GEOM90006 - Spatial Data Analytics

Project Proposal

Group 3

Mingjun Xie (1644304)
Nikki Mo (1423101)
Sarathi Thirumalai Soundararajan (1410995)

Semester 1, 2025

May 14, 2025

Contents

1 Literature Review

There has been a growing concern about air quality issues in recent years. Air pollution poses a significant threat to human health, especially particulate matter ($PM_{2.5}$). Borchers-Arriagada et al. (2024) indicate that most diseases, such as cardiovascular and respiratory diseases, are caused by long-term exposure to high concentrations of $PM_{2.5}$. Although the overall air quality is better in Australia than in other countries, Environment Protection Authority Victoria (2018) pointed out that $PM_{2.5}$ is still the pollutant of greatest concern. The main sources of $PM_{2.5}$ are from natural and anthropogenic sources, such as vehicle exhaust, industrial emissions, dust, and forest fires (Environment Protection Authority Victoria, 2018).

In particular, urban and industrial areas are where its concentration is affected by multiple factors, including high traffic density, dense buildings, and concentrated energy usage. Knibbs et al. (2018) found that $PM_{2.5}$ exposure has obvious spatial heterogeneity, and densely populated urban areas face higher exposure levels based on satellite data and a land use regression model. de Jesus et al. (2020) also found that $PM_{2.5}$ has an upward trend in many Australian capital cities, a trend that is closely related to urban population growth.

International research shows that trees and urban green space can effectively reduce the concentration of $PM_{2.5}$ in the air through blocking and sedimentation effects (Nowak et al., 2006). This research has proved that the density of higher urban vegetation areas tends to have low concentrations of $PM_{2.5}$, especially in traffic-intensive areas.

Knibbs et al. (2018) used 10 km resolution SAT-$PM_{2.5}$ combined with land use data to construct a national-scale exposure model, which successfully reflected the $PM_{2.5}$ exposure levels in the urban-rural transition zone. This method provides strong technical support for the spatial analysis of the relationship between $PM_{2.5}$, population density, and vegetation cover at the SA2 level in this study. However, understanding the spatial relationship between $PM_{2.5}$ and population density and vegetation cover will help formulate more targeted pollution control strategies and provide a scientific basis for urban greening planning.

2 Topic and Research Question

Our group focuses on studying the spatial relationship of $PM_{2.5}$ air quality between urban and rural areas in Victoria, Australia. $PM_{2.5}$ in high concentrations can pose significant risks to human health. Understanding its spatial distribution is essential for public health, environmental management, and sustainable urban planning.

Our primary research question is "How does population density relate to $PM_{2.5}$ air quality?" We aim to explore whether urban areas with higher population density, especially Greater Melbourne, exhibit elevated $PM_{2.5}$ concentrations compared to rural regions. Additionally, we seek to investigate "Is there a spatial relationship between vegetation cover and $PM_{2.5}$ concentrations in Victoria?" This question assesses whether areas with greater vegetation coverage are associated with lower air pollution. Through spatial analysis, our project provides insights into potential mitigating factors that influence air quality in Victoria.

3 Project Scope

This project explores how population density and vegetation cover relate to $PM_{2.5}$ concentrations between selected urban and rural areas. We focus on the state of Victoria, Australia, specifically including the suburbs of Alphington, Footscray, Churchill, Moe, Morwell East, Morwell South, and Traralgon. These locations represent a meaningful contrast between Greater Melbourne (urban) and the LaTrobe Valley Region (rural).

The analysis is conducted at the SA2 spatial resolution, which aligns well with the geographic units of the air quality monitoring sites and the Australian Bureau of Statistics (ABS) 2016 Census boundaries. To ensure temporal consistency, the datasets required to examine the spatial relationship are from 2016.

4 Data Sources

The prime air quality dataset contains point-based hourly average $PM_{2.5}$ measurements from 7 monitoring stations (Environment Protection Authority Victoria, 2019), while the population dataset from the ABS provides population density figures at the SA2 level (Australian Bureau of Statistics, 2017). These two data directly support the first research question. The third dataset is a raster layer representing native vegetation cover across Victoria, provided by (University of Melbourne, 2022). With a resolution of $250\,\mathrm{m} \times 250\,\mathrm{m}$, it will then be spatially aggregated to the SA2 level to quantify vegetation coverage per area and examine its association with $PM_{2.5}$ concentrations.

5 Methodology and Hypothesis

The analysis will begin with data cleaning and reprojection of all datasets to the target coordinate reference system. Followed by handling missing values and subsetting the targeted study area, Exploratory Spatial Data Analysis (ESDA) will be conducted to summarise $PM_{2.5}$, population density, and vegetation cover across SA2 regions, along with plots to show distribution.

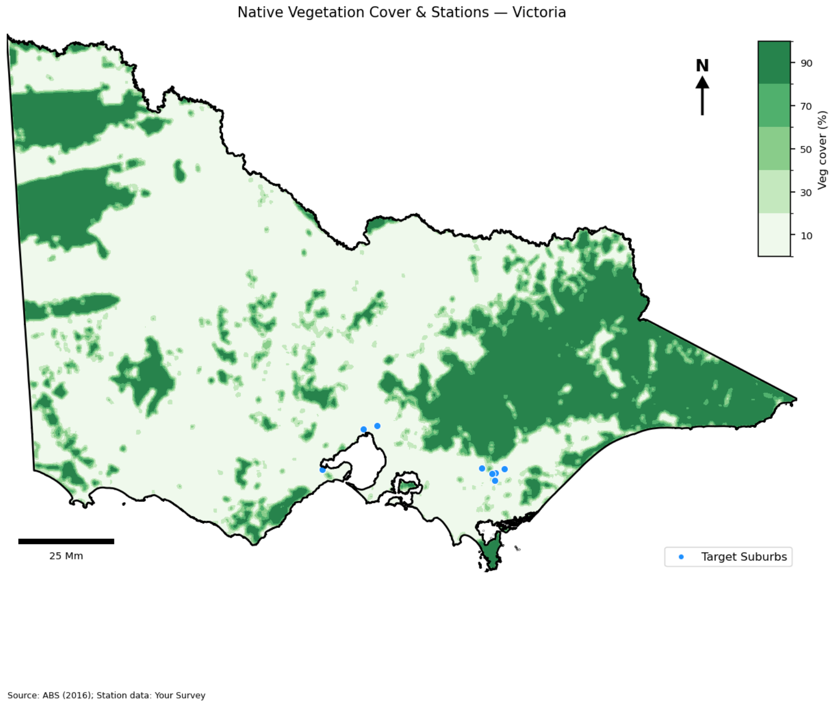
In the next part, SA2 regions will be classified as urban or rural using ABS population density thresholds for grouped analysis. Zonal statistics will also be performed on the vegetation raster to calculate average cover per region. Geographically Weighted Regression (GWR) may be applied to explore how relationships vary spatially. Optionally, Local Moran's I may be computed to identify spatial clusters of $PM_{2.5}$. Python libraries such as Pandas, GeoPandas, Rasterio, and Matplotlib will be used to perform the aforementioned analyses.

This project aims to test the following hypotheses:

H1: $PM_{2.5}$ is higher in urban areas than in rural areas.
H2: $PM_{2.5}$ increases with population density.
H3: $PM_{2.5}$ decreases with higher vegetation cover.

Additional datasets may be sourced if needed to support the spatial analysis.

6 Simple Cartographic-Quality Map



7 Tentative Timeline

Week	Time	Task Description
W1	13 May – 19 May	• Define research questions and variables • Collect and clean datasets
W2	20 May – 26 May	• Perform ESDA • Map and describe variable distributions
W3	27 May – 2 June	• Aggregate vegetation to SA2 • Plot histograms and boxplots
W4	3 June – 9 June	• Compare urban vs rural • Run zonal stats and GWR
W5	10 June – 14 June	• Interpret findings • Draft conclusions and final report

Table 1: Project Timeline

References

Australian Bureau of Statistics (2017). 2016 census of population and housing – general community profile, victoria (sa2), table g01: Selected person characteristics. https://www.abs.gov.au/.

Borchers-Arriagada, N. et al. (2024). Daily $pm_{2.5}$ and seasonal-trend decomposition to identify extreme air pollution events from 2001 to 2020 for continental australia using a random forest model. *Atmosphere*, 15:1341.

de Jesus, A. L., Thompson, H., Knibbs, L. D., Hanigan, I., De Torres, L., Fisher, G., Berko, H., and Morawska, L. (2020). Two decades of trends in urban particulate matter concentrations across australia. *Environmental Research*, 190:110021.

Environment Protection Authority Victoria (2018). Air pollution in victoria: A summary of the state of knowledge (publication no. 1709). https://www.epa.vic.gov.au/.

Environment Protection Authority Victoria (2019). Air quality readings – 2016. https://tinyurl.com/epa-vic-data.

Knibbs, L. D. et al. (2018). Satellite-based land-use regression for continental-scale long-term ambient $pm_{2.5}$ exposure assessment in australia. *Environmental Science & Technology*, 52(21):12445–12455.

Nowak, D. J., Crane, D. E., and Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the united states. *Urban Forestry & Urban Greening*, 4(3–4):115–123.

University of Melbourne (2022). Native vegetation cover raster layer, victoria, australia (2016) [data set]. https://doi.org/10.26188/18095729.v1.