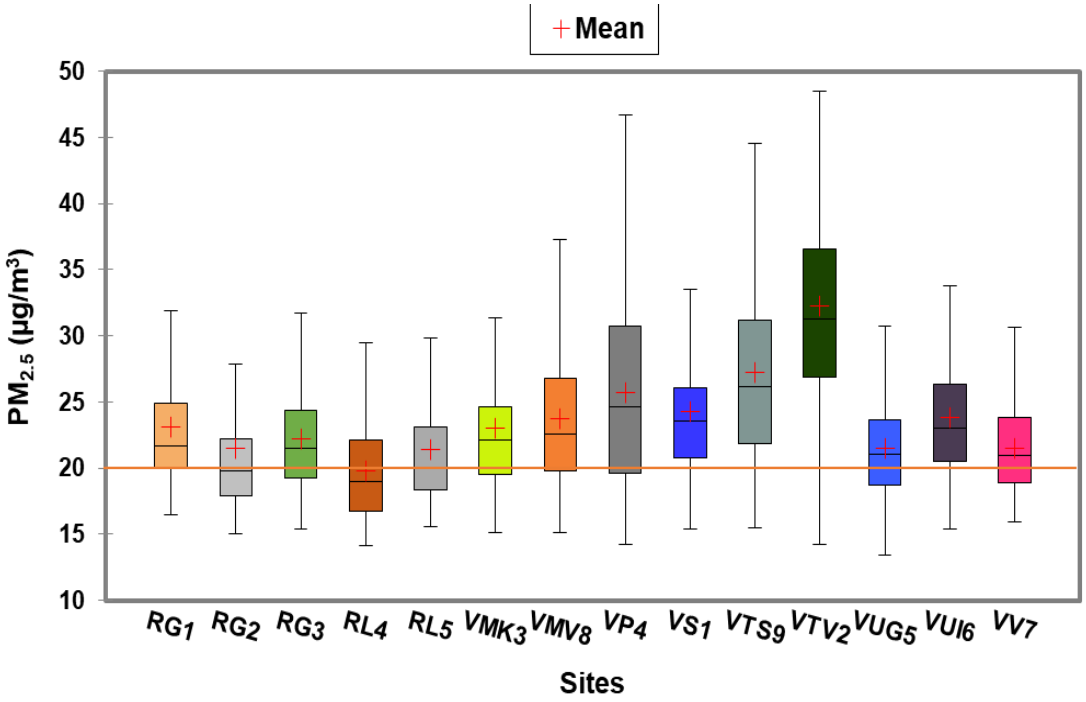



**Figure 4.9:** Annual average $PM_{2.5}$ concentrations at different sites in the study area during the monitoring period. The red line indicates the annual $PM_{2.5}$ threshold limit of 20 µg/m³ set by South Africa's NAAQS.

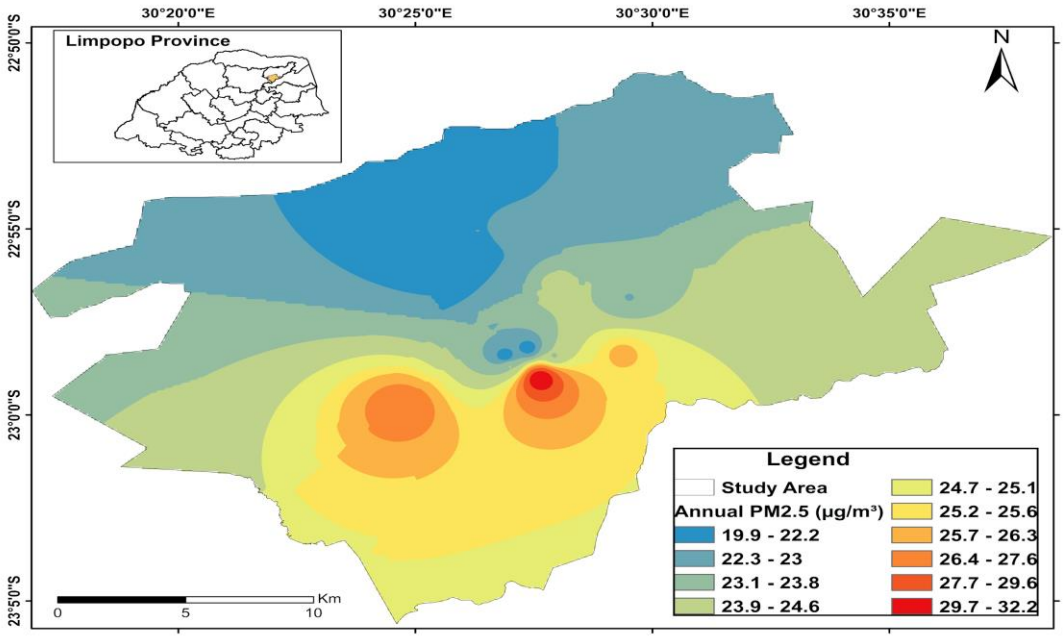



**Figure 4.10:** Spatial distribution map of annual average $PM_{2.5}$ concentrations of the study area during the monitoring period (March 2021 to February 2022).



© University of Venda

### 4.3.3 Seasonal Variability of $PM_{2.5}$ Concentrations

In the southern hemisphere, there are four seasons: summer (DJF - December, January, February), autumn (MAM - March, April, May), winter (JJA - June, July, August), and spring (SON - September, October, November). The $PM_{2.5}$ mass concentrations in the study area exhibited distinct seasonal variations, so during winter, the highest mean concentrations were observed in group VA (33.21 µg/m³), and the lowest of 18.84 µg/m³ was recorded during the summer in group RL (Figure 4.11, Appendix 4.3 and 4.4). This seasonal increase can be attributed to the presence of consistent inversion layers during the winter, which trap pollutants closer to the ground. Notably, although the average $PM_{2.5}$ concentration in the area with traffic and residential sources remained higher than in the residential area alone, both areas demonstrated similar seasonal trends. The observed seasonal variation is consistent with findings from previous studies conducted in Thohoyandou (Edlund *et al*., 2021; Howlett-Downing *et al*., 2023), and other areas (Fu *et al*., 2018; Muyemeki *et al*., 2021; Adesina *et al*., 2022; Wang *et al*., 2022).

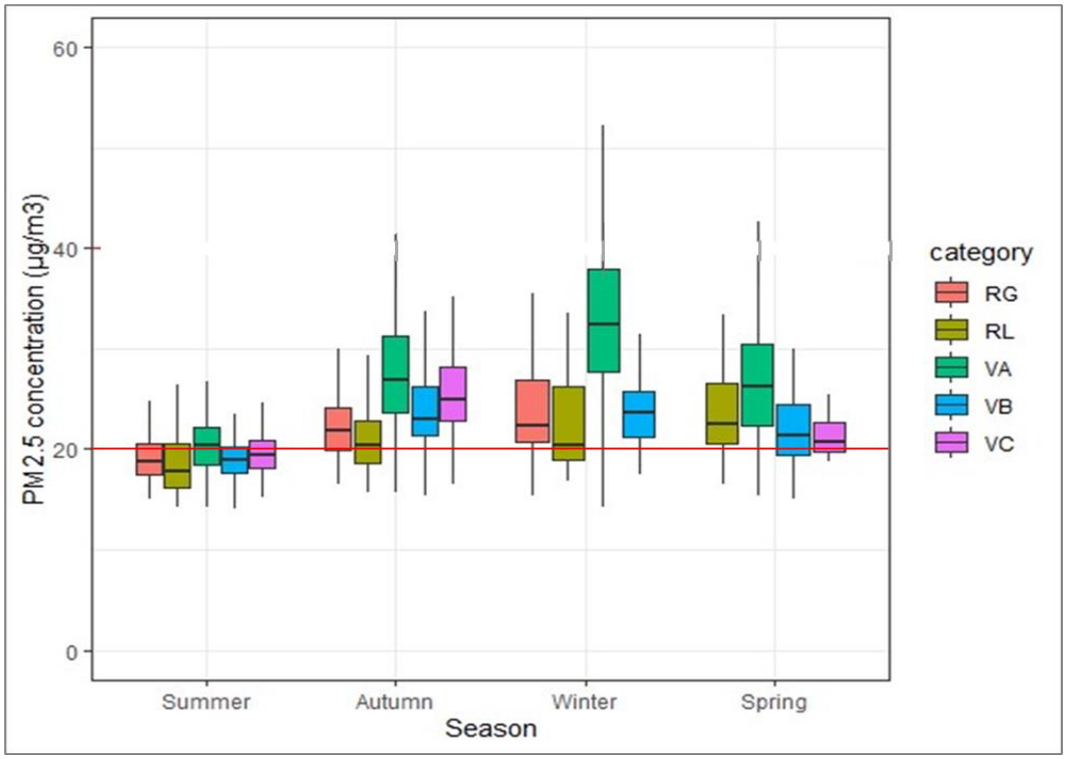



**Figure 4.11:** Seasonal mean concentrations of $PM_{2.5}$ (µg/m$^3$) at different groups across the monitoring area in Thulamela LM for the study period with red line indicating South Africa NAAQS annual $PM_{2.5}$ threshold limit of 20 µg/m³.



© University of Venda

### 4.3.4 Monthly Pattern of $PM_{2.5}$ Pollutant in the Study Domain

Regarding monthly mean $PM_{2.5}$ concentrations across all regions and per site (Figure 4.12 and Appendix 4.5), $PM_{2.5}$ levels displayed a seasonal pattern. The lowest monthly mean $PM_{2.5}$ concentration was recorded during the summer months, ranging from 15.9 µg/m³ at RL to 19.9 µg/m³ at VA. In contrast, the highest concentrations were observed from May through the winter months, ranging from 24.0 µg/m³ at RL to 35.2 µg/m³ at VA. The $PM_{2.5}$ levels at RG and RL shared familiar sources, mainly stemming from domestic fuel burning in rural areas. In contrast, $PM_{2.5}$ at VA, VB, and VC was primarily attributed to road traffic activities and residential fuel burning. The VA exhibited the highest levels (35.2 µg/m³) among the five areas, which might be due to its proximity to the major road and malls. Following this, VC, situated around the taxi rank, showed the next highest levels (29.23 µg/m³), and lastly, VB, located in the University and commercial area, had comparatively lower $PM_{2.5}$ concentrations (27.38 µg/m³). $PM_{2.5}$ levels in RG and RL were primarily due to domestic fuel burning in rural areas and this greatly influenced the RG air quality during the December holiday. The air quality at VA, VB, and VC was primarily impacted by the road traffic activities and residential fuel burning in the urban areas. Interestingly, throughout the monitoring period, the average monthly $PM_{2.5}$ levels at the five monitoring groups consistently remained below the daily $PM_{2.5}$ threshold limit of 40 µg/m³ set by South Africa's NAAQS but exceeded the WHO's daily air quality guideline, set at 15 µg/m³ (WHO, 2021).

The observed monthly variation in $PM_{2.5}$ concentrations across the area, with higher levels during winter months, aligns with findings from previous studies (Scorgie, 2012; Kanchanasuta *et al*., 2020; Moletsane *et al*., 2021; Adeyemi *et al*., 2022). This pattern correlates with increased solid biomass use for energy and lower temperatures during the cold season, as reported in other South African studies (Adesina *et al*., 2022). The results from this study suggest that reliance on wood for cooking and heating, as well as increased vehicle activities in Thulamela LM, South Africa, significantly contribute to local $PM_{2.5}$ sources. These sources influence the overall variability of $PM_{2.5}$ levels within the study area and could potentially extend to other parts of the country. The findings of this study suggest that reliance on wood for cooking and heating, along with increased vehicle



© University of Venda

activity in Thulamela LM, South Africa, significantly contributes to local $PM_{2.5}$ emissions. These sources influence the overall variability of $PM_{2.5}$ levels within the study area but may extend their impact to other parts of the country. This study, hence, emphasises the importance of comprehending and addressing both vehicle emissions and domestic sources in mitigating $PM_{2.5}$ pollution, not only in Thulamela LM but South Africa as a whole (Enitan *et al*. 2022).

The increased $PM_{2.5}$ concentrations in winter months underscore the influence of meteorological conditions on particulate matter concentrations and low ones in summer. This phenomenon is associated with a higher frequency of atmospheric inversions and reduced wind speeds in winter, which hinder the dispersion of particulate matter. Fine particle levels tend to rise during periods of stagnant air or when polluted air is transported into the area by winds, primarily due to the lower temperatures and decreased wind speeds observed in the colder months (Freiman & Piketh, 2003).

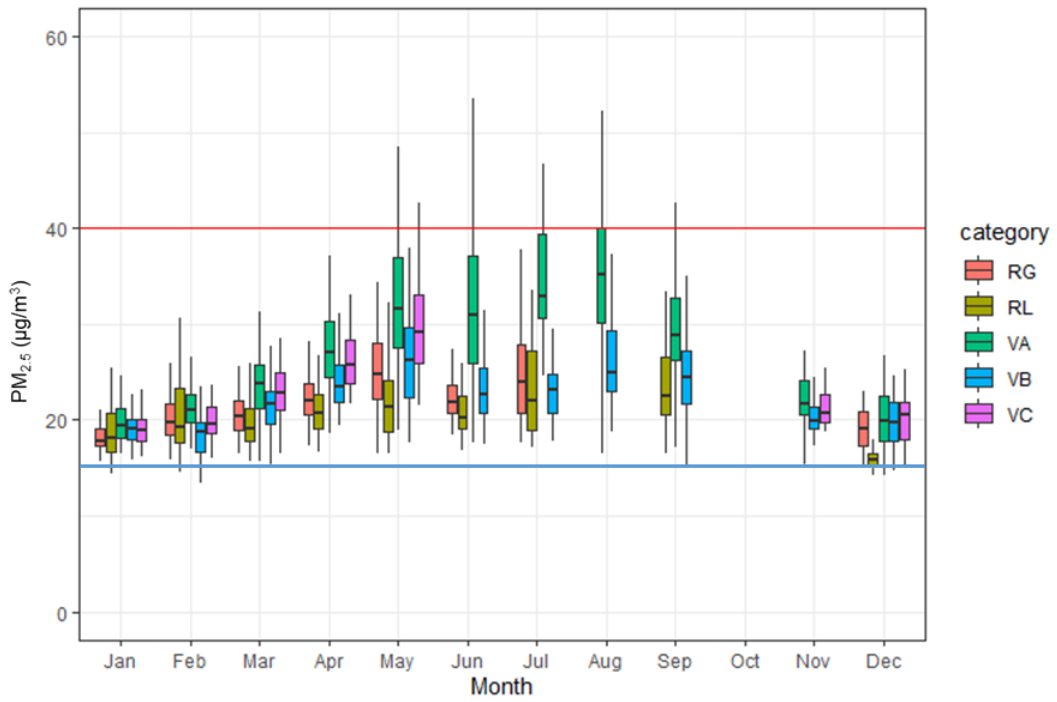



**Figure 4.12:** Monthly variations of $PM_{2.5}$ concentrations, with the bold line indicating the average, in the Thulamela LM during the study period, with the red and blue bold lines indicating the SA NAAQS and WHO daily standard, respectively.



© University of Venda

### 4.3.5 Diurnal and Weekday Pattern of $PM_{2.5}$ Pollutant Concentration

The hourly trend of average $PM_{2.5}$ concentrations ($\mu g/m^3$) across the five groups and all sites displayed bimodal peaks, occurring around 6:00 to 8:00 am and 17:00 to 20:00, with the evening peak surpassing the morning peak as shown in Figure 4.13 (Appendix 4.6). Notably, the evening peak, particularly at RG, showed a substantial increase around 18:00 compared to other locations which is attributed to heightened wood-burning activities in the evening. The results indicate low PM emissions during the late morning and afternoon before 18:00 hour when there are usually no residents at home to cook or do any wood-burning activities (Figures 4.13 and Appendix 4.6).

Additionally, these peak times align with periods of heightened traffic activity during morning and evening rush hours in commercial and residential areas (VA, VB, and VC groups), as well as the prevalent use of solid fuel for cooking and heating by residents in all areas. These factors significantly contribute to increased PM emissions during these periods. This pattern aligns with previous findings in low-income community and urban areas (Wernecke *et al*., 2015; Feig *et al*., 2016; Kanchanasuta *et al*., 2021; Adesina *et al*., 2022; Aniyikaiye *et al*., 2022; Awokola *et al*., 2022). The correlation between peak PM concentrations, domestic biomass burning, and vehicular congestion underscores the substantial contribution of these activities to air pollution in the community. Diurnal $PM_{2.5}$ concentration fluctuations affect air pollution exposure, as overnight concentrations significantly surpassed daytime levels across all sites. Elevated nighttime $PM_{2.5}$ levels may result from temperature inversion or varied source contributions.

The recorded weekdays' mean $PM_{2.5}$ concentration across all sites, within each group during the study period are illustrated in Figure 4.14. Minor variations were observed between different days of the week, with weekday $PM_{2.5}$ concentrations generally slightly higher than those on weekends, across all five groups. Specifically, the highest to lowest weekday $PM_{2.5}$ concentrations were observed in the order of VA > VC > VB > RG > RL. This concentration trend exhibited a notable decrease over the weekends. As shown in Figure 4.14, the daily mean $PM_{2.5}$ level recorded during weekdays varies across all groups, with exceptionally high values recorded in group VA on weekdays and weekends.



© University of Venda

This phenomenon could be attributed to factors such as high traffic volume, proximity to roads and malls, population density, and other human activities occurring in the vicinity that could contribute to these elevated levels.

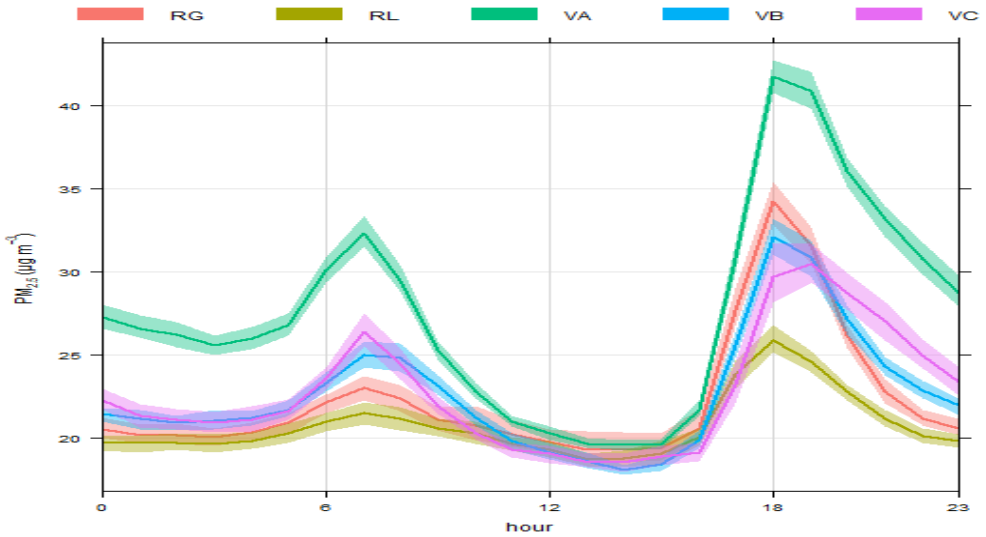



**Figure 4.13:** Diurnal variations of $PM_{2.5}$ concentrations, with the bold line indicating the average levels in the study area from March 2021 to February 2022.

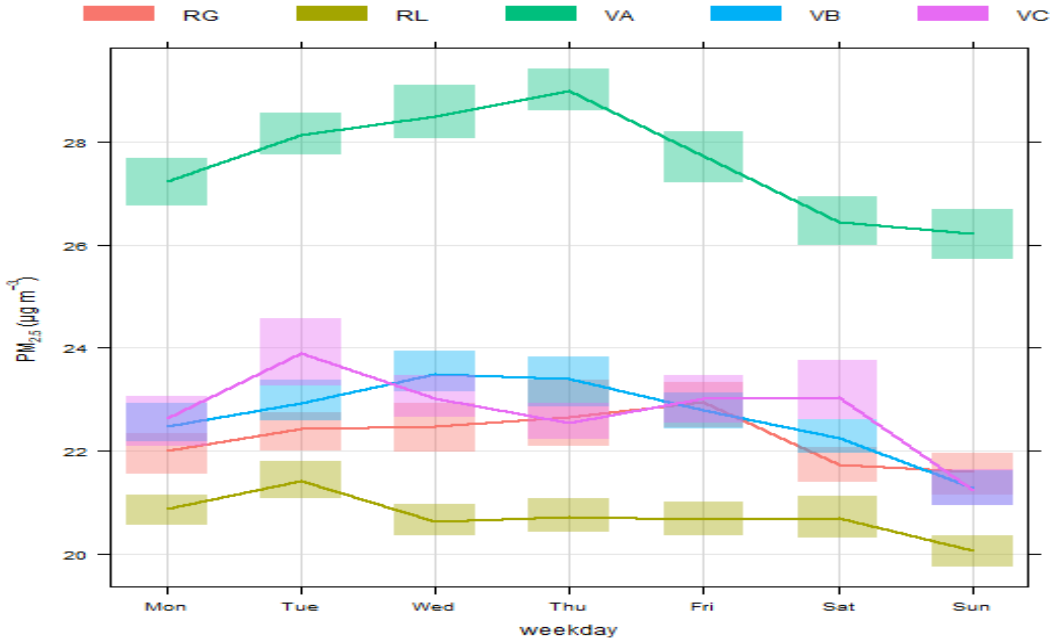



**Figure 4.14:** Weekday variations in the daily mean of $PM_{2.5}$ concentrations across the study area in Thulamela LM.



© University of Venda

### 4.3.6. Pollutants Exceedances

Throughout the study period, the 24-hour South African NAAQS of 40 µg/m³ pollutants standard limit was exceeded at ten sites (71% of all the sites), while the 24-hour WHO standard limit of 15 µg/m³ was exceeded at all the sites for at least 110 days (Table 4.4). The findings align with previous research by Adeyemi *et al*. (2022) at the University of Pretoria, Adesina *et al*. (2022) in Kwazamokuhle community, and Aniyikaiye *et al*. (2023) around brickmaking areas in Thohoyandou, where average daily $PM_{2.5}$ values exceeded both standards in almost all locations. This indicates that ambient $PM_{2.5}$ exposure may increase and negatively impact the health of the residents in the study area, however, the recorded mean $PM_{2.5}$ value in Thulamela LM was higher than the previously reported value for Thohoyandou's suburban study areas (Novela *et al*., 2020). Likewise, our research findings slightly differ from previous studies in other areas within South Africa (Feig *et al*., 2016; Morakinyo *et al*., 2019; Olutola *et al*., 2019; Williams *et al*., 2021), that reported exceeded WHO standard limit for the average daily $PM_{2.5}$ values but compliance with the 24-hour South African NAAQS standard limit.

**Table 4. 4:** Daily $PM_{2.5}$ pollutant exceedances of the NAAQS and WHO air quality standards

| | **$PM_{2.5}$ Pollutant (µg/m³)** | |
|---|---|---|
| | NAAQS 24hrs Standard (Allowed) | WHO 24hrs guidelines |
| **Sites** | 40 (4) | 15 |
| | **Exceedance** | |
| RG1 | 2 | 142 |
| RG2 | 3 | 147 |
| RG3 | 1 | 206 |
| RL4 | 1 | 166 |
| RL5 | 0 | 228 |
| VS1 | 1 | 169 |
| VTV2 | 32 | 208 |
| VMK3 | 2 | 258 |
| VP4 | 7 | 180 |
| VUG5 | 0 | 166 |
| VUI6 | 0 | 110 |
| VV7 | 0 | 144 |
| VMV8 | 2 | 170 |
| VTS9 | 17 | 263 |



© University of Venda

### 4.3.7. Assessment of Inter-Unit Variability of Dylos DC1700 Sensor

Inter-unit variability, also known as reproducibility, refers to the consistency of data produced by different units of the same LCS model. This study evaluated reproducibility using daily data collected from fourteen units of the Dylos DC1700 sensor model. The analysis revealed a moderate to strong positive correlation among the daily average $PM_{2.5}$ concentrations recorded by all sensors, with correlation coefficient values ranging from 0.51 and 0.93. This indicates a synchronous increase in $PM_{2.5}$ levels across sites, suggesting that when $PM_{2.5}$ levels increase at one site, they tend to increase at other sites as well (Figure 4.15). Notably, exceptions were for RL4, exhibiting $R^2$ values ranging from 0.18 to 0.85 compared to other units. These deviations are likely attributed to data inaccuracies stemming from communication issues between the sensors and the environment, alongside equipment battery issues. Such information is crucial for understanding spatial patterns and dependencies of $PM_{2.5}$ pollution and monitoring equipment across different locations. Field tests further demonstrated minimal inter-unit variability, affirming consistent responses among the Dylos units in measuring ambient $PM_{2.5}$ concentrations and their dependencies. These findings of reproducibility and consistency with sensors closely align with previous research on Dylos (AQ-SPEC; Steinle *et al*., 2015; Sousan *et al*., 2016; Hegde *et al*., 2020) and other LCSs (Mykhaylova, 2018; Caseiro *et al*., 2022; Hassani *et al*., 2023). The study demonstrates that the Dylos units provide reliable and consistent measurements of ambient $PM_{2.5}$ concentrations, supporting their use for air quality monitoring purposes. Furthermore, the study highlights their potential application for research studies and citizens' science projects to improve our understanding of air pollution, as well as incorporating Dylos units into existing air quality monitoring networks to facilitate data collection.



© University of Venda

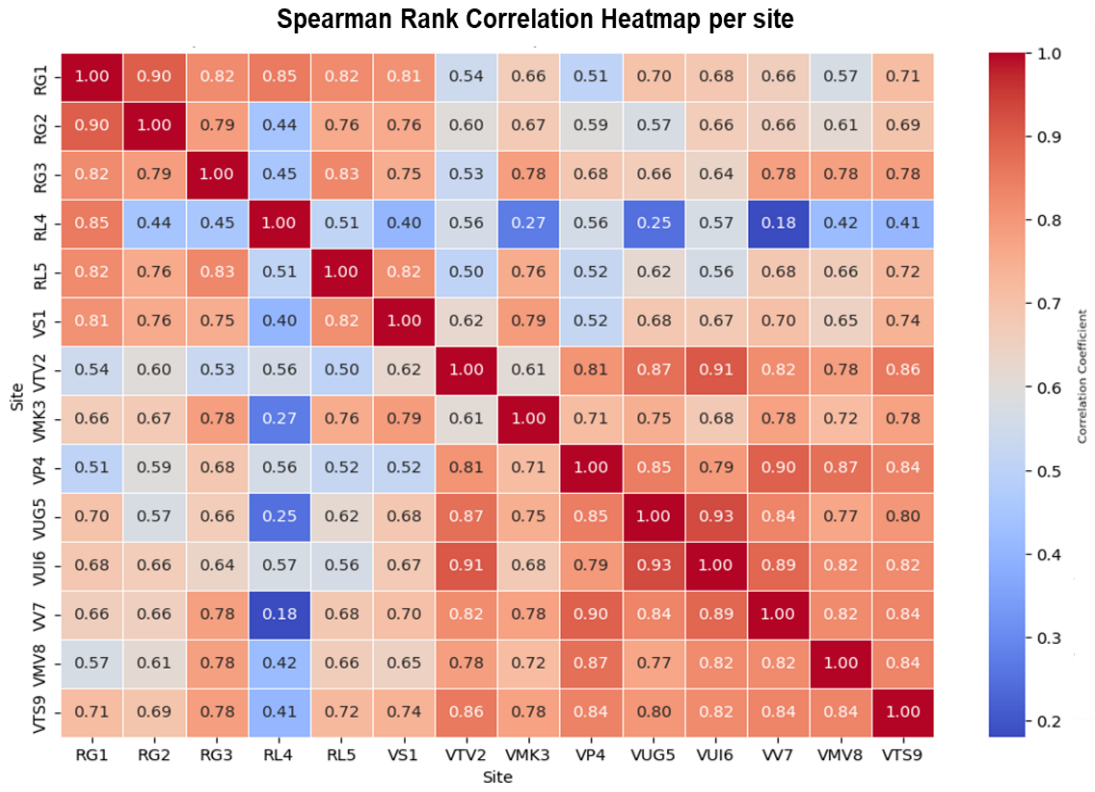



**Figure 4.15:** The correlation matrix heatmap for the daily average concentrations of PM2.5 reported by 14 distinct Dylos DC1700 units.

### 4.3.8 Identification of Potential Sources of Ambient $PM_{2.5}$ Air Pollutants using Bivariate Polar Plot (BVP) Analysis in the Study Area

Applying bivariate polar plots for $PM_{2.5}$ source identification in the Thulamela LM area of South Africa proved insightful. These plots, depict pollutant mixing ratios in polar coordinates based on wind speed and direction, enabling a nuanced analysis of potential sources (Carslaw & Ropkins, 2012; Carslaw & Beevers, 2013). The spatial distribution of the mean ambient $PM_{2.5}$ concentrations across fourteen monitoring locations is more effectively illustrated using Bivariate polar plots (BVP) along with prevailing wind direction frequency (Figures 4.16 and 4.17). BVP analysis identifies major source areas (indicated by red shading) contributing to ambient $PM_{2.5}$ pollution at the respective locations.

In the rural Gondeni and Lunungwi villages, monitoring sites RG1, RG2, RG3, RL4, and RL5 exhibited distinct patterns. Predominant $PM_{2.5}$ sources for RG1 and RL4 included residential wood burning and unpaved roads in the immediate vicinity - southeast and



© University of Venda

southwest areas. RG2, situated close to a road, showed similar patterns identified from the monitoring area and its surroundings, and about 50 meters from the east side, with predominant sources from residential wood burning and unpaved roads. Likewise, RG3 and RL5 exhibit comparable patterns in their immediate surroundings, primarily attributed to residential wood burning (Figures 4.16a and 4.16b). Within the urban and suburban zones (VTV2, VTS9, VP4, VS1, VMK3, VUG5, VUI6, VV7, VMV8), bivariate polar plots illustrate prevailing $PM_{2.5}$ sources in the west, southwest, and southeast directions. Major contributors encompass emissions from road traffic, malls, warehouses, residential wood burning, and industrial activities. In these areas, the primary $PM_{2.5}$ source is consistently identified from the measurement point in the west, southwest, and southeast directions (Figure 4.16c).

Notably, for groupings based on activities and area, group VA which includes sites VTV2, VTS9, VP4, and VS1, the potential sources are within approximately 50 to 300 meters to the west, northwest, and southeast of each site, involving emissions from various sources, including vehicular activities (related to road traffic, malls, and warehouses) and unpaved roads in Thohoyandou, Sibasa, Tshilungoma, Shayandima, and Makumbane in Thulamela Municipality. Further analysis reveals that the southeast direction from these sites corresponds to both residential wood burning and unpaved roads in the Thulamela Municipality. The southwest's predominant source of VTV2 is road traffic emissions and the Shayandima industrial area. These potential sources are near roadsides for each site in the group aligned with the representation of VA (Figure 4.17). Examining sites VMK3, VUG5, and VUI6 (group VB), it is evident that potential $PM_{2.5}$ sources are in the surrounding area near the monitoring station, as well as in the southwest and southeast directions (indicated by red shading in Figure 4.17). This area's red and light-yellow shading corresponds to light traffic and construction activities in the commercial and the University's environment, however, in the southeast of VMK3, the potential source of $PM_{2.5}$ is associated with residential activities, such as cooking with wood and biomass burning. Finally, sites VV7 and VMV8, grouped as VC and situated in the commercial and business centre of Thohoyandou (comprised government offices, departments, warehouses, stores, and hotels), exhibit high traffic density. Taxi ranks, road traffic, and



© University of Venda

warehouse activities characterise the potential source areas of pollution. In contrast, a residential area to the southeast of VMV8 highlights the significant contribution of residential wood burning to $PM_{2.5}$ emissions in this location.

For site grouping (Groups RG, RL, VA, VB, and VC), distinct patterns in $PM_{2.5}$ sources provide a comprehensive understanding of contributing factors across the different areas of Thulamela LM. Residential wood burning and unpaved roads were identified as major contributors to $PM_{2.5}$ levels in RG and RL. Residential wood burning, vehicle activities (road traffic, malls, taxi ranks and construction), as well as unpaved road dust, emerge as dominant sources of ambient $PM_{2.5}$ across monitoring sites in group VA, VB, and VC (Figure 4.17). Likewise, emissions from warehouses and manufacturing facilities may play a significant role. The findings align with previous studies conducted in similar suburban and urban environments in South Africa (Piketh *et al*., 2016; Novela *et al*., 2020; Adesina *et al*., 2022) and other developed countries (Sooktawee *et al*., 2020a; Sooktawee *et al*., 2020b; Demirarslan and Zeybek, 2022; Hodoli *et al*., 2023), reaffirming the persistent challenges of $PM_{2.5}$ pollution in densely populated areas. In conclusion, the bivariate polar plot analysis offers valuable insights into the spatial distribution of $PM_{2.5}$ sources in Thulamela LM. The identified sources underscore the complex interplay of residential, vehicular, and industrial activities contributing to ambient $PM_{2.5}$ pollution in the region, with implications for targeted pollution control measures and urban planning to improve air quality.



© University of Venda

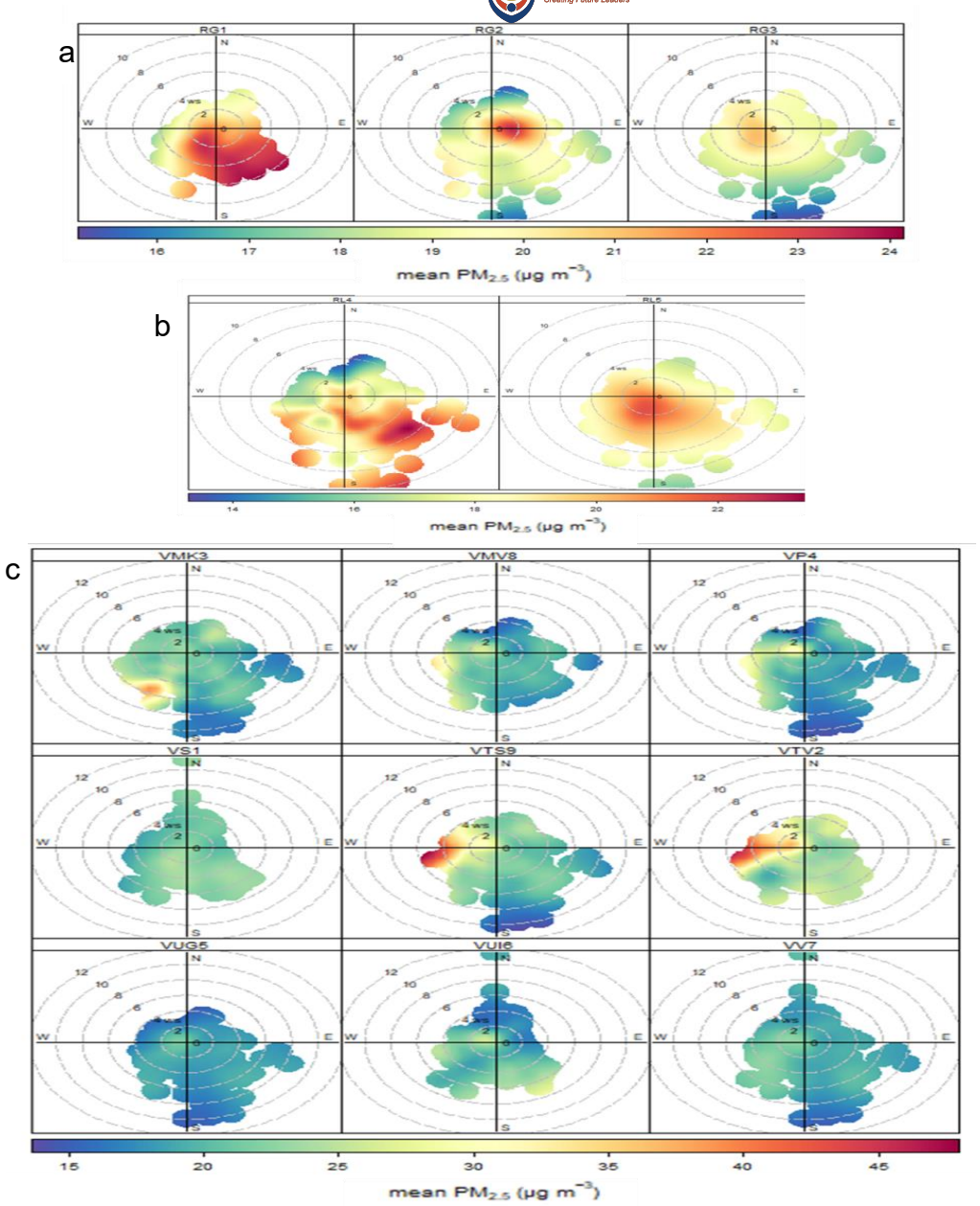



**Figure 4.16:** Bivariate polar plot (BVP) of $PM_{2.5}$ concentrations (µg/m³) per site. The color scale indicates $PM_{2.5}$ concentration and the radial scales show wind speed, which increases radially from the centre. The site grouping is represented in a) for residential site in Gondeni (RG), b) residential site in Lunungwi (RL) and c) vehicular and residential areas (VA-VC) in the study area.



© University of Venda

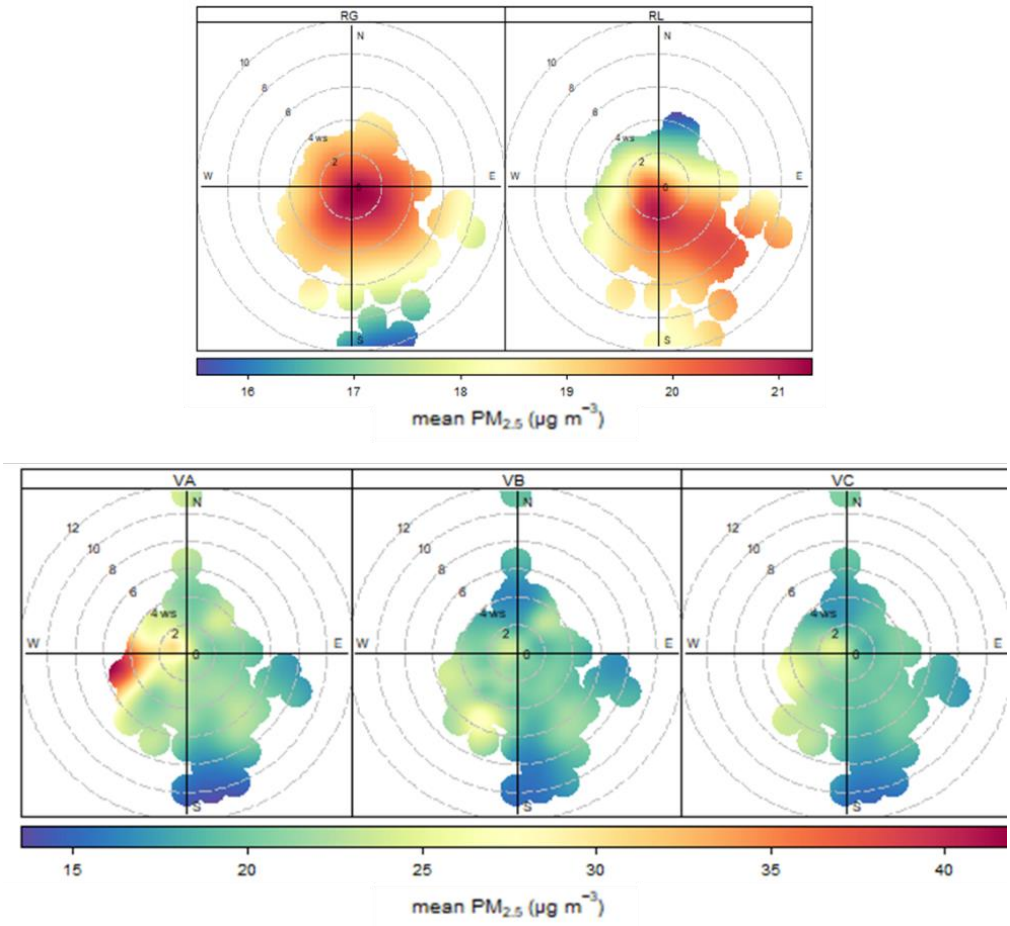



**Figure 4.17:** Bivariate polar plot (BVP) of $PM_{2.5}$ concentrations (µg/m³) per group (RG, RL, VA, VB, and VC). The color scale indicates $PM_{2.5}$ concentration and the radial scales show wind speed, which increases radially from the centre.

## 4.4 Limitations of the Study

Although this study successfully demonstrated the applicability of low-cost sensors (LCS) for air quality monitoring, several limitations were identified:

1. Environmental Sensitivity of the Sensor: The Dylos DC1700's accuracy was influenced by environmental factors such as humidity, temperature variations, and sensor drift over time. This could lead to bias in $PM_{2.5}$ estimations, particularly under extreme weather conditions. Future studies should incorporate real-time correction algorithms to address these interferences.
2. Power Supply Limitations: Continuous data collection was occasionally disrupted due to battery dependence. While solar-powered modifications were introduced,



© University of Venda

power storage efficiency remained suboptimal, affecting long-term deployment of the devices.

3. Potential Underestimation of $PM_{2.5}$ Concentrations: The Dylos DC1700 exhibited a tendency to underestimate $PM_{2.5}$ levels during high-pollution events, likely due to sensor saturation or particle size detection limitations. This suggests the need for sensor-specific correction factors when analysing high-pollution environments.
4. Limited Source Apportionment Analysis: The study primarily identified vehicular emissions and domestic wood burning as major $PM_{2.5}$ contributors, but industrial sources, mining activities, and waste burning were not comprehensively assessed. Future research should integrate chemical speciation and receptor modelling to improve source attribution.
5. Geographical Scope and Generalizability: The study focused on Thulamela Municipality, and while findings are relevant to similar environments, the results may not be fully generalizable to regions with different topography, climatic conditions, and pollution sources. Expanding the study area and incorporating multi-site comparisons, however could enhance applicability.

Despite these limitations, the study provides critical insights into air quality patterns in Thulamela LM and reinforces the role of low-cost sensors as practical tools for environmental monitoring. To further improve sensor deployment and data reliability, future research should focus on refining calibration techniques, integrating additional sensor models, and expanding air quality monitoring networks to cover a broader range of pollution sources.

### 4.5 Conclusion

This study evaluated the performance of the Dylos DC1700 low-cost sensor (LCS) for monitoring ambient $PM_{2.5}$ concentrations in Thulamela Municipality, South Africa. The results demonstrated a strong correlation ($R^2 = 0.75$) between the Dylos DC1700 and the GRIMM EDM180 reference monitor, confirming that the LCS provides reliable $PM_{2.5}$ measurements when appropriately calibrated. The spatial and temporal variability of $PM_{2.5}$ levels revealed that urban areas experienced higher $PM_{2.5}$ concentrations, largely due to vehicular emissions, road dust, and commercial activities, while rural areas were



© University of Venda

more affected by domestic wood burning. Seasonal analysis indicated that $PM_{2.5}$ levels were highest during winter, attributed to temperature inversions and increased fuel combustion, whereas lower levels were recorded in summer due to improved atmospheric dispersion. Diurnal trends indicated two distinct $PM_{2.5}$ peaks, occurring in the morning (6:00–8:00) and evening (17:00–20:00), coinciding with rush-hour traffic and domestic fuel burning. Bivariate Polar Plot (BVP) analysis identified pollution hotspots near high-traffic intersections, commercial hubs, and residential areas, further confirming major $PM_{2.5}$ sources. The study also established that $PM_{2.5}$ levels at most monitoring sites exceeded both the WHO air quality guidelines and South Africa's NAAQS standards, emphasising the urgency of implementing targeted pollution control measures. These findings validate the hypothesis that low-cost sensors can be used for monitoring ambient air quality in the study area. The Dylos DC1700 performed effectively, showing a high correlation with the reference instrument (GRIMM EDM180), however, minor discrepancies, particularly during high-pollution events, suggest that environmental factors may influence measurement accuracy. The sensor, despite this, provided consistent, cost-effective, and scalable air quality data, reinforcing the practicality of LCS technology in regions where conventional air monitoring infrastructure is unavailable. Future applications should incorporate real-time correction models and environmental factor adjustments to enhance accuracy.

## References

Adesina, J. A., Piketh, S. J., Burger, R. P., & Mkhatshwa, G. (2022). Assessment of criteria pollutants contributions from coal-fired plants and domestic solid fuel combustion at the South African industrial highveld. *Cleaner Engineering and Technology*, *6*, 100358.

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), 77–94. https://doi.org/10.1007/s00244-022-00937-4

Almetwally, A. A., Bin-Jumah, M., & Allam, A. A. (2020). Ambient air pollution and its influence on human health and welfare: an overview. *Environmental Science and Pollution Research*, *27*, 24815–24830.

Aniyikaiye, T. E., Piketh, S. J., Enitan, I. T. & Edokpayi, J. N. (2022). Assessment of ambient air quality around University residences in Thohoyandou, South Africa.



© University of Venda

(National Association of Clean Air).

Aniyikaiye, T. E., Piketh, S. J., & Edokpayi, J. N. (2023). Quantification of ambient $PM_{2.5}$ concentrations adjacent to informal brick kilns in the Vhembe District using low-cost sensors. Scientific Reports, 13(1), 22453.

AQ-SPEC: Air Quality Sensor Performance Evaluation Center (n.d). Summary of PM sensor evaluations. South Coast Air Quality Management District. Available online: http://www.aqmd.gov/aq-spec/evaluations/criteria-pollutants/summary-pm. Accessed on 21 February 2024.

Arowosegbe, O. O., Röösli, M., Adebayo-Ojo, T. C., Dalvie, M. A., & de Hoogh, K. (2021). Spatial and temporal variations in $PM_{10}$ concentrations between 2010–2017 in South Africa. *International Journal of Environmental Research and Public Health*, *18*(24). https://doi.org/10.3390/ijerph182413348.

Awokola, B., Okello, G., Johnson, O., Dobson, R., Ouédraogo, A. R., Dibba, B., Ngahane, M., Ndukwu, C., Agunwa, C., Marangu, D., Lawin, H., Ogugua, I., Eze, J., Nwosu, N., Ofiaeli, O., Ubuane, P., Osman, R., Awokola, E., Erhart, A., … Semple, S. (2022). Longitudinal Ambient $PM_{2.5}$ Measurement at Fifteen Locations in Eight Sub-Saharan African Countries Using Low-Cost Sensors. *Atmosphere*, *13*(10). https://doi.org/10.3390/atmos13101593.

Belelie, M. D., Piketh, S. J., Burger, R. P., Venter, A. D., & Naidoo, M. (2019). Characterisation of ambient Total Gaseous Mercury concentrations over the South African Highveld. *Atmospheric Pollution Research*, *10*(1), 12–23. https://doi.org/10.1016/j.apr.2018.06.001.

Broday, D. M., & Collaborators, C.-S. P. (2017). Wireless distributed environmental sensor networks for air pollution measurement—The promise and the current reality. *Sensors*, *17*(10), 2263.

Carslaw, D C. (2019). *The openair manual–open-source tools for analysing air pollution data, Manual for version 2.6–6, University of York*. https://davidcarslaw.com/files/openairmanual.pdf.

Carslaw, David C., & Beevers, S. D. (2013). Characterising and understanding emission sources using bivariate polar plots and k-means clustering. *Environmental Modelling and Software*, *40*, 325–329. https://doi.org/10.1016/j.envsoft.2012.09.005.

Carslaw, David C., & Ropkins, K. (2012). Openair - An r package for air quality data analysis. *Environmental Modelling and Software*, *27–28*, 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008.

Caseiro, A., Schmitz, S., Villena, G., Jagatha, J. V., & von Schneidemesser, E. (2022). Ambient characterisation of PurpleAir particulate matter monitors for measurements to be considered as indicative. *Environmental Science: Atmospheres*, *2*(6), 1400–1410.

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., & Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment International*, *99*, 293–



© University of Venda

302.

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., & Dandona, R. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The Lancet*, *389*(10082), 1907–1918.

Dacunto, P. J., Klepeis, N. E., Cheng, K. C., Acevedo-Bolton, V., Jiang, R. T., Repace, J. L., Ott, W. R., & Hildemann, L. M. (2015). Determining $PM_{2.5}$ calibration curves for a low-cost particle monitor: Common indoor residential aerosols. *Environmental Sciences: Processes and Impacts*, *17*(11), 1959–1966. https://doi.org/10.1039/c5em00365b.

DEAT, (2009). National Environmental Management: Air quality Act 2004 (Act No 39 of 2004)- National Ambient Air Quality, Government gazette No. 32816. Department of Environmental Affairs and Tourism, Pretoria. https://www.environment.gov.za/sites/default/files/legislations/nemaqa_airquality_g32816gon1210.pdf.

DEA, Department of Environmental Affairs, South Africa (2012). National ambient air quality standard for particulate matter with aerodynamic diameter less than 2.5 micrometres($PM_{2.5}$).https://www.gov.za/sites/default/files/gcis_document/201409/35463gon486.pdf.

Demirarslan, K. O., & Zeybek, M. (2022). Conventional air pollutant source determination using bivariate polar plot in Black Sea, Turkey. *Environment, Development and Sustainability*, *24*(2), 2736–2766.

Edlund, K. K., Killman, F., Molnár, P., Boman, J., Stockfelt, L., & Wichmann, J. (2021). Health risk assessment of $PM_{2.5}$ and $PM_{2.5}$-bound trace elements in Thohoyandou, South Africa. *International Journal of Environmental Research and Public Health*, *18*(3), 1359.

Enitan, I. T., Durowoju, O. S., Edokpayi, J. N., & Odiyo, J. O. (2022). A review of air pollution mitigation approach using air pollution tolerance index (APTI) and anticipated performance index (API). Atmosphere, 13(3), 374.

Feig, G. T., Naidoo, S., & Ncgukana, N. (2016). Assessment of ambient air pollution in the Waterberg Priority Area 2012-2015. *Clean Air Journal*, *26*(1), 21–28. https://doi.org/10.17159/2410-972x/2016/v26n1a9.

Franken, R., Maggos, T., Stamatelopoulou, A., Loh, M., Kuijpers, E., Bartzis, J., Steinle, S., Cherrie, J. W., & Pronk, A. (2019). Comparison of methods for converting Dylos particle number concentrations to $PM_{2.5}$ mass concentrations. *Indoor Air*, *29*(3), 450–459. https://doi.org/10.1111/ina.12546.

Freiman, M. T., & Piketh, S. J. (2003). Air transport into and out of the industrial Highveld region of South Africa. *Journal of Applied Meteorology and Climatology*, *42*(7), 994–1002.

Fu, W., Chen, Z., Zhu, Z., Liu, Q., Dang, E., & Dong, J. (2018). *Spatial and Temporal*



© University of Venda

*Variations of Six Criteria Air Pollutants in Fujian Province, China*. *5*. https://doi.org/10.3390/ijerph15122846.

Gao, M., Cao, J., & Seto, E. (2015). A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of $PM_{2.5}$ in Xi'an, China. *Environmental Pollution*, *199*, 56–65. https://doi.org/10.1016/j.envpol.2015.01.013

GRIMM Environmental Dust Monitor #180, User Manual. http://www.lisa.u-pec.fr/~formenti/Tools/Manuals/GRIMM_model1.180_Manual_2004.pdf.

Gupta, S., Agarwal, R., & Mittal, S. K. (2016). Respiratory health concerns in children at some strategic locations from high PM levels during crop residue burning episodes. *Atmospheric Environment*, *137*, 127–134.

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

Hassani, A., Castell, N., Watne, Å. K., & Schneider, P. (2023). Citizen-operated mobile low-cost sensors for urban $PM_{2.5}$ monitoring: field calibration, uncertainty estimation, and application. *Sustainable Cities and Society*, *95*, 104607.

HEI, Health Effects Institute (2018). State of Global Air 2018: Over 7 billion people face unsafe air. https://www.healtheffects.org/announcements/state-global-air-2018-over-7-billion-people-face-unsafe-air.

Hegde, S., Min, K. T., Moore, J., Lundrigan, P., Patwari, N., Collingwood, S., Balch, A., & Kelly, K. E. (2020). Indoor household particulate matter measurements using a network of low-cost sensors. *Aerosol and Air Quality Research*, *20*(2), 381–394. https://doi.org/10.4209/aaqr.2019.01.0046.

Hodoli, C. G., Coulon, F., & Mead, M. I. (2023). Source identification with high-temporal resolution data from low-cost sensors using bivariate polar plots in urban areas of Ghana. *Environmental Pollution*, *317*, 120448.

Howlett-Downing, C., Boman, J., Molnár, P., Shirinde, J., & Wichmann, J. (2023). Health risk assessment of $PM_{2.5}$ and $PM_{2.5}$-bound trace elements in Pretoria, South Africa. *Journal of Environmental Science and Health, Part A*, *58*(4), 342–358.

Jackson-Morris, A., Bleymann, K., Lyall, E., Aslam, F., Bam, T. S., Chowdhury, I., Daouda, E. A., Espinosa, M., Romo, J., & Singh, R. J. (2016). Low-cost air quality monitoring methods to assess compliance with smoke-free regulations: a multi-center study in six low-and middle-income countries. *Nicotine & Tobacco Research*, *18*(5), 1258–1264.

Jovašević-Stojanović, M., Bartonova, A., Topalović, D., Lazović, I., Pokrić, B., & Ristovski, Z. (2015). On the use of small and cheaper sensors and devices for indicative citizen-based monitoring of respirable particulate matter. *Environmental Pollution*,*206*(January 2018),696–704. https://doi.org/10.1016/j.envpol.2015.08.035

Kanchanasuta, S., Sooktawee, S., Bunplod, N., Patpai, A., Piemyai, N., & Ketwang, R. (2021). Analysis of short-term air quality monitoring data in a coastal area. AIMS



© University of Venda

Environmental Science, (6). *8*(6), 517-531.https://doi.org/10.3934/environsci.2021033.

Kanchanasuta, S., Sooktawee, S., Patpai, A., & Vatanasomboon, P. (2020). Temporal Variations and Potential Source Areas of Fine Particulate Matter in Bangkok, Thailand. *Air, Soil and Water Research*, *13*, 1178622120978203.

Khan, K., Su, C.-W., Tao, R., & Hao, L.-N. (2020). Urbanization and carbon emission: causality evidence from the new industrialized economies. *Environment, Development and Sustainability*, *22*, 7193–7213.

Kim, K.-H., Kabir, E., & Kabir, S. (2015). A review on the human health impact of airborne particulate matter. *Environment International*, *74*, 136–143.

Klepeis, N. E., Hughes, S. C., Edwards, R. D., Allen, T., Johnson, M., Chowdhury, Z., Smith, K. R., Boman-Davis, M., Bellettiere, J., & Hovell, M. F. (2013). Promoting smoke-free homes: a novel behavioral intervention using real-time audio-visual feedback on airborne particle levels. *PloS One*, *8*(8), e73251.

Landrigan, P. J., Fuller, R., Acosta, N. J. R., Adeyi, O., Arnold, R., Basu, N. N., Baldé, A. B., Bertollini, R., Fuster, V., Greenstone, M., Haines, A., Hanrahan, D., Hunter, D., Khare, M., Krupnick, A., Lanphear, B., Lohani, B., Preker, A. S., Ramesh, J., … Steiner, A. (2017). *The Lancet Commission on pollution and health*. *6736*(17), 32345. http://www.thelancet-press.com/embargo/pollutioncommission.pdf.

Landrigan, P. J., Fuller, R., Acosta, N. J. R., Adeyi, O., Arnold, R., Baldé, A. B., Bertollini, R., Bose-O'Reilly, S., Boufford, J. I., & Breysse, P. N. (2018). The Lancet Commission on pollution and health. *The Lancet*, *391*(10119), 462–512.

Li, G., Fang, C., Wang, S., & Sun, S. (2016). The effect of economic growth, urbanisation, and industrialization on fine particulate matter ($PM_{2.5}$) concentrations in China. *Environmental Science & Technology*, *50*(21), 11452–11459.

Li, J., Mattewal, S. K., Patel, S., & Biswas, P. (2020). Evaluation of nine low-cost-sensor-based particulate matter monitors. *Aerosol and Air Quality Research*, *20*(2), 254–270.

Liu, D., Zhang, Q., Jiang, J., & Chen, D. R. (2017). Performance calibration of low-cost and portable particle matter (PM) sensors. *Journal of Aerosol Science*, *112* (September), 1–10. https://doi.org/10.1016/j.jaerosci.2017.05.011.

Manikonda, A., Zíková, N., Hopke, P. K., & Ferro, A. R. (2016). Laboratory assessment of low-cost PM monitors. *Journal of Aerosol Science*, *102*, 29–40.

Mei, H., Han, P., Wang, Y., Zeng, N., Liu, D., Cai, Q., Deng, Z., Wang, Y., Pan, Y., & Tang, X. (2020). Field evaluation of low-cost particulate matter sensors in Beijing. *Sensors (Switzerland)*, *20*(16), 1–17. https://doi.org/10.3390/s20164381

Miller, L., & Xu, X. (2018). Ambient $PM_{2.5}$ human health effects-findings in China and research directions. *Atmosphere*, *9*(11), 1–16. https://doi.org/10.3390/atmos9110424.

Moletsane, S. D., Lindeque, F., Nkosi, N. C., Adesina, J. A., Burger, R. P., Mkhatshwa, G., & Piketh, S. J. (2021). Intra-urban variability of $PM_{2.5}$ in a dense, low-income



© University of Venda

settlement on the South African Highveld. *Clean Air Journal*, *31*(1), 1–9.

Morakinyo, O., Mokgobu, M., & Mukhola, S. (2019). Health risk analysis of elemental component of respirable particulate matter in an urban area in South Africa. *Environmental Epidemiology*, *3*, 274.

Muyemeki, L., Burger, R., Piketh, S. J., Language, B., Beukes, J. P., & van Zyl, P. G. (2021). Source apportionment of ambient $PM_{10\text{-}2.5}$ and $PM_{2.5}$ for the Vaal Triangle, South Africa. *South African Journal of Science*, *117*(5), 1–11. https://doi.org/10.17159/SAJS.2021/8617.

Mykhaylova, N. (2018). *Low-cost Sensor Array Devices as a Method for Reliable Assessment of Exposure to Traffic-related Air Pollution*. 229.

Nawaz, M. O., & Henze, D. K. (2020). Premature deaths in Brazil associated with long-term exposure to $PM_{2.5}$ from Amazon fires between 2016 and 2019. *GeoHealth*, *4*(8), e2020GH000268.

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), 1–12. https://doi.org/10.17159/caj/2020/30/2.8735.

Olutola, B., Claassen, N., Voyi, K., & Wichmann, J. (2019). Apparent temperature as a modifier of the effects of air pollution on respiratory disease hospital admissions in Secunda, South Africa. *Environmental Epidemiology*, *3*, 439.

Piketh, S. J., Wernecke, B., & Burger, R. (2016). Household air pollution in South African low-income settlements: a case study. *WIT Transactions on Ecology and the Environment*, *207*, 227–236.

Personal communication, Dylos corporation, 2015.

Rajak, R., & Chattopadhyay, A. (2019). Short and long-term exposure to ambient air pollution and impact on health in India: a systematic review. *International Journal of Environmental Health Research*, *00*(00), 1–25. https://doi.org/10.1080/09603123.2019.1612042.

Scorgie, Y., & Annegarn, H. J. (2012). *Urban Air Quality Management and Planning in South Africa*. Ph.D. thesis at the University of Johannesburg, South Africa.

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics: From air pollution to climate change*. John Wiley & Sons.

Semple, S., Apsley, A., & MacCalman, L. (2013). An inexpensive particle monitor for smoker behaviour modification in homes. *Tobacco Control*, *22*(5), 295–298. https://doi.org/10.1136/tobaccocontrol-2011-050401.

Semple, S., Ibrahim, A. E., Apsley, A., Steiner, M., & Turner, S. (2015). Using a new, Low-Cost air quality sensor to quantify Second-Hand smoke (SHS) levels in homes. *Tobacco Control*, *24*(2), 153–158. https://doi.org/10.1136/tobaccocontrol-2013-051188.

Sooktawee, S, Kanchanasuta, S., Boonyapitak, S., Patpai, A., & Piemyai, N. (2020a). Distinguish potential source areas of $PM_{2.5}$ and $PM_{10}$ by statistical data analysis.



© University of Venda

*IOP Conference Series: Earth and Environmental Science*, *489*(1), 12024.

Sooktawee, Sirapong, Kanabkaew, T., Boonyapitak, S., Patpai, A., & Piemyai, N. (2020b). Characterising particulate matter source contributions in the pollution control zone of mining and related industries using bivariate statistical techniques. *Scientific Reports*, *10*(1), 1–10.

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

South African Weather Service (SAWS). https://www.weathersa.co.za/.

SANAS, South African National Accreditation System (2000). Supplementary Requirements for the Accreditation of Continuous Ambient Air Pollution Monitoring Stations.

Steinle, S., Reis, S., Sabel, C. E., Semple, S., Twigg, M. M., Braban, C. F., Leeson, S. R., Heal, M. R., Harrison, D., Lin, C., & Wu, H. (2015). Personal exposure monitoring of $PM_{2.5}$ in indoor and outdoor microenvironments. *Science of the Total Environment*, *508*, 383–394. https://doi.org/10.1016/j.scitotenv.2014.12.003.

Vanker, A., Barnett, W., Chartier, R., MacGinty, R., & Zar, H. J. (2023). Personal monitoring of fine particulate matter ($PM_{2.5}$) exposure in mothers and young children in a South African birth cohort study–a pilot study. *Atmospheric Environment*, *294*, 119513.

Vercellino, R. J., Sleeth, D. K., Handy, R. G., Min, K. T., & Collingwood, S. C. (2018). Laboratory evaluation of a low-cost, real-time, aerosol multi-sensor. *Journal of Occupational and Environmental Hygiene*, *15*(7), 559–567. https://doi.org/10.1080/15459624.2018.1468565.

Wang, Q., Kwan, M.-P., Zhou, K., Fan, J., Wang, Y., & Zhan, D. (2019). The impacts of urbanization on fine particulate matter ($PM_{2.5}$) concentrations: Empirical evidence from 135 countries worldwide. *Environmental Pollution*, *247*, 989–998.

Wang, Z., Wang, R., Wang, J., Wang, Y., Donahue, N. M., Tang, R., Dong, Z., Li, X., Wang, L., & Han, Y. (2022). The seasonal variation, characteristics and secondary generation of $PM_{2.5}$ in Xi'an, China, especially during pollution events. *Environmental Research*, *212*, 113388.

Wernecke, B., Piketh, S. J., & Burger, R. P. (2015). Indoor and outdoor particulate matter concentrations on the Mpumalanga highveld–A case study. *Clean air journal*, *25*(2), 12-12.

Wernecke, B., Wright, C. Y., Hey, J. V., Piketh, S., Burger, R., Kunene, Z., ... & Baloyi, V. (2021). Opportunities for the application of low-cost sensors in epidemiological studies to advance evidence of air pollution impacts on human health. *Clean Air Journal*, *31*(1), 1-4.

West, S.E., Büker, P., Ashmore, M., Njoroge, G., Welden, N., Muhoza, C., Osano, P., Makau, J., Njoroge, P. and Apondo, W. (2020). Particulate matter pollution in an



© University of Venda

informal settlement in Nairobi: Using citizen science to make the invisible visible. Applied Geography, 114, p.102133.

Williams, J., Petrik, L., & Wichmann, J. (2021). $PM_{2.5}$ chemical composition and geographical origin of air masses in Cape Town, South Africa. *Air Quality, Atmosphere & Health*, *14*, 431–442.

Williams, R., Kaufman, A., Hanley, T., Rice, J., & Garvey, S. (2014). *Evaluation of field-deployed low cost PM sensors. EPA/600/R-14/464, 2014*. *December*, 1–76. www.epa.gov/ord%0Ahttp://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=297517.

WHO (2016). Ambient air pollution: a global assessment of exposure and burden of disease. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

WHO (2021). What are the WHO Air quality guidelines? Improving health by reducing air pollution. https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines.

Wright, C. Y., Matooane, M., Oosthuizen, M. A., & Phala, N. (2014). Risk perceptions of dust and its impacts among communities living in a mining area of the Witwatersrand, South Africa. *Clean Air Journal= Tydskrif Vir Skoon Lug*, *24*(1), 22–27.

Yunesian, M., Rostami, R., Zarei, A., Fazlzadeh, M., & Janjani, H. (2019). Exposure to high levels of PM2. 5 and PM10 in the metropolis of Tehran and the associated health risks during 2016–2017. *Microchemical Journal*, *150*, 104174.



© University of Venda

# CHAPTER FIVE

# ATMOSPHERIC DISPERSION MODELLING OF PARTICULATE MATTER ($PM_{2.5}$) USING AERMOD IN THULAMELA MUNICIPALITY, LIMPOPO PROVINCE, SOUTH AFRICA

## Abstract

*Air pollution is a major environmental and public health concern, with $PM_{2.5}$ recognised as a major contributor to respiratory and cardiovascular diseases. This study applies AERMOD to model $PM_{2.5}$ dispersion from residential wood combustion (RWC) and vehicular emissions in Thulamela Municipality, Limpopo Province, South Africa and assesses compliance with SA NAAQS and WHO guidelines. The results show significant spatial variability, with several areas exceeding regulatory limits. The highest daily $PM_{2.5}$ concentration (49.2 µg/m³) surpassed SA NAAQS (40 µg/m³) and WHO (15 µg/m³) limits, while annual levels (0.5–19.7 µg/m³) remained within SA NAAQS but exceeded the WHO guideline (5 µg/m³). Pollution hotspots were concentrated in densely populated areas, primarily influenced by RWC, vehicular emissions, and warehouse activities. Among emission sources, RWC is the dominant contributor with over 96% of total $PM_{2.5}$ emissions, while traffic had a minor impact. The modelled-to-monitored ratio for $PM_{2.5}$ generally fell within the USEPA recommended range (>0.5 to <2.0), suggesting reasonable model performance. Model validation revealed systematic underestimation, with low correlation ($R^2$ = 0.176) and negative fractional bias (FB) at most receptors, however, localized overestimation was observed at receptor RL (group), where hourly $PM_{2.5}$ levels were higher in the model (47.12 µg/m³) than observed (32.27 µg/m³, FAC2 = 0.25). Diurnal variations showed peak $PM_{2.5}$ levels during the morning (6–8 AM) and evening (6–8 PM) rush hours, due to traffic emissions and domestic combustion activities. Monthly peak patterns observed from May to September were attributed to seasonal changes and increased wood* combustion *during colder months. Notably, a significant reduction occurred in July 2021 during the COVID-19 lockdown, demonstrating the effect of constrained human activities on atmospheric quality; despite AERMOD model limitations, such as not estimating from other sources of emission, such as biomass burning, unpaved roads, warehouses and agricultural activities, the model still efficiently identified pollution hotspots, thus providing valuable insights for policy interventions and regulatory measures. Improving emissions inventories, meteorological inputs, and source representation is essential to improve model accuracy and support air quality management efforts.*

**Keywords:** $PM_{2.5}$, Air Dispersion Modelling, AERMOD, Air Quality Residential Wood Combustion, Vehicular Emissions, Thulamela, South Africa.



© University of Venda

## 5.1 Introduction

Ambient air pollution poses significant environmental and public health challenges, contributing to respiratory infections, cardiovascular diseases, and premature mortality. The World Health Organization (WHO) attributes nearly one-third of deaths from respiratory diseases, heart disease, stroke, and lung cancer to air pollution (WHO, 2021). Among the various pollutants, fine particulate matter ($PM_{2.5}$) particles with a diameter of 2.5 micrometres or less has raised significant concern due to its ability to penetrate deep into the lungs and enter the bloodstream, leading to severe health complications. Air quality management remains a critical concern in South Africa, particularly in regions such as Thulamela Municipality in Limpopo Province, where air pollution arises from both natural and anthropogenic sources. Residential wood combustion (RWC) and vehicular emissions contribute significantly to local $PM_{2.5}$ levels, with their dispersion influenced by meteorological factors, topography, and emission characteristics. Understanding pollutant distribution and assessing population exposure require advanced techniques such as atmospheric dispersion modelling. These techniques aid in determining the spatial spread of $PM_{2.5}$ levels, identifying vulnerable populations, and supporting preventive strategies.

Air dispersion modelling is a mathematical technique used to predict atmospheric pollutants' distribution (Cimorelli *et al*. 2003; Holmes and Morawska, 2006). Among the available models is AERMOD, an acronym for the American Meteorological Society/Environmental Protection Agency Regulatory Model. This model is recommended by the United States Environmental Protection Agency (US EPA) for air regulatory studies and widely recognised for its reliability in simulating air pollutant dispersion. Developed by the American Meteorological Society and the US EPA, AERMOD effectively models pollutant contributions from various emission sources, providing insights into their impact on ambient air quality in local communities (US EPA, 2005; 2019). AERMOD is widely recognised for its accuracy in simulating pollutant dispersion from various sources, making it a valuable tool for evaluating air quality and informing mitigation strategies (Chidhindi, 2022; Rezazadeh *et al.,* 2022).



© University of Venda

AERMOD has been employed to study pollutant dispersion from various sources, including traffic-related emissions (Askariyeh *et al*., 2017; Odediran *et al*., 2023), municipal waste burning (Popoola *et al*., 2023; Adeniran *et al*., 2024), industrial activities (Huang and Guo, 2019) and other sources (Rezazadeh *et al.,* 2022; Yusuf *et al*., 2022). In South Africa, several studies have utilised the AERMOD dispersion model to evaluate particulate matter (PM) emissions from various sources in different locations. Bosman (2019) assessed PM emissions from various sources, including domestic fuel burning and windblown dust in Kwadela, while Chidhindi (2022) and Tshisi (2023) applied the model to examine PM contributions from anthropogenic sources and vehicle emissions in Zamdela. These studies collectively demonstrate the versatility of AERMOD in analysing different PM sources and sizes across multiple locations in South Africa, providing valuable insights into air quality issues in these areas.

Few investigations, however have modelled $PM_{2.5}$ dispersion from RWC and traffic emissions in Limpopo Province, particularly in semi-urban and rural areas, such as Thulamela Municipality, despite the extensive use of AERMOD in air quality studies globally. This research addresses this gap using AERMOD to simulate $PM_{2.5}$ distribution from local emission sources and assess population exposure. This study aims to (i) assess the extent of $PM_{2.5}$ emissions dispersion from the two sources, (ii) determine ground-level pollutant concentrations at selected receptor locations, and (iii) validate AERMOD predictions using monitored data and compare with regulatory standards. By elucidating $PM_{2.5}$ dispersion patterns and their health implications, this study provides valuable insights for policymakers and environmental managers to formulate effective air quality management strategies for Thulamela Municipality. The findings will contribute to a better understanding of air pollution dynamics in semi-urban and rural areas of South Africa, potentially informing future regulatory actions and public health interventions.



© University of Venda

## 5.2: Materials and Methods

This section presents methods for the thesis's third objective of the dispersion modelling of $PM_{2.5}$, its analysis, and interpretation. The methodology flow chart for AERMOD is shown in Figure 5.1.

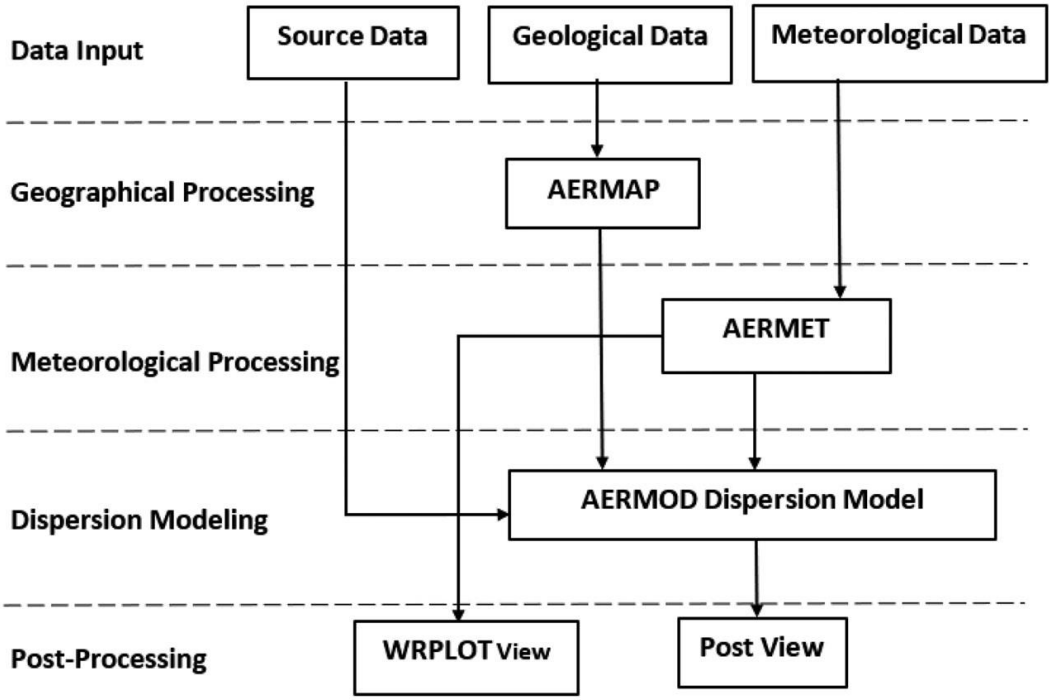



**Figure 5.1:** Flowchart process for AERMOD model (Adopted from Adeniran *et al*., 2024).

### 5.2.1 Quantification of $PM_{2.5}$ Emissions in the study domain

Accurate estimation of $PM_{2.5}$ emissions is crucial for dispersion modelling using the AERMOD model. The bottom-up method was used for estimating these emissions, which is discussed in detail in Subsection 3.2.2 of Chapter 3. Emissions were estimated from two primary sources in the study area: residential wood combustion and vehicular emissions. Relevant emission factors and available data were applied to calculate the total $PM_{2.5}$ emissions. For a thorough description of the $PM_{2.5}$ emission inventory compilation, including source contributions, refer to Subsection 3.2.2. This method ensures that the emission estimates are accurately represented, which is essential for subsequent $PM_{2.5}$ dispersion modelling with AERMOD.



© University of Venda

### 5.2.2 AERMOD Modelling software, components and description

In this study, the air dispersion model was carried out using American Meteorological Society-Environmental Protection Agency Regulatory Model (AERMOD) (Version 4.6.2), which has been established as a reliable tool for estimating ambient concentrations of air pollution. AERMOD is a steady-state air quality dispersion plume model recommended by USEPA. The model incorporates the effects of vertical variations in the planetary boundary layer (PBL) on air pollutants dispersion. The turbulence profiles vary with height, and this determines the plume growth (US EPA 2005: Gulia *et al*., 2015). AERMOD is considered accurate for dispersion modelling at distances not exceeding 50 km from the emission source (Cimorelli, *et al*., 2003; US EPA, 2005). In the model, the dispersion of air pollutants in the atmosphere is modelled mathematically. Using AERMOD, three types of data are required - information on the source (such as emission rates, heights, and locations), meteorological data, as well as topographic data concerning the target area. It incorporates AERMET (meteorological preprocessor) and AERMAP (terrain preprocessor) to divide and streamline plume flow in complex terrain, like a hill. The particulate matter criteria pollutants, $PM_{2.5}$ emitted from area sources are predicted by AERMOD (Figure 5.2).

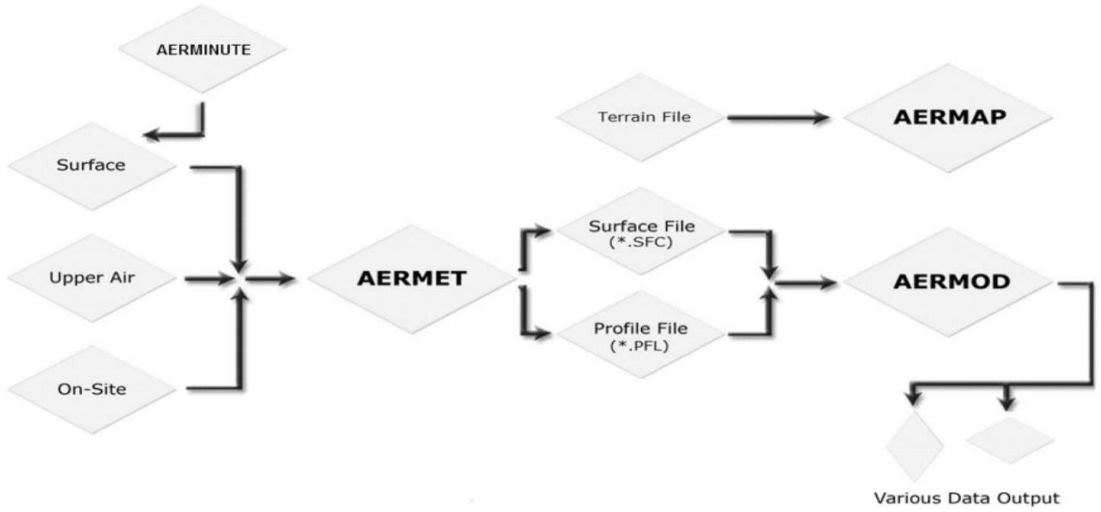



**Figure 5.2:** AERMOD modelling system's data flow (Villalvazo, Davila & Reed G. 2007).



© University of Venda

### 5.2.2.1 AERMET

The meteorology of an area influences the distribution and transport of air pollutants, ultimately controlling their atmospheric concentrations. AERMET is a meteorological pre-processor for AERMOD that converts meteorological data into a planetary boundary layer (PBL) format, making it suitable for dispersion modelling, by incorporating surface properties and meteorological parameters (US EPA, 2016a). Parameters such as albedo, Bowen ratio, cloud cover, and surface roughness, as described by Zou *et al*. (2010), were configured in AERMOD. The input to AERMET included hourly surface meteorological observations, such as wind speed and direction, temperature, dew point, humidity, sea level pressure, and twice-daily upper air soundings. An hourly meteorological dataset for 2019–2022 was acquired from Lakes Environmental Software and processed through AERMET. The processed data provided hourly surface and upper air conditions in a grid cell format for model simulations. The wind rose generated from the meteorological data for Thulamela LM (2021–2022) indicates a predominant southwest, east and northeast wind direction (Figure 5.3). The wind speeds range from 0.50 m/s to above 11.10 m/s, mostly occurring between 3.60 and 8.80 m/s. The average wind speed is 3.09 m/s, while calm conditions account for 0.74%. The wind pattern plays a significant role in influencing the dispersion and transportation of air pollutants in the study area, emphasising the directional impact on pollutant movement.

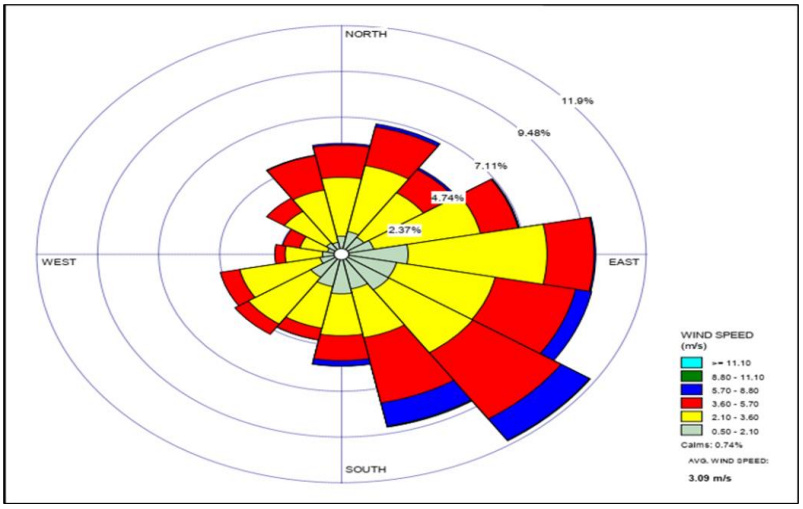



**Figure 5.3**: Modelled Windrose generated for the study domain for 2021 to 2022.



© University of Venda

#### 5.2.2.2 AERMAP

AERMAP provides comprehensive terrain data processing, including the arrangement of receptor locations and sources. It can process a variety of standard data formats, allowing it to create terrain based elevations for specified receptors and sources, as well as a hill height scale value for each (US EPA 2016b). AERMAP uses gridded data to calculate the divided streamline height for the terrain height scale of the receptor site.

### 5.2.3 Data Collection and Preparation

The data preparation process involved the following:

I. Emission Inventory: Emission sources were spatially and temporally allocated based on activity data. Geographic Information System (GIS) tools were used to map source locations for input into the model.
II. Meteorological Data: Meteorological parameters were obtained from the South African Weather Service (SAWS). Key parameters included wind speed, direction, temperature, and atmospheric stability classes, processed using AERMET.
III. Terrain Data: Digital elevation model (DEM) data were sourced from the Shuttle Radar Topography Mission (SRTM) and processed using AERMAP to account for complex terrain features within TLM.

### 5.2.4 AERMOD Model Configuration and Procedures

#### 5.2.4.1 Modelling domain

The modelling domain defines the spatial extent of the $PM_{2.5}$ dispersion simulation using AERMOD. This area was selected to include key emission sources, sensitive receptor sites, and the surrounding areas potentially affected by pollutant transportation. AERMOD can simulate dispersion up to 50 km beyond the study area to account for transboundary effects. The domain was centred on Thohoyandou (Lat.: 22.97653° S, Long.: 30.445° E), covering an area that incorporates emission sources and prevailing meteorological conditions essential for accurate modelling, as shown in Figure 5.4.



© University of Venda

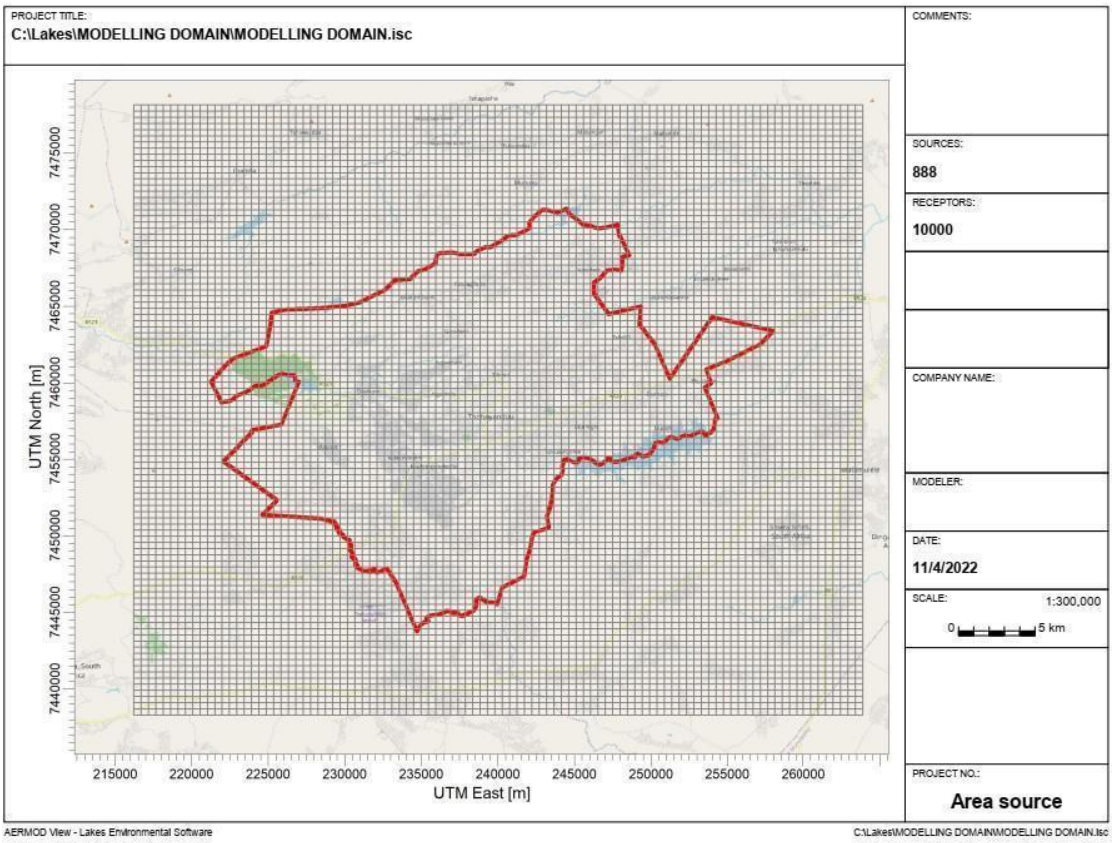



**Figure 5.4**: Study domain within Limpopo Province

### 5.2.4.2 Model Configuration and Setup

This section outlines the steps required to prepare the AERMOD model for execution, ensuring accurate simulation of $PM_{2.5}$ dispersion within the Thulamela Local Municipality (TLM). The key elements of the setup included:

I. Configuring Pathways: Each pathway (control, source, receptor, meteorology, and output) were carefully set up to capture the dispersion dynamics and meet study requirements.

II. Integrating Terrain and Land Use Data: Using AERMAP, terrain features and land-use classifications were incorporated to account for the influence of elevation and surface roughness on pollutant dispersion.

III. Boundary Conditions: Suitable boundary conditions were established to define the limits of the computational domain and ensure realistic modelling of atmospheric flows.



© University of Venda

IV. Computational Domain Alignment: The model's domain was configured to align precisely with the study area, ensuring coverage of all relevant emission sources and sensitive receptor locations.

#### 5.2.4.3 Model Pathways

AERMOD's five main pathways are essential for defining key inputs, simulation parameters, and outputs in the dispersion modelling process. These pathways were configured as follows:

I. Control Pathway: This pathway defines the simulation parameters, including:
   - The selection of the target pollutant ($PM_{2.5}$) due to its relevance to the study's objectives and health implications.
   - The configuration of averaging periods (24-hour and annual averages) in compliance with South African National Ambient Air Quality Standards (NAAQS) and World Health Organization (WHO) guidelines.
   - Dispersion options, where rural dispersion modes were applied to reflect the land-use characteristics within the study domain and ensure accurate representation of local atmospheric conditions.
   - The specification of output formats for graphical and statistical analysis.

II. Source Pathway: Characterised emission sources, type of sources (point, area, and volume), release height, emission rate and variations.

III. Receptor Pathway: Defined the parameters, such as receptor grids at ground level, base elevation, terrain options and sensitive locations such as schools and hospitals as discrete receptors.

IV. Meteorology Pathway: This is used to import the area's integrated processed meteorological data from AERMET, after which the model is operationalised.

V. Output Pathway: The dispersion computation results are presented through the Output Pathway upon completion, along with configured output files for graphical and statistical analysis.



© University of Venda

#### 5.2.4.4 Model Execution Procedure

Following those preparations, the model was executed on a daily and annual scale to simulate $PM_{2.5}$ concentrations from all emission sources within the Thulamela LM study domain, for one year. It was also run iteratively to optimise performance and ensure consistency with monitored or monitoring data.

### 5.2.5 Model Validation

Based on the AERMOD model results for ambient $PM_{2.5}$ in Thulamela Municipality of Limpopo Province, the performance accuracy of the models was validated by comparing the modelled $PM_{2.5}$ output data with the monitored data from the respective monitoring stations in the study domain. Statistical performance metrics such as Fractional Bias (FB), FAC2 and Normalised Mean Square Error (NMSE) were used to assess model accuracy (Chang & Hanna, 2004). This process aimed to validate the simulated results and evaluate the modelled-to-monitored ratio against the acceptable range defined by USEPA standards and considered temporal variations. The distributions of the differences between the observed and predicted values for each monitoring station were also analysed.

### 5.2.6. Data Analysis and Visualization

The model outputs were analysed to identify the spatial and temporal dispersion patterns of $PM_{2.5}$ within Thulamela Local Municipality (TLM). Descriptive statistics were conducted using Microsoft Excel 2016 to summarise key metrics, such as the mean, median, standard deviation, and percentiles for both modelled and observed $PM_{2.5}$ concentrations. Advanced analyses, including time series, evaluations and pollutant rose plots, were performed using Python (version 3.12) and the Openair package in R Studio (version 4.04). Visualization played an essential role in interpreting model outputs, with Geographic Information System (GIS) tools like ArcGIS (version 10.4) employed to create concentration maps and other graphical representations. The findings were effectively communicated through maps, time-series graphs, tables, and charts. The Isopleth plots maps for daily and annual average concentrations were analysed according to the SA NAAQS and WHO standards.



© University of Venda

### 5.2.7. Ethical Considerations

This study adhered to the ethical standards outlined in subsection 3.2.3 of Chapter Three. Additional considerations specific to AERMOD implementation and air quality modelling were addressed, ensuring ethical data collection and analysis practices throughout the research process.

## 5.3 Results and Discussions

### 5.3.1 Emissions estimations for the identified sources

This section presents the $PM_{2.5}$ emissions inventory for residential wood combustion (RWC) and vehicular emissions in 2021. Annual emissions rates were estimated based on activity rates, including vehicle kilometres travelled, vehicle population, and the amount of wood burned, alongside their respective emission factors, as detailed in Chapter Three. The total $PM_{2.5}$ emissions rate for the study domain was 2348.7 t/year, with RWC contributing 96% (2254.65 t/year) and vehicular emissions accounting for 4% (94.05 t/year) (Table 5.1). The increased emissions during the study period were primarily driven by higher residential fuelwood combustion, underscoring the significant impact of household fuel use on the ambient air quality in Thulamela LM, a suburban area in South Africa. RWC and vehicular sources of emissions were represented by 888 sources in AERMOD as shown in the source parameter template (Appendix 5.1).

**Table 5. 1:** Quantified emissions rate of total $PM_{2.5}$ (ton per annum) per sector used by the AERMOD model for Thulamela LM

| Sources | $PM_{2.5}$ | |
|---|---|---|
| | t/year | g/s |
| Residential | 2254.65 | 71.49 |
| Vehicular | 94.05 | 2.98 |
| Total | 2348.70 | 74.48 |

I. Variation of Emissions

The variation in $PM_{2.5.}$ emissions is illustrated in Figure 5.5, showing a diurnal trend in domestic fuel combustion for Thulamela LM during 2021. Emissions peaked during the



© University of Venda

early mornings (7:00) and evenings (19:00), aligning with periods of heightened residential activity, such as cooking and vehicular traffic. Emissions were lowest during late night hours (1:00–5:00), reflecting reduced human activities. These findings emphasise the temporal influence of human behaviour on emission patterns. The dominance of residential wood combustion over vehicular sources underscores the need for targeted interventions for domestic fuel use to mitigate air quality impacts. Also, since the RWC source dominates in the study domains, its emission variation was adopted for the vehicle emission rate and used during the simulation of concentrations in Thulamela LM.

The emission patterns observed in this study align with findings from other South African studies, for instance, Nkosi (2018) identified domestic fuel combustion as a major source of $PM_{2.5}$ in rural areas, with peak emissions during morning and evening hours. Similarly, Chidhindi (2022) highlighted the significant contributions of residential fuel use and vehicle emissions to air pollution in semi-urban settings, and although vehicular emissions play a secondary role, their contribution remains notable in urbanised areas with high traffic volumes. This comparative analysis underscores the critical role of residential emissions in shaping $PM_{2.5}$ levels, particularly in rural and semi-urban regions such as Thulamela LM.

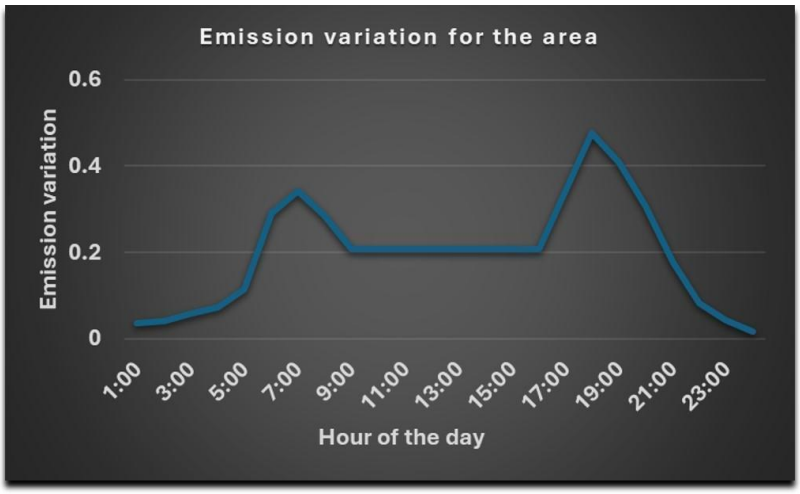



**Figure 5.5:** The diurnal trend in domestic fuel combustion used during the simulation of concentrations from Thulamela LM (This study, 2021).



© University of Venda

### 5.3.2. Spatial Distribution of Ambient $PM_{2.5}$ in Thulamela LM

The results indicate significant variations in pollutant dispersion, with certain areas exceeding both the SA NAAQS and WHO air quality standards. This section provides an overview of the spatial distribution of $PM_{2.5}$ from all identified sources combined, as well as the key individual emission sources, including residential wood combustion and vehicle emissions, highlighting pollution hotspots and regulatory exceedances.

#### 5.3.2.1. Modelled Average Daily and Annual $PM_{2.5}$ Levels from All Sources

The AERMOD modelling results reveal significant variations in $PM_{2.5}$ pollution from all sources across Thulamela LM, with several areas exceeding both national and international air quality standards. Figures 5.6 and 5.7 illustrate the spatial distribution of average daily and annual $PM_{2.5}$ concentrations from all identified sources, revealing critical pollution hotspots and patterns.

Figure 5.7 displays the spatial distribution of the 24-hour $PM_{2.5}$ levels, with values ranging from 2.2 to 49.2 µg/m³. The highest daily concentration, approximately 49.2 µg/m³, significantly exceeds both the SA NAAQS limit of 40 µg/m³ and the WHO guideline of 15 µg/m³. Areas with $PM_{2.5}$ levels between 20 and 49.2 µg/m³, highlighted in yellow, orange, and red contours, are concentrated in densely populated zones where residential wood combustion, traffic-related source emissions, and warehouse activities are prevalent. Several areas within the yellow contour exceed the SA NAAQS daily limit, while most areas surpass the more stringent WHO daily guideline. The spatial distribution of annual average $PM_{2.5}$ levels ranges from 0.5 to 19.7 µg/m³ (Figure 5.7). The highest annual concentration, approximately 19.7 µg/m³, approaches but does not exceed the SA NAAQS annual limit of 20 µg/m³, yet it significantly surpasses the WHO annual guideline of 5 µg/m³. Areas with annual concentrations between 8 and 19.7 µg/m³ (yellow to red contours) are primarily located in residential areas where wood combustion is more prevalent and traffic-related emissions.



© University of Venda

### *Exceedances and Health Implications*

The frequent occurrence of more than four daily exceedances per year, surpassing the SA NAAQS permissible limit, highlights the population's chronic exposure to elevated $PM_{2.5}$ levels, particularly in areas with concentrations exceeding 40 µg/m³. Short-term peaks and annual averages nearing the SA NAAQS limit indicate persistent air quality challenges. This ongoing noncompliance with air quality standards poses significant public health risks, especially for vulnerable, low-income communities that rely on solid fuels for cooking and heating (Figure 5.8).

Similar patterns of $PM_{2.5}$ pollution have been observed in other studies conducted in South African urban and semi-urban areas, including Zamdela, Johannesburg, and Durban, where $PM_{2.5}$ exceedances have been attributed to domestic fuel burning, vehicular emissions, and industrial activities (Lourens *et al*., 2011, 2016; Chidhindi, 2022). These findings highlight the urgent need for targeted emission control measures in Thulamela LM, focusing on major sources such as residential wood combustion, vehicular traffic, and warehouses to reduce $PM_{2.5}$ pollution, in order to achieve air quality standard compliance and safeguard public health.

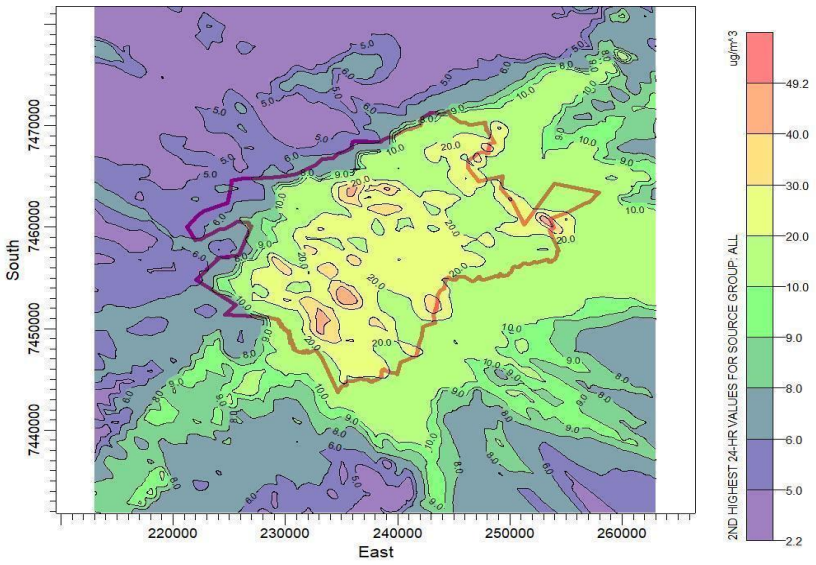



**Figure 5.6:** Spatial distribution of the daily average model $PM_{2.5}$ level within the study domain from all estimated sources.



© University of Venda

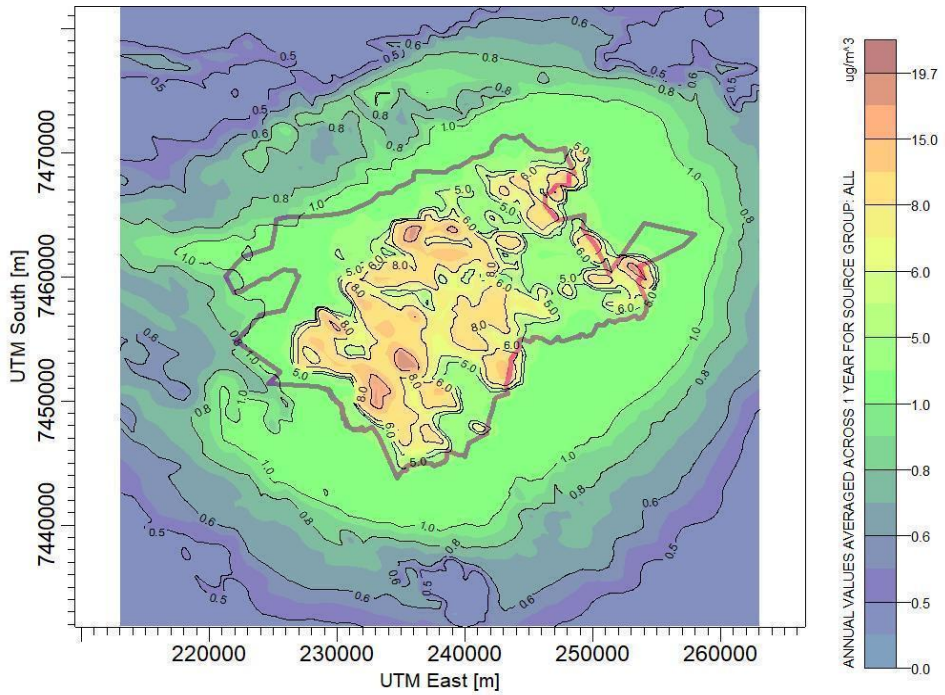



**Figure 5.7:** The spatial distribution of the annual average model $PM_{2.5}$ within the study domain from all estimated sources.

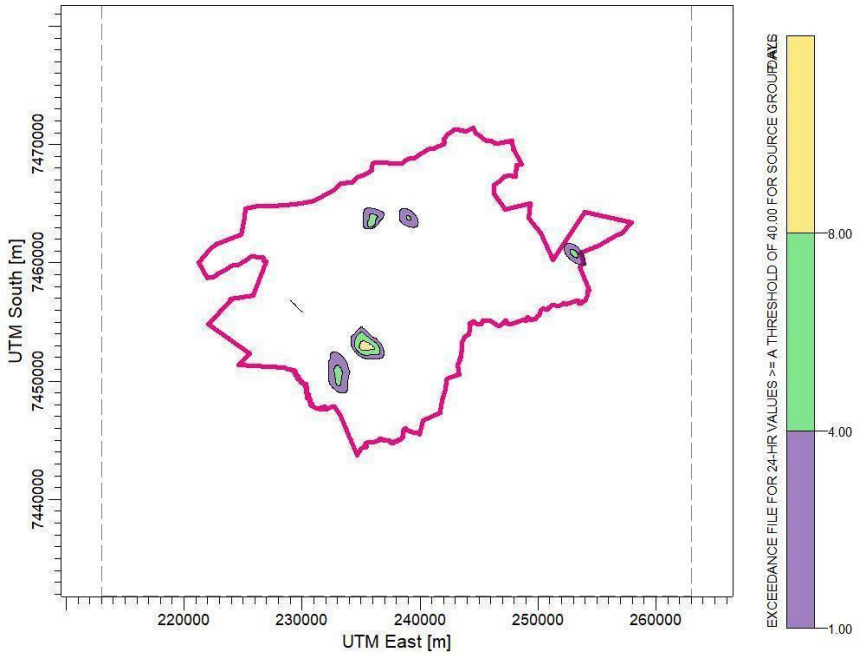



**Figure 5.8**: The spatial distribution of the daily average model $PM_{2.5}$ exceedance within the study domain from all estimated sources.



© University of Venda

#### 5.3.2.2. Modelled Daily and Annual $PM_{2.5}$ from Residential Wood Combustion (RWC) Sources

The AERMOD modelling results highlight residential wood combustion (RWC) as a major contributor to $PM_{2.5}$ pollution within Thulamela LM's semi-urban and rural areas. The spatial distribution of $PM_{2.5}$ from RWC, shown in Figures 5.9 and 5.10, indicates widespread exceedances of both SA NAAQS and WHO standards. These elevated levels suggest a heavy reliance on wood burning for cooking and household tasks, especially in low-income areas where cleaner alternatives remain unaffordable (Piketh *et al*., 2016; Moletsane *et al*., 2021; Chidhindi, 2022). The highest 24-hour $PM_{2.5}$ concentrations from RWC sources reach approximately 48.4 µg/m³ (Figure 5.9), significantly exceeding the SA NAAQS daily standard of 40 µg/m³ and the WHO guideline of 15 µg/m³. Elevated daily concentrations, ranging from 20 to 48.4 µg/m³, are prominent in densely-populated areas and regions with widespread wood burning, as shown by the red and orange contours. These results align with previous research, which identifies biomass burning as a major contributor to air pollution in South African communities and other parts of the world (Naidoo *et al*., 2014; Lin *et al*., 2018; Arku *et al*., 2020; Pauw *et al*., 2022). Annual average $PM_{2.5}$ concentrations from RWC reach a maximum of approximately 19.4 µg/m³ (Figure 5.10), remaining below the SA NAAQS annual limit of 20 µg/m³, but significantly exceeding the WHO annual guideline of 5 µg/m³. The highest annual concentrations, ranging from 8 to 19.4 µg/m³, are concentrated in central residential zones with high wood combustion activity. These results underscore the persistent nature of pollution from domestic fuel consumption and its potential for prolonged exposure and health impacts on the local population.

Figure 5.11 displays the frequency of 24-hour exceedances due to RWC emissions. Several areas recorded more than four exceedances per year, surpassing the maximum permissible threshold set by the SA NAAQS for daily limits. The highest frequency of exceedances is observed in densely populated residential zones where wood-burning is prevalent. This trend indicates prolonged exposure to high pollution levels from domestic fuel consumption, posing significant health risks to the local population, particularly in communities with limited access to cleaner energy alternatives, and highlighting the urgent need for interventions to protect public health.



© University of Venda

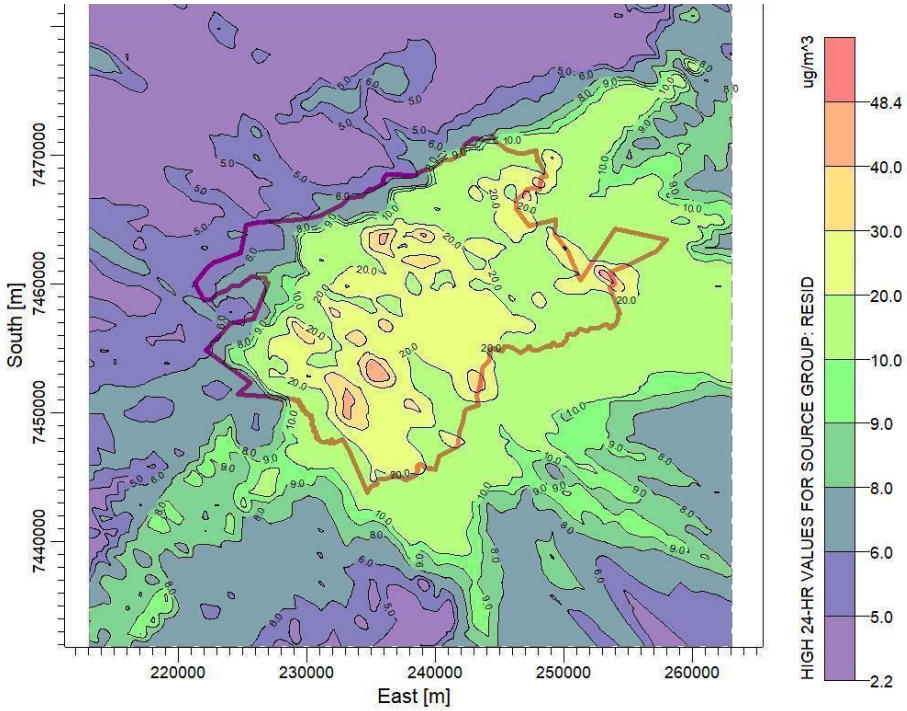



**Figure 5.9**: The spatial distribution of the daily average modelled $PM_{2.5}$ within the study domain from the residential sources.

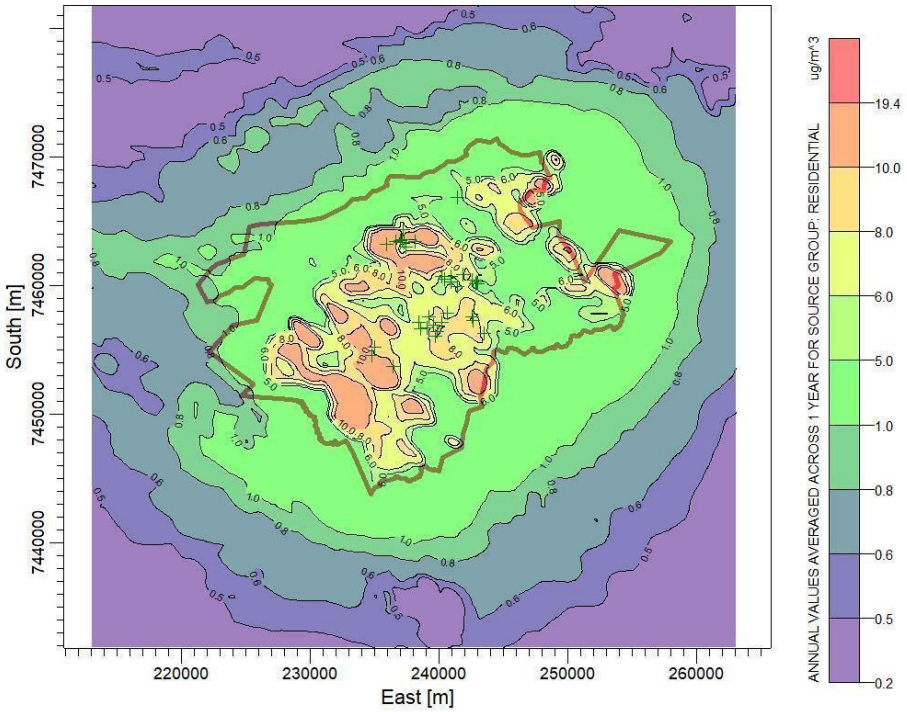



**Figure 5.10:** Spatial distribution of the annual and daily average modelled $PM_{2.5}$ within the study domain from the residential sources.



© University of Venda

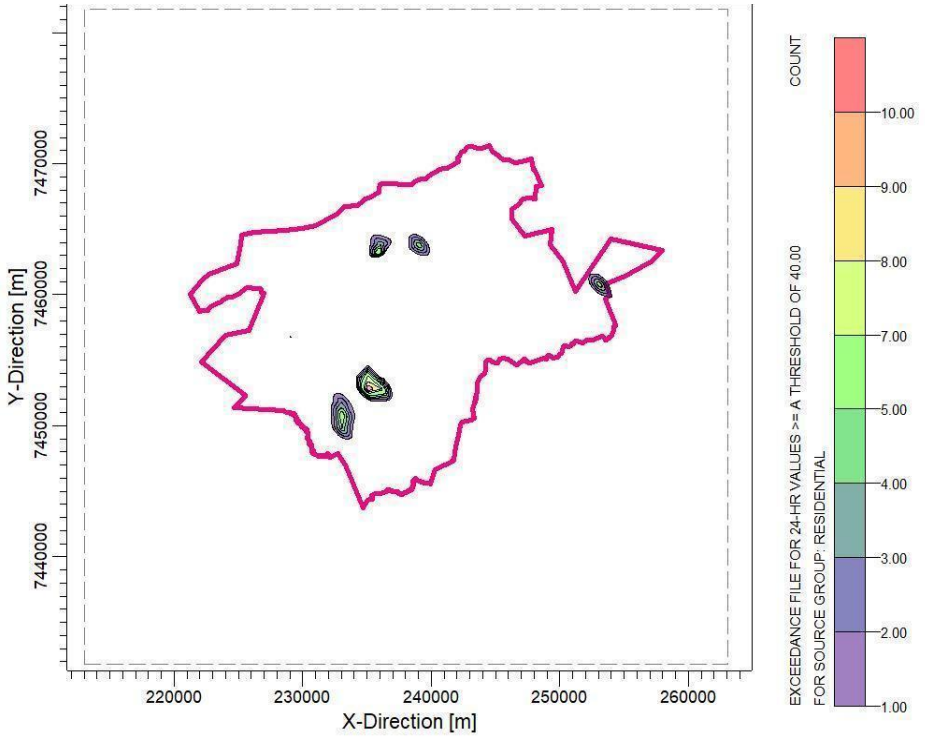



**Figure 5.11:** Spatial distribution of daily average modelled $PM_{2.5}$ exceedances within the study domain from residential sources.

### 5.3.2.3 Modelled average daily and Annual $PM_{2.5}$ Levels from Vehicular Source

The contour plots illustrate the spatial distribution of $PM_{2.5}$ concentrations from vehicle emissions within the study area, highlighting daily and annual averages. The maximum 24-hour $PM_{2.5}$ level from vehicle emissions is approximately 3.05 µg/m³, which is substantially below the SA NAAQS limit of 40 µg/m³ and the more stringent WHO guideline of 15 µg/m³. Elevated concentrations, ranging from 1.0 to 3.0 µg/m³, are observed near major roadways and high-traffic areas, however, despite these localised hotspots, vehicular emissions alone do not significantly contribute to regulatory daily exceedances (Figure 5.12). This suggests that other sources such as residential wood combustion (as previously discussed), or additional activities not covered in the scope of this study, including waste burning, warehouse operations, and agricultural practices, may be playing a more substantial role in the observed exceedances. The annual average $PM_{2.5}$ from vehicle emissions reaches a maximum of approximately 1.29 µg/m³, remaining well below both the SA NAAQS limit of 20 µg/m³ and the WHO guideline of 5 µg/m³. The highest annual concentrations occur in the central suburban areas and along major traffic



© University of Venda

corridors, as shown by the orange-yellow contours in Figure 5.13, however, the annual average remains minimal relative to regulatory standards. Vehicle emissions contribute to ambient $PM_{2.5,}$ although their impact is relatively minor compared to residential fuel combustion sources. The spatial patterns indicate that $PM_{2.5}$ levels from traffic sources are concentrated near major roads, traffic areas, malls, taxi ranks, warehouses and areas with other vehicular activities, but disperse quickly, resulting in low annual averages.

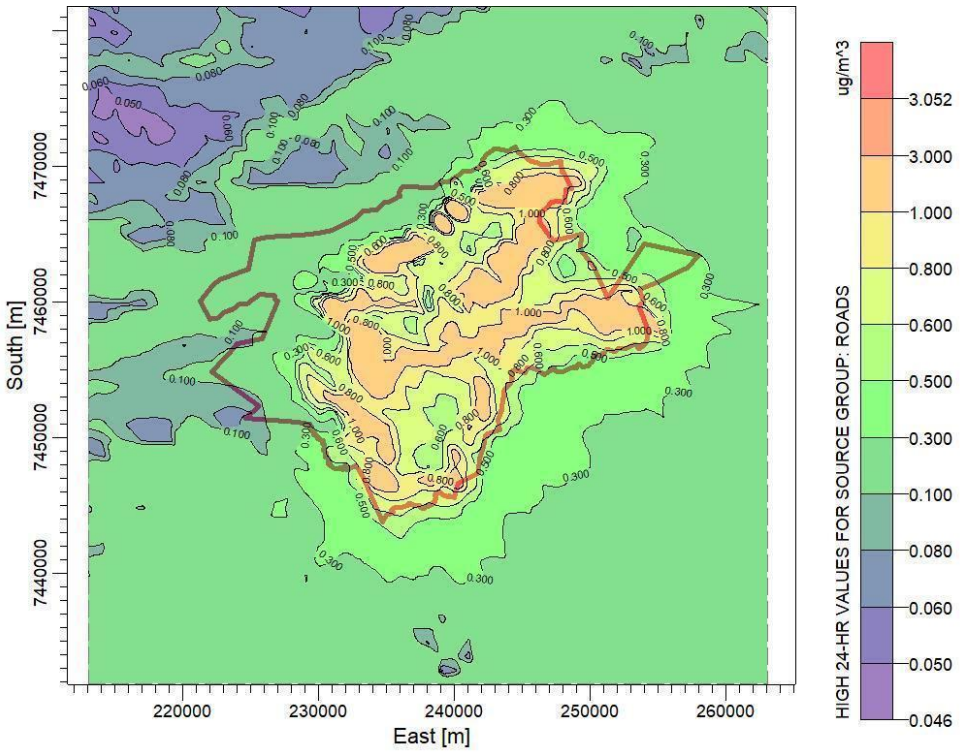



**Figure 5.12:** The spatial distribution of the daily average model $PM_{2.5}$ within the study domain from the vehicular sources.

These findings suggest that while traffic-related emissions are not the primary drivers or contributors to $PM_{2.5}$ exceedances in the study domain, they may exacerbate the overall pollution burden when combined with other sources. Mitigation measures such as improving public transport and promoting cleaner vehicle technologies could further help to reduce localised pollution hotspots.



© University of Venda

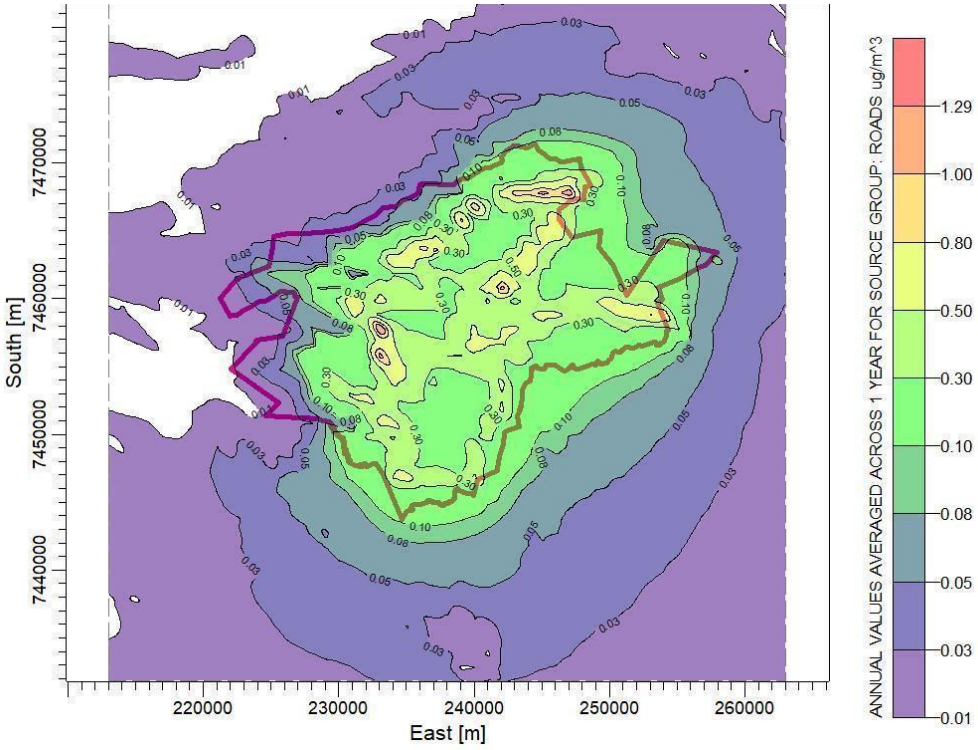



**Figure 5.13:** The spatial distribution of the annual average model $PM_{2.5}$ within the study domain from the vehicular sources.

### 5.3.2.4 Discrete receptor distributions

Discrete sensitive receptors are sites particularly prone to negative impacts of environmental exposure. These areas accommodate individuals more vulnerable to damage from pollutants, hazards, or environmental threats. According to DEA (2014) recommendations, discrete receptors should be included in the spatial distribution of emissions, while schools, public spaces, hospitals and other critical sites should be treated as discrete receptors in air quality modelling. In line with this, 32 locations were selected as discrete receptors, including the 14 monitoring sites and 18 additional sensitive locations. The spatial distribution location of the modelled average PM2.5concentrations for the receptors is illustrated in Figures 5.14a and 5.14b (Appendix 5.2). Likewise, the Average Annual $PM_{2.5}$ modelled level for discrete receptors in Thulamela LM is shown in Table 5.2 at different key locations.



© University of Venda

#### 5.3.2.4.1 Annual Modelled $PM_{2.5}$ Concentrations at Discrete Sensitive Receptors *(Schools, Hospitals, and Public Spaces)*.

For the annual modelled $PM_{2.5}$ concentrations at Monitoring Site Receptors (RG1 to VTS9), the lowest predicted $PM_{2.5}$ levels (below 7 µg/m³) were observed at VMK3 (4.70 µg/m³), RG2 (6.67 µg/m³), VS1 (6.70 µg/m³), and VUG5 (6.85 µg/m³). These sites fall well below the WHO's recommended limit of 15 µg/m³ and SA NAAQS of 40 µg/m³, indicating minimal local pollution sources and relatively clean air quality. In contrast, moderate to high $PM_{2.5}$ concentrations, ranging from 9 to 12.86 µg/m³, were observed at VP4 (9.75 µg/m³), VMV8 (9.21 µg/m³), VTV2 (10.92 µg/m³), RL4 (11.27 µg/m³), and VTS9 (12.86 µg/m³). These locations approach WHO's daily guidelines value, likely due to traffic emissions, commercial activities, and residential wood combustion. Notably, VTS9, situated near Tshilidzini Hospital, along a major express road, is particularly susceptible to pollution from vehicular emissions and nearby residential combustion activities. The highest $PM_{2.5}$ levels were recorded at RL5 (18.86 µg/m³) and VTS9 (12.86 µg/m³), with RL5 exceeding WHO's recommended limit, marking it as a pollution hotspot. Located in Lunungwi village, RL5 is likely affected by residential wood burning, unpaved road dust, and biomass combustion, making it a significantly high-exposure area.

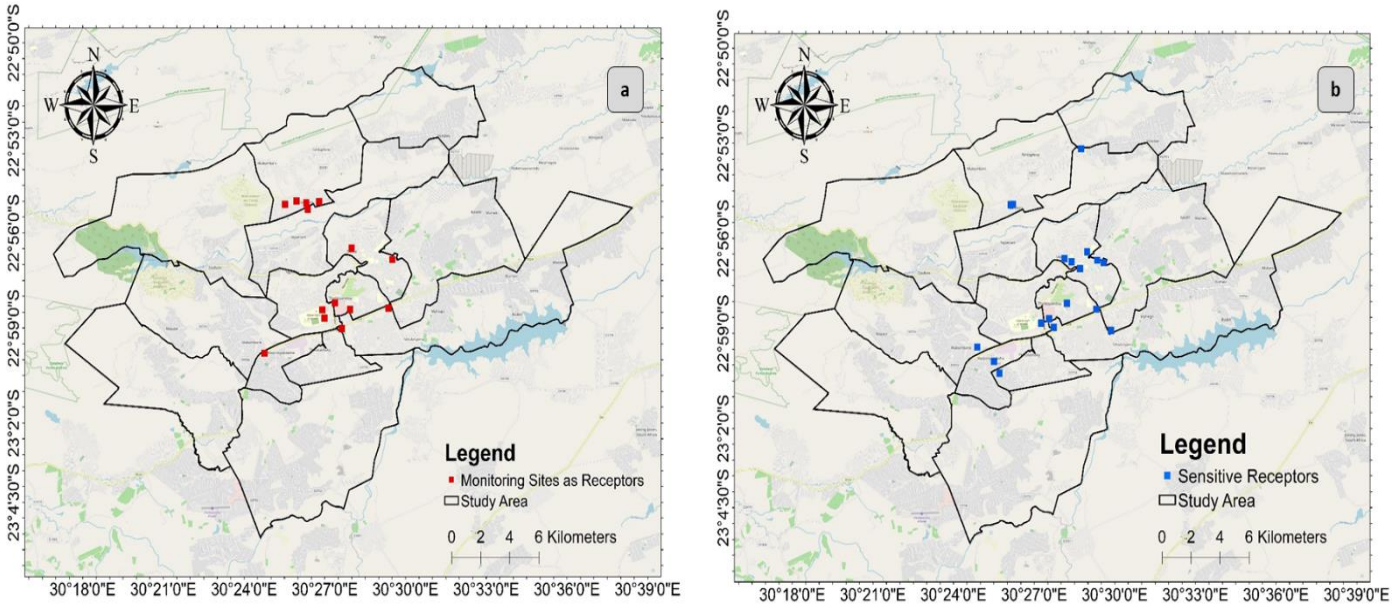



**Figure 5.14:** Map showing the locations used as discrete receptors in the study domain, a) monitoring sites simulated, and b) sensitive receptors.



© University of Venda

Furthermore, a similar pattern was observed at sensitive receptors such as educational institutions, healthcare facilities, and public spaces. At educational institutions, $PM_{2.5}$ concentrations ranged from 5.71 µg/m³ to 10.95 µg/m³. The values followed the order: Gondeni Senior Primary School (10.95 µg/m³), Lunungwi Primary (10.50 µg/m³), Shayandima Secondary (10.24 µg/m³), Tshilidzini Special School (9.40 µg/m³), Vhembe TVET College (8.65 µg/m³), Thohoyandou Secondary (7.51 µg/m³), Thohoyandou Technical High School (7.42 µg/m³), the University of Venda (6.61 µg/m³), and Laerskool Bergvlam Primary School (5.71 µg/m³). These results indicate that air quality at these sites remains below the WHO (15 µg/m³) and South African NAAQS (20 µg/m³) thresholds, suggesting a relatively safer environment for students; however, institutions with concentrations approaching the WHO guideline requires continuous monitoring to safeguard human and environmental health.

At healthcare facilities, Tshilidzini Hospital (9.87 µg/m³) and Tshilidzini Special School (9.40 µg/m³) exhibited slightly elevated $PM_{2.5}$ levels, showing the need for continuous monitoring. In contrast, lower concentrations were observed at Donald Fraser Hospital (4.20 µg/m³), Thohoyandou Health Centre Block G (5.48 µg/m³), and Makwarela Clinic (5.15 µg/m³), which provide a cleaner air environment beneficial for patients and healthcare workers. Commercial and public areas showed variable pollution levels, with higher $PM_{2.5}$ concentrations at Thavhani Mall (10.72 µg/m³), Phangami Mall (9.94 µg/m³), and Mvusuludzo Mall (8.17 µg/m³), likely influenced by vehicular emissions and commercial activities. In comparison, Makwarela Government Offices (8.08 µg/m³) and Sibasa Police Station (6.37 µg/m³) recorded lower levels. These patterns, thus, are consistent with expectations for a low-income urban area, where household combustion emerges as the predominant source of $PM_{2.5}$ pollution, followed by vehicular emissions. These aligned with the monitored and quantified emissions $PM_{2.5}$ data recorded in this study area.

#### 5.3.2.4.2 Daily Modelled $PM_{2.5}$ Concentrations at Sensitive Receptors (Schools, Hospitals, and Public Spaces).

Box plots were used to illustrate the variability in daily modelled $PM_{2.5}$ levels across the Thulamela, enabling a comparative analysis. The modelled spatial distribution of daily



© University of Venda

$PM_{2.5}$ concentrations across the 32 receptors, including monitoring sites and other sensitive locations, as shown in Figure 5.15 (Appendix 5.3 and 5.4), highlights the contribution of various pollution sources at the discrete receptors. The median daily $PM_{2.5}$ values ranged from 4.2 µg/m³ at Donald Fraser Hospital to 19.5 µg/m³ at RL5, highlighting significant variations in air quality. The highest concentrations recorded at RL5 (19.5 µg/m³) in Lunungwi village and VTS9 (17.2 µg/m³) adjacent to Tshilidzini Hospital main gate along the express road, were likely due to household fuel combustion, traffic congestion, vehicle emissions from the main road as well as unpaved roads, and burning of fuel. Moderate pollution levels were observed at Shayandima Secondary School (13.5 µg/m³) and Phangami Mall (12.9 µg/m³), followed by Thavhani Mall (11.8 µg/m³) and Vhembe TVET College (11.2 µg/m³), suggesting contributions from commercial activities, malls and public transport. The lowest $PM_{2.5}$ concentrations were found at Donald Fraser Hospital (4.2 µg/m³) and Makwarela Clinic (5.1 µg/m³), indicating relatively cleaner air. The spatial distribution of the modelled contribution aligns with previous studies where fuel combustion, vehicle emission, traffic-related and commercial activities contributed to modelled $PM_{2.5}$ (Chidhindi, 2022; Odediran *et al.,* 2024; Rezazadeh *et al*., 2022; Salve *et al*., 2021).

The key areas of concern include RL5 and VTS9, where $PM_{2.5}$ levels exceeded WHO's 15 µg/m³ guideline, as well as Shayandima Secondary School, Phangami Mall, Thavhani Mall and Vhembe TVET College, which are also close to the limit, posing health risks to residents, students, and the community. Elevated concentrations near schools, malls, and hospitals suggest pollution from multiple sources, including traffic, domestic combustion, and commercial activities. All locations comply with SA NAAQS (40 µg/m³), however, frequent exceedances of WHO's limit in most areas indicate chronic exposure risks, necessitating improved air quality management, stricter emission controls, and expanded monitoring in high-exposure zones.



© University of Venda

**Table 5.2:** Average annual $PM_{2.5}$ modelled level for discrete receptors in Thulamela LM

| S/N | Discrete Receptors ID and name | $PM_{2.5}$ (µg/m$^3$) |
|---|---|---|
| 1 | RG1 | 8.571 |
| 2 | RG2 | 6.666 |
| 3 | RG3 | 8.954 |
| 4 | RL4 | 11.272 |
| 5 | RL5 | 18.865 |
| 6 | VS1 | 6.702 |
| 7 | VTV2 | 10.925 |
| 8 | VMK3 | 4.701 |
| 9 | VP4 | 9.746 |
| 10 | VUG5 | 6.854 |
| 11 | VUI6 | 7.157 |
| 12 | VV7 | 8.527 |
| 13 | VMV8 | 9.215 |
| 14 | VTS9 | 12.864 |
| 15 | Donald Fraser Hospital | 4.203 |
| 16 | Gondeni Senior Primary School | 10.951 |
| 17 | Laerskool Bergvlam Primary School | 5.7168 |
| 18 | Lunungwi Primary School | 10.495 |
| 19 | Makwarela Clinic | 5.1514 |
| 20 | Makwarela Government Offices | 8.0817 |
| 21 | Mvusuludzo Mall | 8.169 |
| 22 | Phangami Mall | 9.937 |
| 23 | Shayandima Secondary School | 10.240 |
| 24 | Sibasa Police Station | 6.376 |
| 25 | Thavhani Mall | 10.724 |
| 26 | Thohoyandou Health Centre Block G | 5.481 |
| 27 | Thohoyandou Secondary School | 7.510 |
| 28 | Thohoyandou Technical High School | 7.415 |
| 29 | Tshilidzini Hospital | 9.872 |
| 30 | Tshilidzini Special School | 9.400 |
| 31 | University of Venda | 6.608 |
| 32 | Vhembe TVET College Makwarela Campus | 8.648 |



© University of Venda

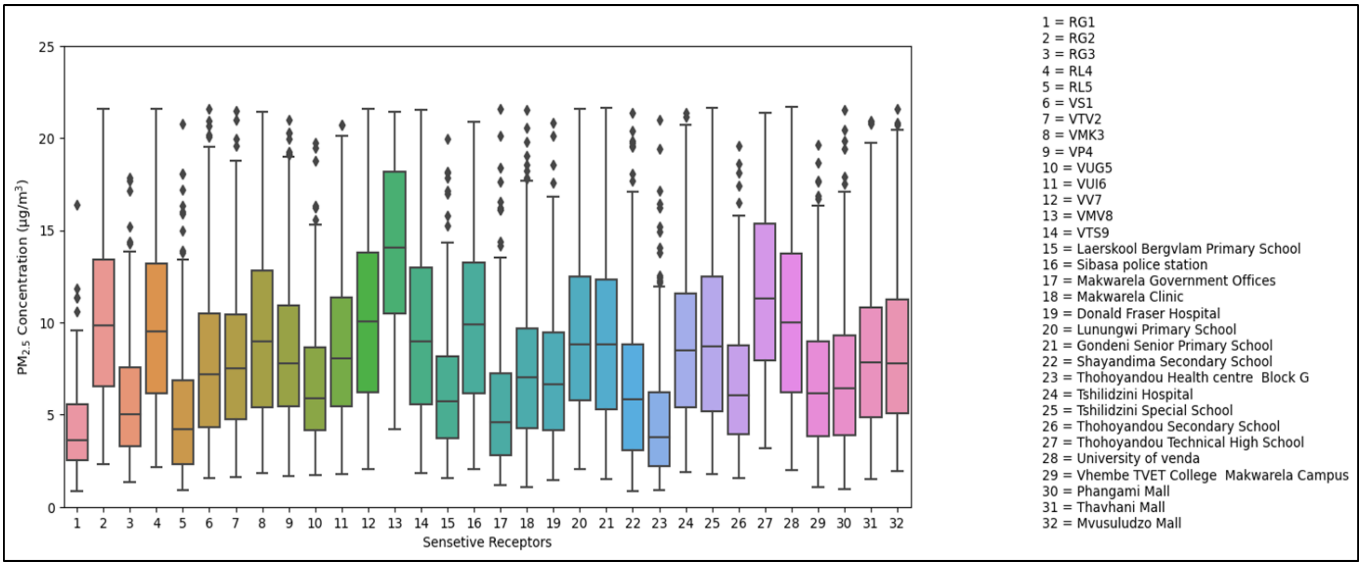



**Figure 5.15:** Box plots showing the modelled $PM_{2.5}$ concentration from all discrete receptors simulated at Thulamela.

### 5.3.3 Ambient $PM_{2.5}$ Dispersion Modelling Results Validation

The simulated data was validated using the average $PM_{2.5}$ concentrations observed at the respective monitoring stations, which were assumed as discrete sensitive receptors in the model setup. While simulated outcomes typically employ constant source emission rates (Hertwig *et al*., 2018), this research incorporated emission rate variation based on hourly and seasonal fluctuations to more accurately represent real-world scenarios. The modelled-to-monitored $PM_{2.5}$ ratio across the study area falls within the US EPA's recommended range (>0.5 to <2.0) for most cases, indicating reasonable model performance in those activities and locations (Table 5.3). Few cases, however, the values fall below 0.5, particularly in annual and some daily estimates, suggesting potential underestimation due to missing other pollution sources. No ratios exceeded 2.0, reinforcing that overestimation is not a significant concern, but refinements in emissions representation and local source contributions are necessary to improve model accuracy.

The analysis of measured versus modelled $PM_{2.5}$ concentrations in Thulamela reveals a consistent underestimation by the model across various timescales. For instance, at receptor RL, the observed hourly concentrations of 32.27 µg/m³ was lower compared to the modelled 47.12 µg/m³ (FAC2 = 0.25), and the daily observed concentration was 28.74 µg/m³, closely aligning with the modelled 27.88 µg/m³ (FAC2 = 1). Other receptors, such



© University of Venda

as VA and VB, show significant underpredictions, with observed hourly concentrations at VA being 53.2 µg/m³ compared to the modelled 31.84 µg/m³, and daily concentrations at VB at 31.10 µg/m³ versus the modelled 13.16 µg/m³. The study area exhibits the same pattern where measured $PM_{2.5}$ concentrations consistently exceed predicted values for hourly, daily, and yearly averages. The negative fractional bias (FB) and low FAC2 values further confirm the model's tendency to underpredict, while variations in normalised mean squared error (NMSE) highlight inconsistencies in model performance.

The scatter plot reveals a positive but weak correlation ($R^2 = 0.176$) between modelled and monitored $PM_{2.5}$ concentrations, indicating that the model struggles to capture the full variability of observed pollution levels (Figure 5.16). The low regression coefficient (0.37) indicates a consistent underprediction of elevated $PM_{2.5}$ levels, possibly resulting from unaccounted for pollution sources, meteorological uncertainties, and disparities in spatial representation. This pattern corresponds with the observed negative fractional bias (FB) in the model performance evaluation, underscoring the necessity for enhanced emissions inventories, improved meteorological data inputs, and the incorporation of missing sources of $PM_{2.5}$ such as biomass burning, unpaved road dust and emission activities related to agriculture and brick making. Although the model's overall performance remains weak in areas with high emissions due to non-estimated emission sources, it can contribute significantly to $PM_{2.5}$ air quality predictions for policy and regulatory purposes. This is particularly evident in areas with fewer emission sources, such as Lunungwi.

Figure 5.17 shows the temporal variation of the modelled results in comparison to the measured data from the monitoring sites in Thulamela, displaying both diurnal and monthly temporal variations. The diurnal variation (left panel) reveals that both the modelled and monitored $PM_{2.5}$ follow a distinct diurnal pattern, with peaks occurring during the morning (6–8 am) and evening (6–8 pm) rush hours; likely influenced by traffic emissions and domestic combustion activities. Similarly, the monthly trend (right panel) shows elevated $PM_{2.5}$ levels between May and September, attributed to seasonal changes, meteorological conditions, and increased biomass burning during colder



© University of Venda

months, with a decline observed in July 2021. This decline corresponds with COVID-19 partial lockdown restrictions on economic activities in South Africa, which reduced traffic and economic activities, thereby contributing to lower $PM_{2.5}$ emissions during that period.

***Table 5.3:*** *Comparison of measured and predicted average $PM_{2.5}$ and Model Performance Evaluation within Thulamela, South Africa*

| Activities | Category | Observed $PM_{2.5}$ (µg/m³) | Modelled $PM_{2.5}$ (µg/m³) | Ratio | FB | NMSE | FAC2 |
|---|---|---|---|---|---|---|---|
| RG | Hourly | 38.28 | 25.21 | 0.66 | -0.90 | 1.94 | 0.09 |
| | Daily | 30.86 | 15.83 | 0.51 | -0.60 | 0.46 | 1.00 |
| | Annual | 23.02 | 8.35 | 0.36 | -0.90 | 1.12 | 0.00 |
| RL | Hourly | 32.27 | 47.12 | 1.46 | -0.30 | 1.18 | 0.25 |
| | Daily | 28.74 | 27.88 | 0.97 | 0.00 | 0.00 | 1.00 |
| | Annual | 21.07 | 15.62 | 0.74 | -0.30 | 0.09 | 1.00 |
| VA | Hourly | 53.20 | 31.84 | 0.60 | -0.90 | 1.75 | 0.13 |
| | Daily | 41.72 | 19.02 | 0.46 | -0.80 | 0.65 | 0.00 |
| | Annual | 28.37 | 10.44 | 0.37 | -0.90 | 1.09 | 0.00 |
| VB | Hourly | 38.41 | 21.24 | 0.55 | -1.10 | 2.79 | 0.05 |
| | Daily | 31.10 | 13.16 | 0.42 | -0.80 | 0.79 | 0.00 |
| | Annual | 23.44 | 6.53 | 0.28 | -1.10 | 1.87 | 0.00 |
| VC | Hourly | 38.41 | 28.48 | 0.74 | -0.90 | 1.64 | 0.07 |
| | Daily | 32.18 | 17.47 | 0.54 | -0.60 | 0.38 | 1.00 |
| | Annual | 24.25 | 9.25 | 0.38 | -0.90 | 1.00 | 0.00 |
| **Average across the estimated sources and activities in the study area** | | | | | | | |
| Study Area | Hourly | 44.57 | 30.48 | 0.68 | -0.90 | 1.80 | 0.11 |
| | Daily | 36.48 | 17.66 | 0.48 | -0.70 | 0.55 | 0.00 |
| | Annual | 24.86 | 9.72 | 0.39 | -0.90 | 0.95 | 0.00 |



© University of Venda

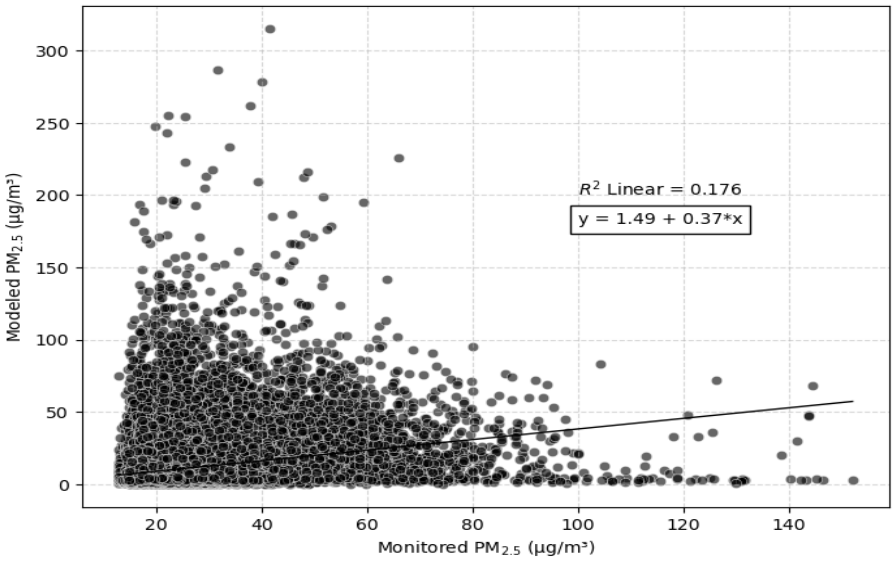



***Figure 5.16**: Scatter plot of monitored against modelled daily average $PM_{2.5}$ concentrations for Thulamela LM, South Africa.*

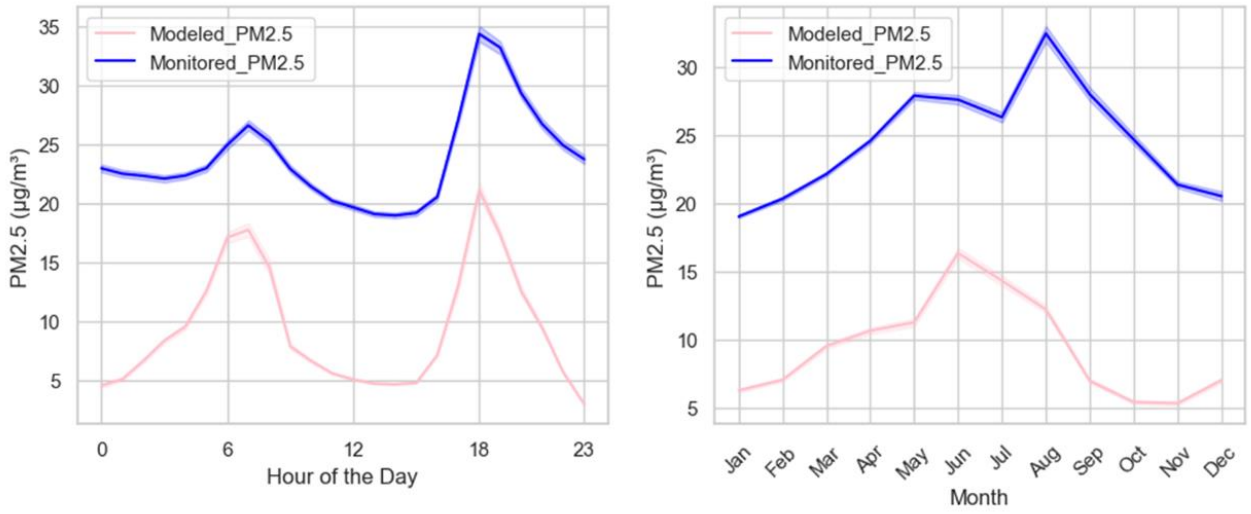



**Figure 5.17:** Temporal variation comparing the modelled and monitored $PM_{2.5}$ concentration in the study area of South Africa.

## 5.4 Limitations of the Study

While the AERMOD model provided valuable insights into $PM_{2.5}$ pollution in Thulamela Municipality, some factors affected its accuracy:

1. Underestimation of $PM_{2.5}$ Levels – The model consistently predicted lower $PM_{2.5}$ concentrations than observed, especially for annual and daily averages. This



© University of Venda

underestimation suggests that some $PM_{2.5}$ sources may not have been fully accounted for in the model except in Lunungwi village (RL group), which was higher or mostly aligned with the monitored $PM_{2.5}$.

2. Missing Pollution Sources – The study only accounted for residential wood combustion and vehicular emissions, leaving out other potential contributors like biomass burning, unpaved road dust, and agriculture and warehouse activities, which may have affected the accuracy of predictions.
3. Meteorological Data Constraints – The upper air meteorological data used in the model came from external sources due to a lack of local data, which may have caused uncertainties in pollutant transport and dispersion.

This study, despite these limitations, is a strong foundation for future research, which can help in improving air quality modelling and management strategies in Thulamela Municipality.

### 5.5 Conclusion

This study applied AERMOD to assess the atmospheric dispersion of $PM_{2.5}$ in Thulamela Municipality, focusing on emissions from residential wood combustion (RWC) and vehicular traffic. The findings highlight significant spatial and temporal variations in $PM_{2.5}$ levels, with pollution hotspots exceeding WHO and SA NAAQS guidelines, particularly in densely populated areas, reliant on solid fuels. The highest modelled daily $PM_{2.5}$ concentration (49.2 µg/m³) exceeded SA NAAQS (40 µg/m³) and WHO (15 µg/m³) standards, while the maximum annual concentration (19.7 µg/m³) approached the SA regulatory limit (20 µg/m³) but far exceeded the WHO standard (5 µg/m³).

Model validation showed that AERMOD's performance was within the USEPA recommended ratio (>0.5 to <2.0), yet the low correlation ($R^2$ = 0.176) and negative fractional bias (FB) confirmed a systematic underestimation of $PM_{2.5}$ levels, particularly in areas with multiple emission sources. The exception was at Lunungwi village (RL receptor group), where modelled $PM_{2.5}$ concentrations were either higher or closely aligned with observed values, suggesting localised overestimation. The underprediction



© University of Venda

of $PM_{2.5}$ levels indicates that some pollution sources were not fully accounted for, such as biomass burning, unpaved road dust, brickmaking, warehouse activities, and construction operations.

AERMOD, despite these limitations, successfully identified pollution hotspots and key emission contributors, offering valuable insights for air quality management, policy interventions, and regulatory actions. Future studies should expand emissions inventories, refine meteorological inputs, and integrate unaccounted pollution sources to improve model accuracy. Given its ability to capture general dispersion patterns, AERMOD remains a valuable tool for air quality assessment, emission reduction planning, and environmental policy development. Effective air quality management and continuous monitoring are essential, particularly in areas with high $PM_{2.5}$ exposure from residential fuel combustion, traffic emissions, and commercial activities, so as to minimize public health risks and ensure regulatory compliance.

**References**

Adeniran, J. A., Aremu, A. S., & Abdulraheem, K. A. (2024). Modelling of air emissions from open burning of municipal waste in Ilorin Metropolis, North Central Nigeria. Environmental Quality Management, 33(4), 795–808.

Arku, R. E., Brauer, M., Ahmed, S. H., AlHabib, K. F., Avezum, Á., Bo, J., Choudhury, T., Dans, A. M. L., Gupta, R., & Iqbal, R. (2020). Long-term exposure to outdoor and household air pollution and blood pressure in the Prospective Urban and Rural Epidemiological (PURE) study. *Environmental Pollution*, *262*, 114197.

Askariyeh, M. H., Kota, S. H., Vallamsundar, S., Zietsman, J., & Ying, Q. (2017). AERMOD for near-road pollutant dispersion: Evaluation of model performance with different emission source representations and low wind options. Transportation Research Part D: Transport and Environment, 57, 392–402.

Chang, J. C., & Hanna, S. R. (2004). Air quality model performance evaluation. Meteorology and Atmospheric Physics, 87(1), 167–196.

Chidhindi, P. (2022). *Using dispersion models as a regulatory tool in South Africa*. North-West University (South Africa).

Cimorelli, A.J., Perry, S.G., Venkatram, A., Weil, J.C., Paine, R.J., Wilson, R.B., Lee, R.F. and Peters, W.D., 2003. AERMOD description of model formulation. US Environmental Protection Agency Rep. EPA 454/R-03-002d. https://gaftp.epa.gov/aqmg/SCRAM/models/preferred/aermod/aermod_mfd.pdf

DEA (Department of Environmental Affairs), 2014. National Environmental Management:



© University of Venda

Air Quality Act, 2004 (Act No. 39 of 2004): Regulations regarding air dispersion modelling. Pretoria: Government Gazette, No. 37804, 11 July 2014.

Gulia, S., Shrivastava, A., Nema, A. K., & Khare, M. (2015). Assessment of Urban Air Quality around a Heritage Site Using AERMOD: A Case Study of Amritsar City, India. *Environmental Modeling and Assessment*, *20*(6), 599–608. https://doi.org/10.1007/s10666-015-9446-6.

Hertwig, D., Soulhac, L., Fuka, V., Auerswald, T., Carpentieri, M., Hayden, P., Robins, A., Xie, Z. T., & Coceal, O. (2018). Evaluation of fast atmospheric dispersion models in a regular street network. Environmental Fluid Mechanics, 18(4), 1007–1044. https://doi.org/10.1007/s10652-018-9587-7.

Holmes, N. S., & Morawska, L. (2006). A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available. Atmospheric Environment, 40(30), 5902–5928.

Huang, D., & Guo, H. (2019). Dispersion modeling of odour, gases, and respirable dust using AERMOD for poultry and dairy barns in the Canadian Prairies. Science of the Total Environment, 690, 620–628. https://doi.org/10.1016/j.scitotenv.2019.07.010

Lin, C., Huang, R.-J., Ceburnis, D., Buckley, P., Preissler, J., Wenger, J., Rinaldi, M., Facchini, M. C., O'Dowd, C., & Ovadnevaite, J. (2018). Extreme air pollution from residential solid fuel burning. *Nature Sustainability*, *1*(9), 512–517.

Lourens, A.S., Butler, T.M., Beukes, J.P., Van Zyl, P.G., Fourie, G.D. and Lawrence, M.G. (2016). Investigating atmospheric photochemistry in the Johannesburg-Pretoria megacity using a box model. South African Journal of Science, 112(1-2), pp.01-11.

Lourens, A.S., Fourie, G.D., Burger, J.W., Pienaar, J.J., Read, C.E., Jordaan, J.H., Van Zyl, P.G. and Beukes, J.P. (2011). Spatial and temporal assessment of gaseous pollutants in the Highveld of South Africa. South African Journal of Science, 107(1), pp.1-8.

Moletsane, S. D., Lindeque, F., Language, B., Nkosi, N. C., Adesina, J. A., Burger, R. P., Mkhatshwa, G., & Piketh, S. J. (2021). Research article Intra-urban variability of $PM_{2.5}$ in a dense, low-income settlement on the South African Highveld. *Clean Air Journal*, *31*(1), 1–9. https://doi.org/10.17159/CAJ/2021/31/1.9413.

Naidoo, S., Piketh, S. J., & Curtis, C. (2014). Quantification of emissions generated from domestic burning activities from townships in Johannesburg. *Clean Air Journal= Tydskrif Vir Skoon Lug*, *24*(1), 34–41.

Nkosi N.C. 2018. Quantifying particulate emissions from domestic burning in kwaDela Township, Mpumalanga. Potchefstroom: North-West University. Dissertation – MSc.

Odediran, E. T., Adeniran, J. A., & Yusuf, R. O. (2024). Assessment of seasonal variations in vehicular emissions at traffic intersections in Ibadan using AERMOD dispersion model. Environmental Quality Management.

Pauw, C. J., Montagu Murray, H., & Howard, M. A. (2022). The use of dirty fuels by low-income households on the South African Highveld. *Clean Air Journal*, *32*(1), 1–17.

Piketh, S. J., Wernecke, B., & Burger, R. (2016). Household air pollution in South African



© University of Venda

low-income settlements: a case study. *WIT Transactions on Ecology and the Environment*, *207*, 227–236.

Popoola, A. O., Jimoda, L. A., Olu-Arotiowa, O. A., Ogunkunle, O., Laseinde, O. T., Adebanjo, S. A., & Raji, W. A. (2023). Dispersion of PM and VOC pollutants from open burning of municipal solid wastes on host communities: emission inventory estimation and dispersion modelling study. Environmental Science: Atmospheres, 3(7), 1090–1109. https://doi.org/10.1039/d3ea00041.

Rezazadeh, A. A., Avami, A., Mashayekhi, M., & Kianbakhsh, A. (2022). Assessing the contribution of different sources in atmospheric dispersion of $PM_{2.5}$ and related health impact in a region of Qazvin, Iran. Air Quality, Atmosphere and Health, 0123456789. https://doi.org/10.1007/s11869-022-01163-6

Salva, J., Vanek, M., Schwarz, M., Gajtanska, M., Tonhauzer, P., & Ďuricová, A. (2021). An assessment of the on-road mobile sources contribution to particulate matter air pollution by AERMOD dispersion model. Sustainability, 13(22), 12748.

South African Weather Service (SAWS). https://www.weathersa.co.za/

Tshisi, L. (2023). Modelling ambient vehicle emissions in Zamdela using AERMOD. July.

U.S. EPA, 2016a. User's Guide for the AERMOD Meteorological Pre-processor (AERMET). Research Triangle Park, NC, Office of Air Quality Planning and Standards.

U.S. EPA, 2016b. User's Guide for the AERMOD Terrain Pre-processor (AERMAP). Research Triangle Park, NC, Office of Air Quality Planning and Standards.

U.S. EPA, 2019. User's Guide for the AMS/EPA Regulatory Model (AERMOD); U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division, Air Quality Modeling Group: Durham, NC, USA, 2019. (modeling).

USEPA: United States Environmental Protection Agency (2005). Guideline on Air

Villalvazo, L., Davila, E., & Reed, G. (2007). Guidance for air dispersion modeling. San Joaquin Valley Air Pollution Control District, 1(07).

World Health Organization. (2021). *Ambient (outdoor) air pollution*. Fact Sheet. Retrieved from https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

Yusuf, R. O., Tiamiyu, A. O., Adeniran, J. A., & Odediran, E. T. (2022). Modelling of ground level concentration of particulate matter in major Nigerian university airshed. LAUTECH Journal of Engineering and Technology, 16(1), 40-51.

Zou, B., Zhan, F. B., Wilson, J. G., & Zeng, Y. (2010). Performance of AERMOD at different time scales. *Simulation Modelling Practice and Theory*, *18*(5), 612–623.



© University of Venda

# CHAPTER SIX

# POTENTIAL HUMAN HEALTH RISKS FROM AMBIENT $PM_{2.5}$ EXPOSURE VIA INHALATION PATHWAY: A CASE STUDY OF THULAMELA MUNICIPALITY, LIMPOPO PROVINCE, SOUTH AFRICA

**Abstract**

*The deterioration of air quality due to ambient particulate matter ($PM_{2.5}$) driven by swift urbanisation, population increase, and heightened human activities, has become a pressing issue. Exposure to ambient $PM_{2.5}$ through inhalation is a significant global health concern, posing both non-carcinogenic and carcinogenic risks, particularly in low- and middle-income countries. High $PM_{2.5}$ levels above the National Ambient Air Quality Standards (NAAQS) limits are often linked to acute and chronic health effects, including persistent chronic respiratory diseases. This study, hence, examined $PM_{2.5}$ exposure and its associated human health risks in the semi-urban and rural areas of Thulamela Municipality, Limpopo Province, South Africa. This research focused on emissions from vehicles, residential fuel combustion, road dust, warehouse operations, and agricultural activities. $PM_{2.5}$ data were gathered from March 2021 to February 2022 and health risks were evaluated using the USEPA's human health risk assessment model. The study revealed that most average $PM_{2.5}$ concentrations, across most groups, exceeded South African National Ambient Air Quality Standards (SA NAAQS) of 20 µg/m³ and WHO's guideline of 5 µg/m³, particularly during winter. Non-carcinogenic risks measured by Hazard Quotient (HQ) values were higher during winter (1.10–0.40) than summer (0.77–0.28), with infants (1.10) and toddlers (0.997) being the most vulnerable. Adults showed the lowest HQs but remained at risk, especially those with pre-existing health conditions. Excess Lifetime Cancer Risk (ELCR) values ranged from $2.61 \times 10^{-3}$ (infants) to $2.35 \times 10^{-1}$ (adults), exceeding the USEPA and WHO acceptable threshold ($10^{-4}$), indicating substantial long-term cancer risks across all ages, with adults showing the highest risks. Seasonal variations, influenced by elevated $PM_{2.5}$ concentrations showed higher ELCR in winter (for example, $3.28 \times 10^{-1}$ for VA) as compared to summer ($2.03 \times 10^{-1}$ for RG and VA). These results emphasise the critical need for strategies to reduce high $PM_{2.5}$ levels and protect public health, especially for susceptible populations and high-risk areas.*

**Keywords**: Ambient air pollution, Particulate matter (PM2.5), Health risks, Exposure assessment, Excess lifetime cancer risk (ELCR), Hazard quotient, inhalation risks, Thulamela Municipality, South Africa.



© University of Venda

## 6.1 Introduction

The World Health Organization (WHO) has identified air pollution as one of the most serious environmental health challenges, with far-reaching consequences for both human health and the environment. Ambient air pollution, particularly fine particulate matter ($PM_{2.5}$), has emerged as a critical global health concern due to its significant contribution to millions of premature deaths and chronic illnesses (WHO, 2018a). The small size of $PM_{2.5}$ allows it to penetrate deep into the respiratory system and enter the bloodstream, causing systemic health effects (WHO, 2018a; WHO, 2021). Exposure to $PM_{2.5}$ has been linked to numerous health problems, ranging from acute respiratory disorders such as asthma and chronic obstructive pulmonary disease (COPD) to chronic conditions such as lung cancer, neurological disorders, and reduced life expectancy due to increased cardiovascular mortality caused by heart attacks and strokes (Heft-Neal *et al*., 2018; Miller & Xu, 2018; HEI, 2018; Costa *et al*., 2020; Vanker *et al*., 2023; Zhang *et al*., 2023; Abbah *et al*., 2024).

In recent years, rapid urbanisation and population growth have led to increased emissions of $PM_{2.5}$, presenting significant risks to human health, especially in terms of respiratory problems. Various human activities, including vehicle emissions, residential fuel combustion, industrial processes, and agricultural activities contribute to the release of fine particulate matter into the atmosphere (Li *et al*., 2017; Harrison, 2020; Zheng *et al*., 2020; Paisi *et al*., 2024). The increasing emission from anthropogenic sources is considered one of the most alarming environmental issues owing to its impact on human health and climate change (Chan & Yao, 2008; Sharma *et al*., 2018). According to the World Health Organization (WHO), air pollution is the leading global cause of death, accounting for 4.2 million fatalities annually from outdoor sources and 3.8 million from household smoke (WHO, 2016; WHO, 2018a; 2018b). Mortality rates are especially high among vulnerable populations, including children, the elderly, and individuals with pre-existing conditions, due to their physiological sensitivity and higher exposure levels relative to body size (Nkosi *et al*., 2016; Li *et al*., 2018; Ibrahim *et al*., 2021; Roche *et al*., 2024; Yu *et al*., 2024).



© University of Venda

Research have also shown that airborne particulate matter negatively impacts urban health and can lead to ecosystem disruption and health issues among the elderly (Pope Iii *et al*., 2002; Brook *et al*., 2004; Nkosi *et al*., 2016). The health effects become more severe with smaller particulate sizes (American Lung Association, 2013), as particles smaller than 10 micrometres can become trapped in the lungs, while those smaller than 2.5 micrometres can enter the bloodstream, causing significant health problems such as respiratory damage, myocardial infarctions, and increased cardiovascular mortality (Nkosi *et al*., 2016; Landrigan, 2017; Lelieveld *et al*., 2020; Yu *et al*., 2024). $PM_{2.5}$ levels vary between urban, semi-urban, and rural areas due to variations in emission sources such as traffic, residential fuel combustion, agricultural practices, warehouse activities, biomass burning, and industrial activities (Feig *et al*., 2016; Williams *et al*., 2021; Adeyemi *et al*., 2022; Roomaney *et al*., 2022). It has been shown that long-range transporting of pollutants from urban centres can also affect rural air quality (Freiman & Piketh, 2003; Piketh & Walton, 2004; Wang, 2007; Novela *et al*., 2020).

Worldwide, numerous studies have highlighted the severity of $PM_{2.5}$ pollution; for instance, $PM_{2.5}$ is linked to 3.3 million premature deaths annually in Asia, with significant health consequences (Lelieveld *et al*., 2015; Heft-Neal *et al*., 2018). In Africa, Safo-Adu *et al*. (2023) emphasised the need for in-depth investigations into the health impacts of particle air pollution, noting that $PM_{2.5}$ and $PM_{10}$ levels in many African cities frequently exceed WHO guidelines. This pollution contributes to respiratory diseases, pulmonary illnesses, and socioeconomic challenges. Sulaymon *et al*. (2023) further identified elevated $PM_{2.5}$ concentrations in Lagos, attributing them to residential and industrial sources, highlighting the need for effective emissions control strategies. This is complemented by Okimiji *et al*. (2021), who uncovered significant correlations between $PM_{2.5}$ and meteorological factors, such as temperature and humidity, in urban slum settlements of Lagos, underscoring the necessity for regular monitoring and evidence-based policy initiatives to mitigate air pollution.

Most studies on $PM_{2.5}$ in South African have focused on urban areas, where concentrations often exceed the recommended annual average limit of 20 µg/m³ under



© University of Venda

the South African National Ambient Air Quality Standards (SA NAAQS) (DEAT, 2009; DEA, 2012), and 5 µg/m³ under the WHO guidelines (WHO, 2021). These standards are particularly frequent in South Africa, where $PM_{2.5}$ levels vary significantly between urban, semi-urban and rural areas due to diverse sources and exposure intensities. Moreover, a survey by Kabudula *et al*. (2017), reported a rise in epidemiological transition of non-communicable diseases that are related to ambient air pollution exposure in rural areas of South Africa. Focusing on the health implications of $PM_{2.5}$, scientists often employ risk assessments to evaluate the health risks associated with exposure to various concentrations of air pollutants. Edlund *et al*. (2021), for example, conducted a health risk assessment on ambient $PM_{2.5}$ air pollution in a rural South African community, revealing substantial health risks despite compliance with SA NAAQS. Similarly, studies conducted in Thulamela Local Municipality (LM), a semi-urban and rural community in South Africa, have highlighted the importance of conducting health risk assessments in these areas to better inform public health policies and interventions (Novela *et al*., 2023; Aniyikaiye *et al*., 2024).

There, however, remains a notable shortage of research and significant gaps in the literature examining $PM_{2.5}$ exposure in semi-urban and rural settings. Consequently, these growing concerns about human health due to $PM_{2.5}$ highlight the need for more studies on the potential exposure and associated risks across different South African regions, including the selected case study area. Such research is essential for developing more effective public health interventions and policies.

This study, hence, addresses one aspect of the research gap by evaluating potential cancer and non-cancer health risks associated with $PM_{2.5}$ exposure in the Thulamela LM region of South Africa. Using the US EPA Human Health Risk Assessment (HHRA) model, the study examined exposure across different age groups, including infants, children, adults, and the elderly, in both commercial and residential areas. Results were compared to regulatory health risk limits, providing critical insights into the health risks faced by residents. Ultimately, the findings aim to inform public health interventions,



© University of Venda

improve air quality management, guide policy decisions, and protect vulnerable populations in South Africa.

## 6.2 Methodology

### 6.2.1 Study Area and the Sampling Sites

A detailed description of the study area is provided in Subsection 4.2.1, with additional information on the sampling sites available in both Subsection 4.2.1 and Appendix 4.1.

### 6.2.2 Data Analysis and Modelling

Detailed information on the data collection process and quality control measures can be found in Subsections 4.2.5 and 4.2.6. The collected $PM_{2.5}$ data were statistically analysed using Excel 2016, XLstat, Openair package in R Studio v.4.04, and ArcGIS v.10.804 (Carslaw & Ropkins, 2012; Carslaw, 2019). Furthermore, the health risks were evaluated using the USEPA Human Health Risk Assessment (HHRA) model (US EPA, 2009; Morakinyo *et al*., 2017). The results were assessed against the South African National Ambient Air Quality Standards (SA NAAQS) and World Health Organization (WHO) guidelines to examine the study area's air quality and regulatory compliance.

### 6.2.3 Human Health Risk Assessment (HHRA) Model of Inhalable Particulate Matter ($PM_{2.5}$)

Health risk assessment is a key predictive tool for evaluating the potential effects of human exposure to specific pollutants using measured $PM_{2.5}$ concentration data to forecast the health impact outcome resulting from pollutant exposure. This study employed the US EPA Risk Assessment (RA) methodology, known as the Human Health Risk Assessment (HHRA) model (US EPA, 1989; US EPA, 2009). The HHRA model consists of four main components: hazard identification, dose-response assessment, exposure assessment, and risk characterisation, as illustrated in Figure 6.1. Exposure to pollutants can occur through three primary pathways: ingestion, inhalation, and dermal contact (US EPA, 2009; Edokpayi *et al*., 2016; Morakinyo *et al*., 2017). This study focused on the inhalation pathway to estimate the health risks associated with $PM_{2.5}$ exposure among residents of Thulamela LM. Inhalation was identified as the predominant route of exposure, with variations in susceptibility across different age groups, considering



© University of Venda

the varying susceptibilities across different age groups. The population was categorised into four age groups: infants (birth to 1 year), toddlers (ages 2 - 5), children (ages 6 - 18), and adults (19 - 75 years), as outlined in Table 6.1.

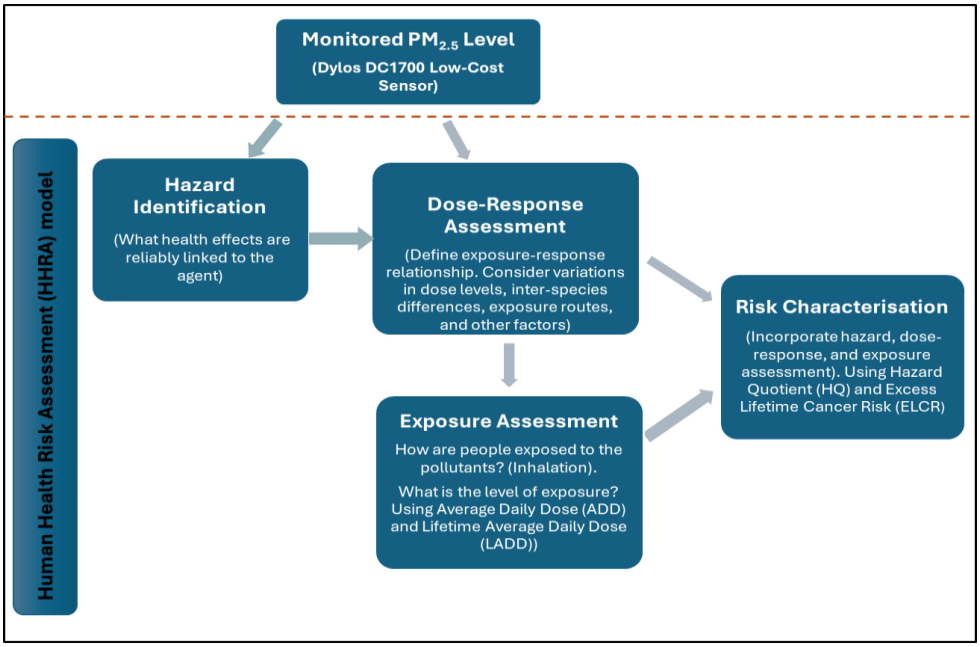



**Figure 6.1:** The flowchart of the human health risk assessment (HHRA) model (Source: This study).

### 6.2.3.1 Hazard identification

Hazard identification involves reviewing existing studies on pollutants to determine whether they are likely to cause harmful health effects in humans and under what environmental conditions these risks may occur. The main goal was to assess whether scientific evidence shows that the pollutant is detrimental to humans and under what conditions it could pose a risk. According to the US EPA (2023), this process also assesses the severity and extent of potential hazards. In this study, particulate matter ($PM_{2.5}$) was found to be harmful to human health through a comprehensive review of the existing research.

### 6.2.3.2 Dose-response assessment

The dose-response assessment examines the relationship between the level of exposure and the severity of possible health effects. This study did not perform specific dose-



© University of Venda

response calculations, instead, the amount of $PM_{2.5}$ inhaled was estimated based on the measured concentration levels and exposure durations. These measured $PM_{2.5}$ levels were then compared against the South African National Ambient Air Quality Standards (SA NAAQS) and WHO standards, which serve as the benchmark reference concentration (RfC) values as used in other similar studies (US EPA, 2011; Morakinyo *et al*., 2017; Mwase, 2020).

#### 6.2.3.3 Exposure assessment (EA)

An exposure assessment (EA) identifies the population exposed to pollutants and evaluates the extent and duration of their exposure. Using the US EPA model (US EPA, 1989 and 1992), exposure to $PM_{2.5}$ via inhalation was estimated in terms of the Average Daily Dose (ADD) (µg/kg/day) using Equation (6.1) (US EPA, 1989 and 1997) and the Lifetime Average Daily Dose (LADD) for carcinogenic exposure (US EPA, 2009; Izhar *et al*., 2016; Roy *et al*., 2019), as presented in Equation (6.2). These estimations considered multiple factors, such as inhalation rates, exposure frequency, body weight, and duration. In this study, the exposure to $PM_{2.5}$ was estimated based on measured ambient concentrations in the study area, and the EA was determined for the four age groups: infants, toddlers, children, and adults (Thabethe *et al*., 2014; Morakinyo *et al*., 2021). The recommended parameter values for calculating exposure doses (ADD and LADD) and the health risks (non-carcinogenic and carcinogenic) from ambient $PM_{2.5}$ via inhalation are shown in Table 6.1.

$$ADD = \frac{C \times IR \times EF \times ED}{BW \times AT_{nc}} \quad (6.1)$$

$$LADD = \frac{C \times EF \times ED \times ET}{AT_c} \quad (6.2)$$

#### 6.2.3.4 Human health risk assessment of $PM_{2.5}$ via inhalation route

Risk characterisation evaluates the potential health risks associated with pollutant exposure (US EPA, 2000; 1989). In this study, the potential health risks from the monitored $PM_{2.5}$ exposure in the ambient air of the study area were estimated. Non-carcinogenic effects were assessed using the hazard quotient (HQ), while the



© University of Venda

carcinogenic risk was evaluated using the Excess Lifetime Cancer Risk (ELCR) (US EPA, 1989 and 2009; Izhar *et al*., 2016; Morakinyo *et al*., 2017; Das *et al*., 2020).

*A. Non-carcinogenic health risk of $PM_{2.5}$ using hazard quotient (HQ)*

The Hazard Quotient (HQ) was used to assess the non-carcinogenic health risks of ambient particulate matter. It is calculated by dividing the Average Daily Dose (ADD) of $PM_{2.5}$ by the specific reference dose (RfD) or reference concentration (RfC), using Equation 6.3 (Thabethe *et al*., 2014; Morakinyo *et al*., 2017; Alfeus *et al*., 2022). The reference concentration (RfC) represents the estimated maximum allowable daily exposure to a pollutant without appreciable risk to human health, including sensitive groups, over a lifetime (Alfeus *et al*., 2022).

$$HQ = \frac{ADD}{RfC} \tag{6.3}$$

An HQ value of 1.0 is considered the safety threshold. If the HQ value is below 1.0, then the possibility of a pollutant having a negative impact on human health is low, even for sensitive individuals. Conversely, an HQ greater than 1 (HQ > 1) indicates a potential risk of adverse effects, particularly in sensitive individuals (US EPA, 2009).

*B. **Carcinogenic health risk of $PM_{2.5}$ using excess lifetime cancer risk (ELCR)***

Excess lifetime cancer risk (ELCR) estimates the increased likelihood of an individual developing cancer over their lifetime owing to cumulative exposure to a potential carcinogen. Based on the US EPA guidelines, the ELCR from human exposure to $PM_{2.5}$ may be determined by calculating the lifetime average daily dose (LADD). The carcinogenic health risk values related to inhalation exposure can be evaluated using either the inhalation unit risk (IUR) method or the cancer slope factor (CSF) method. In this study, the Excess Lifetime Cancer Risk (ELCR) was estimated using the IUR method in Equation (6.4), where the Lifetime Average Daily Dose (LADD) is shown in Equation (6.2). The IUR represents the likelihood of a health effect occurring with a unit increase in contaminant concentration (per µg/m³) (US EPA 2005, 2009; Izhar *et al*., 2016; Das *et al*., 2020; Morakinyo *et al*., 2021; Li *et al*., 2022).

$$ELCR = LADD \times IUR \tag{6.4}$$



© University of Venda

**Table 6.1:** Recommended parameter values for calculating exposure doses (ADD and LADD) and the health risks (Non-Carcinogenic and Carcinogenic) from ambient $PM_{2.5}$ via inhalation

| Parameters | Parameters definition | Units | Value for age Categories | | | | Reference |
|---|---|---|---|---|---|---|---|
| | | | Infants | Toddlers | Child | Adults | |
| C | Mean concentration of $PM_{2.5}$ in ambient air | ($\mu g/m^3$) | | | | | This study |
| InhR | Inhalation rate | ($m^3$/day) | 9.2 | 16.74 | 21.02 | 21.4 | US EPA, (1997) |
| BW | Body weight | kg | 11.3 | 22.6 | 45.3 | 71.8 | (USEPA, 1997; Matooane & Diab, 2003; Morakinyo *et al*., 2021) |
| EF | Exposure frequency | days/year | 350 | 350 | 350 | 350 | (USEPA, 1997; Morakinyo *et al*., 2017) |
| ED | Exposure duration | *years* | 1 | 6 | 12 | 30 | (Matooane & Diab, 2003; USEPA, 1997) |
| ET | *Exposure time* | *hours/day* | 1 | 8 | 6 | 3 | (USEPA, 1997; Matooane & Diab, 2003; Morakinyo *et al*., 2021) |
| $AT_{nc}$ | Averaging time non-carcinogens;<br><br>ATn= ED × 365 days | days | 365 | 2190 | 4380 | 10950 | (USEPA, 1997; Matooane & Diab, 2003) |
| $AT_c$ | Averaging time for carcinogens;<br><br>ATc =70 years × 365 days/year | days | 25550 | | | | (USEPA, 2009; DEA, 2010; Izhar *et al*., 2016) |
| IUR | Inhalation unit risk | $(\mu g/m^3)^{-1}$ | 0.008 | | | | (Greene & Morris, 2006; Odekanle *et al*., 2020) |

## 6.3 Results and Discussion

### 6.3.1 Spatio-Temporal Variability of Ambient $PM_{2.5}$ Levels Per Group in the Study Area

The 14 monitoring sites in the Thulamela LM study area were categorised into five groups based on land-use activities (Appendix 4.1). $PM_{2.5}$ concentration levels in the study area vary spatially, seasonally, monthly, and diurnally across all the groups in the study area,



© University of Venda

as discussed in Subsections 4.3.2 to 4.3.5. These variations in mean $PM_{2.5}$ concentrations were influenced by traffic, vehicular emissions, residential wood combustion (RWC), warehouse activities, and other anthropogenic factors. The seasonal trends showed that the highest $PM_{2.5}$ concentrations displayed distinct seasonal variability, with the highest levels recorded in winter (33.21 µg/m³ in group VA) and the lowest in summer (18.84 µg/m³ in group RL), as summarised in Table 6.2 and discussed in Subsections 4.3.3 and 4.3.4. These seasonal variations underscore the significant influence of meteorological conditions, such as frequent atmospheric inversions and low wind speeds, which restrict pollutant dispersion and exacerbate pollution levels (Freiman & Piketh, 2003). Specifically, atmospheric inversion layers during winter trap pollutants closer to the ground, leading to elevated $PM_{2.5}$ and heightened health risks compared to the summer months. These seasonal observations are consistent with previous studies conducted in Thohoyandou and other regions (Fu *et al*., 2018; Adesina *et al*., 2022; Wang et al., 2022; Howlett-Downing et al., 2023). This seasonal pattern is essential for understanding the potential health risks associated with $PM_{2.5}$. Higher concentrations, particularly in winter, can aggravate chronic health issues over time, especially among vulnerable people such as children, the elderly, and individuals with pre-existing respiratory or cardiovascular conditions.

Mixed-use areas with both traffic and residential wood-burning sources consistently show higher $PM_{2.5}$ levels than residential areas alone, although both follow similar seasonal trends. This pattern highlights persistent air quality issues in the study areas, contributing to increased exposure risks, health complications, and potentially higher hospital admissions during peak pollution periods. These findings emphasise the significant public health implications of high $PM_{2.5}$ levels and air quality issues in the study area. Across all groups, the average $PM_{2.5}$ from all groups exceeded both SA NAAQS of 20 µg/m³ (DEAT, 2009; DEA, 2012) and the WHO guideline of 5 µg/m³ (WHO, 2021), except in summer for groups RL, VB, and VC, where concentrations remained below the SA NAAQS guideline. This suggests that residents in the study area face increased health risks, particularly during winter, when $PM_{2.5}$ levels surpass these standards. Long-term exposure to elevated $PM_{2.5}$ concentrations is associated with serious health complications, including



© University of Venda

respiratory and cardiovascular diseases, and an increased risk of premature mortality, due to the ability of fine particulate matter to penetrate deeply into the lungs and enter the bloodstream (Gupta *et al*., 2016; Yunesian *et al*., 2019; Nawaz & Henze, 2020).

**Table 6.2:** *Seasonal mean of $PM_{2.5}$ Concentrations (µg/m³) by Group in the Study Area*

| Group | Seasons | Number of valid values N (values) | Mean ± Std. | References |
|---|---|---|---|---|
| RG | Autumn | 233 | 22.32 ± 3.48 | This study |
| | Spring | 0 | - | |
| | Summer | 149 | 20.59 ± 10.49 | |
| | Winter | 113 | 24.34 ± 5.53 | |
| | Annual | 4950 | 22.26 ± 6.89 | |
| RL | Autumn | 172 | 21.17 ± 3.97 | |
| | Spring | 24 | 23.61 ± 4.87 | |
| | Summer | 149 | **18.84 ± 3.79** | |
| | Winter | 59 | 22.78 ± 5.04 | |
| | Annual | 404 | 20.69 ± 4.42 | |
| VA | Autumn | 356 | 28.00 ± 6.19 | |
| | Spring | 86 | 26.76 ± 6.03 | |
| | Summer | 177 | 20.56 ± 3.27 | |
| | Winter | 204 | 33.21 ± 7.98 | |
| | Annual | 823 | 27.56 ± 7.55 | |
| VB | Autumn | 259 | 23.94 ± 3.98 | |
| | Spring | 54 | 23.57 ± 8.06 | |
| | Summer | 144 | **19.08 ± 2.41** | |
| | Winter | 84 | 24.35 ± 4.48 | |
| | Annual | 541 | 22.67 ± 4.83 | |
| VC | Autumn | 144 | 25.79 ± 4.59 | |
| | Spring | 26 | 21.35 ± 2.06 | |
| | Summer | 144 | **19.86 ± 2.78** | |
| | Winter | 0 | - | |
| | Annual | 314 | 22.7 ± 4.66 | |
| Study Area (Thulamela LM) | Autumn | 1164 | 24.68 ± 5.4 | |
| | Spring | 190 | 24.71 ± 6.48 | |
| | Summer | 763 | 19.82 ± 5.45 | |
| | Winter | 460 | 28.08 ± 7.98 | |
| | Annual | 2577 | 23.85 ± 6.69 | |
| **Reference Concentration (RfC) for $PM_{2.5}$ pollutant** | | | | |
| SA NAAQS | Daily (µg/m³) | - | 40 | (DEAT, 2009; DEA, 2012) |
| | Annual (µg/m³) | - | 20 | |
| WHO | Daily (µg/m³) | - | 15 | (WHO, 2021) |
| | Annual (µg/m³) | - | 5 | |



© University of Venda

### 6.3.2 Ambient $PM_{2.5}$ Dose-response Assessment

Hazard identification is a crucial step in the human health risk assessment (HHRA) model as it forms the basis or foundation for assessing the health impacts associated with environmental pollutants. Based on the several health hazards identified during the comprehensive literature review, this study implemented a robust assessment framework such as the HHRA model to evaluate the health risks posed by exposure to $PM_{2.5}$ in order to protect the populations in the study area. This is essential for implementing targeted interventions and protecting public health, particularly in vulnerable populations exposed to or affected by ambient air pollution.

In this study, the dose-response assessment was used to evaluate the chronic health implications of long-term $PM_{2.5}$ by comparing observed $PM_{2.5}$ concentrations with the SA NAAQS guideline (Table 6.2), which serves as the primary benchmark value for reference concentration (RfC) for assessing potential health risks, as previously used by other studies (Morakinyo *et al*., 2017; Mwase, 2020). A supplementary comparison with WHO guidelines is also provided in Appendix (6.1), offering additional context on global standards. These comparisons provide critical insights into the potential health risks residents face in the Thulamela LM study area. These findings indicate that average $PM_{2.5}$ levels across most groups generally exceeded both the SA NAAQS and WHO standards, except during summer for groups RL, VB, and VC, which fell below the SA NAAQS standard limit but exceeded the WHO standards. This sustained exposure to elevated $PM_{2.5}$ concentrations is linked to heightened health risks, including respiratory and cardiovascular conditions, and increased mortality rates. These results highlight the urgent need for targeted air quality management strategies to mitigate health risks and reduce $PM_{2.5}$ exposure in the study area.

### 6.3.3 Exposure Assessment and Health Risk Analysis of Ambient $PM_{2.5}$ Inhalation

This section presents the human exposure assessment and health risk characterisation based on measured ambient $PM_{2.5}$ concentrations from the collected monitoring data in the study area, providing a basis for evaluating the actual exposure levels experienced in the Thulamela LM area of Limpopo province. The analysis focused on inhalation as the



© University of Venda

primary exposure pathway, evaluating both non-carcinogenic and carcinogenic risks against established health benchmarks of the SA NAAQS and WHO. This evaluation provides critical insights into public health implications within Thulamela LM by assessing the potential health effects linked to long-term $PM_{2.5}$.

#### 6.3.3.1 Non-carcinogenic risk assessment using average daily dose via inhalation

***A. Average daily dose (ADD) exposure for land-use activity***

The Average Daily Dose ($ADD_{Inh}$) values of $PM_{2.5}$ exposure revealed significant age-related and seasonal differences across the Thulamela Municipality. The Average Daily Dose via inhalation ($ADD_{Inh}$) of $PM_{2.5}$ was estimated for four age groups (infants, toddlers, children, and adults) across regions categorised by land-use activities (RG, RL, VA, VB, and VC) during both winter and summer seasons, as shown in Table 6.3. The results revealed significant variations in $PM_{2.5}$ exposure levels based on age, season, and land use activities (groups), with higher $ADD_{Inh}$ values observed in winter than summer across all groups. This seasonal trend is attributed to increased air pollution driven by greater emissions from residential combustion of solid fuels, vehicle emissions, and reduced atmospheric dispersion due to temperatures during the winter season (Khan *et al*., 2020; Howlett-Downing *et al*., 2023; Room *et al*., 2023; Anbari *et al*., 2024; Paisi *et al*., 2024). Table 6.3 shows the $ADD_{Inh}$ for non-cancer risk via inhalation, expressed as a dose per unit of body weight (Table 6.4). It follows the age group pattern of infant > toddler > child > adult. In winter, infants had the highest ADD values between 25.93–17.78 µg/kg/day, while adults had the lowest value (9.49–6.51 µg/kg/day). Conversely, during the summer, exposure levels were lower across all age groups, with infant values ranging from 16.05 to 14.71 µg/kg/day and adults between 5.88 to 5.38 µg/kg/day.

The ADD pattern also varied according to group and season. In winter, the $ADD_{Inh}$ consistently followed this pattern: VA > VB = RG > RL (except for infants, where VB > RG), with values ranging from 25.93 to 6.51 µg/kg/day. Meanwhile, during summer, the pattern shifted to RG > VA > VC > VB > RL (except for adults, where VA = RG), with $ADD_{Inh}$ values ranging from 16.07 to 5.38 µg/kg/day. Among the groups assessed based on land use activities, the highest winter ADDInh was recorded in the VA group, with



© University of Venda

infants showing the highest exposure at 25.93 µg/kg/day, while adults had the lowest at 9.49 µg/kg/day. This trend reflects elevated concentrations of $PM_{2.5}$ in VA, likely due to intensified activities from traffic, vehicular emissions, dense residential sources, and warehouse operations. In contrast, the RL group exhibited lower $ADD_{Inh}$ values, with infants at 17.78 µg/kg/day and adults at 7.12 µg/kg/day, possibly due to fewer emission sources. During summer, the ADD values decreased across all groups; however, infants still showed the highest exposure levels of 16.05 µg/kg/day in VA, while the lowest was recorded for adults at RL (5.88 µg/kg/day). These disparities highlight the importance of localised exposure assessments in identifying high-risk areas where mitigation efforts should be prioritized.

The consistently higher $ADD_{Inh}$ values for infants and toddlers reflect their increased inhalation rates relative to body weight, making them the most vulnerable to $PM_{2.5}$ exposure (USEPA, 2011; WHO, 2018a; Edlund *et al*., 2021). High exposure levels, particularly during winter, pose significant health risks to younger populations, including an increased susceptibility to asthma, respiratory infections, and long-term lung damage (Rajak & Chattopadhyay, 2019). Adults have lower $ADD_{Inh}$ values, although, they may still face aggravated cardiovascular and respiratory conditions, contributing to increased hospital admissions and healthcare costs; for example, the overall study area during winter recorded an average $ADD_{Inh}$ of 21.92 µg/kg/day for infants, compared to 8.03 µg/kg/day for adults. This trend aligns with findings in Delhi, India, where infants showed elevated ADDInh levels due to their higher ventilation rates relative to body weight and proximity to pollution sources (Roy *et al*., 2019). Similarly, in Cape Town, South Africa, non-carcinogenic risk values for vulnerable groups exceeded thresholds due to the contribution of traffic and residential emissions to ambient $PM_{2.5}$ (Alfeus *et al*., 2022). These findings stress the need to focus health risk evaluations on vulnerable age groups, including adults with pre-existing health conditions.



© University of Venda

**Table 6.3:** Average daily dose (ADD) µg/kg/day of $PM_{2.5}$ exposure via inhalation pathway in the study area

| Group | Season | $ADD_{Inh}$ (µg/kg/day) | | | |
|---|---|---|---|---|---|
| | | Infant | Toddler | Child | Adult |
| RG | Winter | 19.00 | 17.29 | 10.83 | 6.96 |
| | Summer | 16.07 | 14.62 | 9.16 | 5.88 |
| RL | Winter | 17.78 | 16.18 | 10.14 | 6.51 |
| | Summer | 14.71 | 13.38 | 8.38 | 5.38 |
| VA | Winter | 25.93 | 23.59 | 14.78 | 9.49 |
| | Summer | 16.05 | 14.60 | 9.15 | 5.88 |
| VB | Winter | 19.01 | 17.30 | 10.83 | 6.96 |
| | Summer | 14.90 | 13.55 | 8.49 | 5.45 |
| VC | Winter | - | - | - | - |
| | Summer | 15.50 | 14.11 | 8.84 | 5.68 |
| Study area (Average $PM_{2.5}$) | Winter | 21.92 | 19.94 | 12.49 | 8.03 |
| | Summer | 15.47 | 14.08 | 8.82 | 5.66 |
| | Annual | 18.62 | 16.94 | 10.61 | 6.82 |

- ***$ADD_{Inh}$ across the study area and associated potential health implications***

Infants had the highest $ADD_{Inh}$ values, with 21.92 µg/kg/day during winter, decreasing to 15.47 µg/kg/day in summer. The annual average exposure for infants was 18.62 µg/kg/day, whereas for adults, it was significantly lower at 6.82 µg/kg/day, yet, despite this reduced exposure during summer, vulnerable groups such as infants, toddlers, and adults with pre-existing health conditions are still subject to substantial $PM_{2.5}$ levels, which may contribute potentially to cumulative health impacts over time (Landrigan *et al*., 2018; Miller & Xu, 2018). The elevated exposure, particularly in regions such as VA, VB, and RG during winter and RG, VA, and VC in summer, underscores the need for targeted interventions. Implementing localized risk mitigation strategies is crucial to reduce the health burden, especially in high-risk areas, and to improve respiratory health outcomes across the study area, focusing on the most affected age groups.

### *B. Non-carcinogenic risks of $PM_{2.5}$ exposure using hazard quotient (HQ)*

The non-carcinogenic risks of $PM_{2.5}$, via inhalation pathways are illustrated in Figure 6.2 using the South African SA NAAQS standard and the WHO standard (Appendix 6.1). The analysis showed a higher probability of $PM_{2.5}$, causing non-cancer health effects in winter



© University of Venda

than in summer, likely due to increased $PM_{2.5}$ levels from increased combustion activities and stagnant atmospheric conditions during colder months. Across all groups, HQ values followed a consistent age-related vulnerability pattern: **infant > toddler > child > adult (**Figure 6.2). When using the WHO guidelines as RfC (Appendix 6.1), the estimated HQ exceeded 1 for all age groups, suggesting potential health risks from chronic exposure.

Notably, the HQ values based on the WHO guidelines were higher than those calculated using the SA NAAQS standard, highlighting an increased risk when applying a more stringent reference limit. Based on the WHO (2021) RfC limit or benchmark for safe $PM_{2.5}$ levels, the HQs for adults, children, toddlers, and infants in the study area were calculated annually as 1.36, 2.12, 3.39, and 3.72, respectively. These results highlight that $PM_{2.5}$ exposure poses a greater risk for infants, toddlers, and children compared to adults (HQ 1.36) under the same environmental conditions or concentrations. These HQ findings align with and are supported by previous studies conducted in South Africa, which also identified increased susceptibility among children and infants (Thabethe *et al*., 2014; Morakinyo *et al*., 2017; Edlund *et al*., 2021; Alfeus *et al*., 2022). Notably, Edlund *et al*. (2021) reported an HQ of 1.18 for adults in Thohoyandou (a city within this study area) although their analysis was based on the outdated WHO standard of 10 µg/m³. While stricter WHO guidelines suggest an increased likelihood of potential non-carcinogenic risks, the HQ values derived from the SA NAAQS standard are more relevant for assessing health risks within the South African regulatory framework, as they align with local air quality guidelines. The subsequent discussion focuses on the HQ findings based on the SA NAAQS standard.

Infants and toddlers had the highest HQ values, reflecting their increased susceptibility to air pollution; for example, in group VA, the HQ for infants and toddlers was 1.30 and 1.18 in winter **(**Figure 6.2), respectively, compared to 0.80 and 0.73 in summer, indicating a seasonal reduction or decline in exposure risk. In these results, children exhibited higher HQ values than adults, although, their values were lower than those of infants and toddlers, indicating a decreasing trend in risk with age. Adults consistently had the lowest HQ values among the groups. Nevertheless, adults with pre-existing health conditions



© University of Venda

(such as asthma or cardiovascular disease) may still experience worsened health effects, highlighting the importance of targeted interventions for both young children and vulnerable adults. The HQ pattern varied by group and season, with the highest values in winter following this order: VA > VB = RG > RL, ranging from 1.30 to 0.33. In summer, the order shifted to RG = VA > VC > VB > RL, with the HQ values ranging from 0.80 to 0.27, however, no winter data were available for the VC group **(**Figure 6.2).

Previous studies have reported, elevated $PM_{2.5}$ exposure risks with HQ values exceeding 1 in several locations, indicating significant non-carcinogenic health risks. For example, HQ values of 1.52 were recorded in Thohoyandou (Edlund *et al*., 2021), and 1.96 in Cape Town (Alfeus *et al*., 2022) in South Africa. In Tehran, Iran, HQ values of 18.4 (Heydari *et al*., 2019) and 6.1 (2016) and 6.4 (2017) (Yunesian *et al*., 2019) were observed. Additional HQ values of 1.2 in Jharia coalfield, India (Roy *et al*., 2019) and 2.07 in Tangará da Serra, Brazil (2008) (De Oliveira *et al*., 2012) were also reported. In contrast, lower HQ values (below 1), indicating minimal non-carcinogenic health risks, were observed in Gwangju, South Korea (2016) (Kim *et al*., 2019), Kuala Lumpur, Malaysia (2021) (Othman *et al*., 2022) and Yunlin and Changhua, Taiwan (2020) (Chen *et al*., 2021), suggesting that non-carcinogenic effects were unlikely in these areas. The observed higher HQ risk values are attributed to elevated $PM_{2.5}$ emissions from various sources, which significantly contribute to increased non-carcinogenic health risks (Yunesian *et al*., 2019; Galán-Madruga *et al*., 2023; Anbari *et al*., 2024).

Overall, the HQ values in the Thulamela Municipality followed a consistent age group pattern: **infant > toddler > child > adult**, with HQ values ranging from 1.10 to 0.40 in winter and 0.77 to 0.28 in summer. On average, the HQ values for infants and toddlers were 1.10 and 0.997, respectively, in winter, indicating a potential non-carcinogenic health risk, while in summer, the values decreased to 0.77 and 0.70, reflecting even more reduced exposure. The HQ for adults was lower (0.40 to 0.28), although, it still warrants attention, especially for individuals with pre-existing health conditions. The findings of this study align with previous studies which highlight that infants and children are more vulnerable to the adverse effects of pollution compared to adults. This heightened



© University of Venda

vulnerability has been observed in Thohoyandou (Edlund *et al*., 2021; Novela *et al*., 2023) and other regions within and beyond South Africa (Izhar *et al*., 2016; Morakinyo *et al*., 2017; Mwase, 2020; Alfeus *et al*., 2022; Howlett-Downing *et al*., 2023). It is attributed to physiological differences, such as higher inhalation rates, smaller body weight, and larger lung surface areas (Thabethe *et al*., 2014; Alfeus *et al*., 2022). Additionally, individuals' higher air intake relative to body weight and underdeveloped immune systems, among other physiological factors, make them less capable of coping with environmental pollutants (US EPA, 2011). The elevated HQ values for infants and toddlers, particularly in high $PM_{2.5}$ areas such as VA, RG, VC, and VB, indicate substantial potential health risks for these vulnerable populations. Immediate action is required to reduce $PM_{2.5}$ risks, particularly during peak pollution, through strategies like emission reductions, community education, and public health interventions to protect at-risk populations.

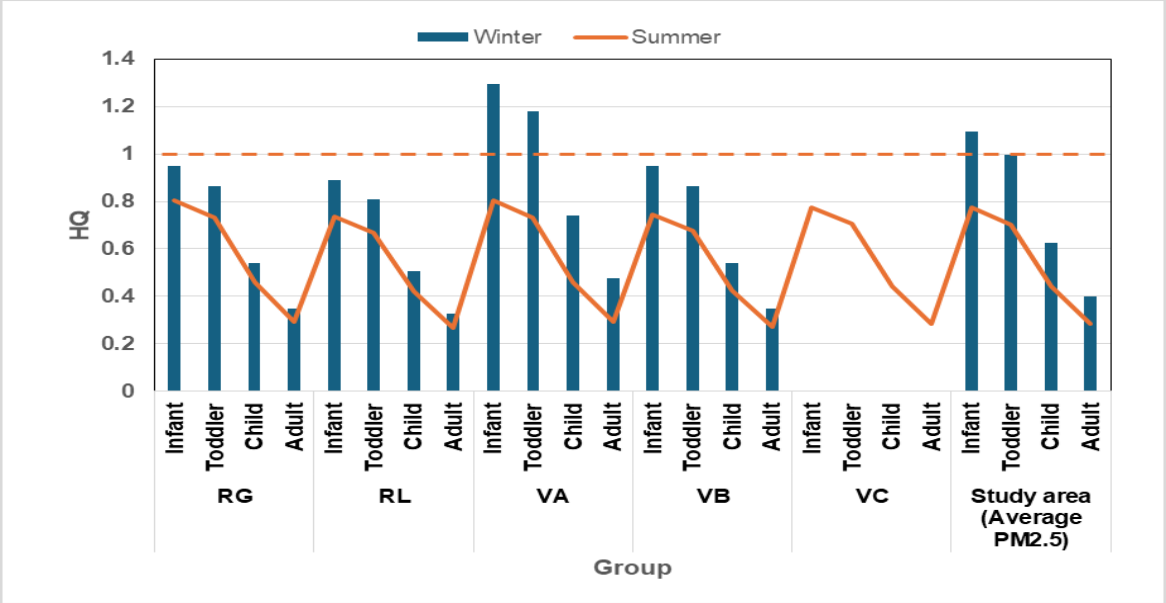



**Figure 6.2:** Hazard Quotient (HQ) for non-carcinogenic risks of $PM_{2.5}$ by land-use activity and across the study area. The blue bar represents the winter, the solid orange line is summer, and the dashed orange line represents the HQ standard limit for all age groups.



© University of Venda

### 6.3.3.2 Lifetime carcinogenic risk assessment of $PM_{2.5}$ exposure via inhalation route

*A. Estimation of lifetime average daily dose (LADD) of inhaling ambient* $PM_{2.5}$

Table 6.5 shows the estimated Lifetime Average Daily Dose (LADD) values of $PM_{2.5}$ exposure, with adults having the highest LADD values across the age groups and seasons in the study area, followed by children, toddlers, and infants. In winter, adults in group RG recorded a LADD of $3.00 \times 10^{1}$ µg/m³, which is significantly higher compared to infants ($3.33 \times 10^{-1}$ µg/m³), toddlers ($1.60 \times 10^{1}$ µg/m³), and children ($2.40 \times 10^{1}$ µg/m³). This trend indicates that adults may experience greater cumulative doses due to prolonged exposure and cumulative inhalation over a lifetime. Seasonal variations were evident, with higher LADD values peaking in winter than summer across all age groups, likely attributed to increased $PM_{2.5}$ concentrations in colder months due to meteorological conditions that trap particulate matter near the ground, consistent with observations in Tehran, and Delhi, Iran (Kushwaha *et al*., 2012; Heydari *et al*., 2019; Das *et al*., 2020; Suradi *et al*., 2021).

In the entire study area, the average LADD for adults was $3.46 \times 10^{1}$ µg/m³ during winter and decreased to $2.44 \times 10^{1}$ µg/m³ in summer (Table 6.5). The pattern of LADD values by group and season followed the order: VA > VB = RG > RL in winter, whereas in summer, it shifted to RG = VA > VC > VB > RL, except for infants and adults, that followed a different variation (where VB = RL for infants and RG > VA for Adult,). Notably, there was no LADD data for the VC group in winter, suggesting either a lower detection limit for $PM_{2.5}$ concentration or insufficient monitoring capacity. These higher LADD values during winter than summer across all age groups mirror the findings from Johannesburg and Pretoria, South Africa, where trace element concentrations in $PM_{2.5}$ were found to increase during colder months due to changes in atmospheric dispersion and human activities (Dubazana, 2020; Morakinyo *et al*., 2020, 2021). Across the study area, the LADD values for infants were consistently lower, ranging from $3.85 \times 10^{-1}$ µg/m³ in winter to $2.72 \times 10^{-1}$ µg/m³ in summer as shown in Table 6.5, yet, despite these lower values, infants are still more vulnerable to $PM_{2.5}$ exposure due to their physiological sensitivity. On the other hand, the annual LADD values across all age groups followed this trend: adults > children > toddlers > infants. This annual average LADD showed a clear age-



© University of Venda

related increase, from 3.27 × $10^{-1}$ µg/m³ for infants to 2.94 × $10^{1}$ µg/m³ for adults, emphasising the importance of accounting for both age-related differences and seasonal variations in exposure when assessing health risks.

**Table 6.4:** Lifetime Average Daily Dose (LADD) µg/m³ for $PM_{2.5}$ exposure in the study area

| Group | Season | LADD (µg/m³) | | | |
|---|---|---|---|---|---|
| | | Infants | Toddlers | Children | Adults |
| RG | Winter | 3.33 x $10^{-1}$ | 1.60 x $10^{1}$ | 2.40 x $10^{1}$ | 3.00 x $10^{1}$ |
| | Summer | 2.82 x $10^{-1}$ | 1.35 x $10^{1}$ | 2.03 x $10^{1}$ | 2.54 x $10^{1}$ |
| RL | Winter | 3.12 x $10^{-1}$ | 1.50 x $10^{1}$ | 2.25 x $10^{1}$ | 2.81 x $10^{1}$ |
| | Summer | 2.58 x $10^{-1}$ | 1.24 x $10^{1}$ | 1.86 x $10^{1}$ | 2.32 x $10^{1}$ |
| VA | Winter | 4.55 x $10^{-1}$ | 2.18 x $10^{1}$ | 3.28 x $10^{1}$ | 4.09 x $10^{1}$ |
| | Summer | 2.82 x $10^{-1}$ | 1.35 x $10^{1}$ | 2.03 x $10^{1}$ | 2.53 x $10^{1}$ |
| VB | Winter | 3.34 x $10^{-1}$ | 1.60 x $10^{1}$ | 2.40 x $10^{1}$ | 3.00 x $10^{1}$ |
| | Summer | 2.61 x $10^{-1}$ | 1.25 x $10^{1}$ | 1.88 x $10^{1}$ | 2.35 x $10^{1}$ |
| VC | Winter | 0 | 0 | 0 | 0 |
| | Summer | 2.72 x $10^{-1}$ | 1.31 x $10^{1}$ | 1.96 x $10^{1}$ | 2.45 x $10^{1}$ |
| Study Area | Winter | 3.85 x $10^{-1}$ | 1.85 x $10^{1}$ | 2.77 x $10^{1}$ | 3.46 x $10^{1}$ |
| | Summer | 2.72 x $10^{-1}$ | 1.30 x $10^{1}$ | 1.95 x $10^{1}$ | 2.44 x $10^{1}$ |
| | Annual | 3.27 x $10^{-1}$ | 1.57 x $10^{1}$ | 2.35 x $10^{1}$ | 2.94 x $10^{1}$ |

*B. Estimation of excess lifetime cancer risk (ELCR) of inhaling ambient $PM_{2.5}$*

The excess lifetime cancer risk (ELCR) characterisation values of $PM_{2.5}$ exposure via inhalation in Thulamela Municipality are shown in Figure 6.3. These values provide a quantitative estimate of the potential long-term carcinogenic risks associated with exposure to ambient $PM_{2.5}$ concentrations. The results revealed significant cancer risks across all age groups, with notable seasonal variations and trends observed in different groups (VA, VB, RG, RL, and VC) based on land-use activities, such as traffic, vehicle emissions, residential wood combustion as well as taxi rank, mall and warehouse activities. Comparable trends were also documented by Allen and Rector (2020) in the Adirondacks of New York, where urban and suburban activities like traffic and residential wood burning were primary contributors to elevated $PM_{2.5}$. This trend is consistent with findings within and outside South Africa, where $PM_{2.5}$ from vehicular and industrial



© University of Venda

sources heightened ELCR values in urban and peri-urban settings (Heydari *et al*., 2019; Yunesian *et al*., 2019; Morakinyo *et al*., 2021). Across the entire study area, the ELCR values followed a consistent age-related vulnerability pattern based on lifetime exposure for adult > children > toddler > infant. Adults consistently exhibited the highest cancer risk based on the ELCR values, which were significantly higher than the threshold of $10^{-4}$ in both winter and summer. In health risk assessment, an ILCR below $10^{-6}$ indicates negligible cancer risk, values between $10^{-6}$ and $10^{-4}$ suggest potential cancer risk, and those above $10^{-4}$ signify significant cancer risk (USEPA,1989; Das *et al*., 2020; Adeniran *et al*., 2024).

In winter, the ELCR values across all groups exceeded the standard threshold of $10^{-4}$, primarily due to elevated $PM_{2.5}$ concentrations (Figure 6.3). Group VA in semi-urban areas had the highest values ($3.28 \times 10^{-1}$), followed by RG in rural areas ($2.40 \times 10^{-1}$), indicating substantial cancer risks from long-term exposure, especially for adults. This age group consistently exhibited the greatest risks across all monitoring groups (Figure 6.3), reflecting the significant health burden associated with localised pollution sources. Infants showed the lowest ELCR values, ranging from $3.64 \times 10^{-3}$ (VA) to $2.50 \times 10^{-3}$ (RL), but these still exceeded the acceptable threshold, emphasising a significant health concern where elevated carcinogenic risks were noted even among the children (Roomaney *et al*., 2022). Seasonal trends showed reduced risks in summer (due to reduced $PM_{2.5}$ concentrations) as compared to winter, although ELCR values for adults in Groups RG and VA remained concerningly high ($2.03 \times 10^{-1}$), while the lowest was observed at RL ($1.86 \times 10^{-1}$). The observed pattern of ELCR values shifted in summer to RG = VA > VC > VB > RL. Notably, elevated risks in RG and VA were likely driven by residential wood combustion, agricultural burning, road dust, and traffic emissions, alongside vehicle activities at malls and warehouse operations in Shayandima's industrial areas, close to the VA group. Additionally, VC (a semi-urban area) exhibited higher ELCR values than VB and RL, possibly due to increased exposure from local taxi rank activities and fuel wood combustion around the monitored sites.



© University of Venda

The recorded ELCR values for the entire study period between March 2021 and February 2022 (Figure 6.3) were $2.61 \times 10^{-3}$ for infants, $1.25 \times 10^{-1}$ for toddlers, $1.88 \times 10^{-1}$ for children, and $2.35 \times 10^{-1}$ for adults. All these values were well above the threshold ($10^{-4}$), indicating a significant cancer risk from prolonged exposure to ambient $PM_{2.5}$. Adults recorded the highest ELCR value ($2.35 \times 10^{-1}$), reflecting the cumulative effect of prolonged exposure, which aligns with other findings in South Africa, Iran (Tehran), India (Delhi), Bangladesh and Malaysia, where urban populations face elevated cancer risks from high PM2.5 levels (Sulong *et al*., 2017; Heydari *et al*., 2019; Roy *et al*., 2019; Morakinyo *et al*., 2020; Saju *et al*., 2023). Similarly, the results resonate with studies in Khulna, Bangladesh, highlighting elevated cancer risks in areas with high $PM_{2.5}$ concentrations due to industrial and vehicular emissions (Saju *et al*., 2023).

This shows that the residents of Thulamela Municipality face a significant lifetime cancer risk due to $PM_{2.5}$ exposure, with adults bearing the highest burden. Infants and toddlers, despite lower absolute exposures, remain vulnerable due to physiological sensitivity, as corroborated by studies from Gwangju, South Korea (Kim *et al*., 2019). The $PM_{2.5}$ ELCR values in the different age groups showed that all of them have exceeded the carcinogenic limit value recommended by the US EPA (Figure 6.3), particularly for adults and children in semi-urban areas (VA), which aligns with findings from previous South Africa studies (Morakinyo *et al*., 2021; Howlett-Downing *et al*., 2023), and in other countries (Kushwaha *et al*., 2012; Izhar *et al*., 2016; Suradi *et al*., 2021). These studies similarly reported elevated cancer risks in regions with high $PM_{2.5}$ concentrations due to its bound metals.

The significant risks identified in this study are attributed to local pollution sources, including vehicular emissions, residential wood combustion, warehouse operations, agricultural burning, and road dust. Identified notable cancer risks in both rural (RG) and semi-urban areas (VA), highlighting the critical role of these local activities in contributing to the overall carcinogenic risks in Thulamela Municipality. These findings underscore the urgent need for targeted risk mitigation strategies to address key local pollution sources, particularly in semi-urban areas like VA and rural zones like RG, to safeguard public health (Suradi *et al*., 2021; Enitan *et al*., 2022; Muyemeki *et al*., 2022).



© University of Venda

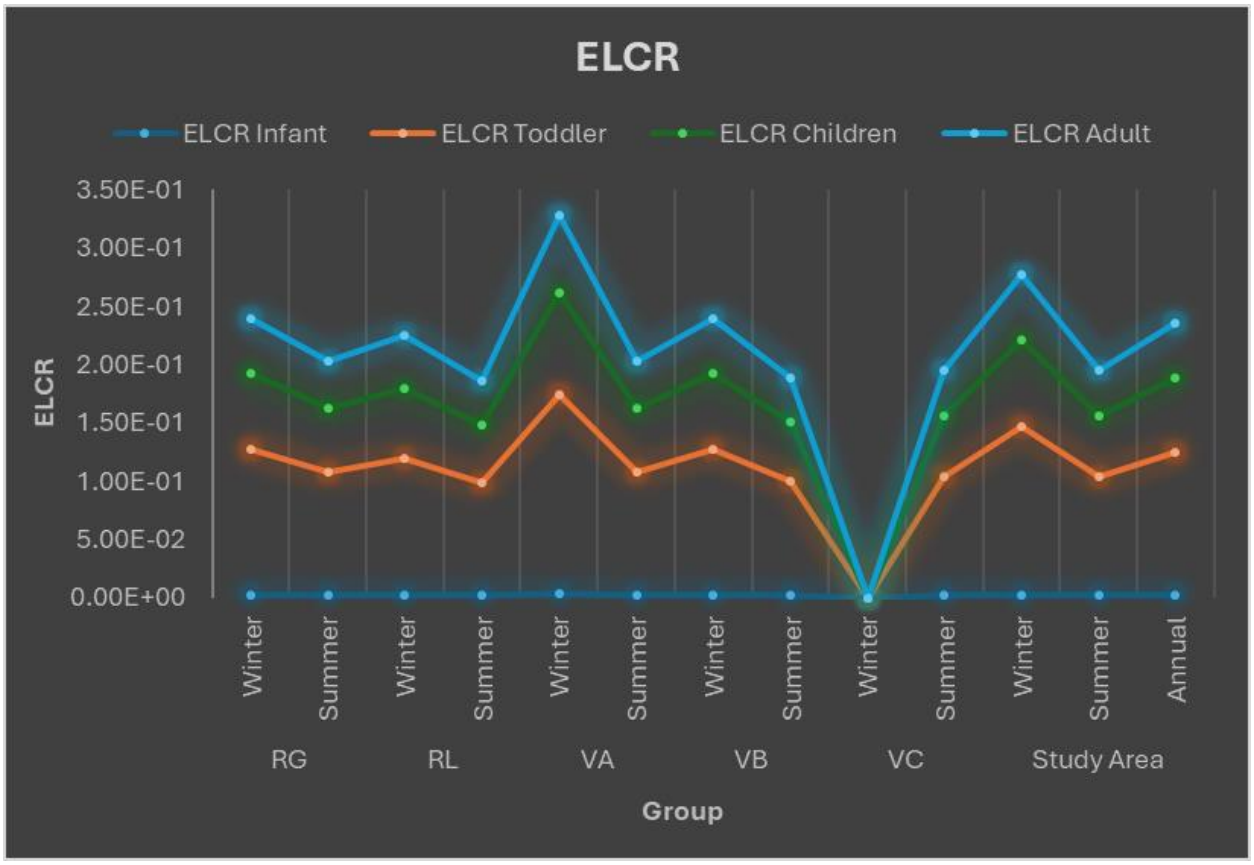



**Figure 6.3:** Excess lifetime cancer risk (ELCR) for $PM_{2.5}$ exposure in Thulamela municipality.

## 6.4 Limitations of the Study

This study is based on the primary assumption that the Dylos DC1700, as an affordable low-cost sensor will accurately measures $PM_{2.5}$ concentrations and reflect exposure levels for residents of Thulamela Local Municipality, however, several constraints were encountered during the investigation. The first limitation was that the collected monitoring data may not fully capture spatiotemporal variations, particularly during localised pollution events or extreme weather conditions. The study's findings are limited by the absence of winter data and incomplete data for some groups due to insufficient sunlight and battery recharging issues. The 14 monitoring sites might not fully reflect personal PM exposure due to variations in other unaccounted anthropogenic activities at sensor placement sites. These factors could have influenced the sensor's response, leading to potential underestimations or overestimations of $PM_{2.5}$ levels in specific periods or locations.



© University of Venda

Furthermore, the HHRA framework may also have overestimated or underestimated the health risks due to uncertainties regarding population activity patterns and a lack of exposure data at the individual level; despite these limitations, the HHRA framework remains a valuable tool for estimating health risks and guiding risk management decisions. Future studies could benefit from source apportionment analyses to better understand pollution source contributions, $PM_{2.5}$ bound metal and to be able to refine mitigation strategies.

### 6.5 Conclusions

$PM_{2.5}$ exposure is strongly associated with various health issues, including respiratory diseases, cardiovascular problems, stroke, increased mortality, and cancer, affecting individuals across all age groups. This study highlights the significant health risks associated with ambient $PM_{2.5}$ exposure in Thulamela Municipality, South Africa, where $PM_{2.5}$ concentrations frequently exceed both SA NAAQS and WHO guidelines, particularly during winter, when levels are highest. The health risk assessments using ADD, LADD, HQ, and ELCR metrics revealed raised non-carcinogenic and carcinogenic risks across all age groups, in the study area. Hazard Quotient (HQ) values were notably higher during winter, posing significant risks, especially for younger populations, such as infants and toddlers, while adults with pre-existing health conditions also faced increased health concerns. Excess Lifetime Cancer Risk (ELCR) values consistently exceeded the threshold of $10^{-4}$ across all age groups, indicating significant severe long-term cancer risks from exposure to $PM_{2.5}$. Adults exhibited the highest cancer risks, followed by children and toddlers, and although infants had the lowest ELCR values, their risks remained elevated due to heightened physiological vulnerability and prolonged exposure. The highest risks were observed in semi-urban and rural areas, driven by pollution sources like traffic emissions, residential fuel combustion, warehouse activities, and agricultural burning. These results emphasise the urgent need for targeted air quality management strategies to reduce $PM_{2.5}$ levels, particularly in high-risk areas associated with these pollution sources. Pollution control measures, such as promoting cleaner fuels, reducing electricity costs, and improving traffic management, are crucial for mitigating carcinogenic and non-carcinogenic health risks. To achieve this, continuous monitoring



© University of Venda

of $PM_{2.5}$ levels coupled with source apportionment studies are vital in identifying the major pollution sources and implementing effective risk management interventions. Additionally, prioritising vulnerable populations and strict enforcement of air quality standards are essential to minimise the health burden of $PM_{2.5}$-related diseases in the Thulamela Municipality of South Africa.

## Reference

Abbah, A. P., Xu, S., & Johannessen, A. (2024). Long-term health effects of outdoor air pollution on asthma and respiratory symptoms among adults in low-and middle-income countries (LMICs): a systematic review and meta-analysis. *Frontiers in Environmental Health*, *3*, 1352786.

Adeniran, J. A., Odediran, E. T., Ogunlade, B. T., Adeagbo, T. O., Akanbi, O. F., & Adesina, O. A. (2024). Assessment of the Pollution Levels, Sources, and Exposure Risks of Polychlorinated Biphenyls (PCBs) in Urban Park Dusts within Lagos Metropolis. *Environmental Quality Management*, *34*(1), 1–11. https://doi.org/10.1002/tqem.22275.

Adesina, J. A., Piketh, S. J., Burger, R. P., & Mkhatshwa, G. (2022). Assessment of criteria pollutants contributions from coal-fired plants and domestic solid fuel combustion at the South African industrial highveld. *Cleaner Engineering and Technology*, *6*(June 2021), 100358. https://doi.org/10.1016/j.clet.2021.100358.

Adeyemi, A., Molnar, P., Boman, J., & Wichmann, J. (2022). Particulate matter (PM2. 5) characterization, air quality level and origin of air masses in an urban background in Pretoria. *Archives of Environmental Contamination and Toxicology*, *83*(1), 77–94.

Alfeus, A., Molnar, P., Boman, J., Shirinde, J., & Wichmann, J. (2022). Inhalation health risk assessment of ambient $PM_{2.5}$ and associated trace elements in Cape Town, South Africa. *Human and Ecological Risk Assessment*, *28*(8), 917–929. https://doi.org/10.1080/10807039.2022.2110449.

Allen, G., & Rector, L. (2020). Characterization of residential woodsmoke $PM_{2.5}$ in the Adirondacks of New York. *Aerosol and Air Quality Research*, *20*(11), 2419–2432.

American Lung Association. (2013). *State of the Air 2013*. American Lung Association. https://www.lung.org/getmedia/338b0c3c-6bf8-480f-9e6e-b93868c6c476/SOTA-2023.pdf.

Anbari, K., Sicard, P., Khaniabadi, Y. O., Naqvi, H. R., Fard, R. F., & Rashidi, R. (2024). $PM_{2.5}$ and $PM_{10}$-related carcinogenic and non-carcinogenic risk assessment in Iran. *Journal of Atmospheric Chemistry*, *81*(1), 10.

Aniyikaiye, T. E., Piketh, S. J., & Edokpayi, J. N. (2024). A spatial approach to assessing $PM_{2.5}$ exposure level of a brickmaking community in South Africa. *Journal of the Air & Waste Management Association*, 74(5), 345-358.

Brook, R. D., Franklin, B., Cascio, W., Hong, Y., Howard, G., Lipsett, M., Luepker, R.,



© University of Venda

Mittleman, M., Samet, J., & Smith Jr, S. C. (2004). Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association. *Circulation*, *109*(21), 2655–2671.

Carslaw, D C. (2019). *The openair manual–open-source tools for analysing air pollution data, Manual for version 2.6–6, University of York*.

Carslaw, David C., & Ropkins, K. (2012). Openair - An r package for air quality data analysis. *Environmental Modelling and Software*, *27–28*, 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008.

Chan, C. K., & Yao, X. (2008). Air pollution in mega cities in China. *Atmospheric Environment*, *42*(1), 1–42.

Chen, C.-R., Lai, H.-C., Liao, M.-I., Hsiao, M.-C., & Ma, H. (2021). Health risk assessment of trace elements of ambient $PM_{2.5}$ under monsoon patterns. *Chemosphere*, *264*, 128462.

Costa, L. G., Cole, T. B., Dao, K., Chang, Y.-C., Coburn, J., & Garrick, J. M. (2020). Effects of air pollution on the nervous system and its possible role in neurodevelopmental and neurodegenerative disorders. *Pharmacology & Therapeutics*, *210*, 107523.

Das, A., Singh, G., Habib, G., & Kumar, A. (2020). Non-carcinogenic and Carcinogenic Risk Assessment of Trace Elements of $PM_{2.5}$ During Winter and Pre-monsoon Seasons in Delhi: A Case Study. *Exposure and Health*, *12*(1), 63–77. https://doi.org/10.1007/s12403-018-0285-y.

De Oliveira, B. F. A., Ignotti, E., Artaxo, P., do Nascimento Saldiva, P. H., Junger, W. L., & Hacon, S. (2012). Risk assessment of $PM_{2.5}$ to child residents in Brazilian Amazon region with biofuel production. *Environmental Health*, *11*, 1–11.

DEAT, (2009), South Africa. National Environmental Management: Air quality Act 2004 (Act No 39 of 2004)- National Ambient Air Quality, Government gazette No. 32816. Department of Environmental Affairs and Tourism, Pretoria. https://www.environment.gov.za/sites/default/files/legislations/nemaqa_airquality_g32816gon1210.pdf.

Department of Environmental Affairs (DEA), (2010). Framework for the Management of Contaminated Land. https://sawic.environment.gov.za/documents/562.pdf. Accessed on 20th of August 2024.

DEA, Department of Environmental Affairs, South Africa (2012). National ambient air quality standard for particulate matter with aerodynamic diameter less than 2.5 micron metres ($PM_{2.5}$). https://www.gov.za/sites/default/files/gcis_document/201409/35463gon486.pdf.

Dubazana, M. (2020). *Inhalation human health risk assessment: case study of a South African international airport*. University of the Witwatersrand, Faculty of Science.

Edlund, K. K., Killman, F., Molnár, P., Boman, J., Stockfelt, L., & Wichmann, J. (2021). Health risk assessment of $PM_{2.5}$ and $PM_{2.5}$-bound trace elements in Thohoyandou,



© University of Venda

South Africa. *International Journal of Environmental Research and Public Health*, *18*(3), 1359.

Edokpayi, J. N., Odiyo, J. O., Popoola, O. E., & Msagati, T. A. M. (2016). Determination and distribution of polycyclic aromatic hydrocarbons in rivers, sediments and wastewater effluents in Vhembe District, South Africa. *International Journal of Environmental Research and Public Health*, *13*(4), 387.

Enitan, I. T., Durowoju, O. S., Edokpayi, J. N., & Odiyo, J. O. (2022). A review of air pollution mitigation approach using air pollution tolerance index (APTI) and anticipated performance index (API). *Atmosphere*, *13*(3), 374.

Feig, G. T., Naidoo, S., & Ncgukana, N. (2016). Assessment of ambient air pollution in the Waterberg Priority Area 2012-2015. *Clean Air Journal*, *26*(1), 21–28. https://doi.org/10.17159/2410-972x/2016/v26n1a9.

Freiman, M. T., & Piketh, S. J. (2003). Air transport into and out of the industrial Highveld region of South Africa. *Journal of Applied Meteorology and Climatology*, *42*(7), 994–1002.

Fu, W., Chen, Z., Zhu, Z., Liu, Q., Konijnendijk van den Bosch, C. C., Qi, J., Wang, M., Dang, E., & Dong, J. (2018). Spatial and temporal variations of six criteria air pollutants in fujian province, china. *International Journal of Environmental Research and Public Health*, *15*(12). https://doi.org/10.3390/ijerph15122846.

Galán-Madruga, D., Cárdenas-Escudero, J., Broomandi, P., Cáceres, J. O., & González, M. D. C. (2023). Evaluating urban indoor and outdoor PM10-bound organochlorine pesticides. Air quality status and health impact. *Building and Environment*, *228*, 109818.

Greene, N. A., & Morris, V. R. (2006). Assessment of public health risks associated with atmospheric exposure to $PM_{2.5}$ in Washington, DC, USA. *International Journal of Environmental Research and Public Health*, *3*(1), 86–97.

Gupta, S., Agarwal, R., & Mittal, S. K. (2016). Respiratory health concerns in children at some strategic locations from high PM levels during crop residue burning episodes. *Atmospheric Environment*, *137*, 127–134.

Harrison, R. M. (2020). Airborne particulate matter. *Philosophical Transactions of the Royal Society A*, *378*(2183), 20190319.

HEI, Health Effects Institute (2018). State of Global Air 2018: Over 7 billion people face unsafe air. https://www.healtheffects.org/announcements/state-global-air-2018-over-7-billion-people-face-unsafe-air.

Heft-Neal, S., Burney, J., Bendavid, E., & Burke, M. (2018). Robust relationship between air quality and infant mortality in Africa. *Nature*, *559*(7713), 254–258.

Heydari, G., Taghizdeh, F., Fazlzadeh, M., Jafari, A. J., Asadgol, Z., Mehrizi, E. A., Moradi, M., & Arfaeinia, H. (2019). Levels and health risk assessments of particulate matters ($PM_{2.5}$ and $PM_{10}$) in indoor/outdoor air of waterpipe cafés in Tehran, Iran. *Environmental Science and Pollution Research*, *26*, 7205–7215.

Howlett-Downing, C., Boman, J., Molnár, P., Shirinde, J., & Wichmann, J. (2023). Health



© University of Venda

risk assessment of $PM_{2.5}$ and $PM_{2.5}$-bound trace elements in Pretoria, South Africa. *Journal of Environmental Science and Health, Part A*, *58*(4), 342–358.

Ibrahim, M. F., Hod, R., Nawi, A. M., & Sahani, M. (2021). Association between ambient air pollution and childhood respiratory diseases in low- and middle-income Asian countries: A systematic review. *Atmospheric Environment*, *256*, 118422. https://doi.org/10.1016/j.atmosenv.2021.118422.

Izhar, S., Goel, A., Chakraborty, A., & Gupta, T. (2016). Annual trends in occurrence of submicron particles in ambient air and health risk posed by particle bound metals. *Chemosphere*, *146*, 582–590. https://doi.org/10.1016/j.chemosphere.2015.12.039.

Kabudula, C. W., Houle, B., Collinson, M. A., Kahn, K., Gómez-Olivé, F. X., Clark, S. J., & Tollman, S. (2017). Progression of the epidemiological transition in a rural South African setting: findings from population surveillance in Agincourt, 1993–2013. *BMC Public Health*, *17*, 1–15.

Khan, S. A., Muhammad, S., Nazir, S., & Shah, F. A. (2020). Heavy metals bounded to particulate matter in the residential and industrial sites of Islamabad, Pakistan: Implications for non-cancer and cancer risks: Evaluation of heavy metals bounded to particulate matter for non-cancer and cancer risks. *Environmental Technology and Innovation*, *19*, 100822. https://doi.org/10.1016/j.eti.2020.100822.

Kim, I., Lee, K., Lee, S., & Kim, S. D. (2019). Characteristics and health effects of $PM_{2.5}$ emissions from various sources in Gwangju, South Korea. *Science of the Total Environment*, *696*, 133890. https://doi.org/10.1016/j.scitotenv.2019.133890.

Kushwaha, R., Lal, H., Srivastava, A., & Jain, V. K. (2012). Human exposure to particulate matter and their risk assessment over Delhi, India. *National Academy Science Letters*, *35*, 497–504.

Landrigan, P. J. (2017). Air pollution and health. *The Lancet Public Health*, *2*(1), e4–e5.

Landrigan, P. J., Fuller, R., Acosta, N. J. R., Adeyi, O., Arnold, R., Baldé, A. B., Bertollini, R., Bose-O'Reilly, S., Boufford, J. I., & Breysse, P. N. (2018). The Lancet Commission on pollution and health. *The Lancet*, *391*(10119), 462–512.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, *525*(7569), 367–371.

Lelieveld, J., Pozzer, A., Pöschl, U., Fnais, M., Haines, A., & Münzel, T. (2020). Loss of life expectancy from air pollution compared to other risk factors: a worldwide perspective. *Cardiovascular Research*, *116*(11), 1910–1917.

Li, J., Liu, H., Lv, Z., Zhao, R., Deng, F., Wang, C., Qin, A., & Yang, X. (2018). Estimation of $PM_{2.5}$ mortality burden in China with new exposure estimation and local concentration-response function. *Environmental Pollution*, *243*, 1710–1718.

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., Zhang, Q., & He, K. (2017). Anthropogenic emission inventories in China: A review. *National Science Review*, *4*(6), 834–866. https://doi.org/10.1093/nsr/nwx150

Li, Y., Munoz-Ibanez, F., Maldonado-Alcaino, A., Jack, D., Yan, B., Xu, L., Acuna, M.,



© University of Venda

Leiva-Guzman, M., Valdes, A., & Caceres, D. D. (2022). Cancer Burden Disease Attributable to $PM_{2.5}$ and Health Risk by PM2. 5-bound Toxic Species in Two Urban Chilean Municipalities. *Aerosol and Air Quality Research*, *22*(12), 220247.

Matooane, M., & Diab, R. (2003). Health risk assessment for sulfur dioxide pollution in South Durban, South Africa. *Archives of Environmental Health*, *58*(12), 763–770. https://doi.org/10.3200/AEOH.58.12.763-770.

Miller, L., & Xu, X. (2018). Ambient $PM_{2.5}$ human health effects—Findings in China and research directions. *Atmosphere*, *9*(11), 424.

Morakinyo, O. M., Mokgobu, M. I., Mukhola, M. S., & Engelbrecht, J. C. (2017). Health risk assessment of exposure to ambient concentrations of benzene, toluene and xylene in Pretoria West, South Africa. *African Journal of Science, Technology, Innovation and Development*, *9*(4), 489–496. https://doi.org/10.1080/20421338.2017.1352123.

Morakinyo, O. M., Mukhola, M. S., & Mokgobu, M. I. (2020). Concentration levels and carcinogenic and mutagenic risks of PM2. 5-bound polycyclic aromatic hydrocarbons in an urban–industrial area in South Africa. *Environmental Geochemistry and Health*, *42*(7), 2163–2178.

Morakinyo, O. M., Mukhola, M. S., & Mokgobu, M. I. (2021). Health risk analysis of elemental components of an industrially emitted respirable particulate matter in an urban area. *International Journal of Environmental Research and Public Health*, *18*(7). https://doi.org/10.3390/ijerph18073653.

Muyemeki, L., Petrus, R., Stuart, B., Piketh, J., Rafaj, P., & Kiesewetter, G. (2022). Integrated assessment of strategies to reduce air pollution in the Vaal Triangle Priority Area, South Africa. *Air Quality, Atmosphere & Health*, *3*, 2157–2169. https://doi.org/10.1007/s11869-022-01242-8.

Mwase, N. S. (2020). *Human Health Risks of Inhalable Exposure to $PM_{2.5}$ in Pretoria, South Africa*. University of Pretoria (South Africa).

Nawaz, M. O., & Henze, D. K. (2020). Premature deaths in Brazil associated with long-term exposure to PM2.5 from Amazon fires between 2016 and 2019. *GeoHealth*, *4*(8), e2020GH000268.

Nkosi, V., Wichmann, J., & Voyi, K. (2016). Comorbidity of respiratory and cardiovascular diseases among the elderly residing close to mine dumps in South Africa: A cross-sectional study. *South African Medical Journal*, *106*(3), 290–297. https://doi.org/10.7196/SAMJ.2016.v106i3.10243.

Novela, Rirhandzu 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), 1–10. https://doi.org/10.17159/caj/2020/30/2.8735.

Novela, Rirhandzu Jeaneth, Gitari, W. M., Mudzielwana, R., & Wichmann, J. (2023). *Temporal Characteristics and Health Risks Assessment of $PM_{2.5}$ in Ambient Air in Thohoyandou, South Africa*.



© University of Venda

Odekanle, E. L., Sonibare, O. O., Odejobi, O. J., Fakinle, B. S., & Akeredolu, F. A. (2020). Air emissions and health risk assessment around abattoir facility. *Heliyon*, *6*(7), e04365. https://doi.org/10.1016/j.heliyon.2020.e04365.

Okimiji, O. P., Techato, K., Simon, J. N., Tope-Ajayi, O. O., Okafor, A. T., Aborisade, M. A., & Phoungthong, K. (2021). Spatial pattern of air pollutant concentrations and their relationship with meteorological parameters in coastal slum settlements of Lagos, southwestern Nigeria. *Atmosphere*, *12*(11), 1426.

Othman, M., Latif, M. T., Naim, N. N. M., Zain, S. M. S. M., Khan, M. F., Sahani, M., Wahab, M. I. A., Sofwan, N. M., Abd Hamid, H. H., & Mohamed, A. F. (2022). Children's exposure to $PM_{2.5}$ and its chemical constituents in indoor and outdoor schools urban environment. *Atmospheric Environment*, *273*, 118963.

Paisi, N., Kushta, J., Pozzer, A., Violaris, A., & Lelieveld, J. (2024). Health effects of carbonaceous $PM_{2.5}$ compounds from residential fuel combustion and road transport in Europe. *Scientific Reports*, *14*(1), 1530.

Piketh, S. J., & Walton, N. M. (2004). Characteristics of atmospheric transport of air pollution for Africa. In *Air Pollution* (pp. 173–195). Springer.

Pope Iii, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama*, *287*(9), 1132–1141.

Roche, I. V., Ubalde-Lopez, M., Daher, C., Nieuwenhuijsen, M., & Gascon, M. (2024). The Health-Related and Learning Performance Effects of Air Pollution and Other Urban-Related Environmental Factors on School-Age Children and Adolescents—A Scoping Review of Systematic Reviews. *Current Environmental Health Reports*, 1–17.

Room, S. A., Lin, C. E., Pan, S. Y., Hsiao, T. C., Chou, C. C. K., & Chi, K. H. (2023). Incremental Lifetime Cancer Risk of PAHs in $PM_{2.5}$ via Local Emissions and Long-Range Transport during Winter. *Aerosol and Air Quality Research*, *23*(3), 1–20. https://doi.org/10.4209/aaqr.220319.

Roomaney, R. A., Cairncross, E., Tesfaye, M., Kapwata, T., Abdelatif, N., Olivier, C., Mathibela, K., Cois, A., Neethling, I., & Botai, J. (2022). Estimating the burden of disease attributable to ambient air pollution (ambient $PM_{2.5}$ and ambient ozone) in South Africa for 2000, 2006 and 2012. *South African Medical Journal*, *112*(8b), 705–717.

Roy, D., Singh, G., & Seo, Y.-C. (2019). Carcinogenic and non-carcinogenic risks from $PM_{10}$-and $PM_{2.5}$-Bound metals in a critically polluted coal mining area. *Atmospheric Pollution Research*, *10*(6), 1964–1975.

Safo-Adu, G., Attiogbe, F., Emahi, I., & Ofosu, F. G. (2023). Outdoor and indoor particle air pollution and its health consequences in African cities: New evidence and an exhortation. *Sustainable Environment*, *9*(1), 2265729.

Saju, J. A., Bari, Q. H., Mohiuddin, K. A. B. M., & Strezov, V. (2023). Measurement of ambient particulate matter ($PM_{1.0}$, $PM_{2.5}$ and $PM_{10}$) in Khulna City of Bangladesh and



© University of Venda

their implications for human health. *Environmental Systems Research*, *12*(1), 42.

Sharma, N., Agarwal, A. K., Eastwood, P., Gupta, T., & Singh, A. P. (2018). Introduction to Air Pollution and Its Control. In *Energy, Environment, and Sustainability*. https://doi.org/10.1007/978-981-10-7185-0_1.

*Health Effects Institute. (2018).* Special Report. Boston, M. E. I. (2018). State of Global Air 2018. *Health Effects Institute*.

Sulaymon, I. D., Zhang, Y., Hopke, P. K., Ye, F., Gong, K., Mao, J., & Hu, J. (2023). Modeling $PM_{2.5}$ During Severe Atmospheric Pollution Episode in Lagos, Nigeria: Spatiotemporal Variations, Source Apportionment, and Meteorological Influences. *Journal of Geophysical Research: Atmospheres*, *128*(13), e2022JD038360.

Sulong, N. A., Latif, M. T., Khan, M. F., Amil, N., Ashfold, M. J., Wahab, M. I. A., Chan, K. M., & Sahani, M. (2017). Source apportionment and health risk assessment among specific age groups during haze and non-haze episodes in Kuala Lumpur, Malaysia. *Science of the Total Environment*, *601*, 556–570.

Suradi, H., Khan, M. F., Sairi, N. A., Rahim, H. A., Yusoff, S., Fujii, Y., Qin, K., Bari, M. A., Othman, M., & Latif, M. T. (2021). Ambient levels, emission sources and health effect of $PM_{2.5}$-bound carbonaceous particles and polycyclic aromatic hydrocarbons in the city of Kuala Lumpur, Malaysia. *Atmosphere*, *12*(5), 549.

Thabethe, N. D. L., Engelbrecht, J. C., Wright, C. Y., & Oosthuizen, M. A. (2014). Human health risks posed by exposure to PM10 for four life stages in a low socio-economic community in South Africa. *The Pan African Medical Journal*, *18*.

US EPA, United States Environmental Protection Agency (1989). Risk Assessment Guidance for Superfund Volume I Human Health Evaluation Manual (Part A). Office of Emergency and Remedial Response U.S. Environmental Protection Agency Washington, D.C. 20450. https://www.epa.gov/sites/default/files/2015-09/documents/rags_a.pdf.

US EPA (1992). EPA's Approach for Assessing the Risks Associated with Chronic Exposure to Carcinogens. Background Document 2, January 17, 1992. https://www.epa.gov/iris/epas-approach-assessing-risks-associated-chronic-exposure-carcinogens. Accessed on 20th July 2024.

US EPA (1997). Exposure factors handbook.http://www. epa.gov/ncea/expofac.htm. Accessed on 24th June 2024.

US EPA. (2000). Office of science policy (OSP)_risk_characterization_handbook_2000. https://www.epa.gov/sites/default/files/201510/documents/osp_risk_characterization_handbook_2000.pdf. Accessed on 24th June 2024.

US EPA (2005). Guidelines for Carcinogen Risk Assessment, Risk Assessment Forum, U.S. Environmental Protection Agency Washington, DC. EPA/630/P-03/001F, March 2005. Accessed on 20th July 2024. https://www.epa.gov/sites/default/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf.

US EPA (2009). Risk Assessment Guidance for Superfund Volume 1: Human Health Evaluation Manual (Part F, Supplemental Guidance for Inhalation Risk Assessment).



© University of Venda

U.S. Environmental Protection Agency Washington, DC. EPA-540-R-070-002. Accessed on 14th July 2024.

US EPA (2011). Exposure Factors Handbook 2011 Edition (Final). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R–09/052F, 2011. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=236252. Accessed on 20th July 2024.

US EPA (2023). Human Health Risk Assessment. U.S. Environmental Protection Agency. https://www.epa.gov/risk/human-health-risk-assessment. Accessed on 12th June 2024.

Vanker, A., Barnett, W., Chartier, R., MacGinty, R., & Zar, H. J. (2023). Personal monitoring of fine particulate matter (PM2. 5) exposure in mothers and young children in a South African birth cohort study–a pilot study. *Atmospheric Environment*, *294*, 119513.

Wang, K. (2007). Long-range transport of the April 2001 dust clouds over the subtropical East Asia and the North Pacific and its impacts on ground-level air pollution: A Lagrangian simulation. *Journal of Geophysical Research: Atmospheres*, *112*(D9).

Wang, Z., Wang, R., Wang, J., Wang, Y., Donahue, N. M., Tang, R., Dong, Z., Li, X., Wang, L., & Han, Y. (2022). The seasonal variation, characteristics and secondary generation of PM2. 5 in Xi'an, China, especially during pollution events. *Environmental Research*, *212*, 113388.

Williams, J., Petrik, L., & Wichmann, J. (2021). $PM_{2.5}$ chemical composition and geographical origin of air masses in Cape Town, South Africa. *Air Quality, Atmosphere & Health*, *14*, 431–442.

WHO, (2016). Ambient air pollution: a global assessment of exposure and burden of disease. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

WHO, (2018a). Air pollution and child health: prescribing clean air: summary. World Health Organization. https://iris.who.int/handle/10665/275545.

WHO, (2018b). Burden of disease from ambient air pollution for 2016. https://cdn.who.int/media/docs/default-source/air-quality-database/aqd-2018/aap_bod_methods_apr2018_final.pdf?sfvrsn=30ac0d62_3.

WHO, (2021). WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and $PM_{10}$), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization. https://iris.who.int/handle/10665/345329.

Yu, W., Xu, R., Ye, T., Abramson, M. J., Morawska, L., Jalaludin, B., Johnston, F. H., Henderson, S. B., Knibbs, L. D., Morgan, G. G., Lavigne, E., Heyworth, J., Hales, S., Marks, G. B., Woodward, A., Bell, M. L., Samet, J. M., Song, J., Li, S., & Guo, Y. (2024). Estimates of global mortality burden associated with short-term exposure to fine particulate matter ($PM_{2\cdot5}$). *The Lancet Planetary Health*, *8*(3), e146–e155. https://doi.org/10.1016/S2542-5196(24)00003-2.

Yunesian, M., Rostami, R., Zarei, A., Fazlzadeh, M., & Janjani, H. (2019). Exposure to



© University of Venda

high levels of $PM_{2.5}$ and $PM_{10}$ in the metropolis of Tehran and the associated health risks during 2016–2017. *Microchemical Journal*, *150*, 104174. https://doi.org/10.1016/j.microc.2019.104174.

Zhang, Y., Wang, Y., Du, Z., Chen, S., Qu, Y., Hao, C., Ju, X., Lin, Z., Wu, W., & Xiao, J. (2023). Potential causal links between long-term ambient particulate matter exposure and cardiovascular mortality: New evidence from a large community-based cohort in South China. *Ecotoxicology and Environmental Safety*, *254*, 114730.

Zheng, M., Yan, C., & Zhu, T. (2020). Understanding sources of fine particulate matter in China. Philosophical Transactions of the Royal Society A, 378(2183), 20190325.



© University of Venda

# CHAPTER SEVEN

# PREDICTION OF AMBIENT $PM_{2.5}$ USING MACHINE LEARNING AND NATURE-BASED SOLUTIONS FOR SUSTAINABLE AIR QUALITY MANAGEMENT IN THULAMELA LOCAL MUNICIPALITY, LIMPOPO, SOUTH AFRICA

## Abstract

*Fine particulate matter ($PM_{2.5}$) remains one of the most significant global threats due to the adverse impacts on both human health and the environment. This study focuses on the role of meteorological factors including atmospheric pressure, humidity, wind direction, rainfall, temperature, and wind speed on the concentration of $PM_{2.5}$ within Thulamela Municipality, South Africa. In this study, $PM_{2.5}$ concentration was predicted in this study through the use of machine learning algorithms like, Multiple Linear Regression (MLR), Polynomial Ridge Regression (PRR), and Random Forest (RF). The models were built using daily site-specific concentrations, calculated area-wide average $PM_{2.5}$, and seasonal patterns across the study area. The model performances were evaluated using 5-fold cross-validation and $R^2$, RMSE, and MSE. The RF model outperformed MLR in all the three scenarios with $R^2$ ranging from 0.49 (RG2) to 0.80 (VP4) for site-specific predictions, whereas MLR's highest $R^2$ was 0.46. For the entire study area, RF outperformed PRR ($R^2$ = 0.297, RMSE = 3.22 µg/m³) and MLR ($R^2$ = 0.144, RMSE = 3.54 µg/m³) with $R^2$ = 0.604 and RMSE = 2.41 µg/m³. Seasonal analysis revealed that RF model performed better in winter ($R^2$ = 0.58, RMSE = 2.10 µg/m³) as compared to PRR ($R^2$ = 0.31, RMSE = 2.43 µg/m³), however, summer predictions were less accurate, with RF ($R^2$ = 0.45, RMSE = 4.72 µg/m³) outperforming PRR ($R^2$ = 0.35). Cross-validation showed comparable RF performance in winter (mean $R^2$ = 0.29 ± 0.21, RMSE = 2.38 ± 0.30 µg/m³), while summer was challenging (mean $R^2$ = -0.37 ± 0.83, RMSE = 5.51 ± 0.69 µg/m³). This study demonstrates the robustness of the RF model in capturing complex, non-linear $PM_{2.5}$ dynamics and outperforming traditional methods under varying meteorological conditions in semi-urban settings like Thulamela Municipality, where direct $PM_{2.5}$ measurements are limited. These results indicate the potential of integrating meteorological factors into air quality models for improved environmental management, particularly in semi-urban regions with limited direct $PM_{2.5}$ measurements. The study further provides constructive suggestions to environmental agencies, urban planners, and policy makers to formulate targeted interventions to mitigate the impacts of semi-urbanisation on air quality, reduce health- and environmentally harmful pollution, leading to a sustainable urban environment.*

**Keywords:** Air pollution, Machine learning, Meteorological conditions, MLR, $PM_{2.5}$, Polynomial Ridge Regression (PRR), Regression, Random Forest.



© University of Venda

## 7.1 Introduction

Particulate matter ($PM_{2.5}$) continues to be a major global environmental and public health concern due to its negative effects on humans and ecosystems. Long-term exposure to $PM_{2.5}$ has been linked to respiratory and cardiovascular diseases, high mortality rate and negative environmental issues such as reduced visibility and ecosystem degradation (Sharma *et al*., 2023). Understanding the factors that influence $PM_{2.5}$ concentrations is critical for developing effective air quality control measures to limit their harmful consequences. Within the African context, Safo-Adu *et al*. (2023) highlighted the urgent need for a deeper investigation into how air pollution affects health outcomes. Their results revealed that $PM_{2.5}$ and $PM_{10}$ concentrations in the most populated African cities frequently exceed the WHO guidelines, which could contribute to respiratory infections, lung diseases, and socioeconomic issues. In the same way, Okimiji *et al*. (2021) and Sulaymon *et al*. (2023) identified elevated levels of $PM_{2.5}$ in Lagos, Nigeria as a result of residential and industrial emissions, underscoring the urgency of effective emission-control strategies. In South Africa, $PM_{2.5}$ pollution is increasingly becoming a problem, especially in semi-urban towns such as Thulamela Municipality in the Vhembe District. This region faces diverse pollution sources, such as vehicular emissions, fuel combustion from domestic households, and wood burning from industry, which is further exacerbated by limited air quality monitoring infrastructure.

Meteorological factors such as wind direction, wind speed, humidity, temperature, and rainfall, have a significant impact on $PM_{2.5}$ levels. Exploring the links between these variables and $PM_{2.5}$ concentrations is critical for developing predictive models and executing targeted interventions to address air quality concerns. Several studies have investigated the relationship between climatic conditions and $PM_{2.5}$ concentrations (Edlund *et al*., 2021; Okimiji *et al*., 2021). Zamani *et al*. (2019) investigated the impacts of temperature and wind speed on $PM_{2.5}$ dispersion in densely populated urban regions. Similarly, Aladag, (2023) noted that seasonal variability in $PM_{2.5}$ levels, particularly during the winter is caused by weather stability, leading to higher pollutant concentrations. Additionally, Sulaymon *et al*. (2023) reported strong correlations between $PM_{2.5}$ and meteorological variables, such as temperature and humidity, in urban slum settlements



© University of Venda

of Lagos. Majority of these studies, however, focused on urban and industrialized areas with strong monitoring systems, leaving semi-urban places like Thulamela Municipality under-represented in the existing literature. This work aims to close that gap by giving insights into $PM_{2.5}$ dynamics in a semi-urban setting.

Machine learning (ML) techniques have become increasingly popular in a variety of disciplines; similarly, they have transformed air quality prediction by providing predictive capabilities that surpass traditional statistical approaches. Ensemble methods based on decision trees, for example, gradient boosting and random forest [RF] and artificial neural networks (ANN), support vector regression (SVR), and polynomial ridge regression (PRR) have become indispensable tools for analysing complex datasets (Shen *et al*., 2020; Lu *et al*., 2021). These approaches have successfully captured nonlinear interactions, handled multicollinearity, and minimised overfitting, making them ideal for air pollution research. Some researchers have successfully used these models to estimate $PM_{2.5}$ levels (Hossain *et al*., 2024; Suthar *et al*., 2024). Traditional statistical models, such as multiple linear regression (MLR) continue to be popular due to their simplicity and interpretability, however, they rely on linear assumptions that limit their ability to model complicated $PM_{2.5}$ dynamics. MLR often includes variables whose influence on $PM_{2.5}$ concentrations has no clear rationale. In contrast, advanced ML techniques such as RF and PRR have been shown to be more effective in capturing complicated, nonlinear relationships between variables (Antai *et al*., 2018; Kayes *et al*., 2019). RF, for example, uses numerous decision trees to average outputs, making it ideal for handling high-dimensional data and nonlinear interactions. A study by Huang *et al*. (2018), for instance, conducted in the provinces of Hebei and Shandong, used the random forest (RF) algorithm to predict $PM_{2.5}$ concentrations and identified a robust link between $PM_{2.5}$ and chronic illnesses.

There is an established success of advanced machine learning techniques, despite that, their application in semi-urban areas like Thulamela Municipality, have not been investigated. This research gap limits our understanding of how meteorological variables affect air quality in regions characterized by multiple pollution sources and limited monitoring infrastructure. To close this significant gap, this study pioneers the use of



© University of Venda

advanced machine learning techniques to estimate $PM_{2.5}$ concentrations in Thulamela Municipality, creating an essential baseline for air quality research in a semi-urban area. In addition, this work focuses on building and comparing accurate, data-driven models to estimate $PM_{2.5}$ concentrations across specific sites, regional, and seasonal variations. This chapter, hence, aims - to examine the relationship between meteorological variables and $PM_{2.5}$ concentrations at multiple sites in Thulamela Municipality; evaluate the predictive ability of MLR, PRR, and RF models in various settings, including site-specific predictions, regional analysis, and seasonal variations, then, provide actionable insights for policymakers, urban planners, and environmental agencies to develop data-driven air quality management strategies. By addressing all these specific objectives, this study would give actionable insights that can assist sustainable urban development while mitigating the negative effects of $PM_{2.5}$ pollution.

## 7.2 Materials and Methods

### 7.2.1 Ambient $PM_{2.5}$ Concentration and Meteorological Data

As explained in section 4.2.1, Dylos DC1700 laser particle counters were used to measure $PM_{2.5}$ concentration at 14 carefully selected monitoring sites across Thulamela Municipality. To ensure spatial heterogeneity during data collection, these sites were chosen to reflect a wide range of pollution sources, such as vehicle emissions, commercial wood burning, and residential fuel combustion. The site selection considered previous studies in the region and consultations with local stakeholders to ensure comprehensive coverage of main pollution hotspots. To maintain reliability and precision throughout the monitoring period, calibration and maintenance of the Dylos DC1700 devices adhered to established protocols as mentioned in Subsection 4.2.2 to 4.2.7.

Meteorological data (atmospheric pressure, humidity, precipitation, temperature, wind speed, and wind direction) were obtained from the South African Weather Service (SAWS). These parameters were selected due to their key role in influencing $PM_{2.5}$ dispersion, transformation, and transport dynamics. The focus on these pollutant variables ensured comprehensive spatial and temporal data coverage and provided a



© University of Venda

solid foundation for subsequent analyses. Figure 7.1 shows the flow chart used in this study.

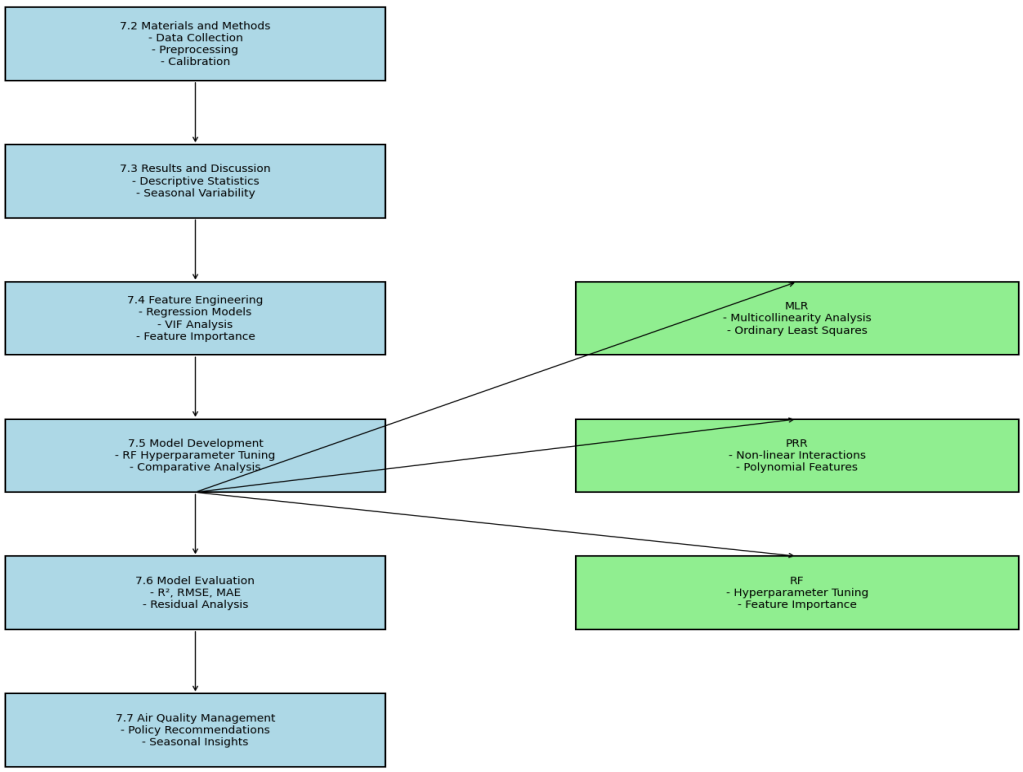



**Figure 7.1***:* Flow chart used for the analysis and prediction of $PM_{2.5}$ in this study.

### 7.2.2 Data Collection, Pre-Processing and Feature Engineering for Modelling

A stringent quality control process was implemented to ensure the accuracy and completeness of the dataset collected from the Dylos DC1700 devices. This approach reduced bias and ensured that the data accurately reflected ambient air quality conditions during the monitoring period. $PM_{2.5}$ and meteorological datasets were pre-processed using Excel 2016, Tableau, R, and Python to ensure compatibility with the statistical and machine learning analyses that followed; this was done to ensure that the selected ML algorithms work properly. All relevant columns were converted to numeric formats, and median imputation was applied to handle missing meteorological values, while rows with missing $PM_{2.5}$ values were excluded from the dataset. The median imputation was selected because of its robustness to outliers, integrity, and preserving of the dataset.



© University of Venda

This approach maintained the stability of models, although, exclusion of rows with missing $PM_{2.5}$ values could cause slightly reduced dataset size, which could affect the statistical power and the variability representation. These preprocessing steps ensured that the dataset was clean and ready for analysis, while the descriptive statistics provided the overview variability, and distribution of each variable. Subsequently, the Shapiro-Wilk test was used to assess the normality of the data. Then, the deviations from normality were further visualized using histograms to highlight kurtosis and skewness. Due to dataset distribution, non-parametric statistical approaches were used, based on Silva *et al*. (2001) suggestion. The Spearman correlation was used to identify significant relationships between meteorological variables and $PM_{2.5}$ levels, offering valuable insights into potential predictors for modelling purposes.

### 7.2.3 Development and Prediction of Ambient $PM_{2.5}$ using Machine Learning Algorithms

In this section, a range of machine learning models were utilised to predict air pollution concentrations based on meteorological data. Figure 7.1 illustrates the selected models, which include multiple linear regression (MLR) (Antai *et al*., 2018), polynomial ridge regression (PRR), and random forest (RF) (Kayes *et al*., 2019). These models were employed to predict seasonal variations, area-wide averages, and daily site-specific $PM_{2.5}$ concentrations throughout the study area. This methodological approach facilitated a comprehensive analysis of meteorological factors that influence $PM_{2.5}$.

#### 7.2.3.1 Estimation of $PM_{2.5}$ concentrations using multiple linear regression and polynomial ridge regression

Python v3.12.3 was used to construct regression models capable of capturing both linear and nonlinear relationships between meteorological predictors (independent variables) and $PM_{2.5}$ concentrations (dependent variable) (Islam *et al*., 2023). The cleaned dataset underwent normalisation and feature scaling to ensure that all predictors contributed equally to the developed model, facilitated coefficient interpretation, and improved output accuracy. The MLR model was fitted to the cleaned data using both the ordinary least squares (OLS) method in the statsmodel's library and the Linear Regression method from the sklearn library. These models were utilized to predict daily and site-specific $PM_{2.5}$



© University of Venda

concentrations across monitoring sites in the Thulamela Municipality. For the Linear Regression model, the dataset for each monitoring site was split into training (80%) and testing (20%) subsets using the train-test-split function. To enhance the model's capacity to capture non-linearity and interactions between predictors for seasonal variations and calculate area-wide averages, polynomial ridge, regression was incorporated. For the PRR framework, third-order polynomial features were employed to model non-linear dependencies more effectively, improving the model's capacity to capture complex seasonal interactions among predictors in the study area. These models adhered to the foundational mathematical equations outlined in Equations (7.1) and (7.2) (polynomial regression model), respectively, enabling precise estimation of $PM_{2.5}$ concentrations based on meteorological inputs.

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_n X_{n,t} + \varepsilon_i \quad (7.1)$$

$$y_i = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t}^2 + \cdots + \beta_n X_{n,t}^n + \varepsilon_i \quad (7.2)$$

Where;

$Y_t$ and $y_i$ represent the pollutant concentration at time t (dependent variable) which are the outputs of the MLR and Polynomial Ridge Regression models, respectively. $X_{1,t}$, $X_{2,t}$, …, $X_{n,t}$ denote the meteorological variables at time t (predictors or independent variables), $X_2$,$X_3$,…, $X_n$ are the polynomial terms of the independent variable (Ws,, Wd, Temp, H, and R, respectively).

In Equation (7.2), $X_{2,t}^2 + X_{3,t}^2 \ldots + X_{n,t}^n$ are the polynomial terms of the meteorological variables, to capture the non-linear relationships. The $\beta_0$ is the model constant (intercept), whereas, $\beta_1$, $\beta_2$, …, $\beta_n$ are regression coefficients representing the contribution of each variable and its polynomial terms to the model. The $\varepsilon_i$ is the random error (residual at time t) representing the difference between the observed and predicted values. This formulation, thus, allowed the models to account for linear and non-linear interactions between meteorological predictors and $PM_{2.5}$, thereby enhancing their predictive performance and interpretability.



© University of Venda

*A. Variance inflation factor (VIF)*

For the assessment of possible multicollinearity between the meteorological predictors, a variance inflation factor (VIF) analysis was performed using the statsmodels library in Python. VIF quantifies the extent to which the variance of a regression coefficient is inflated due to collinearity among the predictors. A VIF value of more than 10 indicates significant multicollinearity, which suggests redundancy between the predictors (Bera *et al*., 2021). Equation (7.3) was used to calculate the VIF:

$$VIF_i = \frac{1}{1-R_i^2} \quad (7.3)$$

Where, $R^2$ is the multiple coefficients of determination for the regression of predictor *i* on all other predictors. An independent regression model was created for each monitoring site, and variance inflation factor (VIF) values were calculated for all predictors to identify and mitigate multicollinearity issues. This approach ensured that the addition of highly correlated variables did not affect the interpretability or reliability of the regression coefficients in both the multiple linear regression (MLR) and polynomial ridge regression (PRR) models (Ebrahimi-Khusfi, *et al*., 2021).

### 7.2.3.2 Random Forest (RF) analysis and model training

Random forest (RF) regression was used because of its robustness in dealing with non-linear relationships and high-dimensional datasets, making it a suitable choice for air quality studies. Compared to other ensemble methods, such as gradient boosting and support vector machines (SVM), RF offers computational efficiency, reduce overfitting and interpretability, particularly for datasets with complex interactions among predictors (Breiman, 2001; Chang *et al*., 2020; Imam *et al*., 2024). Meteorological data were used as predictors to estimate daily site-specific $PM_{2.5}$ concentrations, as well as average and seasonal variations at the monitoring sites. To provide robust model validation and reduce the danger of overfitting, the cleaned dataset was divided into 80% training and 20% testing subsets for each monitoring site. RandomizedSearchCV through hyperparameter optimization was used to optimize key parameters, such as the number of estimators (trees), maximum tree depth, and minimum samples required to split an internal node.



© University of Venda

The optimized models were trained on the training dataset, and predictions were generated for both training and test datasets to comprehensively evaluate the model's accuracy and generalizability of the model. The Random Forest (RF) model was implemented using sklearn.ensemble RandomForestRegressor from the sklearn library. The prediction of the model is expressed as follows (Equation 7.4):

$$y = \frac{1}{N}\sum_{i=1}^{N} h_i(x) \quad (7.4)$$

Where:

- y is the predicted value,
- N is the total number of decision trees in the ensemble,
- $h_i(x)$ is the output of the $i^{th}$ decision tree for input xxx.

The increasing number of decision trees enhances model precision while addressing overfitting concerns and this shows the robustness of the RF approach in predictive tasks. This approach aligns with the study's objective of developing a robust, site-specific, and seasonally variable for $PM_{2.5}$ prediction framework. In this study, RF was selected over other machine-learning models, such as Gradient Boosting and Support Vector Machines, due to its superior ability to handle non-linear interactions and computational efficiency when applied to high-dimensional meteorological datasets (Chang *et al*., 2020). This approach aligns with the study's objective of developing a robust, site-specific, and seasonally variable $PM_{2.5}$ prediction framework, while providing insights into the model's predictive reliability and its alignment with broader study goals.

#### 7.2.3.3 Evaluation of Model Performance and Cross-Validation using Key Metrics

In this study, key statistical metrics such as the coefficient of determination ($R^2$), root mean square error (RMSE), mean squared error (MSE) and sum of squared errors (SSE) were used to gauge the efficacy of the predictive models. Statistical measures were computed for both training and test data sets, based on the algorithm used to evaluate and confirm the prediction models' precision. The MSE was calculated as the mean difference between predicted and observed values, as shown in equation (7.5), while the RMSE, SSR and $R^2$ were derived using equations (7.6) to (7.8), respectively as:



© University of Venda

$$MSE = \frac{1}{N}\sum_{i=1}^{n} \left(Y_{pred,i} - X_{obs,i}\right)^2 \quad (7.5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n} \left(Y_{pred,i} - X_{obs,i}\right)^2} \quad (7.6)$$

$$SSR = \sum_{i=1}^{n} \left(Y_{obs,i} - Y_{pred,i}\right)^2 \quad (7.7)$$

$$R^2 = \frac{Explained\ Variation}{Total\ variation} = \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{n} \left(Y_{obs,i} - Y_{pred,i}\right)^2}{\sum_{i=1}^{n} \left(Y_{obs,i} - Y_{obs,i}\right)^2} \quad (7.8)$$

Where, N represents the total number of measured data points, $Y_{pred,i}$ represents the predicted $PM_{2.5}$ concentration for observation $i$ based on the regression model. $Y_{obs,i}$ denotes the mean of the individual measured $PM_{2.5}$ concentration for observation $i$, and SST is the total sum of squares.

In addition to the above-mentioned metrics, the MLR models were used to assess significant relationships, strength, and explanatory power based on the p-values. These statistical tests were conducted to determine the significance of the correlation coefficients and validate the performance of the predictive models across all folds. These metrics offered information about the generalizability, dependability and robustness of the model across various datasets. This enabled us to draw meaningful conclusions about the meteorological influences and the prediction of air quality. Five-fold cross-validation was used to thoroughly test the robustness and generalizability of the models. This involved dividing the dataset into five equal subsets (folds), with the model being trained sequentially on four folds, while the remaining fold was used for validation purposes. This iterative approach was repeated until all folds were validated, ensuring a thorough assessment of the model's predictive reliability across many subsets. To assess the stability and generalizability of RF models, the mean and standard deviation of key performance metrics, including the RMSE and the $R^2$, were calculated for each fold. After which, the best model for each site was selected based on the results of the cross-validation and the performance of the test set, providing a solid basis for further analysis.



© University of Venda

#### 7.2.3.4 Feature importance and residual analysis

Feature importance analysis was performed to quantify the contribution of each meteorological parameter to $PM_{2.5}$ prediction. For the MLR and PRR models, the significance of each feature was assessed using coefficient estimates and the associated p-values. Predictors achieving p-values < 0.05 were considered statistically significant contributors to $PM_{2.5}$ levels. In the RF models, importance scores were computed to rank the relative influence of each predictor, leading to a better understanding of the factors influencing $PM_{2.5}$ concentrations. These scores allowed a seasonal comparison to determine which meteorological variables have a greater influence in summer and winter, thereafter, the residuals analysis was performed to determine the prediction biases and ensure the robustness of the results. Then, the differences between the measured and predicted $PM_{2.5}$ concentrations were plotted to check for any patterns or biases, providing insight into the model assumptions and highlighting systematic errors or biases in the predictions. The performance of the models was further visualized by comparing measured and predicted $PM_{2.5}$ concentrations for the testing data. Scatter plots were created for each site, displaying the agreement between observed and predicted values.

This adopted methodology ensured a robust framework for analysing site-specific $PM_{2.5}$ variations, incorporating both predictive performance and interpretability through feature importance analysis. By combining feature importance rankings and residual diagnostics, this dual approach ensured that the models were both interpretable and reliable, supporting their use in effective $PM_{2.5}$ prediction. This methodological framework aligns with the study's objective of developing a robust, site-specific, and seasonal variations models for the $PM_{2.5}$ prediction framework while providing insights into the model's predictive reliability and its alignment with broader study objectives. By applying this dual strategy methodology to facilitate effective $PM_{2.5}$ prediction, this study fills critical gaps in air quality modelling, particularly in semi-urban regions with limited monitoring infrastructure, by using novel machine learning techniques. The innovative use of ML techniques highlights the significance of integrating meteorological variables into predictive frameworks, offering a robust foundation in air quality management and policy development. This methodological framework has established a precedent for future research in similar underrepresented settings.



© University of Venda

## 7.3 Results and Discussion

### 7.3.1 Meteorological Variables and their Descriptive Statistics within Thulamela Municipality

Meteorological conditions are critical in determining the transportation, dispersion, and transformation of air pollutants, particularly $PM_{2.5}$. Variables such as temperature, humidity, rainfall, wind speed, and wind direction exhibit significant seasonal variability, influencing air quality dynamics and pollutant behaviour. Seasonal patterns strongly affect pollutant dispersion and chemical transformations. Figure 7.2 highlights the monthly variability of temperature, relative humidity, and rainfall. During the summer months (DJF), higher temperatures and humidity levels were observed, followed by moderate values in autumn (MAM) and spring (SON). Conversely, winter months (JJA) are marked by lower temperatures and humidity, which can lead to reduced $PM_{2.5}$ levels due to limited secondary pollutant formation. Winter's cooler atmospheric condition, however, increases the likelihood of temperature inversions, potentially trapping pollutants near the surface, while higher humidity in summer facilitates the removal of particulate matter through dissolution and wet deposition, resulting in improved air quality. In contrast, lower humidity during winter increases the risk of pollutant accumulation in the atmosphere (Li, *et al*., 2023).

Wind plays a dynamic role in the transport of pollutants, the hourly wind speeds during the study period ranged between 0 and 11.5 m/s, with prevailing southwest wind patterns significantly influencing regional pollutant dispersion, as demonstrated in the wind rose diagram (Figure 7.3a). Seasonal variability in wind speeds was also evident, with summer exhibiting the highest mean wind speed (1.51 m/s), followed by spring (1.35 m/s), winter (1.12 m/s), and autumn (0.87 m/s), as illustrated in Figure 7.3b. The average wind speed throughout the study period was 1.21 m/s, with calm conditions varying from 27.5% to 54.7% monthly (Appendix 7.1). Local emissions from sources such as wildfires and domestic burning significantly impacted pollutant concentrations, especially during the dry season. In contrast, a wet condition during the rainy season cleanses the air by washing pollutants from the atmosphere. Temperature inversions caused by convective heating increased the mixed-layer height, further influencing pollutant dispersion. Previous research highlighted that meteorological conditions could trap pollutants for extended



© University of Venda

periods before it is eventually transported out of the area (Adesina *et al*., 2022; Li, *et al*., 2023), hence, understanding these meteorological factors is crucial for analysis of air quality trends and pollutant dispersion patterns within the Thulamela Municipality.

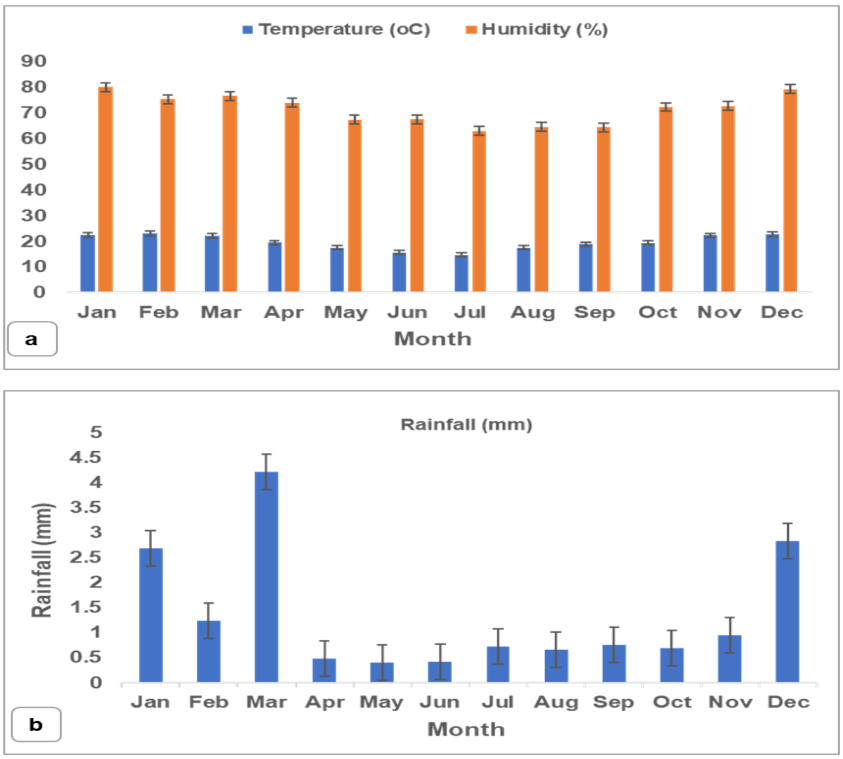



Figure 7.2: Monthly mean relative humidity (%), temperature (°C), and rainfall (mm) for the study area.

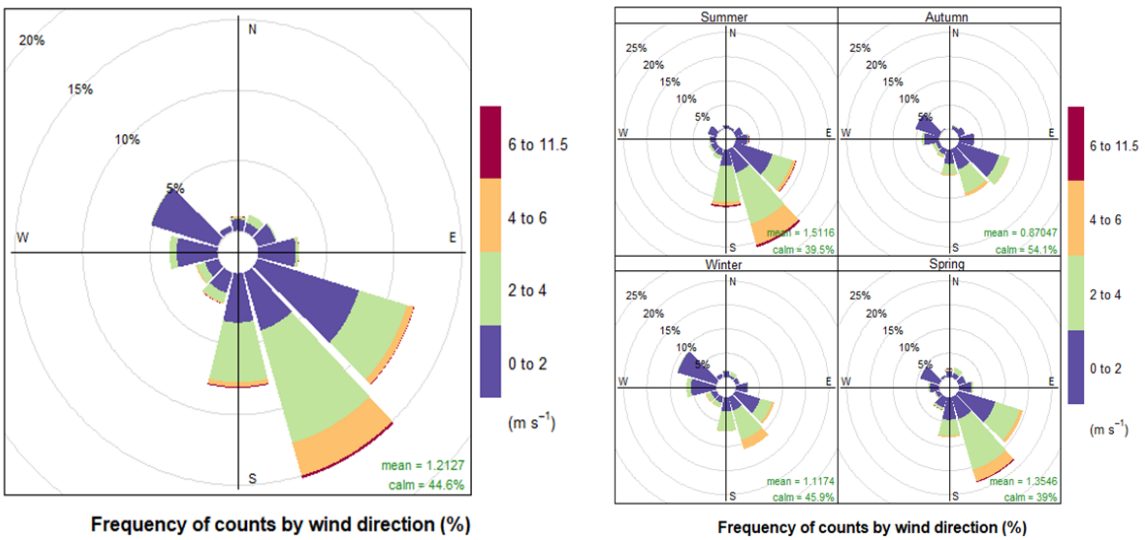



**Figure 7.3**: Windrose depicting the annual and seasonal wind speed, showing the prevailing wind patterns that affect $PM_{2.5}$ transport across the study area, a) annual, and b) by seasons.



© University of Venda

The correlation analysis of daily mean meteorological parameters from March 2021 to February 2022 offered a comprehensive understanding of the interrelationships among various atmospheric variables in the study area. This analysis is crucial for enhancing our understanding of meteorological impacts on $PM_{2.5}$ behaviour and assessing its implications on environmental conditions and public health. Meteorological conditions are critical in estimating $PM_{2.5}$ concentrations, as they affect the formation, transformation, and transport processes of air pollutants (Zhang, 2019; Liu *et al*., 2020). This study, hence, examined five key meteorological variables to gain insight into how these factors collectively modulate $PM_{2.5}$ concentrations in the study sites. Table 7.1 presents the descriptive statistical summary of $PM_{2.5}$ and the associated meteorological parameters. The dataset's normality was assessed using the Shapiro-Wilk test, complemented by visual inspections via histograms (Appendix 7.2). P-values above 0.05, supported by skewed distributions revealed by histograms, indicated non-significant deviation from normality (Appendix 7.3). Given the evidence of skewed data, non-parametric methods, specifically the Spearman correlation test were employed in further analysis. Spearman correlation is a robust, rank-based approach that performs well even when data deviate from normality. This approach has been extensively validated in studies investigating meteorological impacts on air quality (Chen *et al*., 2019; Wang *et al*., 2023a; 2023b). By utilising this method, the analysis ensured the robustness in evaluating the relationships between the selected parameters, while accounting for inherent dataset variability.

**Table 7.1:**: Descriptive statistical analysis of $PM_{2.5}$ and meteorological parameters

| Parameters | Max | Min | Mean | SD | Median |
|---|---|---|---|---|---|
| $PM_{2.5}$ (µg/m$^3$) | 104.99 | 13.45 | 23.85 | 6.69 | 22.26 |
| Wind speed, Ws (m/s) | 4.83 | 0.85 | 2.35 | 0.64 | 2.30 |
| Wind direction, Wd (°) | 358.52 | 1.54 | 151.37 | 67.79 | 139.10 |
| Humidity, H (%) | 97.25 | 28.29 | 67.66 | 14.20 | 69.75 |
| Temperature, T (°C) | 30.89 | 9.34 | 20.40 | 4.28 | 20.29 |
| Rain, R (mm) | 10.90 | 0.20 | 1.01 | 1.62 | 0.43 |
| Atmospheric Pressure, AP | 962.17 | 937.49 | 948.29 | 4.96 | 947.83 |



© University of Venda

### 7.3.2 Association between Changes in Meteorological Parameters within Thulamela Municipality

The Spearman correlation analysis provides critical insights into the interactions between meteorological variables in the study area (Table 7.2). Notably, Ws showed a weak positive correlation with temperature (0.193) and weak negative correlations with humidity (-0.287), rainfall (-0.042), and pressure (-0.041). These findings align with prior studies that associate higher Ws with increased convective activity and surface temperature, while simultaneously reducing relative humidity through evaporation (Jones & Williams, 2020; Liu *et al*., 2021). Humidity exhibited a strong positive correlation with rainfall (0.614) but weak correlations with temperature (-0.153) and pressure (0.123). Temperature displayed a strong negative correlation with atmospheric pressure (-0.653), as well as weak correlations with Ws (0.193) and rainfall (0.005). Rainfall showed a strong positive correlation with humidity (0.614), with weaker correlations observed with pressure (-0.077), Ws (-0.0 42), and temperature (0.005).

These interrelationships highlight the complexity of atmospheric dynamics in the study area. For instance, the inverse relationship between temperature and pressure reflects established principles of thermodynamics, where increased temperatures reduce surface pressure, promoting convective instability (Chen *et al*., 2024). Similarly, the strong positive correlation between humidity and rainfall highlights their shared influence on cloud formation and precipitation patterns, critical for local weather processes (Chen *et al*., 2019). The observed relationships between key variables, such as humidity, rainfall, temperature and pressure, align with established meteorological principles; for instance, the observed strong negative correlation between temperature and atmospheric pressure supports the principle that rising temperatures lead to lower pressure systems, fostering convective mixing and potentially influencing pollutant dispersion (Jones & Williams, 2020).

This dynamic relationship is particularly relevant in regions like the study area, where local airflows and topography may further modulate these effects. Similarly, the strong positive correlation between humidity and rainfall highlights the critical role of atmospheric moisture in precipitation processes and improving air quality. High humidity enhances the



© University of Venda

likelihood of cloud formation and subsequent rainfall, both of which are instrumental in the wet deposition of $PM_{2.5}$ (Chen *et al*., 2019; Zhang *et al*., 2022). This emphasises the need for considering moisture-related variables alongside temperature when analysing precipitation dynamics.

From an air quality management perspective, these findings highlight the multifaceted ways in which meteorological factors such as wind speed, temperature, and humidity influence $PM_{2.5}$ concentrations. For example, higher wind speeds facilitate pollutant dispersion, while lower humidity may inhibit particulate growth, thereby improving air quality. These interdependencies are vital for developing integrated strategies to mitigate air pollution and protect public health. The insights derived from this analysis were explored in subsequent sections of this chapter to:

1. Investigate the specific effects of meteorological variables on $PM_{2.5}$ levels, including their roles in pollutant transport and dispersion.
2. Refine predictive models for $PM_{2.5}$ concentrations to enhance their utility for environmental sustainability.

These findings have significant implications for regions characterised by pronounced seasonal variability, enabling targeted interventions and improved air quality predictions under changing meteorological conditions (Zhang, 2019; Aladag, 2023).

**Table 7.2:** Correlation coefficients between daily mean meteorological parameters in the study area from March 2021 to February 2022.

| Parameters | ws | Wd | Humidity | Temp | Rain | Pressure |
|---|---|---|---|---|---|---|
| ws | 1.000** | 0.138** | -0.287** | 0.193** | -0.042** | -0.041** |
| wd | 0.138** | 1.000** | -0.104** | -0.183** | -0.085** | 0.136** |
| Humidity | -0.287** | -0.104** | 1.000** | -0.153** | 0.614** | 0.123** |
| Temp | 0.193** | -0.183** | -0.153** | 1.000** | 0.005 | -0.653** |
| Rain | -0.042** | -0.085** | 0.614** | 0.005 | 1.000** | -0.077** |
| Pressure | -0.041** | 0.136** | 0.123** | -0.653** | -0.077** | 1.000** |

Note*Correlation coefficient is significant at the 95% confidence level (P <0.05).



© University of Venda

### 7.3.3 Sites-specific Relationship between Meteorological Variables and Particulate Matter ($PM_{2.5}$) using Spearman Approach

The relationship between $PM_{2.5}$ concentrations and meteorological variables in the study area was critically examined to determine how weather conditions influence air quality using the non-parametric Spearman correlation test at a 5% significance level. The heatmap in Figure 7.4 visually illustrates the Spearman correlation coefficients between the meteorological parameters and measured $PM_{2.5}$ concentrations at each monitored site. Most of the p-values associated with the Spearman rank correlation matrix were statistically significant, emphasising the critical impact of meteorological factors on $PM_{2.5}$ dynamics (Appendix 7.4). Humidity consistently exhibited a moderately negative correlation with $PM_{2.5}$ concentrations, particularly at sites such as VTV2 (-0.56), VMV8 (-0.55), and VP4 (-0.51) at 95% confidence interval (CI) (Figure 7.4). This negative correlation suggests that $PM_{2.5}$ levels decrease as humidity increases and vice versa. The hygroscopic nature of particulate matter allows it to absorb moisture, leading to particle enlargement and eventual removal via wet deposition processes such as rainfall. Consequently, humidity acts as a natural air purifier, enhancing overall air quality (Aladag, 2023; Chen *et al*., 2024).

Furthermore, high humidity fosters fog and mist formation, which aids in the capture and removal of airborne particulates. These processes underline humidity's role as an atmospheric cleanser, particularly in regions experiencing high relative humidity levels. The statistical significance of the correlation coefficient at VTV2 ($p = 0.0235$) underscores the necessity of considering humidity in air quality management strategies (Appendix 7.4). This finding is consistent with prior studies that highlight humidity's impact on air pollution dynamics; for instance, Sun *et al*. (2020) demonstrated that $PM_{2.5}$ levels negatively correlated with relative humidity across seasonal variations in China. Similarly, Aladag (2023) found a significant negative correlation between relative humidity and pollutants such as $PM_{10}$, $PM_{2.5}$, $SO_2$, CO, $NO_2$, and $O_3$. The current results are further corroborated by Bose & Roy Chowdhury (2023) and Banerjee *et al*. (2024), who emphasized the role of humidity in enhancing atmospheric cleansing mechanisms.



© University of Venda

Rainfall generally showed negative correlations with $PM_{2.5}$ concentrations, as observed at sites like VMV8 (-0.43) and VTV2 (-0.39). This reinforces the critical role of rainfall as a primary mechanism for removing airborne particulates through wet deposition, thereby significantly reducing $PM_{2.5}$ concentrations (Wang *et al*., 2023a). The efficiency of this process, however, depends on the intensity, frequency, and duration of rainfall, as well as localised environmental factors. Regions with frequent and heavy rainfall often experience marked lower $PM_{2.5}$ levels due to enhanced particle removal. These findings align with prior studies, including Yuan *et al*. (2023), which demonstrated the substantial impact of rainfall on particulate concentrations in urban and semi-urban areas, thus, the consistent negative correlations observed in this study highlight the essential function of rainfalls as a natural atmospheric purifier.

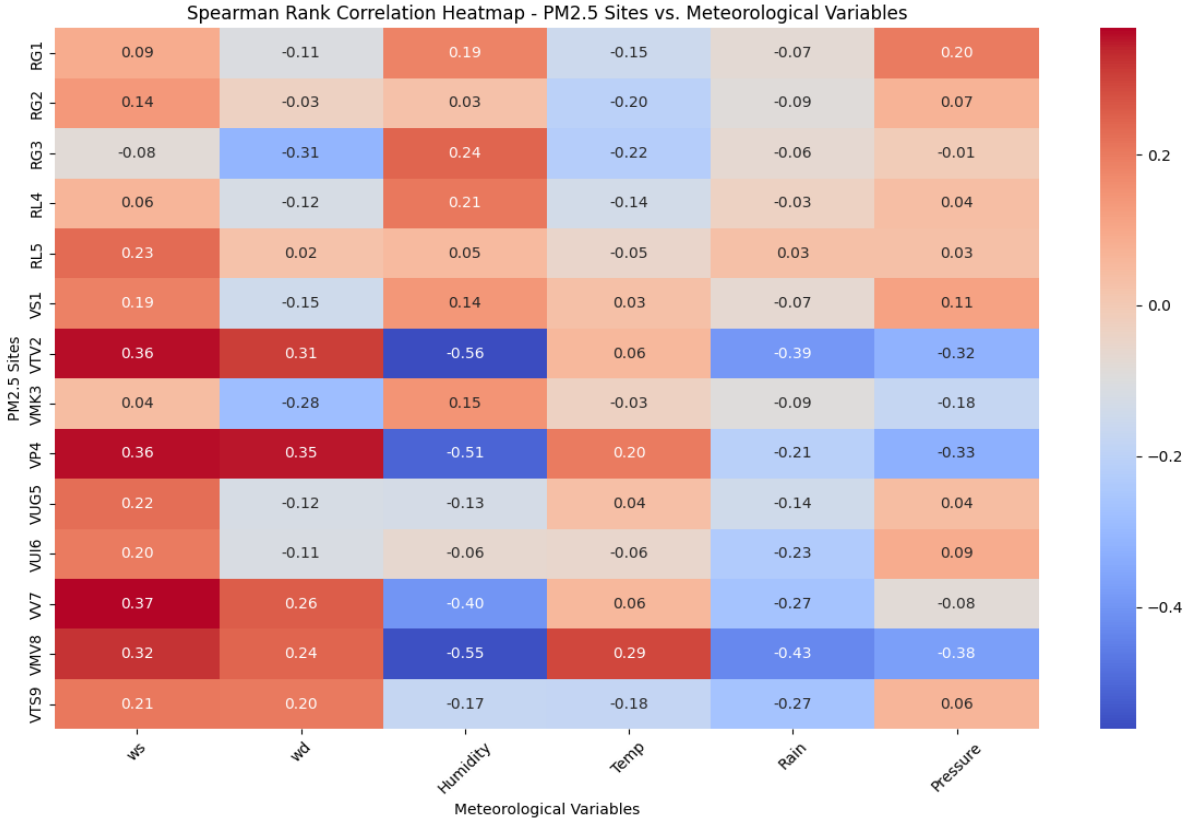



**Figure 7.4**: Spearman rank correlation between the $PM_{2.5}$ pollutant and the meteorological variables.



© University of Venda

The relationship between temperature, atmospheric pressure, and $PM_{2.5}$ concentrations was more nuanced and site-specific (Figure 7.4). For temperature, weak positive correlations were observed at some sites, such as VP4 (0.20) and VMV8 (0.29), while weak negative correlations were noted at RG2 (-0.20), RG3 (-0.22), and VTS9 (-0.18). These mixed findings suggest that temperature influences $PM_{2.5}$ differently based on site-specific geographical and environmental factors (Kallenborn *et al*., 2012). Higher temperatures can enhance atmospheric mixing, facilitating pollutant dispersion and lowering surface-level $PM_{2.5}$ concentrations. Conversely, colder conditions, particularly during winter, often trap pollutants near the surface due to stable atmospheric conditions, leading to elevated $PM_{2.5}$levels (Owusu-Tawiah *et al*., 2022).

Furthermore, elevated temperatures may contribute to the formation of secondary pollutants, such as ozone, which can exacerbate $PM_{2.5}$ concentrations under certain conditions. Likewise, atmospheric pressure exhibited weak correlations with both positive and negative values observed. The negative correlations were observed at sites VTV2 (-0.32), VP4 (-0.33), and VMV8 (-0.38). Lower pressure systems, often associated with unstable weather conditions, enhance vertical mixing and pollutant dispersion, which reduces the concentration of pollutants. Conversely, higher pressure systems, characterised by stable atmospheric conditions, tend to trap pollutants near the surface, leading to increased $PM_{2.5}$ concentrations. These findings are consistent with prior studies highlighting the role of atmospheric stability in air pollution dynamics (Liu *et al*., 2021; Zhang *et al*., 2022). However, the generally weak and statistically insignificant correlations observed in this study suggest that pressure alone may not have a substantial impact on $PM_{2.5}$ levels in this study area.

Wind speeds observed generally exhibit weak to moderate positive associations with $PM_{2.5}$ concentrations, particularly at sites VTV2 (0.36) and VV7 (0.37). This suggests that higher wind speeds may resuspend particulate matter or transport pollutants from surrounding areas (Jones & Williams, 2020). However, this relationship is not consistent across all sites, as factors such as topography, proximity to pollution sources, and local wind patterns significantly influence dispersion dynamics. For instance, regions with complex terrain or nearby pollution sources may experience increased particulate matter



© University of Venda

levels despite elevated wind speeds, as seen in this study. Conversely, wind direction shows weak negative correlations, implying that certain wind directions aid in pollutant dispersion and reduce concentrations. For example, at RG3, a negative correlation (-0.31) indicates that specific wind directions may help disperse pollutants (Figure 7.4), thereby lowering $PM_{2.5}$ concentrations in the area (Zhang *et al*., 2022).

The findings are essential for conducting localised assessments of wind patterns to develop effective air quality management strategies. Humidity and rainfall emerged as significant factors in reducing $PM_{2.5}$ levels, underscoring their role in natural air quality improvement through processes such as wet deposition. In contrast, the multifaceted effects of temperature, wind speed, wind direction, and atmospheric pressure reflect the complexity of air pollution dynamics and highlight the need for tailored interventions that consider local meteorological conditions, hence, this analysis is pivotal for predicting air quality and designing effective pollution mitigation strategies.

The findings further emphasised the necessity of integrating meteorological variables into air quality management strategies and predictive models for $PM_{2.5}$ concentrations. The significant p-values observed at sites VTV2, VP4, and VMV8 indicate that humidity consistently impacts $PM_{2.5}$ levels in these locations (Figure 7.4). This emphasises the need to prioritise humidity as a critical factor in air quality management strategies, particularly in these specific areas. The study also highlights the variability in the influence of other meteorological factors, such as wind speed, rainfall, and atmospheric pressure, suggesting that more detailed and site-specific analyses are required to fully understand their impacts (Liu *et al*., 2020; Liu *et al*., 2024).

### 7.3.4 Unveiling Site-Specific $PM_{2.5}$ Dynamics with Machine Learning Models

The predictive performance of various machine learning models was evaluated to predict $PM_{2.5}$ levels across the different towns surveyed, during the study period. This analysis investigated the variability of ambient $PM_{2.5}$ concentrations to develop predictive models for daily concentrations based on a Multiple Linear Regression (MLR) model (Antai *et al*., 2018). Random Forest (RF) model was developed simultaneously with the same dataset for comparison. The two models were compared to identify the possible impacts of



© University of Venda

meteorological factors (Ws, Wd, Temp, H, R, and AP) on $PM_{2.5}$ concentrations (Tables 7.3 and 7.4). The $R^2$, MSE and RMSE were used as metrics to assess predictive accuracy and model fit.

#### 7.3.4.1 Prediction of daily ambient $PM_{2.5}$ concentrations using MLR model, per site

In this section, the performance evaluation of the MLR predictive model for daily $PM_{2.5}$ concentrations was conducted based on statistical metrics that assess both predictive accuracy and goodness-of-fit (Table 7.3, Appendix 7.5). The predictors varied across sites, although the results indicated that the MLR model aligned reasonably well with the measured data across all study sites. The RMSE values ranged from 3.71 µg/m³ to 9.87 µg/m³, with lower values observed at site, such as RG3 (3.71 µg/m³) and RG1 (4.59 µg/m³). These lower values suggest moderate accuracy, indicating that the model effectively predicts $PM_{2.5}$ concentrations at these locations; however, higher RMSE values at RG2 (9.87 µg/m³) imply larger prediction errors. The MLR model's performance varied across sites, reflecting the differences in the influence of meteorological factors on $PM_{2.5}$ concentration.

The $R^2$ coefficient predicted the $PM_{2.5}$ concentration using multilinear regression analysis (Table 7.3), indicating that the values ranged from 0.048 to 0.485. As depicted in Figure 7.5, the $R^2$ values resulting from MLR were generally low, with the highest values observed at sites VP4 (0.485), VTS9 (0.352), and VMV8 (0.340), indicating the significant meteorological influence on $PM_{2.5}$ variability at high-traffic locations. Conversely, sites RG3, VUI6, VV7, and VUG5 demonstrated moderate $R^2$ values (0.189, 0.242, 0.293, and 0.304, respectively), suggesting that meteorological factors accounted for only 20–30% of the $PM_{2.5}$ variability in both residential and traffic areas. Figure 7.5 further illustrates the model metrics by comparing the predicted and observed $PM_{2.5}$ values. A key limitation of the MLR model was the high condition number of 3.19e x $10^5$, indicating potential multicollinearity among predictors. The VIF values (ranging from 1.062 to 2.125) revealed some collinearity, however, they remained below the acceptable threshold of 10, suggesting that multicollinearity did not significantly affect model reliability. This confirms that model predictions were not biased by redundant meteorological features. Recent atmospheric studies have shown that multicollinearity can inflate standard errors,



© University of Venda

destabilise coefficient estimates, and complicate the interpretation of individual predictors' contributions to $PM_{2.5}$ concentration or dispersion (Xu *et al*., 2024; Talepour *et al*., 2024). Xu *et al*. (2024), for example, emphasised the critical nature of VIF analysis in improving model robustness in air pollution prediction, while Talepour *et al*. (2024) demonstrated its role in feature selection for meteorological predictors in $PM_{2.5}$ concentration studies.

Furthermore, the site-specific MLR model equations (Table 7.3) were developed using ambient daily $PM_{2.5}$ concentrations for each residential and vehicular/traffic study site. These equations provide valuable tools for predicting $PM_{2.5}$ concentrations under varying meteorological conditions, aiding in air quality management. As earlier observed, MLR analysis revealed that meteorological parameters significantly influence $PM_{2.5}$ concentrations across the study area; for instance, wind speed (Ws) and atmospheric pressure (P) positively influenced $PM_{2.5}$ levels at high-traffic sites, such as VP4, VMV8, and VTS9. Conversely, variables such as temperature (Temp) and rainfall (R) exhibited localized impacts, reflecting diverse environmental and anthropogenic influences. This highlights the importance of localized modelling approaches in air quality assessments.

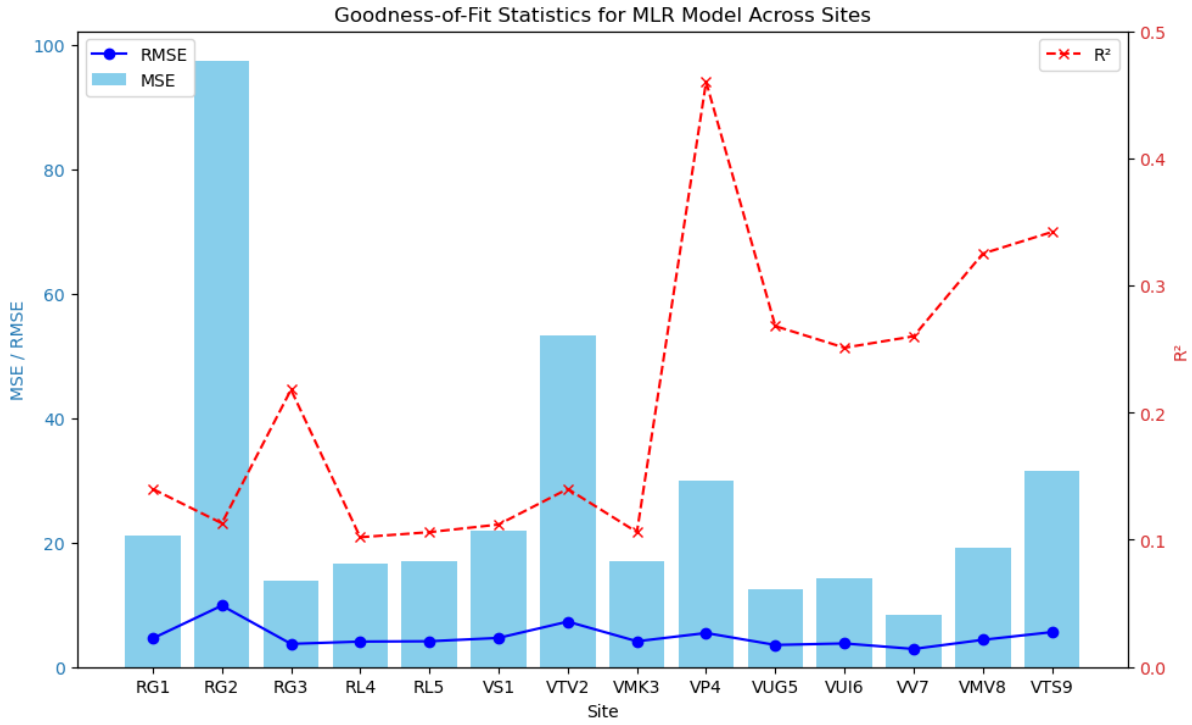



**Figure 7.5:** Predictive accuracy and model fit metrics for each site using multiple linear regression model.



© University of Venda

University of Venda
Creating Future Leaders

**Table 7.3:** Multiple Regression (MLR) Model equations and the goodness of fit statistics for each study site

| **Model Equations** | $R^2$ | MSE | RMSE |
|---|---|---|---|
| RG1 = −219.656 + 1.490 Ws - 0.014 Wd + 0.056 H - 0.127 T + 0.092 R + 0.253 P | 0.146 | 21.078 | 4.591 |
| RG2 = 217.688 + 0212 Ws - 0.040Wd -0.3263 H + 0.450 T + 0.449 R + 0.170 P | 0.148 | 97.342 | 9.866 |
| RG3 = -217.658 - 0854 Ws - 0.004 Wd - 0.042 H - 0.192 T + 0.173 R + 0.263 P | 0.189 | 13.816 | 3.717 |
| RL4 = 244.066 – 0.620 Ws - 0.0057 Wd - 0.075 H - 0.061 T - 0.156 R + 0.289 P | 0.112 | 16.672 | 4.083 |
| RL5 = - 86.942 - 0.240 Ws - 0.006 Wd - 0.039 H - 0.219 T + 0.380 R + 0.123 P | 0.089 | 17.071 | 4.132 |
| VS1 = -70.057 - 0.334 Ws + 0.002 Wd - 0.106 H - 0.330 T - 0.165 R + 0.115 P | 0.126 | 21.854 | 4.675 |
| VTV2 = -120.550 -1.342 Ws + 0.029 Wd - 0.205 H - 0.161 T - 0.221 R + 0.176 P | 0.157 | 53.215 | 7.295 |
| VMK3 = -69.270 - 0.258 Ws - 0.003 Wd - 0.054 H - 0.111 T + 0.114 R + 0.104 P | 0.048 | 17.071 | 4.132 |
| VP4 = -399.265 - 1.225 Ws + 0.031 Wd - 0.258 H - 0.3556 T + 0.169 R + 0.473 P | 0.485 | 29.861 | 5.464 |
| VUG5 = -222.502 - 0.466 Ws + 0.011 Wd - 0.081 H - 0.332 T + 0.208 R + 0.270 P | 0.304 | 12.544 | 3.542 |
| VUI6 = -238.149 - 0.895 Ws + 0.009 Wd - 0.108 H - 0.274 T + 0.178 R + 0.291 P | 0.242 | 14.325 | 3.785 |
| VV7 = - 185.986 - 0.979 Ws + 0.010 Wd - 0.068 H - 0.257 T + 0.430 R + 0.231 P | 0.293 | 8.412 | 2.900 |
| VMV8 = -119.060 - 0.731 Ws + 0.019 Wd - 0.194 H - 0.309 T + 0.048 R + 0.171 P | 0.340 | 19.067 | 4.367 |
| VTS9 = -178.931 - 0.506 Ws + 0.007 Wd - 0.242 H - 0.586 T + 0.256 R + 0.248 P | 0.352 | 31.478 | 5.610 |



© University of Venda

### 7.3.4.2 Prediction of daily ambient $PM_{2.5}$ per site using random forest model

The random forest model was applied to predict daily ambient $PM_{2.5}$ levels across multiple monitoring sites, with hyperparameter tuning conducted for each site using Randomized Search to optimise model performance for each location. The predicted values were plotted against measured values as shown in Figure 7.6, while the goodness of fit metrics, including $R^2$, MSE, SSE, RMSE, and MAE were used to evaluate the model's accuracy during the training and testing phases (Appendix 7.5). The cross-validation was further used to estimate the model's performance in capturing the variance and minimising the error associated with $PM_{2.5}$ predictions for both phases (Chen *et al*., 2021; Zhang *et al*., 2022).

During the training phase, the RF model metrics exhibited highly significant positive correlations between observed and predicted concentrations, with $R^2$ values ranging from 0.53 at RL4 to as high as 0.802 at VP4 (Appendix 7.6). This indicates that RF explains the difference between 53% to 80% variance in $PM_{2.5}$ concentrations across the sites. Notably, sites such as VP4, VTS9, and VUG5 (0.802, 0.786, and 0.748, respectively) achieved higher $R^2$ values, underscoring the model's strength in capturing underlying relationships for $PM_{2.5}$ at these locations. The MSE and SSE values for training datasets further validated the model's accuracy, with minimal error residuals across most sites. Correspondingly, low RMSE values for training (Appendix 7.6), typically between 2 and 8 $\mu g/m^3$ for most sites, further validated the RF model's robustness in handling the training dataset, which agrees with Xu *et al*. (2020) and Liu *et al*. (2021).

At the testing phase, the RF model's performance remained strong at several sites, with RMSE values close to the training RMSEs, indicating good generalisation. Sites such as RL5, VUI6, VP4 and VTS9 showed high $R^2$ values, confirming the RF model's ability to capture complex, non-linear relationships influencing $PM_{2.5}$ concentrations at these locations (Figure 7.6, Appendix 7.6). Figure 7.7 further shows the line plots depicting the predicted $PM_{2.5}$ concentrations against the observed values for some sites (RG3, RL3, VP4, VUI6, VMV8, and VTV9) using the RF model. Certain sites, including RG2, RL4, VS1, and VV7, however, exhibited negative $R^2$ values during testing, suggesting potential



© University of Venda

overfitting or the presence of unique environmental variables affecting $PM_{2.5}$ concentrations that RF could not fully model (Chen *et al*., 2019). The variability in RMSE across sites, ranging from approximately 3 to over 6 µg/m³, further demonstrated how site-specific characteristics played a critical role in predictive performance. This consistency in RMSE values suggests good generalisation capabilities of the RF model. Additionally, MSE and SSE values for training datasets further validated the model's accuracy, with minimal error residuals across most sites (Appendix 7.6).

*A. Model robustness, stability and hyperparameter tuning*

To further assess RF's robustness, cross-validation (CV) results were computed for each site. Mean CV RMSE values ranged from 3.3 µg/m³ (VV7) to over 8.9 µg/m³ (RG2), with standard deviations reflecting data variability (Appendix 7.6). Sites such as VUG5 and RG3 demonstrated lower CV RMSE scores, indicating consistent model performance, while higher standard deviations at sites like RG2 suggested variability in performance, across different data splits. These findings highlight the importance of cross-validation in assessing model stability and identifying potential overfitting or underfitting issues (Wang *et al*., 2023b). Additionally, high mean CV RMSE and standard deviations at certain sites may suggest the presence of unique and complex interactions between meteorological parameters that the RF model could not fully capture. The RF model's flexibility was further illustrated through site-specific hyperparameter optimisation. For example, most sites achieved optimal performance with 200 estimators, a max depth of 30, and a minimum sample split of 2. Constrained models (for example, 100 estimators and a max depth of 10) optimised performance at sites like VUI6 and VTS9, however, suggested that certain locations benefitted from tailored hyperparameters to balance model complexity and accuracy (Safo-Adu *et al*., 2023). These variations indicate that while RF adapts well to various conditions, site-specific hyperparameter tuning is important for maximizing model predictive performance.



© University of Venda

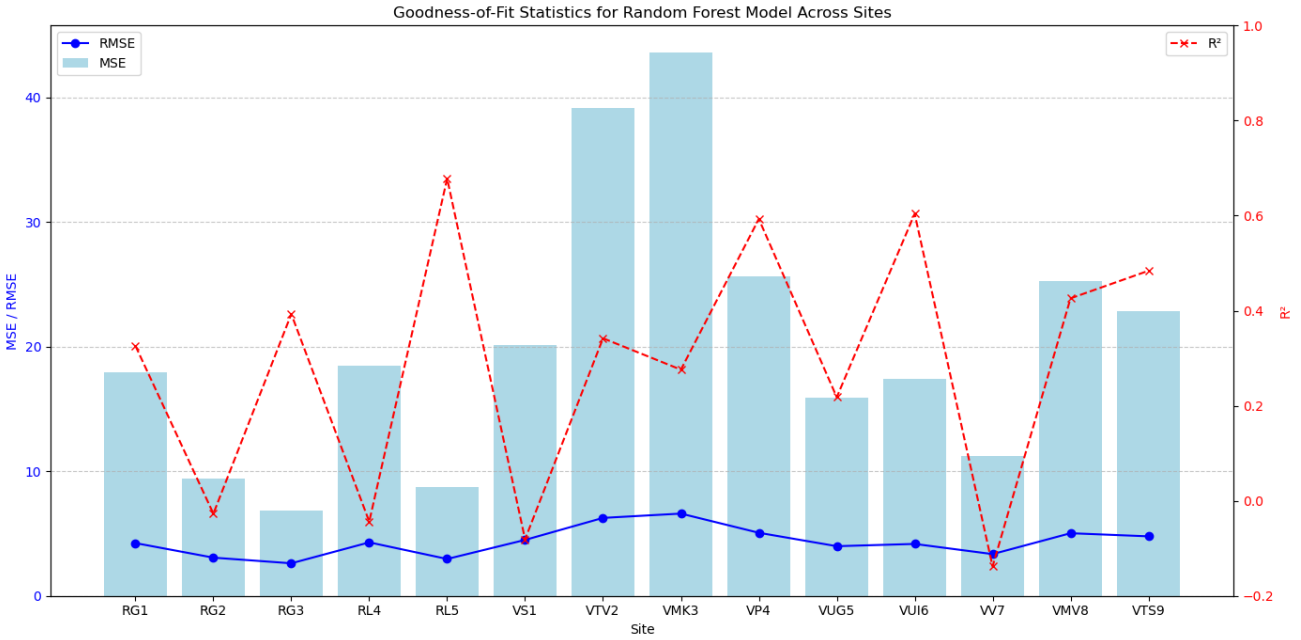



**Figure 7.6:** Predictive accuracy and goodness of fit metrics for each site using random forest model.

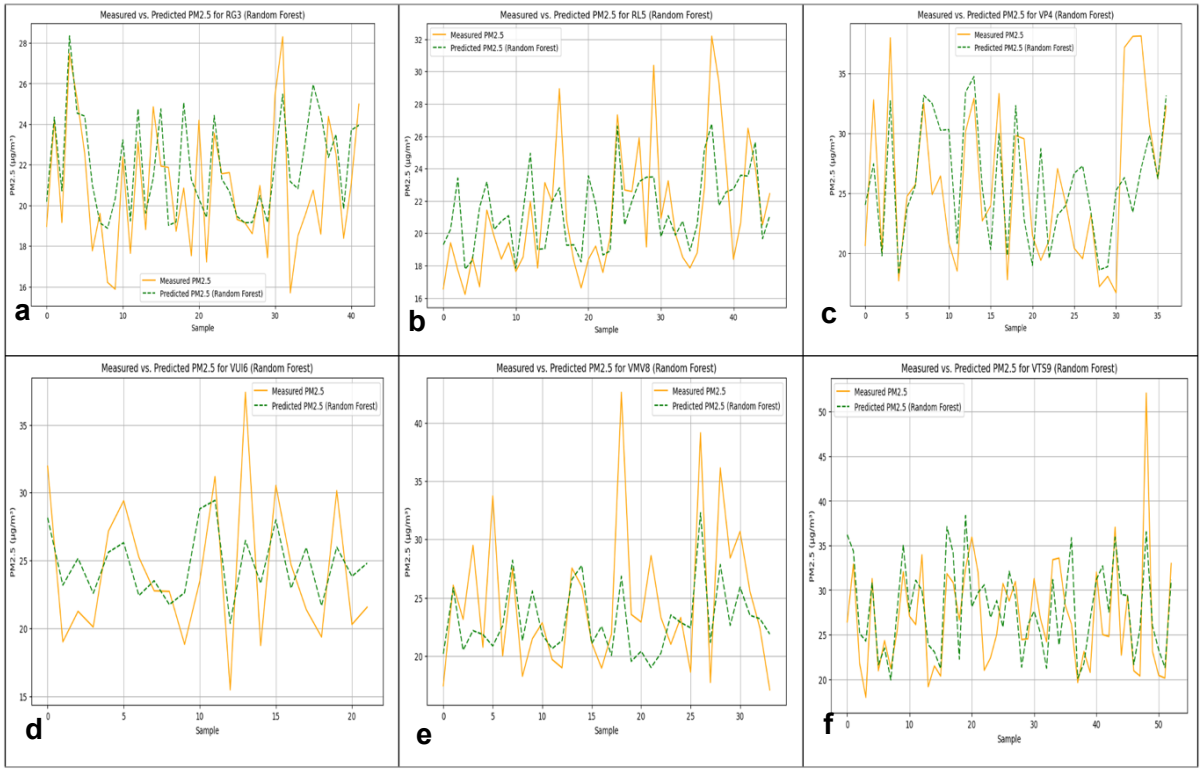



**Figure 7.7:** Relationship between predicted and measured $PM_{2.5}$ concentration ($\mu g/m^3$) for some selected sites using random forest model for a) RG3, b) RL3, c) VP4, d) VUI6, e) VMV8 and f) VTV9.



© University of Venda

***B.*** *Comparative performance between MLR and RF for site specific prediction*

The RF model provided strong predictive capabilities for $PM_{2.5}$ across several sites as compared to MLR model. While MLR exhibited moderate performance at selected sites, the RF model consistently outperformed it across all metrics, demonstrating superior capacity to handle non-linear relationships and complex interactions between meteorological variables (Liu *et al*., 2019; Sulaymon *et al*., 2023). The spatial heterogeneity of $PM_{2.5}$ influences, reflected in varying $R^2$ and RMSE values across sites, highlights the need for site-specific tuning to improve predictive accuracy; for example, RF captured 80% of the variance in $PM_{2.5}$ concentrations at sites like VP4 and VTS9, demonstrating its robust performance. In contrast, the MLR model struggled to account for non-linear interactions, as evidenced by lower $R^2$ values; despite these limitations in generalisation at specific sites (e.g., RG2 and RL4), the RF model's ability to model complex, non-linear relationships provides a significant advantage over linear models like MLR (Hastie *et al*., 2009).

The RMSE values for $PM_{2.5,}$ on the other hand, predictions vary across sites indicating moderate accuracy, while higher errors at RG2 highlight significant prediction challenges. This discrepancy suggests a need for site-specific calibration, where model parameters are adjusted to account for local environmental factors, such as emissions, meteorology, and terrain variability. Calibration enhances the model's robustness and generalizability, ensuring reliable predictions across diverse scenarios and improving overall performance. This was evident in the developed RF-based models that reduce the RMSE values, such as in RG1 reduction from 4.591 to 2.489. This observation in the finding of RF models demonstrates the significant influence and accuracy of prediction as compared to the developed MLR model. The aforementioned decrease in error suggests that the RF models possess the ability to better approximate the true $PM_{2.5}$ concentrations, hence enhancing the reliability of the predictions, as noted by Suchitra, (2024). This author emphasises the utility of RF in identifying problematic sites due to its RMSE's sensitivity to large errors (Suchitra, 2024). This further demonstrates the potential of ML models in environmental monitoring and highlights the importance of continued



© University of Venda

model refinement to address site-specific dynamics, effectively (Xu *et al*., 2024; Talepour *et al*., 2024).

### 7.3.5. Integration of Feature Importance with the Prediction of Average $PM_{2.5}$ Across the Study Areas

This section implemented polynomial ridge regression (PRR) along with multiple linear regression (MLR) and random forest (RF) models in predicting average daily $PM_{2.5}$ concentrations across all monitoring sites, within the study area. The inclusion of PRR aimed to improve prediction accuracy by capturing complex, non-linear atmospheric processes influencing $PM_{2.5}$ levels, including interactions among meteorological variables. Incorporating higher-order terms and interactions via polynomial regression accounts for the non-linear relationships and variations in environmental conditions impacting $PM_{2.5}$ concentrations. Average daily concentrations across all monitoring sites over one-year period were used to evaluate the models' performance in predicting $PM_{2.5}$ levels.

#### 7.3.5.1. Prediction of average $PM_{2.5}$ pollution across the study areas using MLR model along with feature importance

The MLR model yielded an $R^2$ of 0.144, RMSE of 3.545 µg/m³, and MSE of 12.565 µg/m³, highlighting its limitations in modelling complex interactions. These results suggest the inadequacy of the MLR model in capturing non-linear relationships between meteorological parameters and $PM_{2.5}$ levels. The MLR Equation 7.9 developed for $PM_{2.5}$ prediction is:

$$PM_{2.5} = -208.998 - 0.578\,\mathrm{Ws} + 0.019\,\mathrm{Wd} - 0.099\,\mathrm{H} - 0.028\,\mathrm{T} + 0.124\,\mathrm{R} + 0.250\,\mathrm{P} \quad (7.9)$$

This model identifies several meteorological parameters as significant predictors (Figure 7.8). Feature-specific regression analysis identified Ws, Wd, and atmospheric pressure as significant predictors ($p < 0.05$), however, rainfall and pressure showed negligible contributions, indicating that these variables effects may be location-specific or non-linear in nature. The atmospheric pressure demonstrates a significant positive coefficient (0.2501, $p < 0.0001$), which indicates that stable high-pressure systems trap pollutants



© University of Venda

close to the surface, thereby increasing $PM_{2.5}$ levels. Similarly, Wd has a positive coefficient (0.019, $p < 0.0001$), suggesting that certain wind directions contribute to higher $PM_{2.5}$ levels due to the transporting of pollutants from external sources. Likewise, rainfall shows a positive but statistically insignificant coefficient (0.1239, $p = 0.2533$), suggesting that rainfall's influence on $PM_{2.5}$ levels is inconsistent. While rainfall is generally associated with reducing particulate matter through wet deposition, the variability in its impact may be attributed to differences in rainfall intensity, duration, or localised conditions across the study area. Conversely, Ws exhibits a significant negative coefficient (-0.578, $p = 0.045$), indicating that increased wind speed disperses pollutants, leading to reduced $PM_{2.5}$ concentrations. Humidity demonstrates a significant negative coefficient (-0.099, $p < 0.0001$), which supports the hypothesis that higher humidity facilitates particulate matter deposition, thereby lowering airborne concentrations.

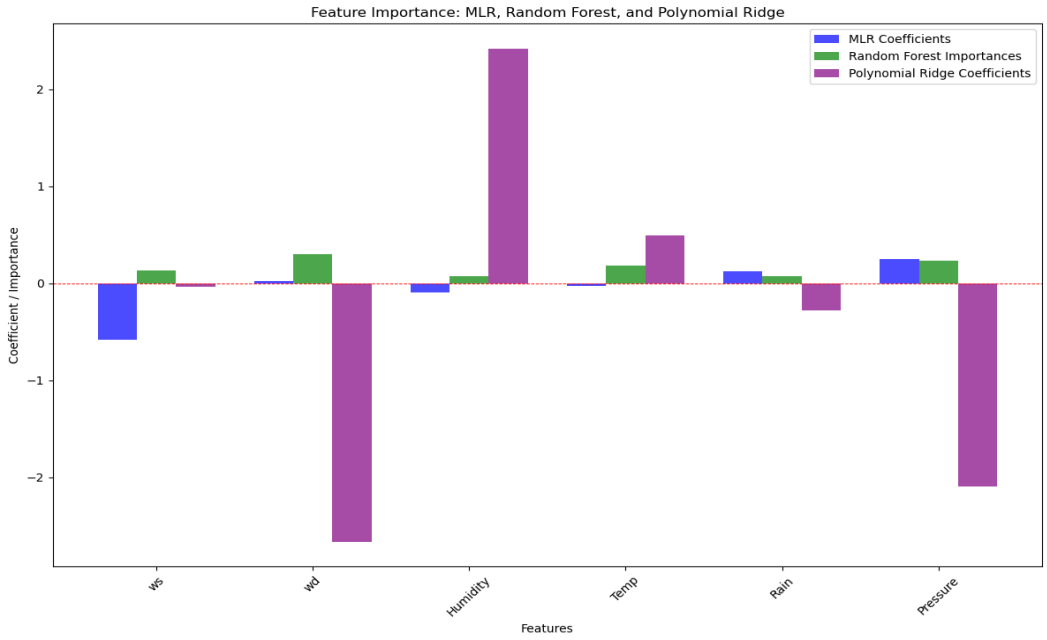



**Figure 7.8:** Bar plot showing the model comparison feature importance of the machine learning models.

Temperature is negligible and statistically insignificant, implying that this parameter does not play a significant linear role in influencing $PM_{2.5}$ levels in this dataset when the MLR algorithm was used. This finding aligns with other studies that emphasise the importance of integrating meteorological and anthropogenic factors for comprehensive air quality



© University of Venda

modelling (Edlund *et al*., 2021; Sulaymon *et al*., 2023; Aruna and Ravindiran, 2024). This does not, however, rule out non-linear or site-specific effects, which may require further investigation. Despite these insights, the model's relatively low $R^2$ value (0.144) suggests that additional non-meteorological factors, such as traffic density, industrial emissions, and land-use characteristics, must be incorporated to improve predictive accuracy.

#### 7.3.5.2 Prediction of average $PM_{2.5}$ pollution across the study areas using the polynomial ridge regression model

The Polynomial Ridge Regression (PRR) model demonstrated notable improvements in predictive accuracy compared to MLR, effectively capturing non-linear interactions and complex atmospheric dynamics. The PRR extends traditional linear regression by incorporating higher-order (for example, quadratic) terms and interaction terms, enabling it to capture non-linear relationships and complex interactions between variables. A detailed statistical evaluation revealed an RMSE of 3.215 µg/m³, $R^2$ of 0.297, MSE of 10.334 µg/m³, and SSE of 1467.35 µg/m³. These metrics indicate that PRR explains nearly 30% of the variance in $PM_{2.5}$ levels, marking a substantial enhancement over the explanatory power of MLR. The PRR Equation (7.10) developed for the $PM_{2.5}$ prediction is:

$$_{Avg}PM_{2.5PRR} = 978.742 - 0.026\ Ws - 2.464\ Wd + 2.816\ H - 0.296\ T - 0.277\ R - 1.946\ P - 0.484\ Ws^2 + 0.013\ Ws*Wd + 0.105\ WsH - 0.223\ WsT - 0.723\ WsR - 0.004\ WsP - 0.0001Wd^2 + 0.002\ WdH + 0.002\ WdT - 0.015\ WdR + 0.0025\ WdP + 0.007\ H^2 + 0.018\ HT - 0.085\ HR - 0.005\ HP - 0.033\ T^2 + 0.0688\ TR + 0.0002\ TP + 0.299\ R^2 + 0.008\ RP + 0.001\ P^2 \quad (7.10)$$

As depicted in Figure 7.8, the feature-specific regression analysis identified Wd, humidity, and temperature as significant predictors ($p < 0.05$). Wind direction exhibited the strongest influence on $PM_{2.5}$ levels, with a negative coefficient (-2.464), highlighting its critical role in pollutant dispersion. This finding aligns with Lu and Fang (2002), who reported an inverse relationship between air pollutant concentrations and wind speed. Wind direction's role is further emphasised as certain wind patterns can significantly alter pollutant distribution, either dispersing pollutants or concentrating them in specific regions. Humidity had a positive coefficient (2.816), underscoring its complex role in $PM_{2.5}$



© University of Venda

dynamics. This relationship is likely tied to hygroscopic particle growth, wherein water vapour condenses onto particulate matter under high humidity conditions, increasing particle size and altering optical and aerodynamic properties. Such changes enhance light scattering, reduce visibility, and influence $PM_{2.5}$ deposition processes, as highlighted by Fu *et al*. (2016) and Amnuaylojaroen *et al*. (2023).

The incorporation of non-linear terms, such as $Ws^2$ and interaction terms like WsWd (interaction between wind speed and direction), allowed the PRR model to effectively capture critical atmospheric processes, including pollutant dispersion, deposition, and transportation, influenced by dynamic meteorological interactions (Lu and Fang, 2002; Hossain *et al*., 2024). Notably, the non-linear terms captured temperature inversions and stagnation events, offering significant insights into $PM_{2.5}$ dynamics as noted by Nejad *et al*. (2023). The temperature inversions occur when a stable layer of warm air traps cooler air below, preventing vertical mixing and concentration of pollutants near the ground, thereby exacerbating air quality issues (Nejad *et al*., 2023). Similarly, stagnation events, characterised by low wind speeds and high atmospheric pressure, restrict horizontal pollutant dispersion and promote accumulation.

By incorporating higher-order and interaction terms, the PRR model provides a more refined depiction of the relationships between meteorological variables and $PM_{2.5}$ concentrations, emphasising its suitability for capturing the intricate environmental dynamics influencing air pollution. These non-linear terms are particularly valuable in environmental modelling, where atmospheric processes often exhibit intricate dependencies; for instance, quadratic terms like wind speed squared ($Ws^2$) account for the non-linear effects of wind turbulence on pollutant dispersion. As wind speeds increase beyond a certain threshold, pollutants may disperse at an accelerated rate, reducing $PM_{2.5}$ concentrations. Interaction terms, such as WsWd, explain how these factors jointly influence pollutant transport. Specific wind directions combined with high wind speeds can either amplify or mitigate pollutant levels, depending on the geographical layout and emission sources.



© University of Venda

Similarly, quadratic terms for humidity ($H^2$) and pressure ($P^2$) model non-linear effects, such as the saturation of wet deposition processes at high humidity levels or the trapping of pollutants under stable high-pressure systems. Interaction terms, such as H*T and R*P, highlight how combined meteorological conditions influence $PM_{2.5}$ concentrations in ways that linear models cannot adequately describe. For instance, high rainfall during stable atmospheric conditions may lead to inconsistent $PM_{2.5}$ reduction due to variability in particle washout efficiency; for although the $R^2$ value may remain relatively low, capturing only 30% of the variance in $PM_{2.5}$ predictions, the PRR model provides valuable insights into the non-linear and complex interactions between meteorological variables and pollutant concentrations. This highlights the model's significance in unveiling intricate atmospheric processes that linear models fail to capture, making it a crucial tool for advancing air quality assessment and informing targeted management and mitigation strategies.

#### 7.3.5.3 Prediction of average $PM_{2.5}$ pollution across the study areas using random forest

The RF model was further applied to better capture non-linear and complex relationships influencing $PM_{2.5}$ concentrations. Unlike the MLR and PRR models, the RF model significantly improved performance with an RMSE of 2.410 µg/m³, an $R^2$ of 0.604, and an MSE of 5.810 µg/m³. These results reveal that the RF model explains approximately 60% of the variance in $PM_{2.5}$ levels, thereby surpassing the explanatory power of the MLR and PRR models. RF effectively manages data variability and extreme values by leveraging on ensemble learning methods, which aggregate predictions from multiple decision trees (Gündoğdu and Elbir, 2024). Each tree identifies unique patterns, and their aggregated predictions provide a more generalised and accurate outcome. This ensemble learning method allows RF to effectively model non-linearities and interactions within the data, thereby enhancing its predictive accuracy (Miller *et al*., 2023). The RF's robustness to outliers and ability to model complex interactions make it the most effective tool for predicting $PM_{2.5}$ concentrations as influenced by a mix of meteorological variables and localised factors. Consequently, this approach not only improves predictive stability but also captures complex, non-linear dependencies that other models struggle to represent.



© University of Venda

These characteristics show the RF's advantage in predicting $PM_{2.5}$ concentrations within dynamic environmental conditions. This aligns with findings from Imam *et al*., (2024) and Panaite *et al*. (2024), who demonstrated the effectiveness of RF in capturing complex environmental dynamics and adapting to diverse datasets, making it a superior choice for air quality modelling.

*A. Integration with feature importance with random forest model prediction*

In addition to the model metrics, feature importance analysis provides valuable insights into the RF model's ability to identify the most influential predictors of $PM_{2.5}$ concentrations. As shown through the bar plots in Figure 7.8, Wd had the highest importance score (0.298). This reflects its key role in air quality dynamics since Wd significantly affects the dispersion and accumulation of pollutants. Specific wind directions are associated with increased $PM_{2.5}$ levels due to pollutant transporting from external sources, particularly in semi-urban areas where emissions from vehicular traffic and industrial activities dominate. The Ws, with an importance score of 0.137, demonstrated moderate importance in the RF model, and although less influential than Wd, Ws plays a critical role in pollutant dispersion and dilution, as higher wind speeds facilitate the reduction of $PM_{2.5}$ concentrations. Similarly, atmospheric pressure, with an importance score of 0.238, highlights its significant role in air quality processes. Stable high-pressure systems have been shown to trap pollutants near the surface by inhibiting vertical mixing, which leads to elevated $PM_{2.5}$ levels. This effect is particularly pronounced in areas with heavy vehicular emissions and stagnant atmospheric conditions, as observed in other studies (Kamińska, 2018; Zhang *et al*., 2022; Imam *et al*., 2024). Humidity exhibited a relatively lower importance score (0.072) in the RF model compared to wind-related factors, although, its contribution aligns with findings from the MLR model. The negative coefficient (-0.099) in MLR highlights humidity's role in facilitating wet deposition processes, thereby reducing $PPM_{2.5}$ concentration. RF analysis also captured interactions between humidity and other variables, such as wind speed and temperature, indicating that humidity's impact on $PM_{2.5}$ is modulated by broader atmospheric conditions. Rainfall and temperature, with an importance score of 0.076 and 0.180, respectively, demonstrated less significance in the RF model, however, their effects were context-



© University of Venda

dependent and exhibited non-linear influences on $PM_{2.5}$ dynamics. RF captured these variable impacts more effectively than MLR and PRR, which highlights the need for further investigation into their roles under diverse meteorological scenarios.

The RF model's ability to rank feature importance and capture complex non-linear interactions makes it the most robust tool for predicting $PM_{2.5}$ concentrations in this study. Shared features among MLR, PRR, and RF such as Ws, Wd, atmospheric pressure, humidity, rainfall, and temperature, demonstrated their consistent relevance across models; RF's capacity to model non-linearities and site-specific interactions provided deeper insights into the dynamics of air pollution. For instance, while wind direction was consistently ranked as the most influential feature across all models, RF's higher importance score (0.298) highlighted its critical role in pollutant transport and accumulation. Wind speed and atmospheric pressure are similarly significant across models. They are contextualised more effectively in RF through their interactions with other variables, such as humidity and rainfall. These findings align with prior studies, such as those by Kallenborn *et al*. (2012) and Edlund *et al*. (2021), which discuss the role of meteorological parameters in shaping $PM_{2.5}$ levels and the importance of incorporating advanced ML techniques into air quality assessments (Okimiji *et al*., 2021; Safo-Adu *et al*., 2023; Sulaymon *et al*., 2023). By effectively capturing the intricate interplay of meteorological factors, RF provides a detailed understanding of the processes driving $PM_{2.5}$ concentrations, offering valuable insights for policy development and targeted mitigation strategies. Research has further shown how variations in meteorological conditions, such as temperature, humidity, and wind patterns, are among the strongest predictors, and can significantly influence the formation, dispersion, and concentration of $PM_{2.5}$ particles in the atmosphere (Li *et al*., 2023).

Additionally, Zhang, (2019) discussed the importance of wind direction in redistributing particulate matter across urban and peri-urban areas. Their findings showed that specific prevailing wind directions amplify pollutant transport from external sources, exacerbating local air pollution levels. Similarly, the current findings resonate with the systematic reviews by Seinfeld and Pandis (2016) and Edlund *et al*. (2021), which underline the



© University of Venda

necessity of incorporating meteorological patterns, such as elevated humidity and wind interactions, in air quality assessments. Similarly, Safo-Adu *et al*. (2023) focus on the significant influence of localised conditions, such as vehicular emissions and biomass burning, on air quality, particularly in semi-urban regions with limited regulatory frameworks. These studies show that integrating meteorological parameters into advanced predictive models offers deeper insights into the mechanisms driving particulate matter dynamics, thereby supporting more effective environmental policies and mitigation strategies.

#### 7.3.5.4. Comparative analysis of model performance across machine learning algorithms

*A. Comparison based on feature selection*

Based on the three approaches used in the study, RF emerged as the most robust and accurate model, ranking above PRR and MLR. Its ability to dynamically model complex non-linear relationships and interactions between variables ensures superior predictive accuracy. By building and averaging multiple decision trees, RF handles variability, missing data, and outliers effectively, offering unmatched robustness for diverse datasets. Its feature importance rankings highlight key predictors - Wd (0.298), atmospheric pressure (0.238), and Ws (0.137) - that are pivotal to understanding $PM_{2.5}$ dynamics. In contrast, PRR ranked second, demonstrating improved accuracy over MLR by incorporating quadratic and interaction terms. These terms allowed PRR to model non-linear relationships, such as the effects of wind speed squared ($Ws^2$) on pollutant dispersion and wind direction's interaction with other variables, however, PRR's reliance on pre-specified terms, limits its adaptability, making it less versatile, than RF, in uncovering data-driven patterns.

The MLR ranked lowest, constrained by its inability to model non-linear dynamics, while it identified essential features like wind direction and atmospheric pressure, its linear assumptions restricted its performance, particularly when modelling complex atmospheric interactions. The increasing order of performance for the prediction of $PM_{2.5}$ concentrations was RF > PRR > MLR. This shows that RF's outperformed ranking is



© University of Venda

attributed to its consistent performance across cross-validation folds due to its ability in treating missing data without normalization, handling many input variables without overfitting, and feature-importance identification. Additionally, RF is flexible in capturing non-linear interactions between variables, such as the combined effects of humidity, rainfall, and temperature on $PM_{2.5}$ concentrations making it a leading algorithm for better prediction outputs.

This flexibility enables RF to contextualise meteorological dynamics more effectively than PRR or MLR; moreover, several studies that compared different ML algorithms for detecting peak $PM_{2.5}$ values found that RF emerged as the most effective choice (Ameer *et al*., 2019; Ghahremanloo *et al*., 2021), as further confirmed in Essamlali *et al*. (2024). Furthermore, Kumar *et al*. (2021) provided compelling evidence on the utility of advanced ML models like RF in accurately predicting $PM_{2.5}$ concentrations, while Agbehadji *et al*. (2024) identified the advantage of combining ML techniques as hybrid models to enhance the predictive accuracy of air quality models. By effectively ranking feature importance and capturing complex non-linearities, RF models outperformed traditional statistical methods, emphasising critical meteorological factors, such as Wd, atmospheric pressure, and humidity in shaping air quality dynamics.

*B. Comparison of predicted vs. actual $PM_{2.5}$ concentrations across models*

The comparison of predicted versus actual $PM_{2.5}$ concentrations, as illustrated in Figure 7.9 and Appendix 7.7, indicates the predictive performance of the three models - MLR, PRR, and RF in estimating $PM_{2.5}$ concentrations. These figures comprehensively demonstrate the varying capabilities of the models to align with actual $PM_{2.5}$ measurements. Figure 7.9 showcases scatter plots comparing predicted versus actual $PM_{2.5}$ concentrations for each model. Appendix 7.7 compared the training and test performance of the algorithms for $PM_{2.5}$ concentration predictions. Among the models, RF exhibits the most robust performance, with data points tightly clustered along the diagonal line, indicating a strong alignment between predictions and actual values (Zamani *et al*., 2019). This highlights the model's ability to capture the underlying relationships between predictors and $PM_{2.5}$ concentrations. In comparison, PRR performs



© University of Venda

moderately well, with some improvement in capturing non-linear trends, although its predictions still exhibit notable deviations from the actual values. Conversely, MLR reveals a much broader spread of data points, reflecting its weaker performance and its inability to model the complex, non-linear interactions inherent in $PM_{2.5}$ dynamics.

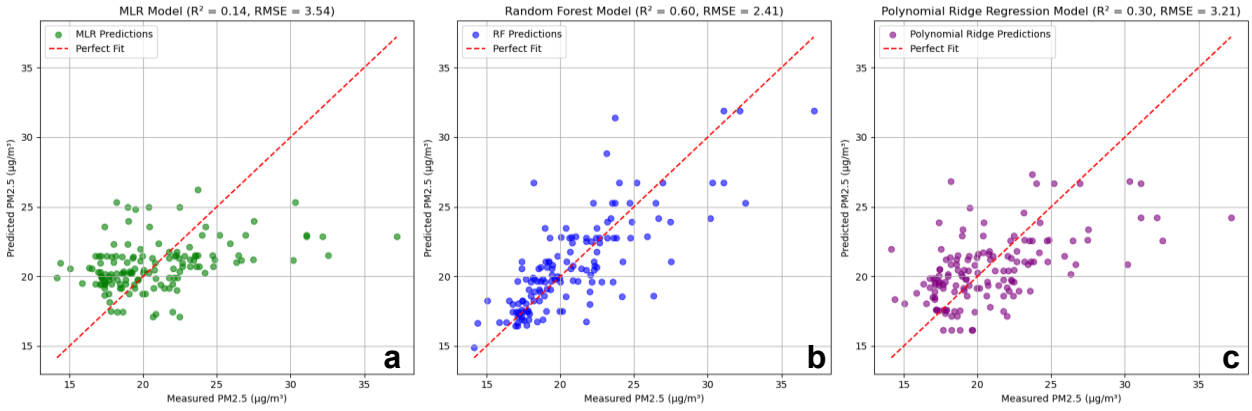



**Figure 7.9:** Scatter plot visualizing the predicted vs. actual $PM_{2.5}$ in the study area using the three selected machine learning models. a) MLR, b) RF and c) PRR models.

Figures 7.10 and 7.11 provide a deeper analysis of the comparative performance and stability of the three models. Figure 7.10 demonstrates the clustering of predictions and their alignment with actual measurements across the dataset, highlighting RF's superior performance and PRR's moderate alignment compared to MLR. The temporal variation in predictions underscores each model's ability to capture $PM_{2.5}$ dynamics. In addition to this depiction, Figure 7.11 further corroborates these findings through a time-series representation of the actual and predicted $PM_{2.5}$ concentrations. RF predictions closely follow the actual $PM_{2.5}$ values (yellow line), effectively capturing the temporal fluctuations across the dataset. The PRR model demonstrates improved predictive alignment compared to MLR, which consistently under- or over-estimates $PM_{2.5}$ levels. This Figure illustrates RF's ability to closely track actual $PM_{2.5}$ fluctuations over time, showcasing its effectiveness in generalising on unseen data; for example, on March 11, 2021, the actual concentration was 24.00 µg/m³, with RF predicting 26.72 µg/m³, PRR predicting 26.67



© University of Venda

μg/m³, and MLR predicting 22.98 μg/m³. Such results underscore RF's superior ability to model the intricate dynamics of $PM_{2.5}$ compared to the other models.

Figure 7.12 offers a detailed breakdown of the models' performance across key metrics, including RMSE, $R^2$, and MSE, as well as their stability during cross-validation. The bar plots in Figure 7.12a reveal that RF consistently outperformed the other models, achieving the lowest RMSE and MSE values, alongside the highest $R^2$. This can be attributed to RF's ensemble learning framework, which effectively models complex, non-linear relationships and variable interactions. PRR, while demonstrating improvements over MLR, falls short of RF in accuracy, particularly in scenarios involving highly dynamic meteorological conditions. The MLR model exhibits the weakest performance, as evidenced by its higher RMSE (3.54 μg/m³) and MSE (12.54 μg²/m⁶), and lower $R^2$ (0.14), emphasising its limited capability to capture the complexities of $PM_{2.5}$ dynamics. The box plots in Figure 7.12b further highlight the stability of the models. RF displays the narrowest range of RMSE values (2.41 μg/m³ across folds), signifying its robustness and reliability. PRR shows moderate variability with an RMSE of 3.21 μg/m³, indicating partial success in addressing non-linearities but with limitations in handling more complex interactions. MLR, on the other hand, exhibits the highest variability and RMSE (3.54 μg/m³), reflecting its susceptibility to fluctuations in the data and reduced predictive reliability. The findings from Figures 7.9, 7.10, 7.11 and 7.12 collectively show the importance of employing comparative and advanced ML techniques for air quality monitoring. RF's ability to effectively model non-linear relationships, variable interactions, and temporal dynamics makes it a valuable tool for predicting $PM_{2.5}$ concentrations. This enhanced predictive capability provides critical insights into atmospheric processes, enabling the development of more accurate and targeted air quality management strategies. Policymakers and environmental managers can leverage these insights to address air pollution more effectively, particularly in regions with complex and diverse emission sources.



© University of Venda

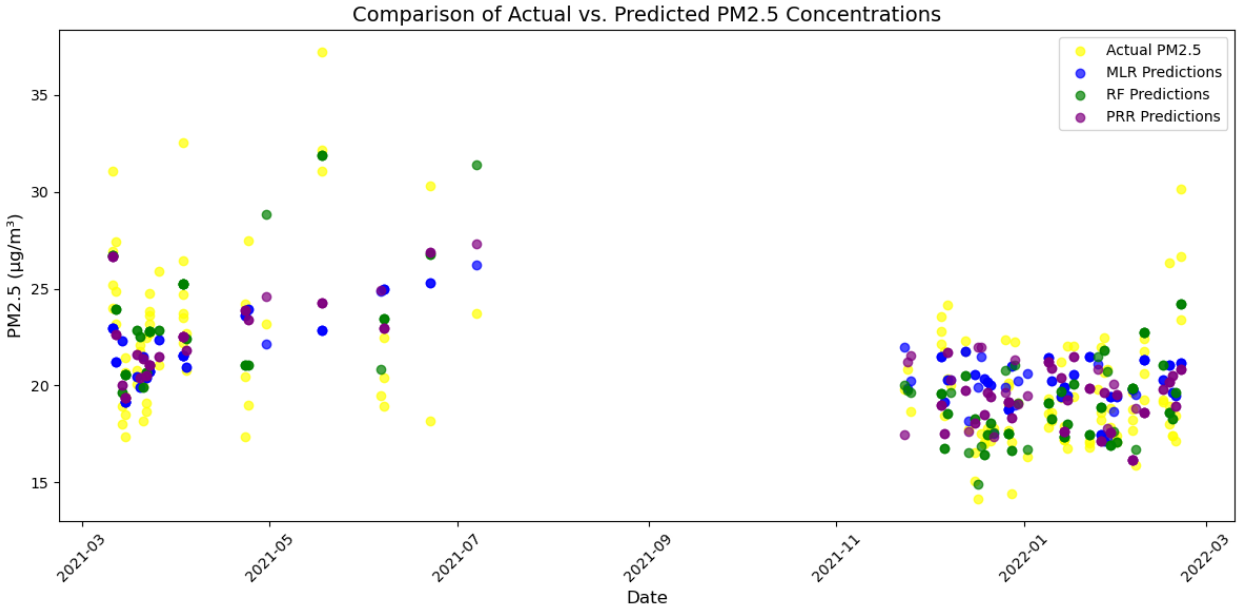



**Figure 7.10:** Scatter plot illustrating the comparison between actual $PM_{2.5}$ concentrations (yellow dots) and predictions made by the three machine learning models: MLR (blue dots), RF (green dots), and PRR (purple dots) over time.

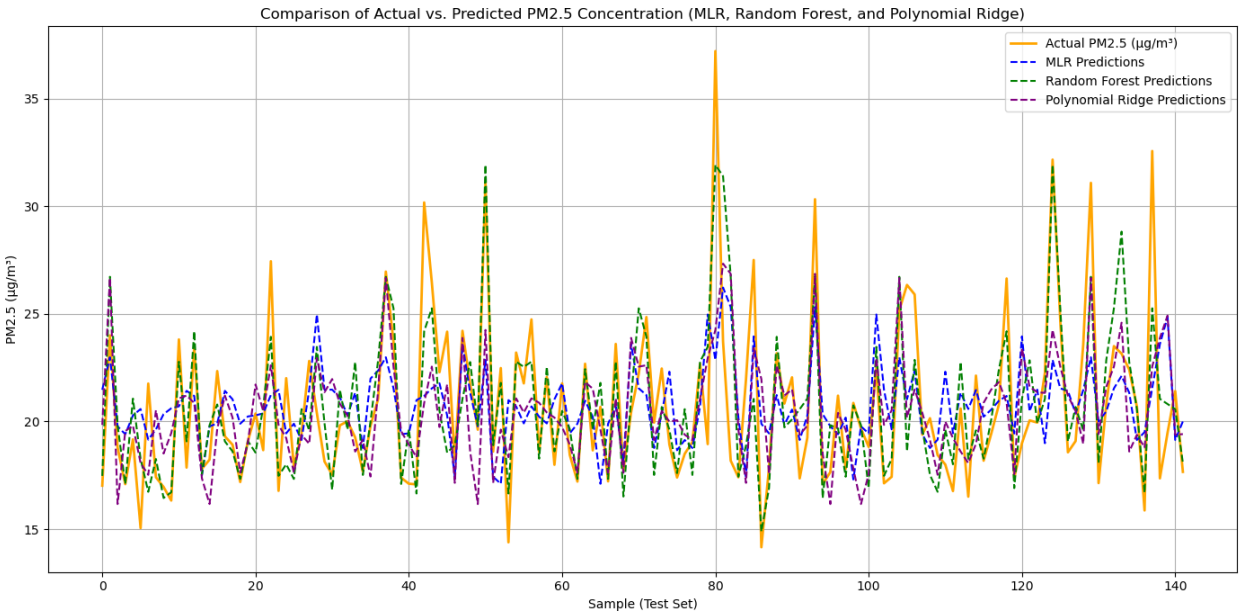



**Figure 7.11:** Actual $PM_{2.5}$ values against the predicted values for the machine learning models. The yellow line represents the actual values, blue dashed line represents the MLR predictions, green dashed line represents the random forest prediction while purple colour represents the polynomial ridge regression model.



© University of Venda

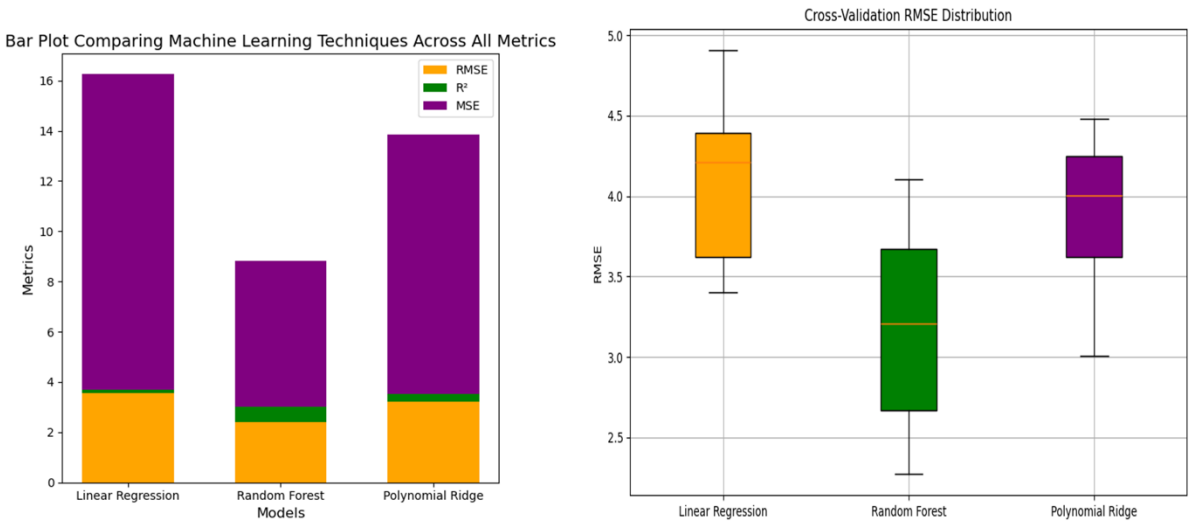



**Figure 7.12:** Bar plot comparing the machine learning techniques across all metrics (a), box plot showing the distribution of the model metrics, and (b) visualization of cross-validation results for better insights into the stability and consistency of the models across key metrics.

The results of the present study are juxtaposed with findings from prior research conducted in diverse geographical locations and under varying meteorological conditions (Table 7.4). This comparative analysis shows the influence of geographical diversity, dataset characteristics, and input parameters on model performance. The RF model in the current study demonstrates its strength by achieving an $R^2$ of 0.604 and RMSE of 2.410 µg/m³, outperforming MLR and PRR in predictive accuracy. These findings align with Kaminska *et al*. (2018) and Panaite *et al*. (2024), who reported RF's robustness in handling seasonal and pollutant-specific variations, however, the lower $R^2$ in this study compared to Panaite *et al*. (2024) may be attributed to the exclusion of parameters such as $NO_2$ and CO. Safo-Adu *et al*. (2023) also revealed the influence of regional meteorological variability, a factor similarly reflected in the current study's findings. The differences in $PM_{2.5}$ dynamics, particularly during biomass burning or vehicular emissions, were region-specific.

Moreover, Lolli *et al*. (2020) demonstrated that simpler models, such as MLR, cannot model complex atmospheric conditions effectively, a limitation corroborated by the lower performance metrics observed in this study's MLR model. The comparative analysis



© University of Venda

highlights the variability introduced by geographical context and input parameters. The higher $R^2$ reported by Panaite *et al*. (2024) could be linked to their inclusion of pollutant-specific parameters. Similarly, Zhang *et al*. (2022) found repeated fold cross-validations to outperform traditional RF model due to its ability to handle intricate data relationships. This analysis underscores the need for customizing predictive models to the unique characteristics of the study area. Integrating pollutant-specific parameters and adopting advanced ML techniques, as demonstrated in studies like Panaite *et al*. (2024), can further enhance the predictive accuracy of air quality models.

*C. Residual analysis of models' performance*

This study subsequently conducted a residual analysis to evaluate the predictive performance and fit of the models (Figure 7.13). Residuals, defined as the difference between observed and predicted $PM_{2.5}$ concentrations, provide valuable insights into model accuracy and the underlying patterns not captured by the model (Hastie *et al*, 2009; Lolli *et al*., 2020; Miller *et al*., 2023). The residuals for the MLR model, plotted against the predicted values, reveal some important characteristics of the model's performance. In an ideal scenario, residuals should be randomly distributed around zero without discernible patterns, suggesting a model that adequately captures the relationship between predictors and the target variable (Meyer *et al*., 2021).

The residuals for the MLR model, however exhibited a "fan-shaped" pattern, with increasing variance at higher predicted $PM_{2.5}$ concentrations. This observation indicates that the MLR model struggles to capture the relationship between meteorological parameters and $PM_{2.5}$ accurately, particularly for non-linear interactions and complex interactions, resulting in larger prediction errors at higher $PM_{2.5}$ levels. Furthermore, several large residuals, both positive and negative, demonstrate the MLR model's limitations in accurately predicting certain data points. This evidence further supports the notion that the linear model is insufficient in capturing the complexities of the $PM_{2.5}$ data, especially in the presence of outliers and non-linearities.



© University of Venda

**Table 7.4:** Comparative analysis of model metrics in current and previous studies

| Study Reference | Region/Location | Model Used | $R^2$ | RMSE (µg/m³) | Key Findings |
|---|---|---|---|---|---|
| Current Study | Thulamela, South Africa | RF | 0.604 | 2.410 | Highlighted RF's strength in capturing non-linear interactions, achieving better predictive accuracy compared to MLR and PRR. |
| | | PRR | 0.297, | 3.215 | |
| | | MLR | 0.144 | 3.544 | |
| Zamani et al. (2019) | Tehran, Iran | ~~RF~~ | 0.78 | 14.54. | Predict $PM_{2.5}$ concentration in Tehran's urban area. |
| Zhang et al. (2022) | Chengdu, China | RF | 0.877 | 9.212 | Model could improve the accuracy of health effect estimates in city scale epidemiological studies, considering dense populations based on accurate estimation of $PM_{2.5}$ concentration |
| Hasnain et al. (2024) | Xuancheng | RF | 0.74 | 10.60 | Effectiveness of RF among other time series models in forecasted ambient $PM_{2.5}$ concentrations along with meteorological data. |
| Xiang et al. (2024) | Chinese cities | RF | 0.6350 | 0.1085 | |
| Panaite et al. (2024) | University of Petrosani | RF | 0.828. | N/A | Air-Quality Predictions on University Campuses using machine learning. Pollutant-specific parameters like $NO_2$ and CO to significantly impact $PM_{2.5}$ concentrations. |
| William et al. (2023) | . Beijing | RF | N/A | 0.0725 | Forecasting pollution in data sets of different size and location with a variety of features |
| Mustapha et al. (2024) | Abeokuta, Nigeria | $MLR_{(DS)}$ | 0.157 | 111.414 | Use for seasonal prediction of $PM_{2.5}$ and Influence of Meteorological Parameters on Air Pollution Propagation in an Active Dumpsite |
| | | $MLR_{(WS)}$ | 0.319 | 46.745 | |
| Kayes et al. (2019) | Dhaka | MLR | 0.7027 | N/A | Forecast accuracy of air pollution under variable meteorological parameters considering seasonal fluctuations. |
| Kaminska (2018) | Wrocław, Poland | RF | S: 0.268<br>W: 0.521 | N/A | Showed RF's effectiveness in addressing seasonal variations in $PM_{2.5}$ concentration. |



© University of Venda

University of Venda
Creating Future Leaders

| Study Reference | Region/Location | Model Used | $R^2$ | RMSE ($\mu g/m^3$) | Key Findings |
|---|---|---|---|---|---|
| | | | CS: 0.306 | | |
| Suthar et al. (2024) | Bengaluru, India | RF | ST: 0.88<br>STT: 0.85<br>WT 0.89<br>WTT: 0.87 | S: 4.13<br>W: 3.87 | Predicting land surface temperature and examining its relationship with air pollution and urban parameters |
| Miller et al. (2023) | Poland | RF | 0.88 | 0.30 | Emphasizing ML capability to capture non-linear relationships and intricate patterns in the data, which traditional models might overlook. |
| | | MLR | 0.78 | 0.45 | |

D = Dry Season; S = Summer; W = Winter; CS = Combined season; ST = Summer; Training; STT = Summer Testing; W= Wet Season; WT = Winter; Training; WTT = Winter Testing; N/A = Not Applicable



© University of Venda

In contrast, the RF model's residuals were randomly distributed around zero, with smaller magnitudes and no discernible patterns. This randomness indicates a better model fit and confirms that RF effectively captures the non-linear relationships present in the dataset. Unlike the MLR model, its residuals are more consistently distributed across the range of predicted $PM_{2.5}$ values. Moreover, the magnitude of the residuals is generally smaller compared to the MLR model. This indicates that the RF predictions are closer to the observed $PM_{2.5}$ concentrations, leading to smaller errors overall and an improved performance. The residual statistics provide further evidence of the superior performance of the RF model in predicting higher $PM_{2.5}$ concentrations (Zamani *et al.*, 2019). The mean of the residuals for both models is close to zero, which is desirable and indicates that the models, on average, do not systematically over- or under-predict $PM_{2.5}$ concentrations; however, the standard deviation of residuals was significantly lower for RF (±2.419) compared to PRR (±3.206) and MLR (±3.557), reinforcing RF's consistency and precision in predictions. This signifies that RF Model's predictions were more consistent and closer to the observed values, as shown in most of the sites. This reinforces the conclusion that the RF model provides a more accurate and robust prediction of $PM_{2.5}$ concentrations compared to the MLR model (Panaite *et al.*, 2024).

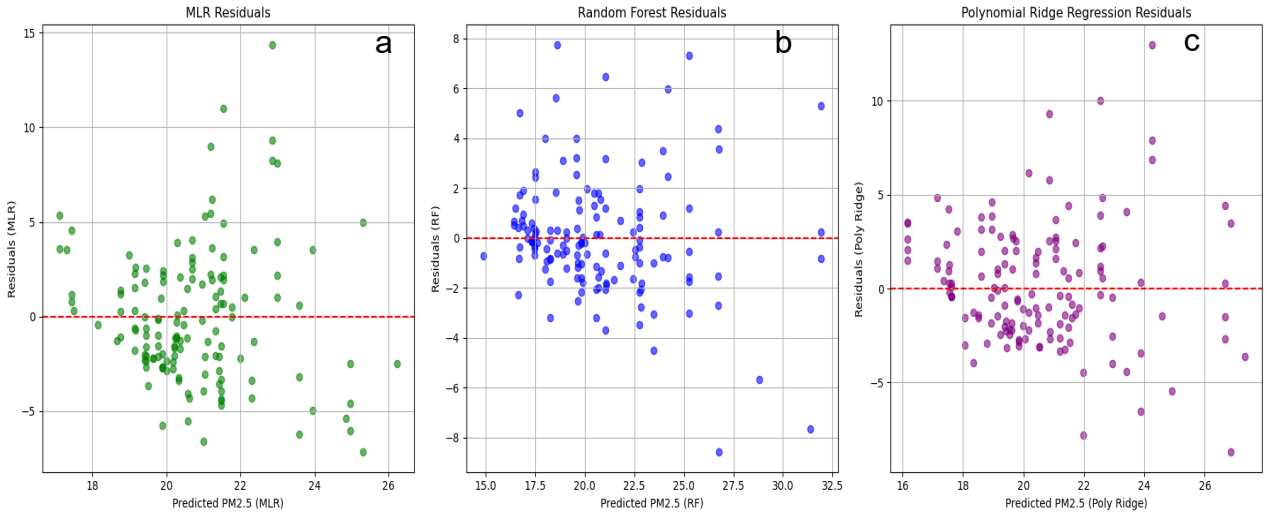



**Figure 7.13:** Residual analysis of a) MLR, b) random forest and c) purple colour represents the polynomial ridge regression.



© University of Venda

### 7.3.6. Modelling and Prediction of Seasonal Variations in $PM_{2.5}$ Pollution using Polynomial Ridge Regression and Random Forest Models

In the seasonal analysis, polynomial Ridge regression and random forest models were selected to predict $PM_{2.5}$ concentration during the winter and summer seasons. These methods were chosen over standard multiple linear regression due to their ability to capture non-linear relationships and interactions among seasonal environmental changes.

#### 7.3.6.1 Prediction of seasonal variations using polynomial ridge regression (PRR)

*A. Winter prediction and PRR performance*

In the winter model, the polynomial ridge regression approach included both first order (linear) and second order (quadratic) terms to capture direct and interaction effects among selected predictors. The resulting winter model equation (Equation 7.11) utilised to predict pollutant concentrations for the sampling region is as follows:

$$PM2.5_{\_winter} = -197.3507 - 0.0070ws - 0.0217wd + 0.0112H - 0.0121T - 0.0182R + 0.0185P + 0.6738ws^2 + 0.0039ws*wd + 0.1615\ ws*H + 0.2810\ ws*T - 0.3613ws*R - 0.0259ws*P + 0.0002\ wd^2 - 0.0031wd*H - 0.0028wd*T - 0.0066\ wd*R + 0.0003\ wd*P + 0.0153\ H^2 + 0.0859\ H*T + -0.1543\ H*R - 0.0045\ H*P + 0.0153T^2 - 0.7878T*R - 0.0070T*P + 0.1503R^2 + 0.0342R*P + 0.0005P^2 \quad (7.11)$$

Larger coefficients, such as for $Ws^2$ (0.6738) and T·R (-0.7878), illustrate significant interactions that influence $PM_{2.5}$ levels during winter; for example, the quadratic term for wind speed ($ws^2$) indicates that higher wind speeds significantly contribute to the dispersion of pollutants, whereas the interaction between temperature and rainfall (T·R) reflects how combined effects of these factors can drive $PM_{2.5}$ variability under certain atmospheric conditions, although, the model's performance metrics reveal its limitations in fully explaining $PM_{2.5}$ variations during winter.

The model demonstrates an overall goodness of fit with an $R^2$ value of 0.31 and an adjusted $R^2$ of 0.29, suggesting the model accounts for only 31% of the variance in $PM_{2.5}$ concentrations. Additionally, the MSE (5.89) and RMSE (2.43 μg/m³) indicate that predictions deviate by approximately 2.43 units from measured $PM_{2.5}$ values on average. The winter scatter plot (Figure 7.14a) of predicted, versus, actual $PM_{2.5}$ concentrations



© University of Venda

reveals considerable scattered data points around the perfect fit line, emphasizing the model's limited accuracy in winter predictions. The residual spread suggests that while the model captures some seasonal dynamics, challenges remain in modelling $PM_{2.5}$ levels during winter. These arise from non-linearities and site-specific factors that are not fully captured by the PRR model (Wa *et al*., 2023; Alkhodaidi *et al*., 2024).

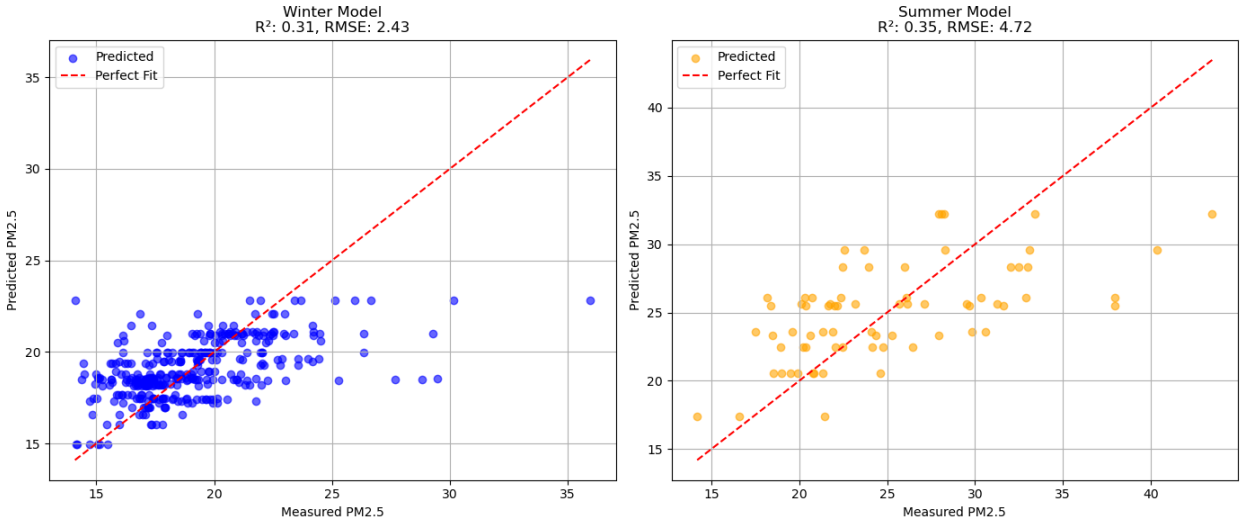



**Figure 7. 14:** Relationship between predicted and measured average $PM_{2.5}$ concentration during the (a) winter, and (b) summer using polynomial ridge regression model

*B. Summer prediction and PRR performance*

The summer model includes similar terms to the winter model, however, most coefficients near zero indicate weak or inconsistent relationships between $PM_{2.5}$ and the selected environmental features. Some interaction terms, such as *Ws·Wd, wd·T,* and *Wd·P,* have non-zero values, suggesting moderate influences, yet, with these interactions, the model's $R^2$ of 0.35 and adjusted $R^2$ of 0.29, demonstrate modest explanatory power, accounting for 35% of the variability in $PM_{2.5}$ levels. The MSE of 22.23 and RMSE of 4.72 μg/m³ reveal higher prediction errors in summer, likely due to increased $PM_{2.5}$ variability or unmodeled seasonal factors. The scatter plot for summer (Figure 7.14b) reveals wider dispersion from the perfect fit line, indicating lower prediction accuracy. The summer model equation (7.12) was utilised to predict pollutant concentrations:



© University of Venda

$$\mathbf{PM_{2.5summer}}=294.9898 - 0.0000\cdot ws - 0.0001\cdot wd + 0.0000\cdot H + 0.0000\cdot T + 0.0000\cdot R - 0.0000\cdot P + 0.0001\cdot ws2 - 0.0226\cdot(ws\cdot wd) - 0.0006\cdot(ws\cdot H) + 0.0006\cdot(ws\cdot T) + 0.0003\cdot(ws\cdot R) - 0.0032\cdot(Ws\cdot P) + 0.0006\cdot Wd^2 - 0.0071\cdot(Wd\cdot H) - 0.0727\cdot(Wd\cdot T) - 0.0068\cdot(Wd\cdot R) + 0.0015\cdot(Wd\cdot P) + 0.0153\cdot H^2 - 0.0122\cdot(H\cdot T) - 0.0056\cdot(H\cdot R) - 0.0010\cdot(H\cdot P) + 0.0063\cdot T^2 + 0.0013\cdot(T\cdot R) + 0.0164\cdot(T\cdot P) - 0.0000\cdot R^2 - 0.0007\cdot(R\cdot P) - 0.0005\cdot P^2 \quad (7.12)$$

*C. Importance of interaction terms and seasonal adaptability using PRR model*

The inclusion of interaction terms, such as those between wind speed and humidity or temperature and rainfall, proves crucial for environmental data analysis, as these interactions influence air quality in ways a linear model cannot capture. Accounting for non-linear relationships is essential in $PM_{2.5}$ prediction, and the polynomial model achieves this by incorporating these complex interactions. For instance, humidity might impact $PM_{2.5}$ differently depending on temperature, and the relationship captured by the polynomial model could be overlooked by a simple linear model. By integrating these interaction terms, the model adapts to seasonal variability, such as higher humidity in winter, for example, from snow, or increased temperatures in summer.

The polynomial ridge regression model captures some seasonality in $PM_{2.5}$ levels, although, the result indicates that certain features play stronger roles in either winter or summer; however, the seasonal models exhibit limited explanatory power and predictive accuracy. These findings emphasise the inherent complexity of $PM_{2.5}$ dynamics and suggest that advanced modelling techniques or additional variables, such as traffic density, industrial emissions, or land-use patterns may be required to enhance predictive performance (Kamińska, 2018; Alkhodaidi *et al*., 2024). Furthermore, the seasonal differences in feature importance demonstrate the need for adaptive air quality management strategies. Addressing these seasonal variations through targeted interventions can lead to more effective mitigation measures and enhanced long-term air quality management.

### 7.3.6.2 Prediction of seasonal variation using machine learning algorithms: Random Forest (RF) model

In response to the limitations of the polynomial Ridge model and based on the previous RF model's performance for site-specific and overall study predictions, RF was applied to



© University of Venda

analyze seasonal shifts in $PM_{2.5}$ concentration. This approach aims to mitigate overfitting caused by the increased complexity of higher-order terms while introducing a penalty for overly complex models to ensure generalizability and robustness, particularly, for environmental datasets that often exhibit high variance and outliers. Additionally, the ridge regularization inherent in the RF method addresses multicollinearity issues, where high correlations between features can distort coefficient estimates, resulting in more stable and interpretable outcomes.

As shown in Figure 7.15a, the RF model notably outperformed the PRR model for winter $PM_{2.5}$ predictions, achieving an $R^2$ of 0.58, indicating that 58% of $PM_{2.5}$ variance is explained by the selected environmental variables. This aligns with the findings of Kamińska (2018), Zhang *et al*. (2022), and Suthar *et al*. (2024). This improvement demonstrates the model's ability to account for the non-linear interactions that dominate in winter dynamics (Figure 7.15a), especially under lower pollutant dispersion conditions. This observation is consistent with Zhang *et al*. (2022), who reported higher $PM_{2.5}$ level in winter as compared to lower concentrations and minimum variation in summer. The MSE of 4.42 and RMSE of 2.10 µg/m³ further highlight relatively accurate winter predictions, with deviations of approximately 2.10 units from the observed $PM_{2.5}$ values. The 5-fold cross-validation results (mean RMSE of 2.38 µg/m³ ± 0.30) further confirmed the model's stability, despite a moderate cross-validated $R^2$ score of 0.29 ± 0.21, suggesting that additional factors might influence winter $PM_{2.5}$ levels. In contrast, the RF model for summer achieved an $R^2$ of 0.45, MSE of 22.27, and RMSE of 4.72 µg/m³, reflecting weaker predictive power compared to winter.

The reduced explanatory power during summer, where only 45% of $PM_{2.5}$ variability was explained, may be attributed to the complex interactions and unmodeled external factors, such as heightened human activities or varying meteorological influences. Cross-validation results, with a mean $R^2$ score of -0.37 (±0.83) and a mean RMSE of 5.51 µg/m³ (±0.69), further indicate the model's challenges in capturing summer $PM_{2.5}$ variability (Figure 7.15b). These findings emphasise the need to incorporate additional seasonal or external variables to enhance predictive performance.



© University of Venda

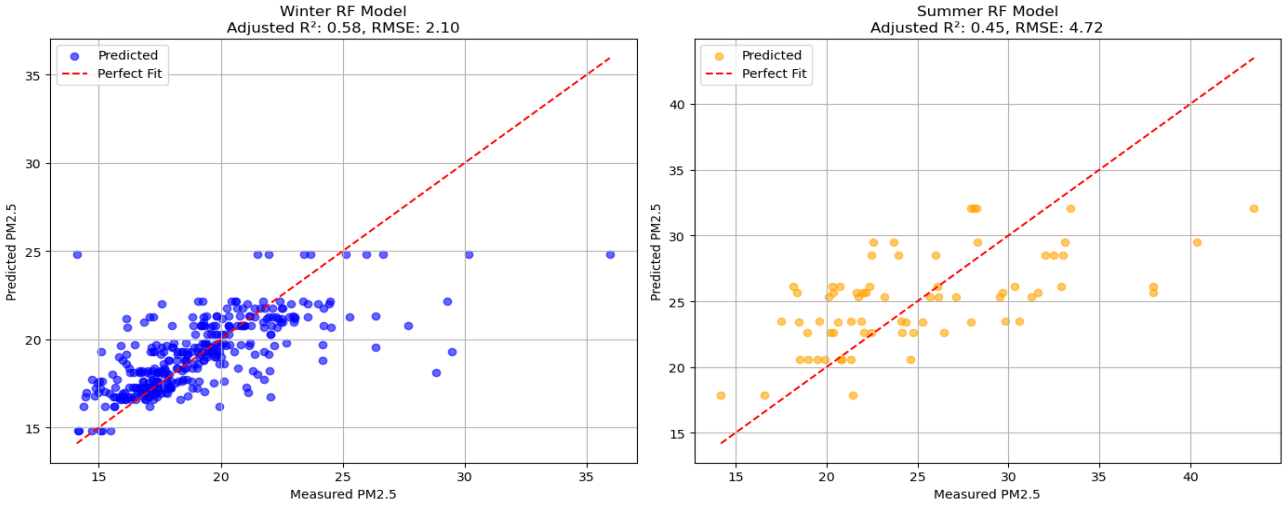



**Figure 7.15:** Relationship between predicted and measured average $PM_{2.5}$ concentration during the winter and summer seasons using RF Model, a) winter and b) summer prediction.

*A. Feature importance and residual analysis of RF model*

The RF model's feature importance analysis indicates distinct seasonal dependencies in $PM_{2.5}$ determinants (Figure 7.16). In winter, temperature (importance score: 0.2914) plays a dominant role, likely due to temperature inversions trapping pollutants (Zhang *et al*., 2022). Wind speed (0.229) and wind direction (0.176) also significantly influence $PM_{2.5}$ levels, as lower wind speeds and specific directions limit pollutant dispersion, resulting in higher $PM_{2.5}$ concentrations. In summer, wind speed emerges as the most influential feature (importance score: 0.3277), reflecting its critical role in dispersing pollutants under warmer conditions, followed by Temp (0.2007).

Table 7.5 further shows the shift in feature importance, underscores RF's adaptability in capturing seasonal variations in $PM_{2.5}$ levels. Residual analysis further illustrates the RF model's performance (Figure 7.17). In winter, residuals are tightly clustered around zero, indicating relatively unbiased predictions and strong model alignment with observed data. In contrast, the summer residual plot shows wider dispersal, indicating greater prediction variability and lower reliability, most likely, due to the complex interactions of environmental factors specific to summer. The RF model's predicted values align closely with observed data in winter, evidenced by smaller residuals and a tight distribution around the perfect fit line. Conversely, summer predictions show significant deviation from observed values, aligning with the broader residual distribution and the lower $R^2$ scores; while the RF model demonstrates moderate predictive ability, particularly in winter, its



© University of Venda

summer limitations suggest the influence of unmodeled environmental factors on $PM_{2.5}$ levels, which aligns with Suthar *et al*. (2024).

The seasonal dependency in feature importance and residual plots indicates that $PM_{2.5}$ determinants vary throughout the year, with temperature being more impactful in winter and wind speed more influential in summer. While the RF model improves performance over the polynomial Ridge model, its summer limitations suggest a need for additional features or more advanced ensemble approaches, such as boosting techniques or hybrid models. These adjustments could enhance the model's predictive accuracy across seasons, ultimately supporting more targeted air quality management strategies.

**Table 7.5:** Seasonal random forest feature importance

| Season | Ws | Temp | Humidity | Wd | Pressure | Rain |
|---|---|---|---|---|---|---|
| Summer | 0.3277 | 0.2007 | 0.1544 | 0.1544 | 0.1403 | 0.0225 |
| Winter | 0.2289 | 0.2913 | 0.1720 | 0.1755 | 0.0656 | 0.0665 |

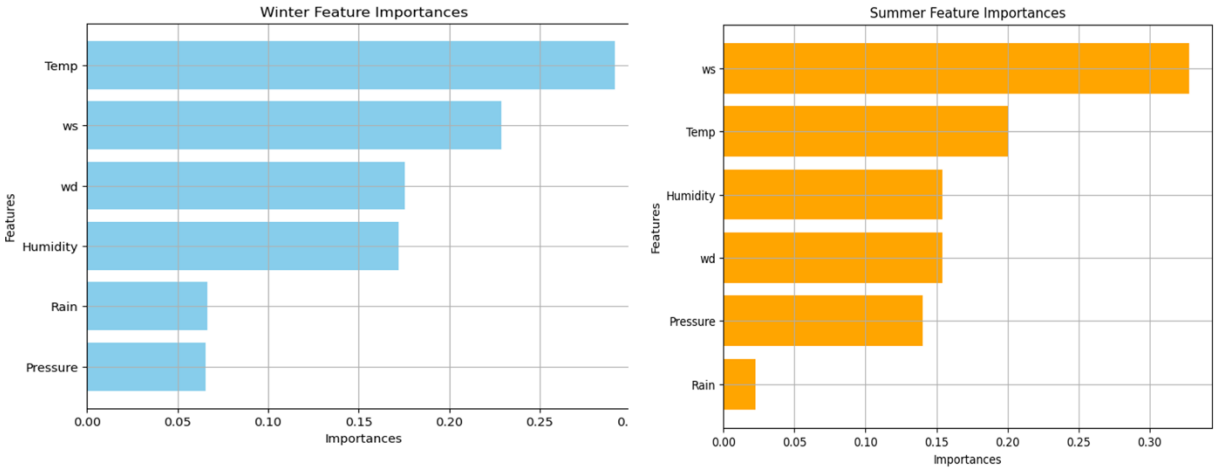



**Figure 7.16:** Bar plot showing the winter and summer model comparison feature Importance for RF using average $PM_{2.5}$ concentration.



© University of Venda

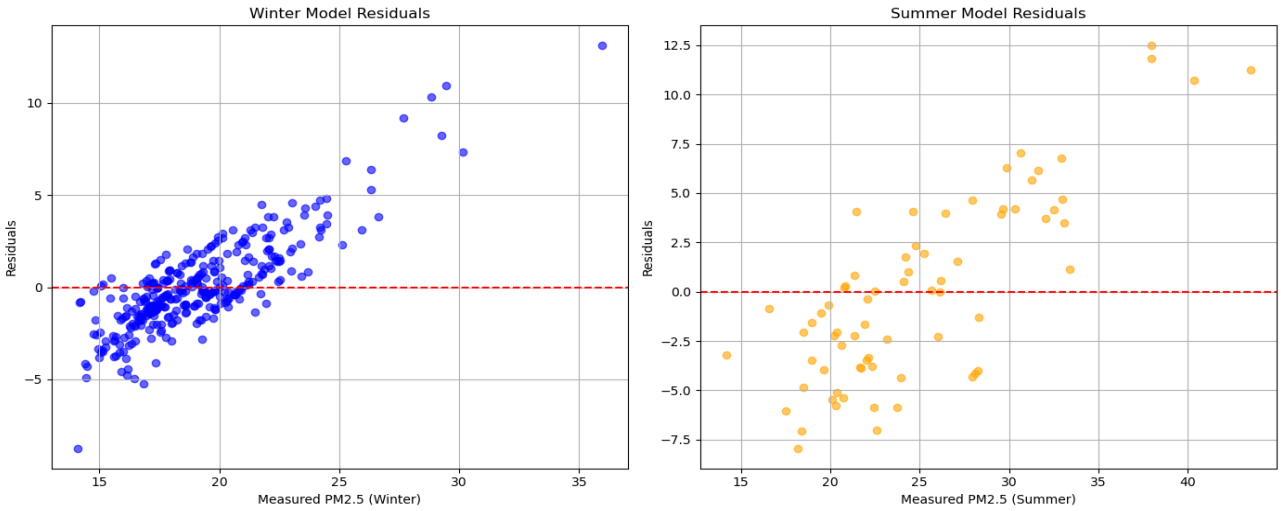



**Figure 7.17:** Residual analysis for random forest model using average $PM_{2.5}$ concentration.

## 7.4 Implications for Air Quality Management

The RF model's ability to integrate meteorological variables and accurately predict $PM_{2.5}$ concentrations highlights its potential as a valuable decision-support tool for air quality management. By identifying critical factors such as wind direction, wind speed, and atmospheric pressure, the RF model provides actionable insights to guide pollution mitigation strategies, including traffic flow optimisation, emission controls, and localised interventions tailored to prevailing meteorological conditions. The model's superior performance across site-specific, overall study area, and seasonal variations demonstrates its adaptability to diverse environmental settings. Its robustness in capturing non-linear relationships makes it particularly effective where traditional models fall short. To further improve its accuracy, incorporating additional variables such as industrial emissions, land-use patterns, and topography is recommended. These enhancements would allow for a more comprehensive understanding of $PM_{2.5}$ dynamics, supporting better-informed decisions for urban planning and environmental monitoring.

The improved precision of RF-based predictions enables reliable air quality prediction in order to facilitate targeted public health interventions that will protect vulnerable populations. Policymakers and environmental agencies can use these insights to prioritize pollution hotspots and implement effective strategies for reducing air pollution. Overall, the RF model not only advances $PM_{2.5}$ prediction but also offers broader applicability for other environmental and health-related challenges, supporting evidence-



© University of Venda

based decision-making to safeguard human health and promote environmental sustainability.

### 7.5 Policy Recommendations for Air Quality Improvement: Lessons from APTI and API Studies

Building on the above findings, which explored the application of machine learning regression models for predicting ambient particulate matter ($PM_{2.5}$) in Thulamela Local Municipality, this subheading shifts focus on actionable policy recommendations for improving air quality as reviewed in Enitan *et al*. (2022). By leveraging insights from scientific studies on the Air Pollution Tolerance Index (APTI) and the Anticipated Performance Index (API), this review explores eco-friendly and sustainable approaches to air quality management (Appendix 7.8). It highlights the role of phytoremediation and the selection of indigenous plant species as effective tools for mitigating air pollution (Che *et al*., 2024). By integrating these indices into urban planning and environmental management strategies, this chapter aims to bridge the gap between scientific research and practical implementation. The insights presented in Enitan *et al*. (2022) are intended to guide policymakers, environmental agencies, stakeholders, and the public with science-driven strategies for reducing pollution levels, improving public health, and promoting sustainable practices that will foster a coordinated approach to air quality management and pollution control.

### 7.6 Limitations of the Study

This study, although it demonstrated the effectiveness of RF, PRR, and MLR in predicting $PM_{2.5}$ concentrations, several limitations should be acknowledged to contextualise the findings. The models relied exclusively on six meteorological variables, excluding other significant environmental and anthropogenic factors such as industrial emissions, traffic density, biomass burning, and land-use patterns. Including these variables in future research could provide a more holistic understanding of $PM_{2.5}$ dynamics and enhance model performance. Seasonal data gaps, particularly during summer, posed challenges in fully capturing the complex interactions between meteorological factors and pollutant behaviour. This limitation reduced the models' predictive accuracy under highly variable conditions, emphasizing the importance of expanding datasets to ensure consistent seasonal representation. Additionally, site-specific variations in predictive performance



© University of Venda

reflected the heterogeneous nature of $PM_{2.5}$ sources and meteorological influences, which were not fully addressed by the models. RF consistently outperformed MLR and PRR, however, its computational intensity remains a limitation for real-time forecasting in resource-constrained regions with limited infrastructure. Furthermore, the assumption of stationarity that the relationship between meteorological factors and $PM_{2.5}$ concentrations will remain constant over time may not hold under dynamic conditions influenced by new emission sources, policy interventions, or climate variability. To address these limitations, future research should integrate additional predictor variables, ensure comprehensive seasonal and spatial data coverage, and explore advanced ensemble learning methods, hybrid modelling approaches, and real-time data integration. These efforts will enhance the reliability and applicability of predictive models, ultimately improving air quality management strategies.

## 7.7 Conclusion

This thesis chapter employed statistical and machine learning techniques to study the effect of meteorological parameters on $PM_{2.5}$ concentration in Thulamela Municipality, Vhembe District, South Africa. Three machine learning techniques—Multiple Linear Regression (MLR), Polynomial Ridge Regression (PRR), and Random Forest (RF) were employed to estimate $PM_{2.5}$ concentration for three scenarios: site-specific, aggregated, and seasonal patterns. The results demonstrated the prescriptive power of RF in effectively modelling complicated, non-linear meteorological relationships among variables. Results indicated that RF was better than both PRR and MLR and also achieved the highest accuracy, particularly in winter as meteorological stability boosted prediction stability. Summer seasonality caused instability and lowered accuracy through greater meteorological variability. Cross-validation validated RF's performance in semi-urban areas with scarce direct measurements of $PM_{2.5}$ and consequently the possibility of its application to air quality monitoring. Second, the article suggested the integration of the Air Pollution Tolerance Index (APTI) and Anticipated Performance Index (API) into the model to respond to machine learning-based predictions with an environmentally friendly, phytoremediation-based intervention. By the integration of data-driven predictive modelling and ecological interventions, the study provides an integrative framework for policymakers and environmental regulators to design situational interventions that meet



© University of Venda

the requirements of technical upgradation and green city planning. The study is a significant step towards data-driven air quality management, with actionable suggestions and hands-on tools to address the complex problems of air pollution in South Africa and other comparable semi-urban regions globally. By embracing these scientifically informed policies, the Thulamela Local Municipality could be an example of sustainable development, reducing pollution, improving public health, and supporting global environmental goals.

## References

Adesina, J.A., Piketh, S.J., Burger, R.P. & Mkhatshwa, G. (2022). Assessment of criteria pollutants contributions from coal-fired plants and domestic solid fuel combustion at the South African industrial highveld. *Cleaner Engineering and Technology*, *6*, p.100358.

Agbehadji, I. E., & Obagbuwa, I. C. (2024). Systematic Review of Machine Learning and Deep Learning Techniques for Spatiotemporal Air Quality Prediction. *Atmosphere*, *15*(11), 1352.

Aladag, E. (2023). The influence of meteorological factors on air quality in the province of van, Turkey. water air soil Pollut 234, 259. https://doi.org/10.1007/s11270-023-06265-0.

Alkhodaidi, A., Attiah, A., Mhawish, A. & Hakeem, A. (2024). The role of machine learning in enhancing particulate matter estimation: A Systematic Literature Review. *Technologies*, *12*(10), p.198.

Ameer, S., Shah, M.A., Khan, A., Song, H., Maple, C., Islam, S.U. & Asghar, M.N. (2019). Comparative analysis of machine learning techniques for predicting air quality in smart cities. *IEEE access*, *7*, pp.128325-128338.

Amnuaylojaroen, T., Kaewkanchanawong, P., and Panpeng, P. (2023). Distribution and meteorological control of $PM_{2.5}$ and its effect on visibility in Northern Thailand. *Atmosphere*, *14*(3), 538.

Antai, R.E., Osuji, L.C., Obafemi, A.A & Onojake, M.C. (2018). Prediction and modeling of dry seasons air pollution changes using multiple linear models: A case study of Port Harcourt and its environs, Niger Delta Nigeria. *International Journal of Environment, Agriculture and Biotechnology* (IJEAB). 3(3). 899 – 915.

Aruna, S.K and Ravindiran, G. (2024). Integrating machine learning techniques for Air Quality Index forecasting and insights from pollutant-meteorological dynamics in sustainable urban environments. *Earth Science Informatics*, *17*(4), 3733-3748.

Banerjee, B., Kundu, S., Kanchan, R. *et al*. (2024). Examining the relationship between atmospheric pollutants and meteorological factors in Asansol city, West Bengal, India, using statistical modelling. *Environmental Science and Pollution Research*, https://doi.org/10.1007/s11356-024-33608-z.



© University of Venda

Bera, B., Bhattacharjee, S., Sengupta, N. & Saha, S. (2021). $PM_{2.5}$ concentration prediction during COVID-19 lockdown over Kolkata metropolitan city, India using MLR and ANN models. Environmental Challenges, 4, 100155.

Bose, A., Roy Chowdhury, I. (2023). Investigating the association between air pollutants' concentration and meteorological parameters in a rapidly growing urban center of West Bengal, India: a statistical modeling-based approach. Model. Earth Syst. Environ. 9, 2877–2892. https://doi.org/10.1007/s40808-022-01670-6.

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

Chang, Y.S., Abimannan, S., Chiao, H.T., Lin, C.Y. & Huang, Y.P., (2020). An ensemble learning based hybrid model and framework for air pollution forecasting. Environmental Science and Pollution Research, 27, pp.38155-38168.

Chen, J., Shen, H., Li, T., Peng, X., Cheng, H., & Ma, C. (2019). Temporal and spatial features of the correlation between $PM_{2.5}$ and $O_3$ concentrations in China. Int J Environ Res Public Health 16(23):4824.

Chen, C. C., Wang, Y. R., Yeh, H. Y., Lin, T. H., Huang, C. S., & Wu, C. F. (2021). Estimating monthly $PM_{2.5}$ concentrations from satellite remote sensing data, meteorological variables, and land use data using ensemble statistical modeling and a random forest approach. *Environmental Pollution*, *291*, 118159.

Chen, D., Yan, J., Sun, N., Sun, W., Zhang, W., Long, Y., & Yin, S. (2024). Selective capture of $PM_{2.5}$ by urban trees: The role of leaf wax composition and physiological traits in air quality enhancement. *Journal of Hazardous Materials*, *478*, 135428.

Choi, H., Park, S., Kang, Y., Im, J. & Song, S. (2023). Retrieval of hourly $PM_{2.5}$ using top-of-atmosphere reflectance from geostationary ocean color imagers I and II. Environmental Pollution, 323, p.121169.

Ebrahimi-Khusfi, Z., Taghizadeh-Mehrjardi, R., & Mirakbari, M. (2021). Evaluation of machine learning models for predicting the temporal variations of dust storm index in arid regions of Iran. Atmospheric Pollution Research, 12(1), 134-147.

Edlund, K., Killman, F., Molnár, P., Boman, J., Stockfelt, L., & Wichmann, J. (2021). Health risk assessment of $PM_{2.5}$ and $PM_{2.5}$-bound trace elements in Thohoyandou, South Africa. *International Journal of Environmental Research and Public Health*, 18.

Enitan, I. T., Durowoju, O. S., Edokpayi, J. N., & Odiyo, J. O. (2022). A review of air pollution mitigation approach using air pollution tolerance index (APTI) and anticipated performance index (API). *Atmosphere*, *13*(3), 374.

Essamlali, I., Nhaila, H., & El Khaili, M. (2024). Supervised machine learning approaches for predicting key pollutants and for the sustainable enhancement of urban air quality: A systematic review. *sustainability*, *16*(3), 976.

Fu, X., Wang, X., Hu, Q., Li, G., Ding, X., Zhang, Y., *et al*. (2016). Changes in visibility with $PM_{2.5}$ composition and relative humidity at a background site in the Pearl River Delta region. *Journal of environmental sciences*, *40*, 10-19.



© University of Venda

Ghahremanloo, M., Choi, Y., Sayeed, A., Salman, A. K., Pan, S., & Amani, M. (2021). Estimating daily high-resolution $PM_{2.5}$ concentrations over Texas: Machine Learning approach. *Atmospheric Environment*, *247*, 118209.

Gündoğdu, S., & Elbir, T. (2024). Elevating hourly $PM_{2.5}$ forecasting in Istanbul, Türkiye: Leveraging ERA5 reanalysis and genetic algorithms in a comparative machine learning model analysis. *Chemosphere*, *364*, 143096.

Hasnain, A., Hashmi, M.Z., Khan, S., Bhatti, U.A., Min, X., Yue, Y., He, Y. & Wei, G. (2024). Predicting ambient $PM_{2.5}$ concentrations via time series models in Anhui Province, China. *Environmental Monitoring and Assessment*, *196*(5), p.487.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D., & Liu, Y. (2018). Predicting monthly high-resolution $PM_{2.5}$ concentrations with random forest model in the North China plain. Environmental Pollution, 242, 675–683.

Hossain, M. L., Shams, S. N., & Ullah, S. M. (2024). Comparative study of machine learning algorithms for wind speed prediction in Dhaka, Bangladesh. *Sustainable Energy Research*, *11*(1), 23.

Imam, M., Adam, S., Dev, S. & Nesa, N. (2024). Air quality monitoring using statistical learning models for sustainable environment. Intelligent Systems with Applications, 22, p.200333.

Islam, A.R.M.T., Al Awadh, M., Mallick, J., Pal, S.C., Chakraborty, R., Fattah, M.A., Ghose, B., Kakoli, M.K.A., Islam, M.A., Naqvi, H.R. & Bilal, M., (2023). Estimating ground-level $PM_{2.5}$ using subset regression model and machine learning algorithms in Asian megacity, Dhaka, Bangladesh. Air Quality, Atmosphere & Health, 16(6), 1117-1139.

Jones, M. & Williams, D. (2020). Meteorological dynamics: Linking wind speed and air temperature patterns. *Environmental Science Journal, 45*(4), 567-582.

Kallenborn, Roland, Crispin Halsall, Maud Dellong, & Pernilla Carlsson (2012). The influence of climate change on the global distribution and fate processes of anthropogenic persistent organic pollutants." *Journal of Environmental Monitoring* 14(11) 2854-2869.

Kaminńska, J.A., 2018. The use of random forest in modelling short-term air pollution effects based on traffic and meteorological conditions: A case study in Wrocław. *Journal of Environmental Management*, *217*, 164-174.

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.

Kumar, V., & Sahu, M. (2021). Evaluation of nine machine learning regression algorithms for calibration of low-cost $PM_{2.5}$ sensors. *Journal of Aerosol Science*, *157*, 105809.



© University of Venda

Li, D., Zheng, J., Yang, M., Meng, Y., Yu, X., Zhou, H., Tong, L., Wang, K., Li, Y.F., Wang, X. & Xiao, H. (2023). Atmospheric wet deposition of trace metal elements: Monitoring and modelling. *Science of the Total Environment*, *893*, p.164880.

Liu, B., Yu, W., Wang, Y., Lv, Q., & Li, C. (2021). Research on data correction method of micro air quality detector based on combination of partial least squares and random forest regression. *IEEE Access*, *9*, 99143-99154.

Liu, J., Wang, S., Zhu, K., Hu, J., Li, R., & Song, X. (2024). Spatial patterns of the diurnal variations of $PM_{2.5}$ and their influencing factors across China. *Atmospheric Environment*, *318*, 120215.

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), 14518.

Liu, Z., Wang, H., Shen, X., Peng, Y., Shi, Y., Che, H., & Wang, G. (2019). Contribution of meteorological conditions to the variation in winter $PM_{2.5}$ concentrations from 2013 to 2019 in Middle Eastern China. *Atmosphere*, *10*(10), 563.

Lolli, S., Chen, Y. C., Wang, S. H., & Vivone, G. (2020). Impact of meteorological conditions and air pollution on COVID-19 pandemic transmission in Italy. *Scientific reports*, *10*(1), 16213.

Lu, D., Mao, W., Zheng, L., Xiao, W., Zhang, L., & Wei, J. (2021). Ambient $PM_{2.5}$ estimates and variations during COVID-19 Pandemic in the Yangtze River delta using machine learning and big data. Remote Sens, 13, 1423. https:// doi. org/ 10. 3390/ rs130 81423.

Lu, H.C. & Fang, G.C. (2002). Estimating the frequency distributions of $PM_{10}$ and $PM_{2.5}$ by the statistics of wind speed at Sha-Lu, Taiwan. Science of the total environment, 298(1-3), 119-130.

Meyer, H. & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, *12*(9), 1620-1633.

Miller, T., Durlik, I., Adrianna, K., Kisiel, A., Cembrowska-Lech, D., Spychalski, I., & Tuński, T. (2023). Predictive modeling of urban lake water quality using Machine Learning: A 20-Year Study. *Applied Sciences*, *13*(20), 11217.

Mustapha, A. T., Sobowale, A., Adeosun, O. J., & Badejo, A. A. (2024). Influence of meteorological parameters on air pollution propagation in an active dumpsite in Abeokuta, Nigeria. *Bima Journal of Science and Technology*, 8(2A), 39-49.

Nejad, M. T., Ghalehteimouri, K. J., Talkhabi, H., & Dolatshahi, Z. (2023). The relationship between atmospheric temperature inversion and urban air pollution characteristics: a case study of Tehran, Iran. *Discover Environment*, *1*(1), 17.

Okimiji, O., Techato, K., Simon, J., Tope-Ajayi, O., Okafor, A., Aborisade, M., & Phoungthong, K. (2021). Spatial pattern of air pollutant concentrations and their relationship with meteorological parameters in coastal slum settlements of Lagos, Southwestern Nigeria. *Atmosphere*. https://doi.org/10.3390/atmos12111426.



© University of Venda

Ostertagová, E. (2012). Modelling using polynomial regression. Procedia engineering, 48, 500-506.

Owusu-Tawiah, V., Westervelt, D. M., & Annor, T. (2022, October). Relationships between meteorological parameters and $PM_{2.5}$ in Accra. In International Conference on air quality in Africa (pp. 81-83). Cham: Springer Nature Switzerland.

Panaite, F. A., Rus, C., Leba, M., Ionica, A. C., & Windisch, M. (2024). Enhancing air-quality predictions on university campuses: A Machine-Learning approach to PM2.5 forecasting at the University of Petroșani. *Sustainability*, *16*(17), 7854.

Safo-Adu, G., Attiogbe, F., Emahi, I., & Ofosu, F. (2023). Outdoor and indoor particle air pollution and its health consequences in African cities: New evidence and an exhortation. *Sustainable Environment*, 9.

Shah, V., Jagupilla, S. C. K., Vaccari, D. A., & Gebler, D. (2021). Non-linear visualization and importance ratio analysis of multivariate polynomial regression ecological models based on river hydromorphology and water quality. Water, 13(19), 2708.

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons.

Sharma, A.K., Sharma, M., Sharma, A.K. & Sharma, M., (2023). Mapping the impact of environmental pollutants on human health and environment: A systematic review and meta-analysis. *Journal of Geochemical Exploration*, p.107325.

Shen, J., Valagolam, D., & McCalla, S. (2020). Prophet forecasting model: A machine learning approach to predict the concentration of air pollutants ($PM_{2.5}$, $PM_{10}$, $O_3$, $NO_2$, $SO_2$, CO) in Seoul. South Korea. PeerJ, 8, e9961. https://doi. org/ 10. 7717/ peerj. 9961

Silva, C., Pérez, P. & Trier, A., (2001). Statistical modelling and prediction of atmospheric pollution by particulate material: two nonparametric approaches. Environmetrics: The official journal of the International Environmetrics Society, 12(2), pp.147-159.

South African Weather Service (SAWS). https://www.weathersa.co.za/.

Suchitra, S. (2024, April). unveiling hidden air quality patterns: A data-driven predictive regression analysis for sustainable air management. In *2024 2nd International Conference on Networking and Communications (ICNWC)* (pp. 1-10). IEEE.

Sulaymon, I. D., Zhang, Y., Hopke, P. K., Ye, F., Gong, K., Mao, J., & Hu, J. (2023). Modeling $PM_{2.5}$ during severe atmospheric pollution episode in Lagos, Nigeria: spatiotemporal variations, source apportionment, and meteorological influences. *Journal of Geophysical Research: Atmospheres*, *128*(13), e2022JD038360.

Sun, X., Zhao, T., Liu, D., Gong, S., Xu, J., & Ma, X. (2020). Quantifying the influences of $PM_{2.5}$ and relative humidity on change of atmospheric visibility over recent winters in an urban area of East China. Atmosphere, 11(5), 461.

Suthar, G., Kaul, N., Khandelwal, S. & Singh, S. (2024). Predicting land surface temperature and examining its relationship with air pollution and urban parameters in Bengaluru: A machine learning approach. *Urban Climate*, *53*, 101830.



© University of Venda

Talepour, N., Birgani, Y. T., Kelly, F. J., Jaafarzadeh, N., & Goudarzi, G. (2024). Analyzing meteorological factors for forecasting $PM_{10}$ and $PM_{2.5}$ levels: a comparison between MLR and MLP models. *Earth Science Informatics*, *17*(6), 5603-5623.

Wa, S., Lu, X., & Wang, M. (2023). Regression model and method settings for air pollution status analysis based on air quality data in Beijing (2017–2021). *International Journal of Data Science and Analytics*, 1-21.

Wang, R., Zhou, J., & Li, F. (2023b). Cross-validation approaches for air quality prediction models: An empirical study. *Journal of Environmental Management*, 325, 116546. DOI: 10.1016/j.jenvman.2023.116546.

Wang, Y., Sun, Y., Zhu, Q., *et al*. (2023a). Rainfall and its scavenging effects on particulate matter in diverse climatic regions. *Science of the Total Environment, 866*, 161217.

William, P., Paithankar, D.N., Yawalkar, P.M., Korde, S.K., Rajendra, A. and Rakshe, D.S., (2023), January. Divination of air quality assessment using ensembling machine learning approach. In *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)* (pp. 1-10). IEEE.

Xiang, K., Shi, D. & Xiang, X. (2024). Machine learning analysis of socioeconomic drivers in urban ozone pollution in Chinese cities. *Environmental Monitoring and Assessment*, *196*(3), p.314.

Xu, Chunyu, Dongqun Xu, Zhe Liu, Yunpu Li, Ning Li, Ryan Chartier, Junrui Chang, Qin Wang, Yaxi Wu, & Na Li. (2020). Estimating hourly average indoor $PM_{2.5}$ using the random forest approach in two megacities, China." *Building and Environment* 180 107025.

Xu, Z., Liu, B., Wang, W., Zhang, Z. & Qiu, W., 2024. Assessing the impact of straw burning on $PM_{2.5}$ using explainable machine learning: A case study in Heilongjiang Province, China. *Sustainability*, *16*(17), p.7315.

Yuan, Z., Pei, C., Li, H., Lin, L., Liu, S., Hou, R., *et al*. (2023). Atmospheric microplastics at a southern China metropolis: occurrence, deposition flux, exposure risk and washout effect of rainfall. *Science of the Total Environment*, *869*, 161839.

Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B., & Talebiesfandarani, S. (2019). $PM_{2.5}$ predictions based on random forest, XGBoost, and deep learning using multisource remote sensing data. *Atmosphere*, *10*(7), 373.

Zhang, Y. (2019). Dynamic effect analysis of meteorological conditions on air pollution: A case study from Beijing. *Science of the Total Environment*, *684*, 178-185.

Zhang, Y., Zhai, S., Huang, J., Li, X., Wang, W., Zhang, T., Yin, F. & Ma, Y. (2022). Estimating high-resolution $PM_{2.5}$ concentrations in the Sichuan Basin using a random forest model with data-driven spatial autocorrelation terms. *Journal of Cleaner Production*, *380*, p.134890.



© University of Venda

# CHAPTER EIGHT

# CONCLUSIONS AND RECOMMENDATIONS

## 8.1 Conclusions

This study provides a comprehensive assessment of ambient air pollution monitoring; its dispersion modelling, and human health risks in Thulamela Municipality, Vhembe District, South Africa; integrating emission inventories of criteria pollutants ($SO_2$, $NO_x$, CO, $PM_{2.5}$ and $PM_{10}$), $PM_{2.5}$ real-time monitoring; its dispersion modelling, machine learning-based predictive modelling, as well as health risk assessments. The findings emphasise the significant impact of air pollution on environmental and human health, highlighting the need for targeted air quality management strategies.

The emission inventory revealed that residential wood combustion (RWC) and vehicular emissions are the dominant sources of $PM_{2.5}$, $PM_{10}$, $SO_2$, $NO_x$, and CO in Thulamela, with RWC contributing over 96% of total $PM_{2.5}$ emissions. The spatial emission patterns revealed that densely populated suburban and rural areas experienced elevated pollution levels, especially during the winter, due to increased wood combustion, seasonal changes, and temperature inversions. This comprehensive inventory fills a critical data gap in understanding pollution contributions, providing a foundation for future air quality control strategies. The study also developed a first-time 10-year vehicle emission inventory for Thulamela, LM.

Dylos DC 1700 Low-cost sensor (LCS) monitoring demonstrated strong agreement ($R^2$ = 0.75) with reference instruments (GRIMM EDM180), confirming the feasibility and good performance of Dylos DC1700 LCS technology for real-time air quality assessment. The results showed significant spatial and temporal variations in $PM_{2.5}$ concentrations, with urban areas experiencing higher pollution levels due to vehicular emissions, road dust, and commercial activities. In contrast, rural areas were more affected by domestic fuel burning. The seasonal analysis confirmed higher $PM_{2.5}$ levels in winter, while diurnal variations showed peak concentrations during the morning (6–8 am) and evening (6–8 pm) rush hours, linked to traffic and fuel combustion.

AERMOD dispersion modelling identified pollution hotspots exceeding both SA NAAQS and WHO guidelines, with the highest daily $PM_{2.5}$ concentration (49.2 µg/m³) exceeding



© University of Venda

the SA NAAQS (40 μg/m³) and WHO (15 μg/m³) limits. Model validation showed a systematic underprediction ($R^2$ = 0.176, negative fractional bias), particularly in areas with multiple emission sources such as biomass burning, unpaved roads, brickmaking, warehouses, and agricultural activities, which were not included in the model; despite these limitations, AERMOD effectively captured $PM_{2.5}$ dispersion patterns, offering valuable insights for air quality management, regulatory compliance, and policy interventions.

Health risk assessment (HHRA) confirmed substantial human health risks associated with $PM_{2.5}$ exposure, particularly among infants, toddlers, and individuals with pre-existing health conditions. Hazard Quotient (HQ) values exceeded safe limits (1.10 in winter), while Excess Lifetime Cancer Risk (ELCR) values ($2.35 \times 10^{-1}$ for adults) surpassed the acceptable threshold ($10^{-4}$), indicating severe long-term health concerns. These findings highlight the urgent need for mitigation strategies, such as promoting cleaner fuels, reducing electricity costs, and improving traffic management to protect vulnerable populations.

Machine learning-based predictive modelling demonstrated the superior performance of the Random Forest (RF) model, achieving higher accuracy ($R^2$ = 0.80 for site-specific predictions, 0.604 for area-wide estimates) compared to Multiple Linear Regression (MLR) and Polynomial Ridge Regression (PRR). The model effectively captured seasonal variations, performing best in winter ($R^2$ = 0.58, RMSE = 2.10 μg/m³) but with lower accuracy in summer due to meteorological variability. These results emphasize the importance of integrating meteorological factors into air quality models for improved environmental management.

This study establishes a clear association between ambient air pollution and human health risks in Thulamela Municipality, providing a robust framework for regulatory enforcement, air quality improvement, and public health protection. Through the integration of scientific research and practical application, this study offers a comprehensive approach to addressing air pollution via monitoring, modelling, prediction, and mitigation strategies. By combining academic research, advanced technology, and policy recommendations, the study aims to enhance air quality management and safeguard community well-being. Future efforts should focus on strengthening predictive



© University of Venda

models, broadening monitoring networks, and advancing policy-driven interventions to foster cleaner air and healthier environments.

Together, the findings presented across the five manuscripts contribute toward the development of an evidence-based air quality management plan tailored to the specific needs of Thulamela Municipality. This thesis bridges key gaps in air quality research for semi-urban African contexts by unifying emission inventory development, real-time monitoring, dispersion modelling, machine learning prediction, and health risk assessment into an integrated framework. The combined results provide a holistic understanding of the region's air pollution landscape and serve as a coherent evidence base to inform local air quality planning, regulatory actions, and public health strategies in similarly data-scarce regions. This synthesis underscores the thesis's unique contribution to advancing interdisciplinary, data-driven approaches to air quality research and supporting sustainable environmental policy development and implementation.

### 8.2 Recommendations

Based on the findings of this research, the following recommendations are proposed to improve air quality management in Thulamela Municipality:

1. Future studies should incorporate unaccounted sources, such as biomass burning, unpaved roads, warehouses, and agricultural activities to improve the accuracy of dispersion modelling.
2. Continuous monitoring of air pollutants and meteorological conditions is essential for timely interventions and policy adjustments.
3. Source apportionment and PM chemical composition analysis for epidemiological studies should be conducted.
4. Further research into targeted site-specific mitigation strategies should be conducted for the abatement of $PM_{2.5}$ exposure in high-risk areas.
5. Deployment of the developed machine learning model for predicting $PM_{2.5}$ for real-time monitoring.
6. Future studies should incorporate AI driven technology, including deep learning techniques (CNNs, LSTMs), to improve the predictive accuracy of $PM_{2.5}$ dispersion under varying meteorological conditions.



© University of Venda

7. The establishment of an evidence-based policies incorporating real-time pollution trends, climate variability, and emerging environmental challenges to ensure long-term sustainability should be further explored.

### 8.3 Contributions to Knowledge

This study makes significant contributions to the field of air quality management and environmental health as outlined in the sections below.

1. This study developed the first-ever 10-year vehicle emission inventory (VEI) for criteria pollutants ($SO_2$, $NO_x$, CO, $PM_{2.5,}$ and $PM_{10)}$ in Thulamela LM, South Africa.
2. The study identified and quantified the major sources of air pollution in Thulamela Local Municipality, South Africa, highlighting the dominant contribution of residential wood combustion (RWC) and vehicular emissions to concentrations of criteria pollutants ($SO_2$, $NO_x$, CO, $PM_{2.5,}$ and $PM_{10}$).
3. It also developed the first comprehensive emission inventory for Thulamela Local Municipality for the year 2021, providing critical baseline data to support air quality management and future mitigation strategies.
4. The Dylos DC1700 was validated against reference-grade instruments during this study and confirms its feasibility for cost-effective air quality monitoring in resource-limited settings.
5. The study established how meteorological conditions including temperature, humidity, wind speed, and pressure influence $PM_{2.5}$ concentrations in a peri-urban environment in South Africa.
6. By utilising correlation analysis, machine learning, and AI-driven approaches, the study provided new insights into $PM_{2.5}$ pollution dynamics and improved prediction accuracy.
7. Pollution distribution in the study area was reported via the developed heatmaps, bivariate polar plots, and spatial dispersion maps. These can facilitate decision-making for environmental management.
8. Pollution hotspots were identified for the first time in the study area through the application of the AERMOD dispersion model.



© University of Venda

### 8.4 Areas for Further Research

Based on the findings of this study, future research should focus on:

1. An investigation of how $PM_{2.5}$ levels and meteorological conditions evolve over time, particularly in response to climate change and anthropogenic activities.
2. Quantifying the contributions of additional emission sources, including unaccounted industrial activities, unpaved road dust, agricultural burning, and warehouse emissions.
3. Developing hybrid modelling approaches integrating machine learning and atmospheric dispersion models to enhance $PM_{2.5}$ forecasting accuracy.
4. Evaluating the effectiveness of present air pollution policies in reducing $PM_{2.5}$ concentrations and suggest evidence-based improvements.
5. Assess public awareness, perceptions, and involvement in citizen-science initiatives to encourage grassroots air pollution mitigation efforts.
6. Investigate the potential of remote sensing, sensor networks, and big data analytics for real-time $PM_{2.5}$ monitoring and pollution control.
7. Conduct long-term studies on the health effects of $PM_{2.5}$ exposure, focusing on respiratory and cardiovascular diseases across different demographics.



© University of Venda

# APPENDICES

**Appendix 3.1:** Sampling locations for vehicle activities

| ID | Area | Latitude (S) | Longitude E | Altitide (m) |
|---|---|---|---|---|
| 1 | Mvusoludzo taxi rank | 22° 58' 45. 9" | 030° 28' 01. 6" | 772 |
| 2 | Venda taxi rank | 22° 58' 12. 2" | 030° 27' 27. 2" | 601 |
| 3 | Thulamela taxi rank | 22° 58' 12. 5" | 030° 27' 35. 1" | 604 |
| 4 | Sibasa Taxi Rank | 22° 57' 01. 6" | 030° 28' 15. 3" | 706 |
| 5 | University of venda | 22° 58' 38. 24" | 030° 26' 39. 81" | |

**Appendix 3.2:** Sampling point coordinates for residential wood measurement

| Sampling Area | House Code | Code | Coordinates | | Elevation |
|---|---|---|---|---|---|
| | | | Latitudes (S) | Longitudes (E) | |
| LUNUNGWI | HL1 | L1 | 22° 55' 23.9" | 030° 25' 46.9" | 716 |
| | HL2 | L2 | 22° 55' 22.6" | 030° 25' 49.8" | 746 |
| | HL 3 | L3 | 22° 55' 22.0" | 030° 25' 51.8" | 732 |
| | HL 4 | L4 | 22° 55' 19.8" | 030° 25' 56.0" | 712 |
| | HL 5 | L5 | 22° 55' 14.8" | 030° 25' 56.0" | 712 |
| | HL 6 | L6 | 22° 55' 16.3" | 030° 25' 56.7" | 726 |
| | HL 7 | L7 | 22° 55' 13.8" | 030° 25' 54.0" | 707 |
| | HL 8 | L8 | 22° 55' 08.2" | 030° 25' 55.6" | 685 |
| | HL 9 | L9 | 22° 55' 06.9" | 030° 25' 57.2" | 716 |
| | HL 10 | L10 | 22° 55'08.0" | 030° 25' 59.4" | 699 |
| | HL 11 | L11 | 22° 55' 06.2" | 030° 25' 55.0" | 693 |
| | HL 12 | L12 | 22° 55' 10.6" | 030° 26' 05.9" | 729 |
| | HL 13 | L13 | 22° 55′ 3.16" | 030° 26′ 18.51" | 728 |
| | HL 14 | L14 | 22° 54′ 59.36" | 030° 25′ 56.67" | 714 |
| | HL 15 | L15 | 22° 55′ 13.86" | 030° 25′ 51.71" | 690 |
| GONDENI | HG 16 | G1 | 22°55'14.4" | 030 26' 15. 3" | 723 |
| | HG17 | G2 | 22 55'09. 6" | 030 26' 23. 0" | 719 |
| | HG 18 | G3 | 22 55' 07. 9" | 030 26' 21. 8" | 720 |
| | HG 19 | G4 | 22 55' 03. 2" | 030 26' 18. 3" | 690 |
| | HG 20 | G5 | 22 54' 59. 9" | 030 26' 18. 6" | 678 |
| | HG 21 | G6 | 22 55' 12.7" | 030 26' 27.3" | 733 |
| | HG 22 | G7 | 22 55' 11. 8" | 030 26' 29.0" | 735 |
| | HG 23 | G8 | 22 55'12. 8" | 030 26' 31. 9" | 732 |
| | HG 24 | G9 | 22 55'09. 9" | 030 26' 33. 9" | 725 |
| | HG 25 | G10 | 22 55' 10.2" | 030 26' 39. 3" | 731 |
| | HG 26 | G11 | 22 55' 08.1" | 030 26' 38. 7" | 729 |
| | HG 27 | G12 | 22 55' 05. 6" | 030 26' 35. 8" | 723 |
| | HG 28 | G13 | 22 55' 07. 3" | 030 26' 28. 8" | 719 |
| | HG 29 | G14 | 22 55' 21. 4" | 030 26' 27. 2" | 734 |
| | HG 30 | G15 | 22° 55′ 14.88" | 30° 26′ 22.14" | 739 |

*HL for House number in Lunungwi and HG for house number at Gondeni



© University of Venda

**Appendix 3.3:** Vehicle Kilometre Travelled (VKT) Research Questionnaire

**Vehicle Kilometre Travelled (VKT) Research Questionnaire**

Dear Respondents,

Thank you in advance for using part of your productive time to complete this questionnaire. This exercise is meant to estimate vehicle emissions and their impact on the ambient air quality of Thulamela Municipality within Vhembe District, Limpopo Province. Kindly assist in answering the following questions as the confidentiality of your response is guaranteed. I will be grateful for your cooperation.

**SECTION 1. AREA AND VEHICLE IDENTIFICATION**

Please tick the appropriate box

1. What is the name of your resident city? .....
2. What Municipality is your area located in? [ ] Thulamela [ ] Makhado [ ] Others: .....
3. Type of vehicle: [ ] Car [ ] Quantum [ ] Bakkie [ ] Bus [ ] Truck [ ] Others: .....
5. What is the model year of your vehicle? .....
6. What is your vehicle's current service date? .....
7. What is your vehicle registration date? .....
8. What is the purpose of your vehicle? [ ] Personal [ ] Commercial
9. How long have you been using the vehicle for that purpose? [ ] 1 year [ ] 2 years [ ] 3 years [ ] 4 years [ ] 5 years [ ] Others: .....
10. Which area is covered by your vehicle? Click all that apply: [ ] Sibasa [ ] Thohoyandou [ ] Shayandima [ ] Tshakhuma [ ] Others: .....
11. What type of fuel do you use? [ ] Petrol [ ] Diesel [ ] Others: .....
12. How many litres of fuel do you use per day? .....
13. How much fuel do you use per week? .....
14. What is your fuel tank size in litre? .....
15. How many kilometres do you cover with a filled-up tank? .....
16. How many passengers does your vehicle take? [ ] 1-6 [ ] 7-12 [ ] 13-18 [ ] 19-24 [ ] 25-30 [ ] Above 30
17. What are the likely sources of air pollution in your area? [ ] Vehicular emission [ ] Domestic burning [ ] Industrial emission [ ] Dust [ ] Others: .....
18. What is the condition of the atmosphere in your area? [ ] Good [ ] Fair [ ] Bad
Explain: .....
19. What is your vehicle engine capacity? .....

**SECTION 2. VEHICLE KILOMETRE TRAVELLED (VKT) DETAILS**

Please fill in the table below

| Day | Average accumulative mileage (km) Odometer reading Before going out | Average accumulative mileage (km) Odometer reading End of the day | Fuel consumption (Litres) |
|---|---|---|---|
| Monday | | | |
| Tuesday | | | |
| Wednesday | | | |
| Thursday | | | |
| Friday | | | |
| Saturday | | | |
| Sunday | | | |



© University of Venda

**Appendix 3.4:** Residential Wood Burning Field Questionnaire

## Residential Wood Burning Field Questionnaire

Dear Respondents,

Thank you in advance for using part of your productive time to complete this questionnaire. This exercise is meant to estimate Residential fuelwood burning and its impact on the ambient air quality of Gondeni and Lunungwi villages in Thulamela Municipality within Vhembe District, Limpopo Province. Kindly assist in answering the following questions as the confidentiality of your response is guaranteed. I will be grateful for your cooperation.

### SECTION 1. VILLAGE AND HOUSE IDENTIFICATION

1) Sex: [ ] Male [ ] Female

2) Age (years): [ ] 11–20 [ ] 21–30 [ ] 31–40 [ ] 41–50 [ ] Above 50

3) Educational attainment: [ ] No formal education [ ] Primary education [ ] High school [ ] Tertiary education

4) Are-you: [ ] Full-time employed [ ] Part-time employed [ ] Self-employed [ ] Unemployed

5) How many people stays in your house now? [ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] Above 5

6) How many bedrooms does your house have? [ ] 1-2 [ ] 3-4 [ ] 5-6 [ ] Above 7

### SECTION 2. FUEL IDENTIFICATION

7) How frequently do you use this type of fuel for household activities? Tick

| Types of fuel | Frequently | Not frequently | Not applicable |
|---|---|---|---|
| Gas | [ ] | [ ] | [ ] |
| Coal | [ ] | [ ] | [ ] |
| Wood | [ ] | [ ] | [ ] |
| Paraffin | [ ] | [ ] | [ ] |
| Candle | [ ] | [ ] | [ ] |
| Electricity | [ ] | [ ] | [ ] |

8) What type of fuel do you use in the summer season (December – February), for

8.1) Lighting? [ ] Candle [ ] Electricity [ ] Kerosene/Paraffin [ ] Electricity and candle [ ] Electricity and paraffin

8.2) Heating? [ ] Electricity [ ] Wood [ ] Electricity and wood

8.3) Cooking? [ ] Electricity [ ] Wood [ ] Electricity and wood Specify if others……..

9) What type of fuel do you use in the autumn season (March – May), for

9.1) Lighting? [ ] Candle [ ] Electricity [ ] Kerosene/Paraffin [ ] Electricity and candle [ ] Electricity and paraffin



© University of Venda

9.2) Heating? [ ] Electricity [ ] Wood [ ] Electricity and wood

9.3) Cooking? [ ] Electricity [ ] Wood [ ] Electricity and wood Specify if others……..

10) What type of fuel do you use in the winter season (June - August), for

10.1) Lighting? [ ] Candle [ ] Electricity [ ] Kerosene/Paraffin [ ] Electricity and candle [ ] Electricity and paraffin

10.2) Heating? [ ] Electricity [ ] Wood [ ] Electricity and wood

10.3) Cooking? [ ] Electricity [ ] Wood [ ] Electricity and wood Specify if others……..

11) What type of fuel do you use in the spring season (September - November), for

11.1) Lighting? [ ] Candle [ ] Electricity [ ] Kerosene/Paraffin [ ] Electricity and candle [ ] Electricity and paraffin

11.2) Heating? [ ] Electricity [ ] Wood [ ] Electricity and wood

11.3) Cooking? [ ] Electricity [ ] Wood [ ] Electricity and wood Specify if others……..

## SECTION 3: WOOD FUEL USAGE

12) What type of woodstove or fireplace do you use for wood combustion?

[ ] Traditional stove [ ] Three-legged stand [ ] Open fireplace [ ] Closed fireplace [ ] Braai stand

13) In the summer season (December – February),

13.1) What is the average number of days of wood use during the week? [ ] 1 day [ ] 2 days [ ] 3 days [ ] 4 days [ ] 5 days [ ] 6 days [ ] 7 days

13.2) What is the average number of daily hours of wood use? [ ] <1 hour [ ] 1-2 hours [ ] 2-3 hours [ ] 3-4 hours [ ] 4-6 hours [ ] 6-8 hours [ ] >8 hours

13.3) How do you get wood? [ ] Buy [ ] Harvest [ ] Buy and harvest

13.4) How often do you get or buy wood? [ ] Daily [ ] Weekly [ ] Monthly [ ] As needed

14) In the winter season (June - August),

14.1) What is the average number of days of wood used during the week? [ ] 1 day [ ] 2 days [ ] 3 days [ ] 4 days [ ] 5 days [ ] 6 days [ ] 7 days

14.2) What is the average number of daily hours of wood use? [ ] <1 hour [ ] 1-2 hours [ ] 2-3 hours [ ] 3-4 hours [ ] 4-6 hours [ ] 6-8 hours [ ] >8 hours

14.3) How do you get wood? [ ] Buy [ ] Harvest [ ] Buy and harvest



© University of Venda

14.4) How often do you get or buy wood? [ ] Daily [ ] Weekly [ ] Monthly [ ] As needed

15) Type of fuelwood or tree (mark one or more boxes for the following question)?

| S/No | Local name | Common name | 15.1 How often (F/NF/NA) | 15.2 How do you get it? (1/2/3) |
|---|---|---|---|---|
| 1 | Muafukhada | Avocado | | |
| 2 | Mugwiti | Mopane | | |
| 3 | Muluwa | Matumi | | |
| 4 | Mupani | Flame thorn | | |
| 5 | Murenzhe | Mitzeerie | | |
| 6 | Mutulume | Velvet bushwillow | | |
| 7 | Munngo | Mango | | |
| 8 | Munzere | Sickle bush | | |
| 9 | Muvhula | Mobola-plum | | |

16) What is the main reason to use fuelwood burning for domestic purposes (mark one box)?

[ ] Abundant [ ] Economic purpose [ ] Taste [ ] No electricity [ ] Tradition [ ] Economic purpose and abundant [ ] Economic purpose and Taste

17) Are you satisfied in general with the fuel wood burning system used? [ ] Very much [ ] A little [ ] Not at all

18) Are you aware that fuel wood burning can contribute to air pollution? [ ] Yes [ ] No

19) What is the condition of the atmosphere in your area? [ ] Good [ ] Fair [ ] Bad

20) What is the name of your village or area? [ ] Lunungwi [ ] Gondeni



© University of Venda

**Appendix 3.5:** Ethical Clearance Certificate

RESEARCH AND INNOVATION
OFFICE OF THE DIRECTOR

NAME OF RESEARCHER/INVESTIGATOR:

**Ms IT Enitan**

**Student No:**
**18021500**

PROJECT TITLE: **Ambient air pollution modelling and human health risk within Capricorn and Vhembe District, Limpopo Province, South Africa.**

PROJECT NO: **SES/19/ERM/20/0511**

SUPERVISORS/ CO-RESEARCHERS/ CO-INVESTIGATORS

| NAME | INSTITUTION & DEPARTMENT | ROLE |
|---|---|---|
| Prof JO Odiyo | University of Venda | Promoter |
| Dr JN Edokpayi | University of Venda | Co - Promoter |
| Ms IT Enitan | University of Venda | Investigator – Student |

ISSUED BY:
UNIVERSITY OF VENDA, RESEARCH ETHICS COMMITTEE

Date Considered: November 2019
Decision by Ethical Clearance Committee Granted
Signature of Chairperson of the Committee: .....
Name of the Chairperson of the Committee: Senior Prof. G.E. Ekosse

UNIVERSITY OF VENDA
DIRECTOR
RESEARCH AND INNOVATION
2019 -11- 06
Private Bag X5050
Thohoyandou 0950

University of Venda
PRIVATE BAG X5050, THOHOYANDOU, 0950, LIMPOPO PROVINCE, SOUTH AFRICA
TELEPHONE (015) 962 8504/8313 FAX (015) 962 9060
*"A quality driven financially sustainable, rural-based Comprehensive University"*



© University of Venda

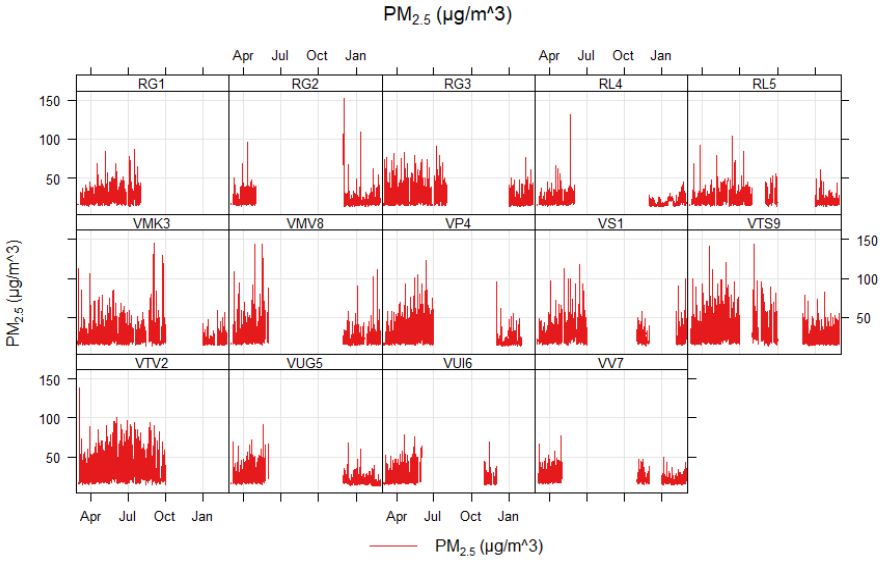
**Appendix 4.1:** Sampling locations of the monitoring sites in the study area using Dylos DC1700

| S/N | Land use Category | Proximity | Sampling Area | Sampling Location | Group Code | Site code | Coordinate | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Latitude (S) | Longitude (E) |
| 1 | Residential | | Gondeni | Within the villages | RG | RG1 | -22.916986 | 30.446544 |
| 2 | | | | | | RG2 | -22.920799 | 30.439484 |
| 3 | | | | | | RG3 | -22.917543 | 30.438475 |
| 4 | | | Lunungwi | | RL | RL4 | -22.916489 | 30.432409 |
| 5 | | | | | | RL5 | -22.918311 | 30.425297 |
| 6 | Vehicular/ Traffic/ Residential | Roadsides and Malls | Sibasa | Sibasa | VA | VS1 | -22.9414 | 30.466546 |
| 7 | | | Thohoyandou | Thavhani mall and Library area | | VTV2 | -22.983681 | 30.460576 |
| 8 | | | Tshilungoma | Phangami mall area | | VP4 | -22.972857 | 30.489417 |
| 9 | | | Makumbane | Adjacent Tshilidzini Hospital main Gate (Makumbane) | | VTS9 | -22.996598 | 30.412734 |
| 10 | | University and commercial | Makwarela | Behind Makwarela clinic (Along the road) | VB | VMK3 | -22.947363 | 30.491878 |
| 11 | | | Thohoyandou | University of Venda Main gate | | VUG5 | -22.973743 | 30.448464 |
| 12 | | | | University of Venda campus residence | | VUI6 | -22.978062 | 30.449829 |
| 13 | | Taxi ranks and roadsides | Thohoyandou | Opposite Venda Taxi rank | VC | VV7 | -22.970142 | 30.456357 |
| 14 | | | | Around Mvusuludzo Taxi rank | | VMV8 | -22.973632 | 30.465607 |

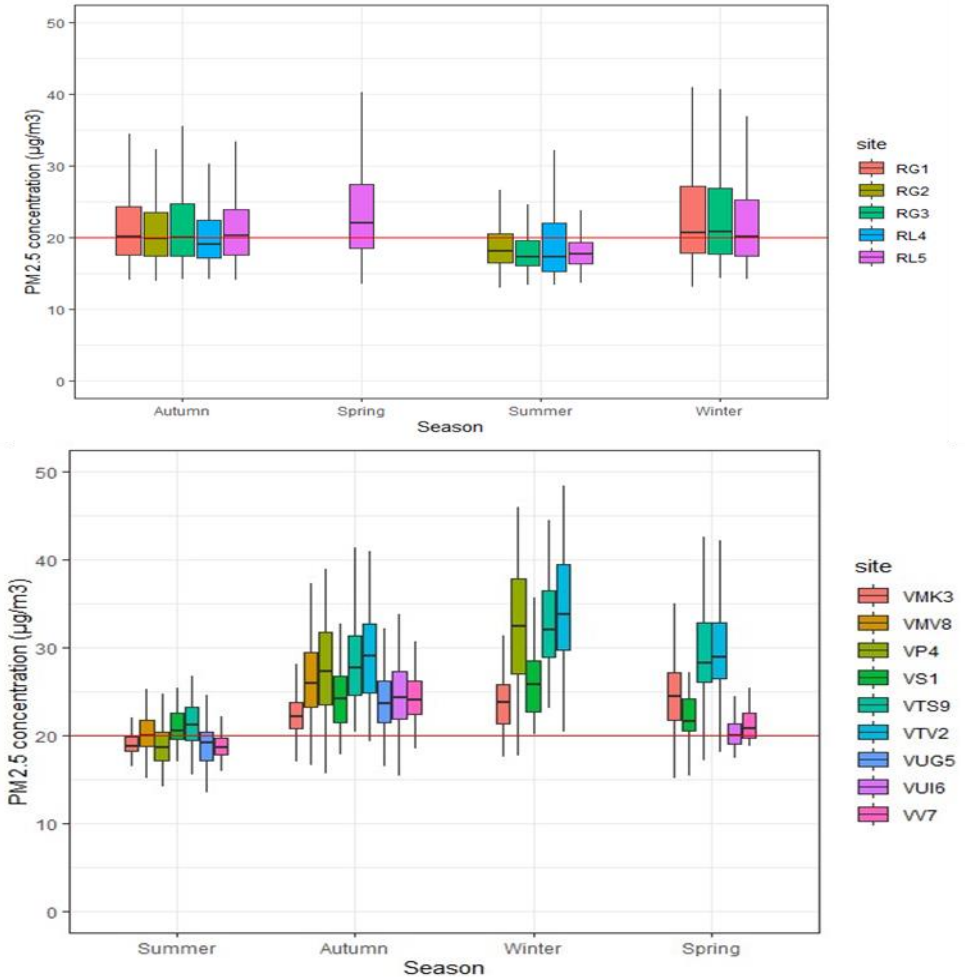


© University of Venda

**Appendix 4.2:** Time series plot of $PM_{2.5}$ concentrations using hourly averages for all the sites.




**Appendix 4.3**: Seasonal concentrations of $PM_{2.5}$ per site in Thulamela LM for the study period



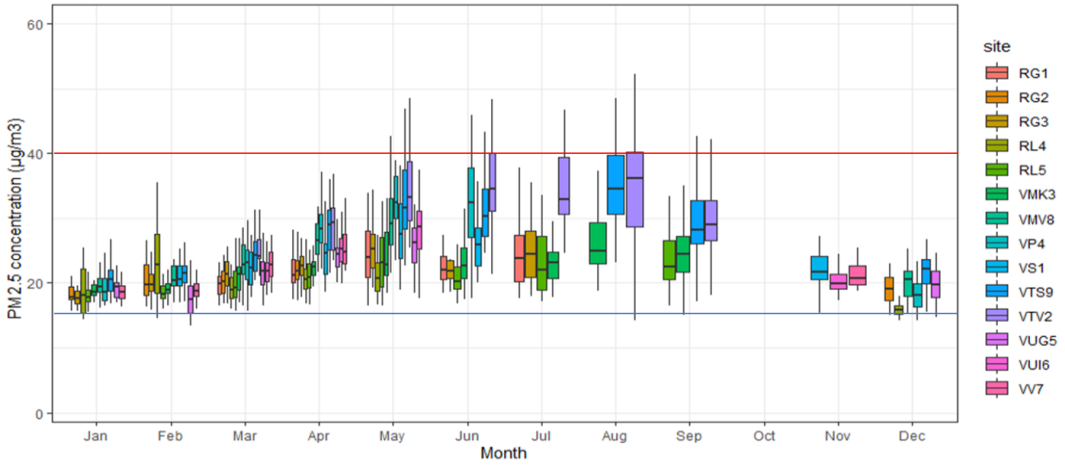



© University of Venda

**Appendix 4.4:** Descriptive statistics for seasonal mean concentrations of $PM_{2.5}$ (µg/m³) by group across the monitoring area

| Group | Seasons | Mean ± Std. | Number of valid values N(values) |
|---|---|---|---|
| RG | Autumn | 22.32 ± 3.48 | 233 |
| | Winter | 24.34 ± 5.53 | 113 |
| | Spring | - | 0 |
| | Summer | 20.59 ± 10.49 | 149 |
| | Annual | 22.26 ± 6.89 | 495 |
| RL | Autumn | 21.17 ± 3.97 | 172 |
| | Winter | 22.78 ± 5.04 | 59 |
| | Spring | 23.61 ± 4.87 | 24 |
| | Summer | **18.84 ± 3.79** | 149 |
| | Annual | 20.69 ± 4.42 | 404 |
| VA | Autumn | 28 ± 6.19 | 356 |
| | Winter | 33.21 ± 7.98 | 204 |
| | Spring | 26.76 ± 6.03 | 86 |
| | Summer | 20.56 ± 3.27 | 177 |
| | Annual | 27.56 ± 7.55 | 823 |
| VB | Autumn | 23.94 ± 3.98 | 259 |
| | Winter | 24.35 ± 4.48 | 84 |
| | Spring | 23.57 ± 8.06 | 54 |
| | Summer | **19.08 ± 2.41** | 144 |
| | Annual | 22.67 ± 4.83 | 541 |
| VC | Autumn | 25.79 ± 4.59 | 144 |
| | Winter | - | - |
| | Spring | 21.35 ± 2.06 | 26 |
| | Summer | **19.86 ± 2.78** | 144 |
| | Annual | 22.7 ± 4.66 | 314 |

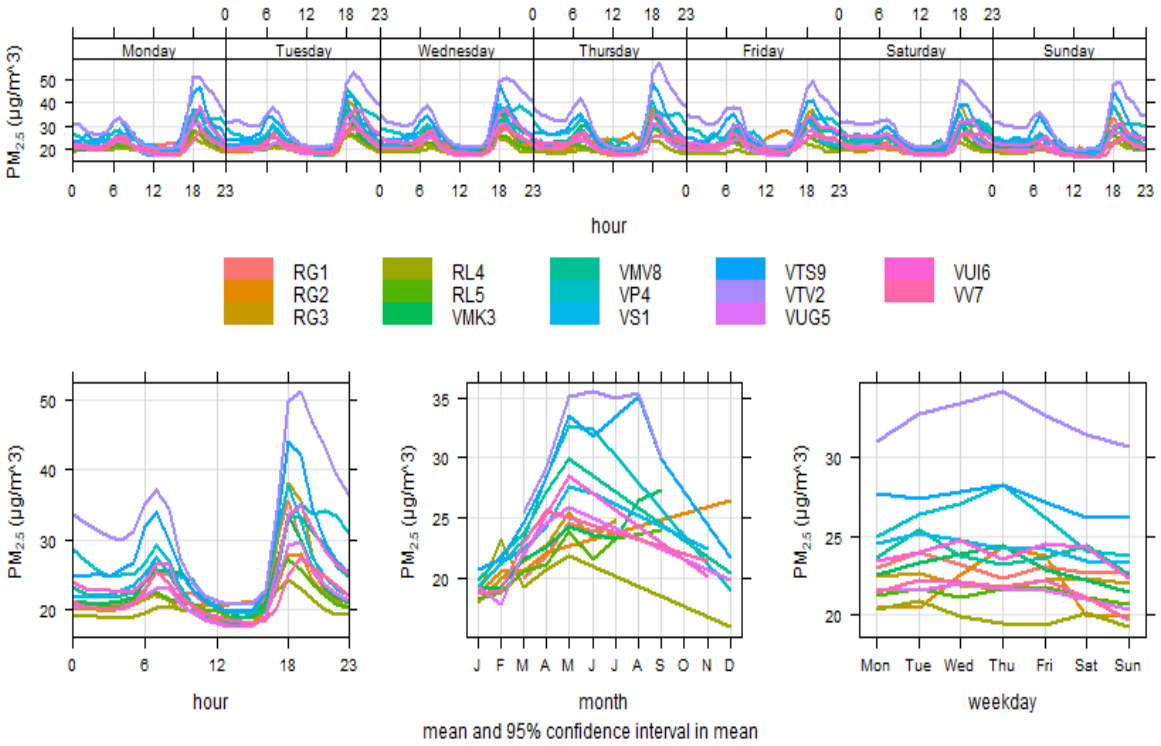
**Appendix 4.5:** Monthly variations of $PM_{2.5}$ concentrations in the study area, with the red and blue bold line indicating the SA NAAQS and WHO daily standard






© University of Venda

**Appendix 4.6:** Time variations analysis of $PM_{2.5}$ concentrations per site in the study area.






© University of Venda

## **Appendix 5.1:** Source parameter template for AERMOD

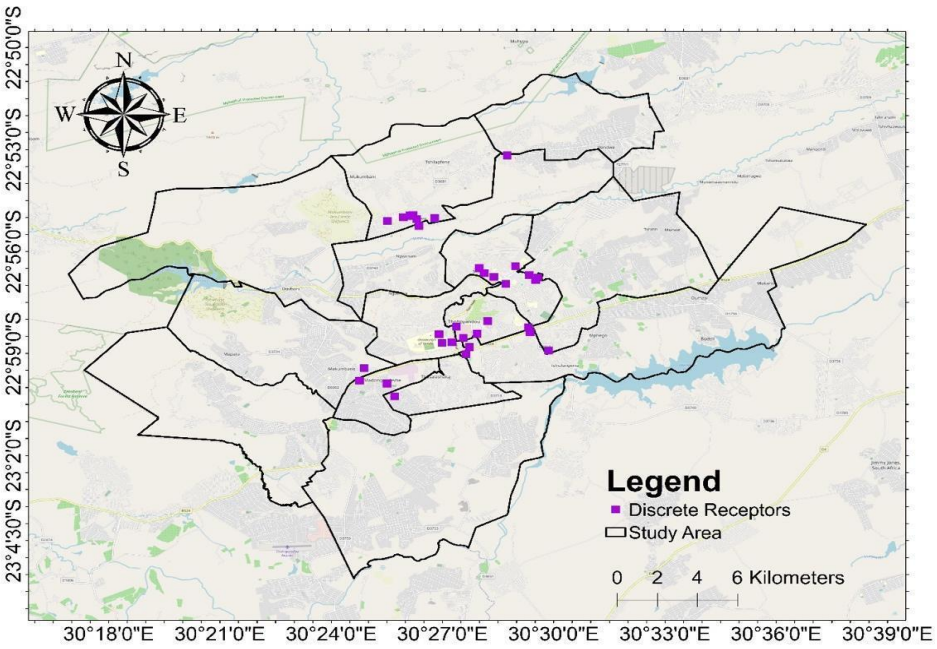
**AERMOD View - Source Parameters**
**MS Excel - Lakes Format: Version 2.0**
**Supported Source Types: Point, Rectangular Area, Circular Area, Polygon Area, Volume, Open Pit, Line Volume, Line Area**

| | Parameters | Units | Description |
|---|---|---|---|
| | Type | | POINT, AREA, AREA_CIRC, AREA_POLY, VOLUME, OPEN_PIT, LINE, LINE_VOLUME, LINE_AREA |
| | ID | | Source ID up to 12 characters |
| | Desc | | Optional description |
| | SourceID_Prefix | | Text prefix up to 4 characters long for generated LINE_VOLUME and LINE_AREA sources |
| | Base_Elev | [m] | Source base elevation above mean sea level |
| | Height | [m] | Release height above ground |
| | Diam | [m] | Inner stack diameter (POINT) or circular area radius (AREA_CIRC) |
| | Exit_Vel | [m/s] | Exit velocity (POINT only) |
| | Exit_Temp | [K] | Exit temperature (POINT only) |
| | Release Type | | VERTICAL, HORIZONTAL, CAPPED (POINT only) - HORIZONTAL and CAPPED are non-default beta options |
| | SigmaY | [m] | Initial sigma Y (VOLUME only) |
| | SigmaZ | [m] | Initial sigma Z (AREA, AREA_CIRC, AREA_POLY, VOLUME, LINE, and LINE_AREA only; optional for AREA, AREA_CIRC, AREA_POLY, and LINE) |
| | Length_X | [m] | X side length (AREA, VOLUME, OPEN PIT, and LINE_AREA only; optional for VOLUME, will be used to calculate SigmaY) |
| | Length_Y | [m] | Y side length (AREA and OPEN PIT only); width for LINE sources |
| | Rotation_Angle | [degrees] | Clockwise rotation from North of Y side (AREA and OPEN PIT only) |
| | Pit_Volume | [m^3] | Volume of the open pit (OPEN PIT only) |
| | Emission_Rate | [g/s or g/s/m2] | Emission rate (g/s for POINT, VOLUME, and LINE_VOLUME; g/s/m2 for AREA, AREA_CIRC, AREA_POLY, OPENPIT, LINE, and LINE_AREA) |
| | Configuration | | LINE_VOLUME configuration: Separated, Adjacent or Separated2W |
| | LineVolumeHeight | [m] | Plume Height or Building Height for LINE_VOLUME source |
| | PlumeWidth | [m] | Plume width for LINE_VOLUME source |
| | LineVolumeType | | LINE_VOLUME type: None, Surface-Based, Elevated, Elevated Building |
| | LineArea_Ratio1 | | Ratio 1 for LINE_AREA sources |
| | Num_Coords | | Number of coordinate pairs (POINT, AREA, AREA_CIRC, VOLUME, OPENPIT = 1; AREA_POLY >= 3; LINE = 2; LINE_AREA, LINE_VOLUME >=2) |
| | X1 | [m] | X coordinate of source location [m] |
| | Y1 | [m] | Y coordinate of source location [m] |
| | X2 | [m] | Secondary X coordinate of source location [m] (AREA_POLY, LINE, LINE_VOLUME, LINE_AREA sources only) |
| | Y2 | [m] | Secondary Y coordinate of source location [m] (AREA_POLY, LINE, LINE_VOLUME, LINE_AREA sources only) |
| | X3 | [m] | Additional X coordinate of source location [m] (AREA_POLY, LINE_VOLUME, LINE_AREA only) |
| | Y3 | [m] | Additional Y coordinate of source location [m] (AREA_POLY, LINE_VOLUME, LINE_AREA only) |
| | X4 | [m] | Additional X coordinate of source location [m] (AREA_POLY, LINE_VOLUME, LINE_AREA only) |
| | Y4 | [m] | Additional Y coordinate of source location [m] (AREA_POLY, LINE_VOLUME, LINE_AREA only) |
| | Base_Elev_m | [m] | Base Elevation for LINE_VOLUME, LINE_AREA Nodes |
| | Rel_Height_m | [m] | Release height for LINE_VOLUME, LINE_AREA Nodes |

NOTE: you may keep adding additional coordinate pairs for an AREA_POLY or LINE_VOLUME sources, be sure to add the headers as well (eg. X5, Y5, etc)

| Type | ID | Base_Elev | Height | Length_X | Length | Emission | X1 | Y1 | Num_Coords | SigmaY | SigmaZ | Rotation_Angle | Pit_Volume |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | [m] | [m] | [m] | [m] | | [m] | [m] | | [m] | [m] | [deg] | [m^3] |
| Area | Res1 | 690.4 | 0.5 | 1000 | 1000 | 4E-08 | 247393 | 7470384.2 | 1 | | | | |
| Area | Res2 | 690.4 | 0.5 | 1000 | 1000 | 4E-07 | 248317 | 7470399.6 | 1 | | | | |
| Area | Res3 | 690.4 | 0.5 | 1000 | 1000 | 2E-07 | 242794 | 7469310.2 | 1 | | | | |
| Area | Res4 | 690.4 | 0.5 | 1000 | 1000 | 3E-07 | 244641 | 7469341.5 | 1 | | | | |
| Area | Res5 | 690.4 | 0.5 | 1000 | 1000 | 4E-07 | 245564 | 7469357.1 | 1 | | | | |
| Area | Res6 | 690.4 | 0.5 | 1000 | 1000 | 1E-07 | 246487 | 7469372.6 | 1 | | | | |
| Area | Res7 | 690.4 | 0.5 | 1000 | 1000 | 9E-07 | 247410 | 7469388 | 1 | | | | |
| Area | Res8 | 690.4 | 0.5 | 1000 | 1000 | 3E-06 | 248333 | 7469403.4 | 1 | | | | |
| Area | Res9 | 690.4 | 0.5 | 1000 | 1000 | 2E-07 | 240042 | 7468266.6 | 1 | | | | |
| Area | Res10 | 690.4 | 0.5 | 1000 | 1000 | 1E-06 | 240965 | 7468282.5 | 1 | | | | |
| Area | Res11 | 690.4 | 0.5 | 1000 | 1000 | 2E-06 | 241888 | 7468298.3 | 1 | | | | |

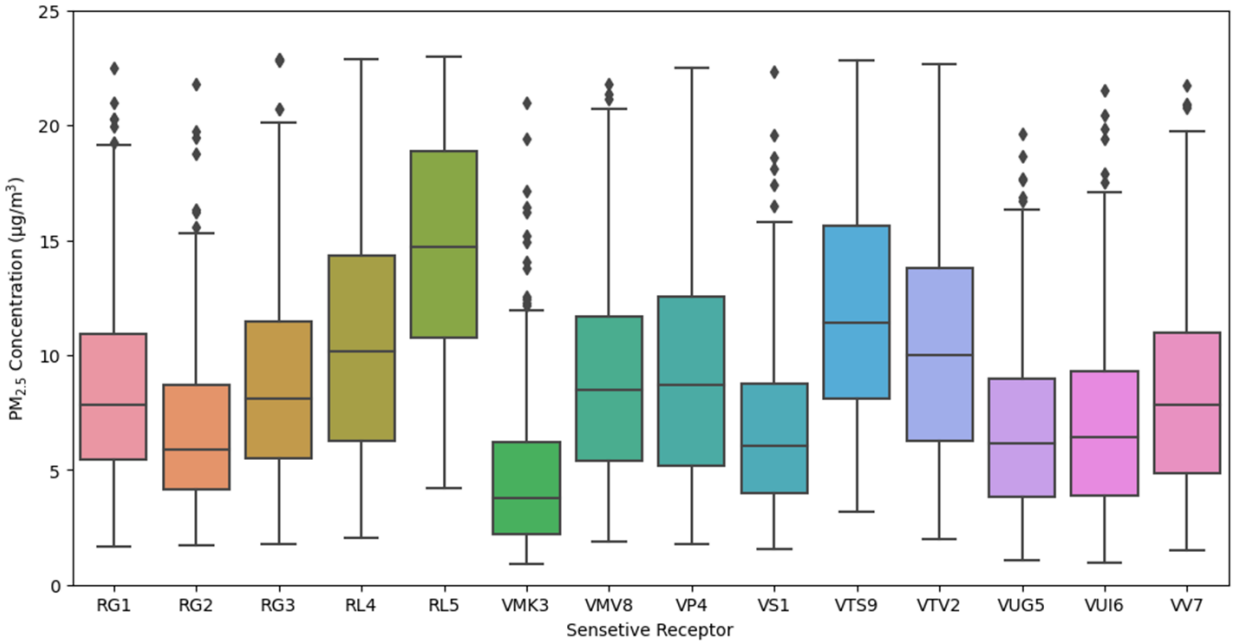
## **Appendix 5.2:** Map showing the locations used as discrete receptors in the study domain



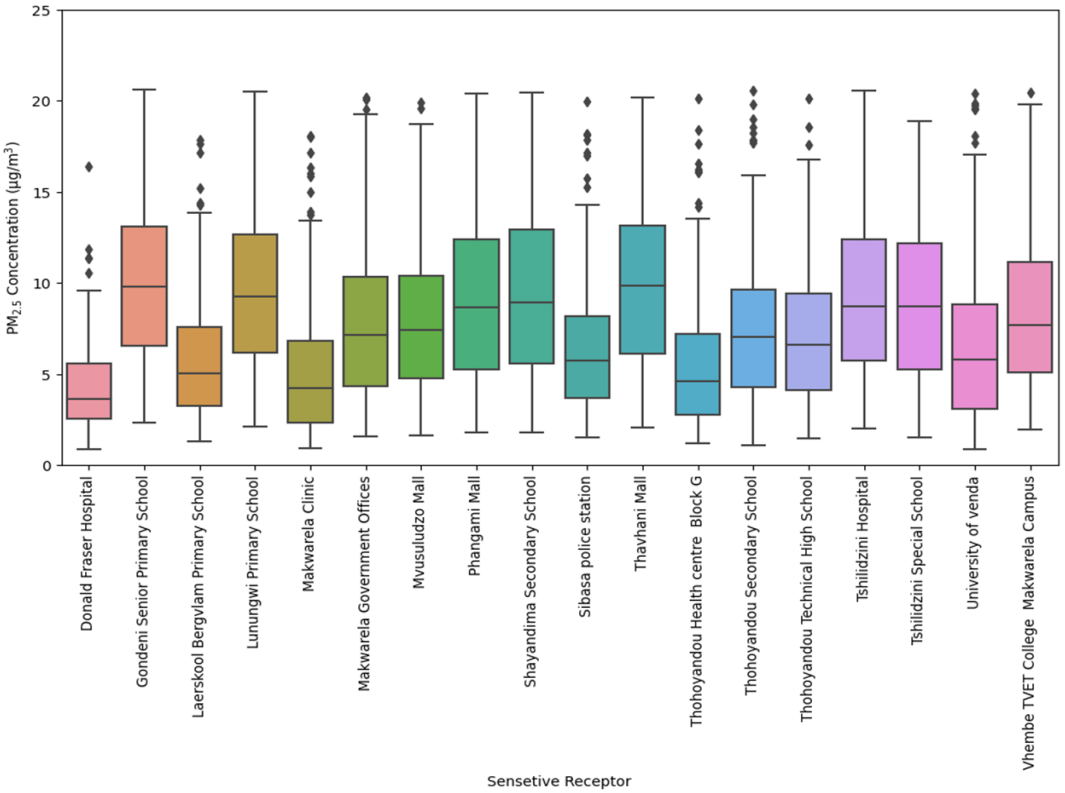



© University of Venda

**Appendix 5.3.** Box plots showing the modelled $PM_{2.5}$concentration from the monitoring location receptors simulated at Thulamela LM area.




**Appendix 5.4:** Box plots showing the estimated $PM_{2.5}$ concentration in selected sensitive locations of the Thulamela LM area.






© University of Venda

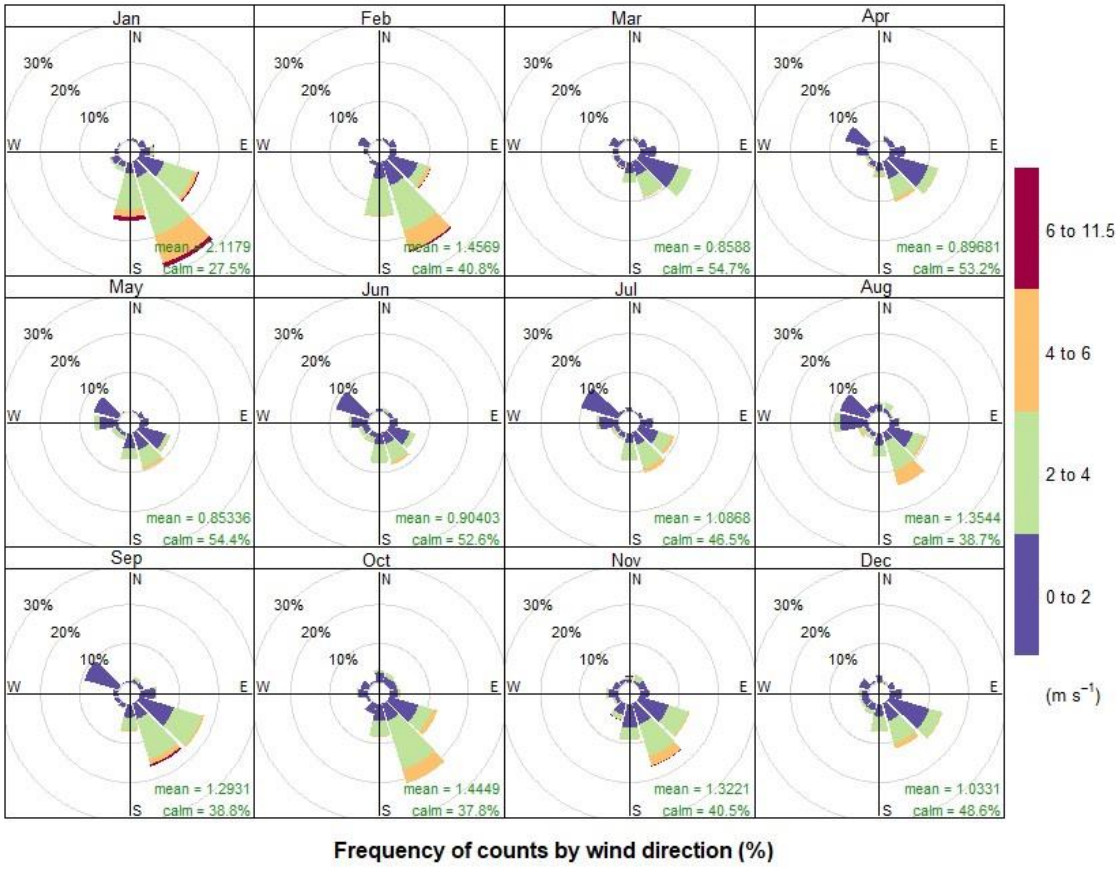
**Appendix 6.1:** Hazard Quotient (HQ) for non-carcinogenic risks of $PM_{2.5}$ *by land*-use activity *and across the study area* in comparison to World Health Organization (WHO) Reference concentration (RfC) standard

| Group | Season | WHO | | | |
|---|---|---|---|---|---|
| | | HQ | | | |
| | | Infant | Toddler | Child | Adult |
| RG | Winter | 3.80 | 3.46 | 2.17 | 1.39 |
| | Summer | 3.21 | 2.92 | 1.83 | 1.18 |
| RL | Winter | 3.56 | 3.24 | 2.03 | 1.30 |
| | Summer | 2.94 | 2.68 | 1.68 | 1.08 |
| VA | Winter | 5.19 | 4.72 | 2.96 | 1.90 |
| | Summer | 3.21 | 2.92 | 1.83 | 1.18 |
| VB | Winter | 3.80 | 3.46 | 2.17 | 1.39 |
| | Summer | 2.98 | 2.71 | 1.70 | 1.09 |
| VC | Winter | - | - | - | - |
| | Summer | 3.10 | 2.82 | 1.77 | 1.14 |
| $PM_{2.5}$ for Study area | Winter | 4.38 | 3.99 | 2.50 | 1.61 |
| | Summer | 3.09 | 2.82 | 1.76 | 1.13 |
| | Annual | 3.72 | 3.39 | 2.12 | 1.36 |



© University of Venda

**Appendix 7.1:** The monthly windrose for the study period showing wind speed and prevailing direction frequencies



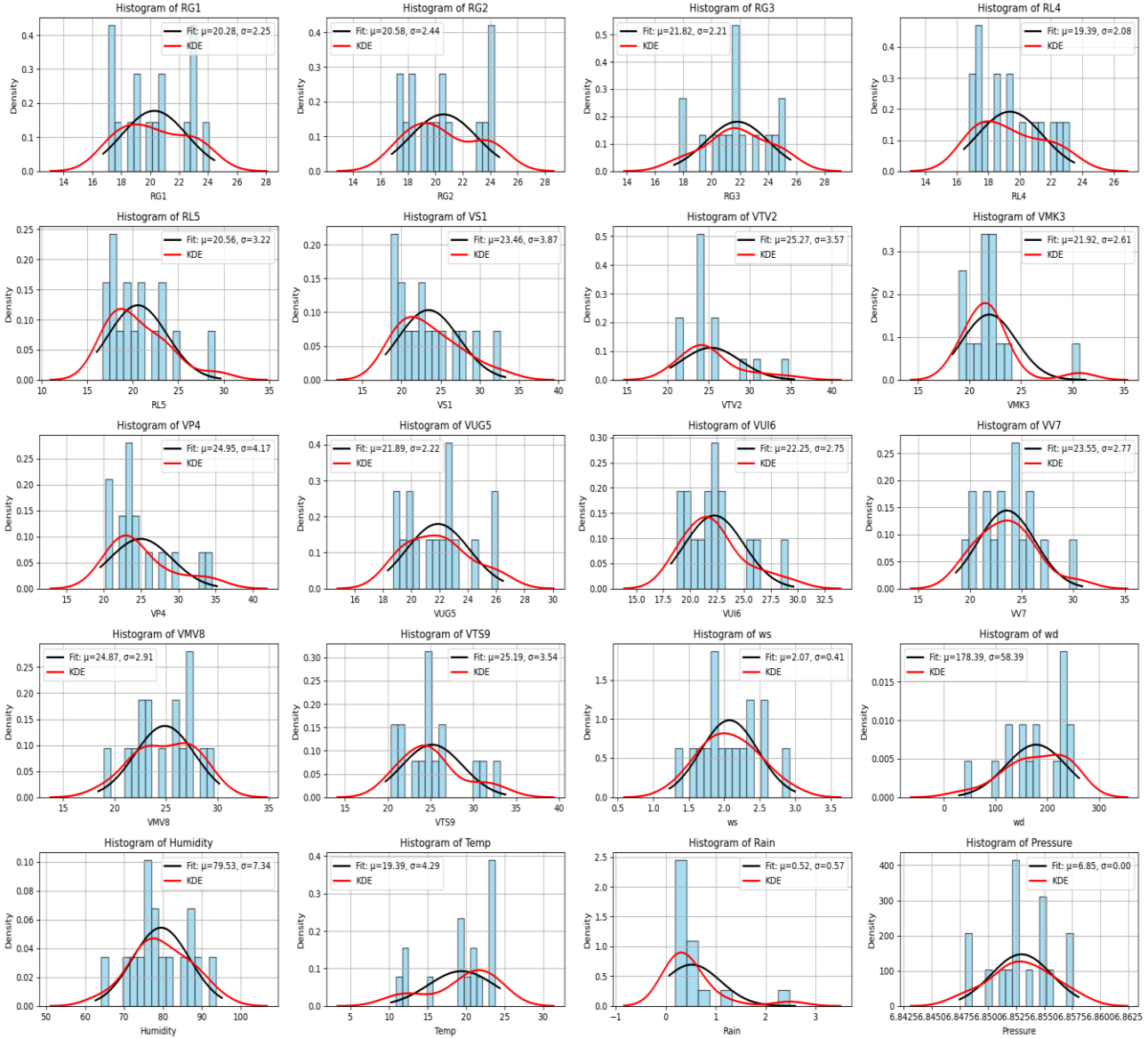

**Appendix 7.2:** Shapiro-Wilk analysis of criterion air pollutants

| Column | Test Statistic | P-value | Normality |
|---|---|---|---|
| RG1 | 0.9290 | 0.2348 | Normal |
| RG2 | 0.9000 | 0.0804 | Normal |
| RG3 | 0.9515 | 0.5138 | Normal |
| RL4 | 0.9156 | 0.1433 | Normal |
| RL5 | 0.8973 | 0.0727 | Normal |
| VS1 | 0.9251 | 0.2041 | Normal |
| VTV2 | 0.8536 | 0.0154 | Not Normal |
| VMK3 | 0.7571 | 0.0008 | Not Normal |
| VP4 | 0.8543 | 0.0158 | Not Normal |
| VUG5 | 0.9414 | 0.3665 | Normal |
| VUI6 | 0.9130 | 0.1303 | Normal |
| VV7 | 0.9674 | 0.7951 | Normal |
| VMV8 | 0.9641 | 0.7355 | Normal |
| VTS9 | 0.9101 | 0.1169 | Normal |
| ws | 0.9887 | 0.9983 | Normal |
| wd | 0.9249 | 0.2023 | Normal |
| Humidity | 0.9864 | 0.9947 | Normal |
| Temp | 0.8391 | 0.0095 | Not Normal |
| Rain | 0.4120 | 0.0000 | Not Normal |
| Pressure | 0.9599 | 0.6602 | Normal |



© University of Venda

**Appendix 7.3:** Visualization of skewness of $PM_{2.5}$ concentrations and meteorological factors from the study site using histogram from March 2021 to February 2022






© University of Venda

**Appendix 7.4:** Spearman correlation with confidence interval

| Site | Ws | Wd | Humidity | Temperature | Rain | Pressure |
|---|---|---|---|---|---|---|
| RG1 | 0.09 (p=0.7370, CI=[-0.45, 0.63]) | -0.11 (p=0.6963, CI=[-0.58, 0.50]) | 0.19 (p=0.4921, CI=[-0.36, 0.64]) | -0.15 (p=0.5717, CI=[-0.69, 0.40]) | -0.07 (p=0.7862, CI=[-0.52, 0.50]) | 0.20 (p=0.4645, CI=[-0.37, 0.71]) |
| RG2 | 0.14 (p=0.6097, CI=[-0.46, 0.67]) | -0.03 (p=0.9139, CI=[-0.49, 0.48]) | 0.03 (p=0.9139, CI=[-0.49, 0.51]) | -0.20 (p=0.4510, CI=[-0.74, 0.38]) | -0.09 (p=0.7430, CI=[-0.63, 0.53]) | 0.07 (p=0.7950, CI=[-0.52, 0.59]) |
| RG3 | -0.08 (p=0.7700, CI=[-0.60, 0.43]) | -0.31 (p=0.2352, CI=[-0.67, 0.14]) | 0.24 (p=0.3682, CI=[-0.36, 0.67]) | -0.22 (p=0.4117, CI=[-0.67, 0.32]) | -0.06 (p=0.8125, CI=[-0.61, 0.45]) | -0.01 (p=0.9741, CI=[-0.52, 0.53]) |
| RL4 | 0.06 (p=0.8118, CI=[-0.54, 0.62]) | -0.12 (p=0.6564, CI=[-0.58, 0.50]) | 0.21 (p=0.4443, CI=[-0.34, 0.62]) | -0.14 (p=0.6174, CI=[-0.61, 0.42]) | -0.03 (p=0.9012, CI=[-0.54, 0.51]) | 0.04 (p=0.8882, CI=[-0.51, 0.60]) |
| RL5 | 0.23 (p=0.3865, CI=[-0.40, 0.73]) | 0.02 (p=0.9311, CI=[-0.51, 0.61]) | 0.05 (p=0.8541, CI=[-0.52, 0.53]) | -0.05 (p=0.8541, CI=[-0.64, 0.47]) | 0.03 (p=0.9101, CI=[-0.54, 0.57]) | 0.03 (p=0.9053, CI=[-0.50, 0.58]) |
| VS1 | 0.19 (p=0.4851, CI=[-0.45, 0.71]) | -0.15 (p=0.5868, CI=[-0.62, 0.43]) | 0.14 (p=0.6020, CI=[-0.43, 0.64]) | 0.03 (p=0.9053, CI=[-0.57, 0.54]) | -0.07 (p=0.8037, CI=[-0.60, 0.45]) | 0.11 (p=0.6723, CI=[-0.47, 0.58]) |
| VTV2 | 0.36 (p=0.1723, CI=[-0.18, 0.71]) | 0.31 (p=0.2445, CI=[-0.13, 0.64]) | -0.56 (p=0.0235, CI=[-0.87, -0.04]) | 0.06 (p=0.8371, CI=[-0.47, 0.57]) | -0.39 (p=0.1321, CI=[-0.87, 0.25]) | -0.32 (p=0.2260, CI=[-0.72, 0.25]) |
| VMK3 | 0.04 (p=0.8711, CI=[-0.49, 0.59]) | -0.28 (p=0.2893, CI=[-0.65, 0.36]) | 0.15 (p=0.5868, CI=[-0.47, 0.71]) | -0.03 (p=0.9139, CI=[-0.62, 0.53]) | -0.09 (p=0.7344, CI=[-0.67, 0.51]) | -0.18 (p=0.5133, CI=[-0.57, 0.40]) |
| VP4 | 0.36 (p=0.1723, CI=[-0.13, 0.73]) | 0.35 (p=0.1839, CI=[-0.22, 0.79]) | -0.51 (p=0.0441, CI=[-0.89, 0.05]) | 0.20 (p=0.4645, CI=[-0.35, 0.74]) | -0.21 (p=0.4309, CI=[-0.72, 0.38]) | -0.33 (p=0.2128, CI=[-0.72, 0.13]) |
| VUG5 | 0.22 (p=0.4053, CI=[-0.29, 0.65]) | -0.12 (p=0.6564, CI=[-0.55, 0.42]) | -0.13 (p=0.6329, CI=[-0.63, 0.47]) | 0.04 (p=0.8968, CI=[-0.55, 0.56]) | -0.14 (p=0.6019, CI=[-0.69, 0.44]) | 0.04 (p=0.8711, CI=[-0.55, 0.61]) |
| VUI6 | 0.20 (p=0.4645, CI=[-0.35, 0.67]) | -0.11 (p=0.6803, CI=[-0.50, 0.40]) | -0.06 (p=0.8202, CI=[-0.57, 0.48]) | -0.06 (p=0.8371, CI=[-0.59, 0.52]) | -0.23 (p=0.3844, CI=[-0.69, 0.38]) | 0.09 (p=0.7452, CI=[-0.44, 0.59]) |
| VV7 | 0.37 (p=0.1612, CI=[-0.20, 0.73]) | 0.26 (p=0.3388, CI=[-0.28, 0.66]) | -0.40 (p=0.1248, CI=[-0.78, 0.10]) | 0.06 (p=0.8371, CI=[-0.57, 0.62]) | -0.27 (p=0.3172, CI=[-0.79, 0.34]) | -0.08 (p=0.7700, CI=[-0.59, 0.55]) |
| VMV8 | 0.32 (p=0.2216, CI=[-0.20, 0.72]) | 0.24 (p=0.3682, CI=[-0.28, 0.76]) | -0.55 (p=0.0263, CI=[-0.85, -0.03]) | 0.29 (p=0.2688, CI=[-0.33, 0.76]) | -0.43 (p=0.0966, CI=[-0.89, 0.14]) | -0.38 (p=0.1506, CI=[-0.71, 0.13]) |
| VTS9 | 0.21 (p=0.4443, CI=[-0.38, 0.67]) | 0.20 (p=0.4510, CI=[-0.28, 0.60]) | -0.17 (p=0.5204, CI=[-0.65, 0.32]) | -0.18 (p=0.5061, CI=[-0.64, 0.32]) | -0.27 (p=0.3172, CI=[-0.78, 0.36]) | 0.06 (p=0.8118, CI=[-0.53, 0.61]) |



© University of Venda

**Appendix 7.5:** Model Summary for MLR site-specific prediction

### OLS Regression Summary for RG1

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | RG1 | R-squared: | 0.146 |
| Model: | OLS | Adj. R-squared: | 0.097 |
| Method: | Least Squares | F-statistic: | 3.008 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 0.00937 |
| Time: | 05:50:22 | Log-Likelihood: | -335.31 |
| No. Observations: | 113 | AIC: | 684.6 |
| Df Residuals: | 106 | BIC: | 703.7 |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -219.6558 | 118.669 | -1.851 | 0.067 | -454.929 | 15.617 |
| ws | 1.4895 | 1.091 | 1.365 | 0.175 | -0.674 | 3.653 |
| wd | -0.0143 | 0.009 | -1.594 | 0.114 | -0.032 | 0.003 |
| Humidity | 0.0561 | 0.047 | 1.192 | 0.236 | -0.037 | 0.149 |
| Temp | -0.1271 | 0.166 | -0.765 | 0.446 | -0.457 | 0.202 |
| Rain | 0.0920 | 0.333 | 0.277 | 0.783 | -0.567 | 0.751 |
| Pressure | 0.2530 | 0.123 | 2.063 | 0.042 | 0.010 | 0.496 |

| | | | |
|---|---|---|---|
| Omnibus: | 45.707 | Durbin-Watson: | 1.973 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 110.181 |
| Skew: | 1.581 | Prob(JB): | 1.19e-24 |
| Kurtosis: | 6.660 | Cond. No. | 2.52e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.52e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

### OLS Regression Summary for RG2

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | RG2 | R-squared: | 0.148 |
| Model: | OLS | Adj. R-squared: | 0.102 |
| Method: | Least Squares | F-statistic: | 3.186 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 0.00638 |
| Time: | 05:50:24 | Log-Likelihood: | -443.81 |
| No. Observations: | 117 | AIC: | 901.6 |
| Df Residuals: | 110 | BIC: | 921.0 |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | 217.6878 | 341.079 | 0.638 | 0.525 | -458.251 | 893.627 |
| ws | 0.2122 | 1.814 | 0.117 | 0.907 | -3.383 | 3.807 |
| wd | -0.0397 | 0.020 | -1.976 | 0.051 | -0.079 | 0.000 |
| Humidity | -0.4492 | 0.130 | -3.464 | 0.001 | -0.706 | -0.192 |
| Temp | 0.1152 | 0.346 | 0.333 | 0.740 | -0.571 | 0.801 |
| Rain | 0.5688 | 0.736 | 0.773 | 0.441 | -0.890 | 2.028 |
| Pressure | -0.1703 | 0.355 | -0.480 | 0.632 | -0.874 | 0.533 |

| | | | |
|---|---|---|---|
| Omnibus: | 146.119 | Durbin-Watson: | 2.051 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 3480.256 |
| Skew: | 4.594 | Prob(JB): | 0.00 |
| Kurtosis: | 28.089 | Cond. No. | 3.19e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.19e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

### OLS Regression Summary for RG3

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | RG3 | R-squared: | 0.189 |
| Model: | OLS | Adj. R-squared: | 0.158 |
| Method: | Least Squares | F-statistic: | 6.115 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 8.85e-06 |
| Time: | 05:50:26 | Log-Likelihood: | -457.53 |
| No. Observations: | 164 | AIC: | 929.1 |
| Df Residuals: | 157 | BIC: | 950.8 |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -217.6577 | 79.896 | -2.724 | 0.007 | -375.467 | -59.849 |
| ws | -0.8541 | 0.610 | -1.399 | 0.164 | -2.060 | 0.351 |
| wd | -0.0035 | 0.006 | -0.548 | 0.584 | -0.016 | 0.009 |
| Humidity | -0.0419 | 0.029 | -1.424 | 0.156 | -0.100 | 0.016 |
| Temp | -0.1915 | 0.100 | -1.909 | 0.058 | -0.390 | 0.007 |
| Rain | 0.1734 | 0.250 | 0.694 | 0.489 | -0.320 | 0.667 |
| Pressure | 0.2625 | 0.083 | 3.179 | 0.002 | 0.099 | 0.426 |

| | | | |
|---|---|---|---|
| Omnibus: | 33.782 | Durbin-Watson: | 1.672 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 52.279 |
| Skew: | 1.079 | Prob(JB): | 4.44e-12 |
| Kurtosis: | 4.730 | Cond. No. | 2.46e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.46e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

### OLS Regression Summary for RL4

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | RL4 | R-squared: | 0.112 |
| Model: | OLS | Adj. R-squared: | 0.072 |
| Method: | Least Squares | F-statistic: | 2.807 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 0.0133 |
| Time: | 05:50:28 | Log-Likelihood: | -395.64 |
| No. Observations: | 140 | AIC: | 805.3 |
| Df Residuals: | 133 | BIC: | 825.9 |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -244.0656 | 116.075 | -2.103 | 0.037 | -473.657 | -14.474 |
| ws | -0.6202 | 0.645 | -0.961 | 0.338 | -1.897 | 0.656 |
| wd | -0.0057 | 0.007 | -0.843 | 0.401 | -0.019 | 0.008 |
| Humidity | -0.0752 | 0.037 | -2.014 | 0.046 | -0.149 | -0.001 |
| Temp | -0.0614 | 0.122 | -0.504 | 0.615 | -0.303 | 0.180 |
| Rain | -0.1558 | 0.265 | -0.589 | 0.557 | -0.679 | 0.367 |
| Pressure | 0.2885 | 0.121 | 2.388 | 0.018 | 0.050 | 0.527 |

| | | | |
|---|---|---|---|
| Omnibus: | 77.456 | Durbin-Watson: | 1.858 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 408.467 |
| Skew: | 1.945 | Prob(JB): | 2.01e-89 |
| Kurtosis: | 10.409 | Cond. No. | 3.15e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.15e+05. This might indicate that there are
strong multicollinearity or other numerical problems.



© University of Venda

## OLS Regression Summary for RL5

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | RL5 | R-squared: | 0.089 |
| Model: | OLS | Adj. R-squared: | 0.058 |
| Method: | Least Squares | F-statistic: | 2.851 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 0.0113 |
| Time: | 05:50:30 | Log-Likelihood: | -523.08 |
| No. Observations: | 182 | AIC: | 1060. |
| Df Residuals: | 175 | BIC: | 1083. |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -86.9423 | 97.909 | -0.888 | 0.376 | -280.177 | 106.292 |
| ws | -0.2402 | 0.591 | -0.406 | 0.685 | -1.407 | 0.927 |
| wd | -0.0060 | 0.006 | -1.042 | 0.299 | -0.017 | 0.005 |
| Humidity | -0.0388 | 0.028 | -1.381 | 0.169 | -0.094 | 0.017 |
| Temp | -0.2191 | 0.119 | -1.847 | 0.066 | -0.453 | 0.015 |
| Rain | 0.3796 | 0.302 | 1.259 | 0.210 | -0.216 | 0.975 |
| Pressure | 0.1229 | 0.101 | 1.211 | 0.227 | -0.077 | 0.323 |

| | | | |
|---|---|---|---|
| Omnibus: | 44.456 | Durbin-Watson: | 1.884 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 72.997 |
| Skew: | 1.276 | Prob(JB): | 1.41e-16 |
| Kurtosis: | 4.766 | Cond. No. | 2.92e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.92e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

## OLS Regression Summary for VS1

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | VS1 | R-squared: | 0.126 |
| Model: | OLS | Adj. R-squared: | 0.085 |
| Method: | Least Squares | F-statistic: | 3.082 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 0.00750 |
| Time: | 05:50:31 | Log-Likelihood: | -402.63 |
| No. Observations: | 135 | AIC: | 819.3 |
| Df Residuals: | 128 | BIC: | 839.6 |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -70.0569 | 117.127 | -0.598 | 0.551 | -301.813 | 161.699 |
| ws | -0.3342 | 0.790 | -0.423 | 0.673 | -1.897 | 1.229 |
| wd | 0.0021 | 0.007 | 0.285 | 0.776 | -0.012 | 0.017 |
| Humidity | -0.1056 | 0.042 | -2.485 | 0.014 | -0.190 | -0.022 |
| Temp | -0.3295 | 0.136 | -2.415 | 0.017 | -0.599 | -0.060 |
| Rain | -0.1647 | 0.280 | -0.588 | 0.558 | -0.719 | 0.390 |
| Pressure | 0.1148 | 0.122 | 0.941 | 0.348 | -0.127 | 0.356 |

| | | | |
|---|---|---|---|
| Omnibus: | 44.720 | Durbin-Watson: | 1.865 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 100.963 |
| Skew: | 1.378 | Prob(JB): | 1.19e-22 |
| Kurtosis: | 6.218 | Cond. No. | 2.68e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.68e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

## OLS Regression Summary for VTV2

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | VTV2 | R-squared: | 0.157 |
| Model: | OLS | Adj. R-squared: | 0.125 |
| Method: | Least Squares | F-statistic: | 4.956 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 0.000111 |
| Time: | 05:50:33 | Log-Likelihood: | -570.46 |
| No. Observations: | 167 | AIC: | 1155. |
| Df Residuals: | 160 | BIC: | 1177. |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -120.5543 | 149.158 | -0.808 | 0.420 | -415.127 | 174.019 |
| ws | -1.3422 | 1.197 | -1.121 | 0.264 | -3.707 | 1.023 |
| wd | 0.0275 | 0.009 | 2.939 | 0.004 | 0.009 | 0.046 |
| Humidity | -0.2045 | 0.055 | -3.723 | 0.000 | -0.313 | -0.096 |
| Temp | -0.1605 | 0.222 | -0.723 | 0.471 | -0.599 | 0.278 |
| Rain | -0.2208 | 0.536 | -0.412 | 0.681 | -1.279 | 0.838 |
| Pressure | 0.1762 | 0.155 | 1.135 | 0.258 | -0.130 | 0.483 |

| | | | |
|---|---|---|---|
| Omnibus: | 80.434 | Durbin-Watson: | 2.084 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 412.633 |
| Skew: | 1.736 | Prob(JB): | 2.50e-90 |
| Kurtosis: | 9.873 | Cond. No. | 2.48e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.48e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

## OLS Regression Summary for VMK3

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | VMK3 | R-squared: | 0.048 |
| Model: | OLS | Adj. R-squared: | 0.019 |
| Method: | Least Squares | F-statistic: | 1.676 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 0.129 |
| Time: | 05:50:34 | Log-Likelihood: | -600.64 |
| No. Observations: | 206 | AIC: | 1215. |
| Df Residuals: | 199 | BIC: | 1239. |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -69.2696 | 86.877 | -0.797 | 0.426 | -240.587 | 102.048 |
| ws | -0.2577 | 0.542 | -0.476 | 0.635 | -1.326 | 0.810 |
| wd | -0.0026 | 0.006 | -0.449 | 0.654 | -0.014 | 0.009 |
| Humidity | -0.0535 | 0.028 | -1.919 | 0.056 | -0.108 | 0.001 |
| Temp | -0.1112 | 0.111 | -0.998 | 0.320 | -0.331 | 0.109 |
| Rain | 0.1142 | 0.345 | 0.331 | 0.741 | -0.566 | 0.794 |
| Pressure | 0.1040 | 0.090 | 1.161 | 0.247 | -0.073 | 0.281 |

| | | | |
|---|---|---|---|
| Omnibus: | 85.655 | Durbin-Watson: | 1.891 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 316.271 |
| Skew: | 1.680 | Prob(JB): | 2.10e-69 |
| Kurtosis: | 8.055 | Cond. No. | 2.65e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.65e+05. This might indicate that there are
strong multicollinearity or other numerical problems.



© University of Venda

## OLS Regression Summary for VP4

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | VP4 | R-squared: | 0.485 |
| Model: | OLS | Adj. R-squared: | 0.462 |
| Method: | Least Squares | F-statistic: | 21.62 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 8.29e-18 |
| Time: | 05:50:36 | Log-Likelihood: | -452.93 |
| No. Observations: | 145 | AIC: | 919.9 |
| Df Residuals: | 138 | BIC: | 940.7 |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -399.2649 | 136.030 | -2.935 | 0.004 | -668.238 | -130.292 |
| ws | -1.2248 | 0.921 | -1.330 | 0.186 | -3.045 | 0.596 |
| wd | 0.0312 | 0.010 | 3.258 | 0.001 | 0.012 | 0.050 |
| Humidity | -0.2576 | 0.046 | -5.543 | 0.000 | -0.350 | -0.166 |
| Temp | -0.3556 | 0.147 | -2.423 | 0.017 | -0.646 | -0.065 |
| Rain | 0.1694 | 0.345 | 0.491 | 0.624 | -0.512 | 0.851 |
| Pressure | 0.4728 | 0.142 | 3.338 | 0.001 | 0.193 | 0.753 |

| | | | |
|---|---|---|---|
| Omnibus: | 17.038 | Durbin-Watson: | 1.689 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 26.030 |
| Skew: | 0.615 | Prob(JB): | 2.23e-06 |
| Kurtosis: | 4.673 | Cond. No. | 2.80e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.8e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

## OLS Regression Summary for VUG5

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | VUG5 | R-squared: | 0.304 |
| Model: | OLS | Adj. R-squared: | 0.272 |
| Method: | Least Squares | F-statistic: | 9.534 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 1.10e-08 |
| Time: | 05:50:38 | Log-Likelihood: | -366.11 |
| No. Observations: | 138 | AIC: | 746.2 |
| Df Residuals: | 131 | BIC: | 766.7 |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -222.5015 | 95.636 | -2.327 | 0.022 | -411.693 | -33.310 |
| ws | -0.4655 | 0.547 | -0.851 | 0.396 | -1.548 | 0.617 |
| wd | 0.0109 | 0.005 | 2.013 | 0.046 | 0.000 | 0.022 |
| Humidity | -0.0805 | 0.035 | -2.324 | 0.022 | -0.149 | -0.012 |
| Temp | -0.3316 | 0.099 | -3.333 | 0.001 | -0.528 | -0.135 |
| Rain | 0.2084 | 0.225 | 0.925 | 0.357 | -0.237 | 0.654 |
| Pressure | 0.2704 | 0.100 | 2.717 | 0.007 | 0.074 | 0.467 |

| | | | |
|---|---|---|---|
| Omnibus: | 1.903 | Durbin-Watson: | 1.754 |
| Prob(Omnibus): | 0.386 | Jarque-Bera (JB): | 1.789 |
| Skew: | 0.277 | Prob(JB): | 0.409 |
| Kurtosis: | 2.943 | Cond. No. | 3.06e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.06e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

## OLS Regression Summary for VUI6

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | VUI6 | R-squared: | 0.242 |
| Model: | OLS | Adj. R-squared: | 0.186 |
| Method: | Least Squares | F-statistic: | 4.308 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 0.000806 |
| Time: | 05:50:39 | Log-Likelihood: | -236.16 |
| No. Observations: | 88 | AIC: | 486.3 |
| Df Residuals: | 81 | BIC: | 503.7 |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -238.1491 | 119.389 | -1.995 | 0.049 | -475.695 | -0.603 |
| ws | -0.8950 | 0.831 | -1.077 | 0.285 | -2.548 | 0.758 |
| wd | 0.0086 | 0.007 | 1.270 | 0.208 | -0.005 | 0.022 |
| Humidity | -0.1078 | 0.049 | -2.219 | 0.029 | -0.204 | -0.011 |
| Temp | -0.2737 | 0.138 | -1.978 | 0.051 | -0.549 | 0.002 |
| Rain | 0.1787 | 0.286 | 0.625 | 0.534 | -0.390 | 0.748 |
| Pressure | 0.2909 | 0.125 | 2.323 | 0.023 | 0.042 | 0.540 |

| | | | |
|---|---|---|---|
| Omnibus: | 2.406 | Durbin-Watson: | 1.747 |
| Prob(Omnibus): | 0.300 | Jarque-Bera (JB): | 1.917 |
| Skew: | 0.137 | Prob(JB): | 0.384 |
| Kurtosis: | 3.669 | Cond. No. | 2.92e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.92e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

## OLS Regression Summary for VV7

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | VV7 | R-squared: | 0.293 |
| Model: | OLS | Adj. R-squared: | 0.253 |
| Method: | Least Squares | F-statistic: | 7.446 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 1.10e-06 |
| Time: | 05:50:41 | Log-Likelihood: | -285.70 |
| No. Observations: | 115 | AIC: | 585.4 |
| Df Residuals: | 108 | BIC: | 604.6 |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -185.9864 | 92.312 | -2.015 | 0.046 | -368.965 | -3.007 |
| ws | -0.9792 | 0.470 | -2.083 | 0.040 | -1.911 | -0.048 |
| wd | 0.0096 | 0.005 | 1.778 | 0.078 | -0.001 | 0.020 |
| Humidity | -0.0675 | 0.036 | -1.854 | 0.066 | -0.140 | 0.005 |
| Temp | -0.2572 | 0.098 | -2.615 | 0.010 | -0.452 | -0.062 |
| Rain | 0.4302 | 0.201 | 2.136 | 0.035 | 0.031 | 0.830 |
| Pressure | 0.2311 | 0.097 | 2.374 | 0.019 | 0.038 | 0.424 |

| | | | |
|---|---|---|---|
| Omnibus: | 16.274 | Durbin-Watson: | 2.120 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 18.366 |
| Skew: | 0.931 | Prob(JB): | 0.000103 |
| Kurtosis: | 3.607 | Cond. No. | 3.17e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.17e+05. This might indicate that there are
strong multicollinearity or other numerical problems.



© University of Venda

## OLS Regression Summary for VMV8

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | VMV8 | R-squared: | 0.340 |
| Model: | OLS | Adj. R-squared: | 0.310 |
| Method: | Least Squares | F-statistic: | 11.09 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 5.96e-10 |
| Time: | 05:50:42 | Log-Likelihood: | -385.27 |
| No. Observations: | 136 | AIC: | 784.5 |
| Df Residuals: | 129 | BIC: | 804.9 |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -119.0596 | 117.582 | -1.013 | 0.313 | -351.699 | 113.580 |
| ws | -0.7317 | 0.650 | -1.126 | 0.262 | -2.017 | 0.554 |
| wd | 0.0194 | 0.007 | 2.767 | 0.006 | 0.006 | 0.033 |
| Humidity | -0.1941 | 0.041 | -4.786 | 0.000 | -0.274 | -0.114 |
| Temp | -0.3091 | 0.117 | -2.631 | 0.010 | -0.542 | -0.077 |
| Rain | 0.0483 | 0.283 | 0.171 | 0.865 | -0.511 | 0.607 |
| Pressure | 0.1711 | 0.122 | 1.397 | 0.165 | -0.071 | 0.413 |

| | | | |
|---|---|---|---|
| Omnibus: | 39.750 | Durbin-Watson: | 1.708 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 75.657 |
| Skew: | 1.304 | Prob(JB): | 3.73e-17 |
| Kurtosis: | 5.559 | Cond. No. | 3.12e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.12e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

## OLS Regression Summary for VTS9

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | VTS9 | R-squared: | 0.352 |
| Model: | OLS | Adj. R-squared: | 0.333 |
| Method: | Least Squares | F-statistic: | 18.35 |
| Date: | Fri, 03 Jan 2025 | Prob (F-statistic): | 5.35e-17 |
| Time: | 05:50:44 | Log-Likelihood: | -664.41 |
| No. Observations: | 210 | AIC: | 1343. |
| Df Residuals: | 203 | BIC: | 1366. |
| Df Model: | 6 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -178.9313 | 122.723 | -1.458 | 0.146 | -420.907 | 63.044 |
| ws | -0.5058 | 0.737 | -0.686 | 0.493 | -1.960 | 0.948 |
| wd | 0.0072 | 0.007 | 1.007 | 0.315 | -0.007 | 0.021 |
| Humidity | -0.2416 | 0.035 | -6.886 | 0.000 | -0.311 | -0.172 |
| Temp | -0.5863 | 0.149 | -3.921 | 0.000 | -0.881 | -0.291 |
| Rain | 0.2562 | 0.334 | 0.767 | 0.444 | -0.402 | 0.915 |
| Pressure | 0.2476 | 0.127 | 1.953 | 0.052 | -0.002 | 0.498 |

| | | | |
|---|---|---|---|
| Omnibus: | 37.690 | Durbin-Watson: | 1.919 |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 57.526 |
| Skew: | 1.009 | Prob(JB): | 3.22e-13 |
| Kurtosis: | 4.581 | Cond. No. | 2.94e+05 |

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.94e+05. This might indicate that there are
strong multicollinearity or other numerical problems.



© University of Venda

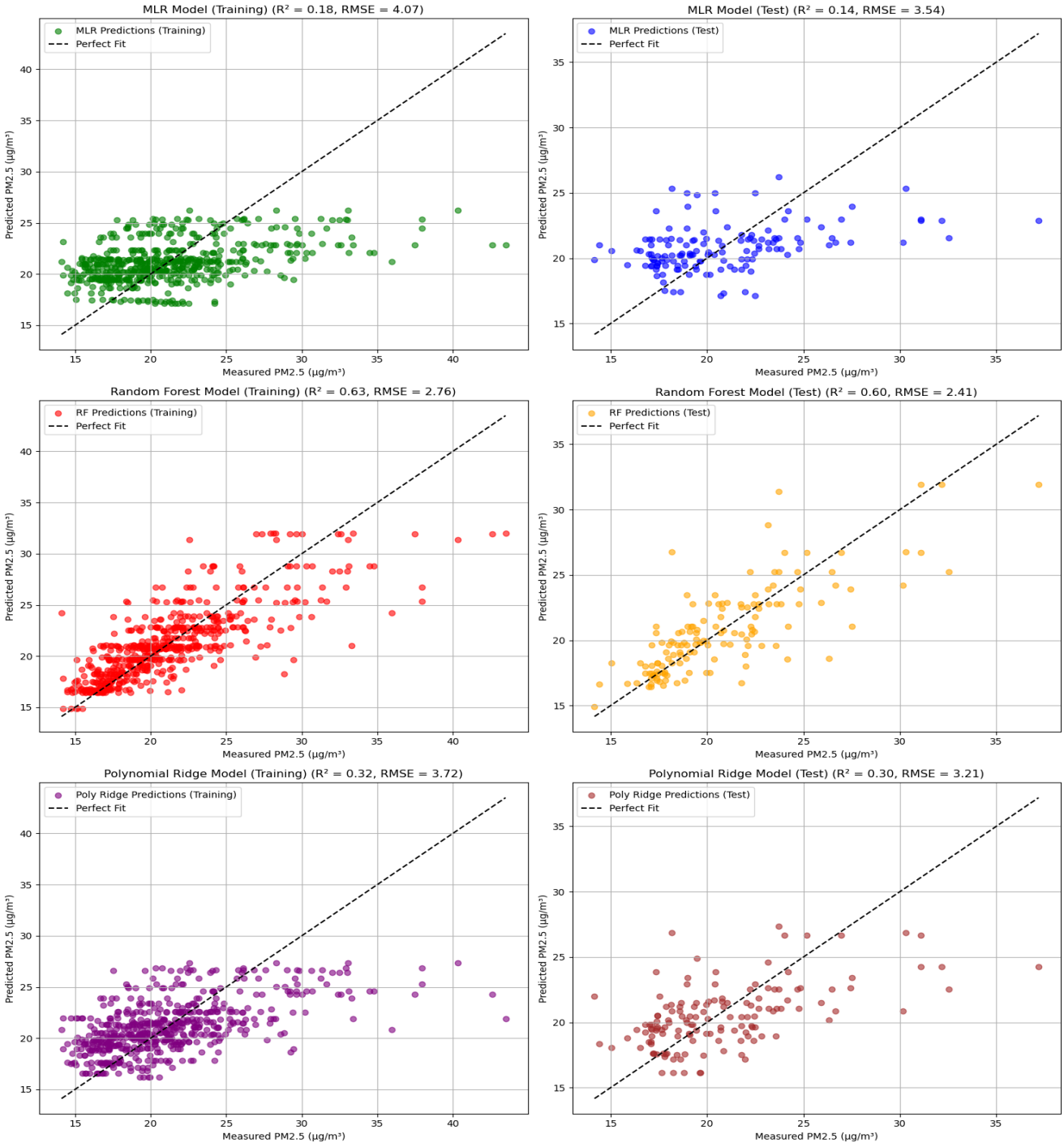
**Appendix 7.6:** Performance evaluation per site for the random_forest_models with_hyperparameters for daily ambient $PM_{2.5}$ ($\mu g/m^3$).

| | Model-Based | | | | | | | | 5-fold CV | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Training | | | | Testing | | | | | |
| Site | RMSE_ Train | $R^2$_Train | MSE_Train | SSE_Train | RMSE_Test | $R^2$_Test | MSE_Test | SSE_Test | Mean CV RMSE | Std CV RMSE |
| RG1 | 3.166 | 0.613 | 10.026 | 1132.947 | 4.234 | 0.326 | 17.923 | 519.755 | 5.478 | 1.489 |
| RG2 | 8.316 | 0.490 | 69.160 | 8091.636 | 3.068 | -0.027 | 9.415 | 282.446 | 8.890 | 7.478 |
| RG3 | 2.546 | 0.661 | 6.481 | 1062.809 | 2.614 | 0.393 | 6.834 | 287.011 | 4.209 | 0.849 |
| RL4 | 2.973 | 0.530 | 8.840 | 1237.618 | 4.293 | -0.044 | 18.430 | 663.497 | 4.716 | 1.711 |
| RL5 | 2.647 | 0.652 | 7.009 | 1275.647 | 2.961 | 0.678 | 8.766 | 403.231 | 4.270 | 0.764 |
| VS1 | 3.075 | 0.638 | 9.457 | 1276.664 | 4.490 | -0.081 | 20.159 | 685.404 | 5.262 | 1.566 |
| VTV2 | 4.833 | 0.637 | 23.354 | 3900.094 | 6.256 | 0.342 | 39.133 | 1643.600 | 8.048 | 1.413 |
| VMK3 | 2.833 | 0.617 | 8.025 | 1653.241 | 6.601 | 0.276 | 43.572 | 2265.722 | 5.627 | 1.828 |
| VP4 | 3.402 | 0.803 | 11.571 | 1677.773 | 5.062 | 0.593 | 25.627 | 948.259 | 6.251 | 1.922 |
| VUG5 | 2.064 | 0.749 | 4.259 | 587.758 | 3.986 | 0.218 | 15.888 | 556.073 | 4.041 | 0.841 |
| VUI6 | 2.182 | 0.712 | 4.762 | 419.059 | 4.168 | 0.604 | 17.372 | 382.181 | 4.439 | 0.863 |
| VV7 | 1.744 | 0.745 | 3.040 | 349.625 | 3.352 | -0.138 | 11.235 | 325.822 | 3.296 | 1.195 |
| VMV8 | 2.860 | 0.681 | 8.182 | 1112.794 | 5.027 | 0.426 | 25.269 | 859.134 | 4.557 | 0.758 |
| VTS9 | 3.290 | 0.786 | 10.820 | 2272.303 | 4.777 | 0.484 | 22.815 | 1209.183 | 6.398 | 0.919 |



© University of Venda

**Appendix 7.7:** Scatter plots comparing the training and test performance of Multiple Linear Regression (MLR), Random Forest (RF), and Polynomial Ridge Regression (PRR) models for $PM_{2.5}$ concentration predictions. The plots on the left side represent training performance, while the plots on the right depict test performance.






© University of Venda

University of Venda
Creating Future Leaders

**Appendix 7.8:** Policy recommendations for air quality improvement: Lessons from air pollution tolerance index (APTI) and anticipated performance index (API) studies

atmosphere

MDPI

*Review*

# A Review of Air Pollution Mitigation Approach Using Air Pollution Tolerance Index (APTI) and Anticipated Performance Index (API)

**Ibironke Titilayo Enitan [1,*], Olatunde Samod Durowoju [2], Joshua Nosa Edokpayi [2] and John Ogony Odiyo [3]**

[1] Department of Geography and Environmental Science, University of Venda, Private Bag X5050, Thohoyandou 0950, South Africa
[2] Department of Earth Sciences, University of Venda, Private Bag X5050, Thohoyandou 0950, South Africa; olatunde.durowoju@univen.ac.za (O.S.D.); joshua.edokpayi@univen.ac.za (J.N.E.)
[3] Office of The DVC Research Innovation Commercialisation and Internationalisation (RICI), Vaal University of Technology, Vanderbijlpark 1901, South Africa; john.odiyo@univen.ac.za
* Correspondence: ronkmodel2003@gmail.com; Tel.: +27-071-073-6656

**Abstract:** Air pollution is a global environmental issue, and there is an urgent need for sustainable remediation techniques. Thus, phytoremediation has become a popular approach to air pollution remediation. This paper reviewed 28 eco-friendly indigenous plants based on both the air pollution tolerance index (APTI) and anticipated performance index (API), using tolerance level and performance indices to evaluate the potential of most indigenous plant species for air pollution control. The estimated APTI ranged from 4.79 (*Syzygium malaccense*) to 31.75 (*Psidium guajava*) among the studied indigenous plants. One of the selected plants is tolerant, and seven (7) are intermediate to air pollution with their APTI in the following order: *Psidium guajava* (31.75) > *Swietenia mahogany* (28.08) > *Mangifera indica* L. (27.97) > *Ficus infectoria* L. (23.93) > *Ficus religiosa* L. (21.62) > *Zizyphus Oenoplia Mill* (20.06) > *Azadirachta indica A. Juss.* (19.01) > *Ficus benghalensis* L. (18.65). Additionally, the API value indicated that *Mangifera indica L.* ranges from best to good performer; *Ficus religiosa* L. and *Azadirachta indica A. Juss.* from excellent to moderate performers; and *Cassia fistula* L. from poor to very poor performer for air pollution remediation. The Pearson correlation shows that there is a positive correlation between API and APTI (R2 = 0.63), and this implies that an increase in APTI increases the API and vice versa. This paper shows that *Mangifera indica* L., *Ficus religiosa* L., and *Azadirachta indica A. Juss.* have good potential for sustainable reduction in air pollution for long-term management and green ecomanagement development.

**Keywords:** air pollution; API; APTI; indigenous plant; environmental pollution; sustainable environmental–pollution nexus; green ecomanagement development

check for updates

**Citation:** Enitan, I.T.; Durowoju, O.S.; Edokpayi, J.N.; Odiyo, J.O. A Review of Air Pollution Mitigation Approach Using Air Pollution Tolerance Index (APTI) and Anticipated Performance Index (API). *Atmosphere* **2022**, *13*, 374. https://doi.org/10.3390/atmos13030374

Academic Editor: Roberto Bellasio

Received: 30 December 2021
Accepted: 19 February 2022
Published: 23 February 2022



© University of Venda