School of Design and the Built Environment

Assessing sustainable development in industrial regions towards smart built environment management using Earth observation big data

Zehua Zhang

ORCID: 0000-0003-3462-4025

This thesis is presented for the Degree of

Doctor of Philosophy

of

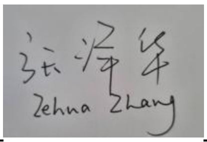
Curtin University

February 2023

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledge has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: ______

Date: ____25-Feb-2023______

Supervisor: **Dr. Yongze Song**, Curtin University

Co-supervisor: **Prof. Peng Wu**, Curtin University

Abstract

The development of geographic information system (GIS) and spatial methods for sustainable built environment planning is an important issue in the construction management and spatial science fields. The application of GIS and spatial statistics to Earth observation data can provide information and help advise guidelines for sustainable built environment design and management at a macro level. Moreover, relevant spatial analysis results are valuable to urban and regional governance for designing and maintaining phases of the construction management lifecycle. Thus, innovation in spatial statistics methods and improvements in GIS applications can stimulate built environment management efficiency and support smart urban design. Currently, the properties of the built environment of industrial regions supporting mining, manufacturing, and utility supply and waste services require more research efforts, considering their significant role in the national economy, environmental degradation, and potential social costs. In Australia, these three industries contribute to approximately 25% of the national GDP and over 95% of air pollutant emissions, including nitrogen dioxide and sulphur dioxide, as reported by the National Pollutant Inventory (NPI). Thus, it is necessary to monitor and assess the development of the built environment, including specific regions and infrastructures supporting these three key industries, for smart built environment management and sustainable city design purposes.

Currently, sustainability assessments, smart built environment management, and smart cities can be supported by Earth observation big data and spatial analysis methods. Earth observation data reveal various remote sensing (RS) properties on Earth's surface, and these properties are indicators of sustainable development. Spatial analysis methods are tools that explore the values of Earth observation. Spatial methods have been applied to sustainability assessment and scientific decision-making for urban construction and the built environment, despite some

drawbacks. Some of the drawbacks are from spatial methodology design. The geographical detector is an innovative method quantifying the association among factors, which has been applied in environmental assessment and construction sustainability analysis fields. However, this geospatial method is sensitive to statistical distribution of variables, and an improved method is needed for general cases. Furthermore, some of the spatial phenomena are not simple and may be related to the status of complex. Thus, it is also beneficial to extend the understanding of complexity from current spatial theories. The other drawbacks are from the application of Earth observation data. The capacity of Earth observation data to represent remote sensing properties of industrial lands at a continental level remains to be shown. Considering the need for industrial sustainability analysis, the capability of Earth observation data, and current geospatial method limitations, this study was designed to assess the sustainable development of nationwide industrial regions toward smart built environment management using Earth observation big data. This study comprised four key steps to assess industrial sustainability, from environmental and socio-economic perspectives, to add value to industrial built environment management.

First, the environmental sustainability of nationwide industrial regions was assessed by exploring the spatial disparity of the factors affecting air pollutants. This step involved a pilot study on environmental sustainability and laid the foundation for the subsequent steps. Geographical boundaries of industrial regions were identified using a spatial method, and RS and spatial properties of these regions were accessed from multiple sources. Spatial disparities in factors affecting air pollutants in industrial regions were analysed by geographically weighted regression with standardised coefficients. The spatial patterns were further determined at a higher level of spatial granularity. Although road density and meteorological factors are influential in general, the spatial disparity of factors determining air pollutants in nationwide industrial regions is evident. In major cities, natural factors dominate changes in air pollutants. However, human activities are more relevant to the air pollutant density in remote areas.

Second, an innovative spatial method was developed to support industrial sustainability analysis and built environment management. A robust geographic detector (RGD) was developed to overcome the limitations of previous spatially stratified heterogeneity methods. By introducing an optimisation algorithm to determine spatial zones, a RGD can analyse spatial data with accurate and robust results. This new spatial method was further applied to indicate the spatial association between air pollutants and influential factors in industrial regions. The RGD indicated that human activities and meteorological factors were strongly associated with air pollutant densities in industrial regions. The design of this new spatial method can be further applied to various studies involving spatially stratified heterogeneity.

Third, the sustainable development of industrial regions was analysed from a socio-economic perspective via further urban exploration. The development of industrial regions is tied to cities, and the pace of industrial development follows the pace of the scaling rule. We identified the scaling pace of development in Australian cities with a key focus on industrial features. The characteristics of Australian cities were demonstrated from multiple perspectives, based on urban scaling theories. The spatial association between the pace of industrial development and other urban indicators was identified using a RGD. In general, this study validated the consistency of scaling development among Australian cities using power-law theory and the similarity of scaling disparity features among top-populated cities. Specifically, the urban innovation indicator and income level were predominantly and positively associated with industrial companies and employees, indicating that innovation growth and economic development in Australian cities would stimulate the performance of industrial companies and the employment scale. We noted the synergy between urban innovation and industrial company performance to be particularly significant in major capital cities. The developed spatial models have broad potential to address the spatial and scaling characteristics of industrial features.

Fourth, the impact of industrial development on economic inequality was investigated. The concept of geocomplexity was proposed and quantified to measure the spatial impact of selected variables on nationwide economic inequality at a higher spatial granularity. Geocomplexity, while accounting for the geography law of autocorrelation, was quantified using spatial local complexity. The concept of geocomplexity has been applied to explain the errors associated with traditional spatial and aspatial models. The results showed that industrial features and relevant spatial impacts influence local economic inequality. The consideration of geocomplexity can improve the performance of the traditional spatial and aspatial models.

This Ph.D. thesis explores the sustainability of nationwide industrial regions from environmental and socio-economic perspectives by processing Earth observation big data using innovative methods. The research outcomes yielded upon completion of the thesis contribute to innovative methodology design and advice toward smart built environment management. In terms of innovation in new methodology development, we developed a RGD to overcome the limitations of previous spatially stratified heterogeneity models and proposed a spatial concept of geocomplexity to explain errors from traditional models and improve estimation accuracy. These new methods can improve the spatial planning and management by incorporating geographical data-based smart decision-making with reliable outcomes. In terms of advancement of smart built environment management, this study provides detailed scientific results on sustainability analysis for industrial regions and infrastructures. The results of this study provide beneficial, practical, and supportive insights for urban and regional governance, with respect to both design and maintenance at the macro level.

Acknowledgment

I wish to express my appreciation to my primary supervisor, Dr. Yongze Song, who advised and guided my Ph.D. thesis. My research path is different from others, and it would be hard to move forward without the assistance and excellent supervision provided by Ph.D. advisers. I have a bachelor's degree in mathematics and a master's degree in information technology. I then found my research interests in spatial analysis and modelling and enrolled in the Ph.D. course in construction management. A diverse study experience may not provide a perfect background for specific research. Dr. Song helped me lay a solid foundation in spatial data mining and provided directions for my future career path. I gained writing skills, an understanding of quantitative analysis methodologies, and academic research experiences during my study at Curtin University. I appreciate Dr. Song's academic supervision and support. I am also grateful to my associate supervisor, Professor Peng Wu, for his supervision and help. Professor Wu provided academic help in preparation for my Ph.D. candidature and was the primary supervisor at that time. His advice and help have been beneficial to my academic studies and career path.

I am grateful to my parents for their support in my study abroad. I would like to express my gratitude for their encouragement, advice, and support. I, being the only son, has not been back to my hometown for four years due to research exchange and the pandemic. I also feel sorry for their sacrifice.

Finally, I would like to express my thanks to the researchers and industry partners who joined the co-authorship of my studies. The academic assistance provided by Peng Luo at the Technical University of Munich, Xiaochi Liu at Macquarie University, Dr. Mingshu Wang at the University of Glasgow, and Nicholas Archer from the Main Roads is greatly appreciated.

Listed publications below are from the research outcomes of the Ph.D. thesis. The publications are accessible with gold open access, and thus myself, the thesis

committee, and co-authors of listed publications hold the copyrights of articles, or the author and our thesis committee have the right to include these in this Ph.D. thesis. Attribution statements have been made clear and transparent during the publication process. For each publication, the author order represents the significance of contribution, and all co-authors have confirmed that the CRediT authorship contribution statement or relevant attribution statements are true and correct. Contributions of all co-authors are fairly reflected. The thesis author and the thesis committee preserve the evidence of attribution statements confirmation.

Related publications

Zhang, Z., Song, Y., & Wu, P. (2022). Robust geographical detector. *International Journal of Applied Earth Observation and Geoinformation: ITC Journal*, *109*(102782), 102782. https://doi.org/10.1016/j.jag.2022.102782

Zhang, Z., Song, Y., Luo, P., & Wu, P. (2022). Earth observation for Land Cover and Human-Environment Interactions. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W5-2022.* https://doi.org/10.5194/isprs-archives-XLVIII-4-W5-2022-211-2022

Zhang, Z., Song, Y., Archer, N., & Wu, P. (2023) Spatial disparity of urban performance from a scaling perspective: a study of industrial features associated with economy, infrastructure, and innovation. *GIScience & Remote Sensing, 60:1.* http://doi.10.1080/15481603.2023.2167567

Zhang, Z., Song, Y., Luo, P., Wu, P., Liu, X., Wang, M. (2023). Elucidation of spatial disparities of factors that affect air pollutant concentrations in industrial regions at a continental level. *International Journal of Applied Earth Observation and Geoinformation: ITC Journal*, 117(103221), 103221. https://doi.org/10.1016/j.jag.2023.103221

Zhang, Z., Song, Y., Luo, P., & Wu, P. (2023). Geocomplexity explains spatial errors. *International Journal of Geographical Information Science*. https://doi.org/10.1080/13658816.2023.2203212

Table of content

Chapter 1. Introduction

1.1 Background

The construction industry has served to change the world through infrastructure development. These infrastructures include roads, bridges, utility supply systems, buildings, industrial infrastructure, and the built environment that supports modern societies (Bansal, 2011). Construction can be categorised into agricultural, residential, commercial, industrial, and environmental. The industrial built environment and constructions, including infrastructure for mining, manufacturing, waste services, and utility supplies, are critical to the Australian society. Industrial regions and relevant infrastructure are common sources of air pollution. In the Australian industry, factories and infrastructure from three key industries (i.e. manufacturing, mining, and utility supply and waste services) are the main causes of air pollutant emissions, including carbon monoxide (CO), ozone (O_3), nitrogen dioxide (NO_2), and sulphur dioxide (SO_2) emissions. More than 95% of NO_2 and SO_2 pollutants are emitted from relevant industrial infrastructure (Department of the Environment and Energy, Australian Government, 2021). Spatial studies also imply that air pollutant investigations with specific industrial land uses as geographical boundaries deserve more attention (Satterthwaite, 2008). Furthermore, current knowledge about the impact of industrial features on residential living quality from a socio-economic perspective is still limited. Considering the importance of the economic contribution and environmental risk, it is necessary to investigate and assess the sustainable construction of these three industries. The investigation of sustainable development of three key industries from the perspective of infrastructure and construction design can be supported by geospatial tools and Earth observation big data.

Understanding and recognising real problems, establishing models, and providing feasible solutions are the key stages of construction management. These key stages can be further broken down into multiple tasks from various disciplines, including but not limited to spatial engineering, environmental engineering, civil engineering, and human research (Chiu and Russell, 2011). Each task requires specific tools to provide solutions. A geographic information system (GIS) is an integrated system designed to store and analyse spatial information using statistical and spatial science strategies. GIS software includes multiple computational functionalities, including database management, spatial data visualisation, spatial data computation, construction scheduling, environmental modelling, and safety planning at a macro level. One of the goals of the GIS application is to provide scientific decision-making for stakeholders (Worboys and Duckham, 2021). Therefore, the functionality of GIS software meets the demands of smart built environment management in the construction industry, and GIS has been applied in construction and built environment management during the planning, designing, and construction phases. Improvements and innovations in GIS technologies can stimulate decision-making in built environment construction management (Cheng and Chen, 2002).

Although GIS software has been utilised in built environment construction management from multiple perspectives, three limitations need to be addressed (Bansal, 2012). GIS software and relevant strategies are tools for a sustainable built environment. However, this application has not been fully developed because of the complexity of the interactions among multiple influential factors. Furthermore, the concept of sustainability is general and inclusive, and it can be redefined in different scenarios. Second, risk assessment is a critical part of the construction process, and the utility of GIS functionality in construction risk assessment has not been fully developed. Risk assessment from a sustainable development perspective can be further analysed using GIS with advanced functions. Third, the popularity of GIS software and spatial functions in construction management may be hampered by educational background, work experience, and other social issues. Construction management is a multi-disciplinary field, and not all construction professionals are familiar with GIS and its latest changes.

Earth observations based on remote sensing (RS) techniques can be an effective tool for measuring factors that influence sustainability issues. Earth observation techniques offer disclosure of physical and chemical properties on the Earth's surface, from spectral information to physical compositions (Ibrahim et al., 2018). These advanced technologies have been applied from space to the ground and include topics in human-environment interactions (HEIs). Numerous sustainability issues, spanning biology, environmental engineering, spatial engineering, urban planning, and social science, are discussed based on datasets generated from Earth observation techniques.

Considering the importance of sustainable industrial development and the capability of GIS and RS applications in built environment management, this thesis is designed to assess a sustainable built environment for three industries using Earth observation data and GIS techniques.

1.2 Research scope

This study focused on assessing sustainable development in Australian industrial regions toward smart built environment management based on Earth observation data analysis. Sustainability is investigated from socio-economic and environmental perspectives, with the scope explained as follows. From the perspective of environmental sustainability, air pollutants and influential factors in industrial regions are the focus. Industrial regions are developing at the cost of environmental degradation, particularly air pollution. Although industrial regions and infrastructure are acknowledged as the main sources of air pollutant emissions, spatial disparities of factors affecting air pollutants indicating the internal properties of industrial regions remain to be discovered. Current knowledge of spatial patterns from the relationship between human or environmental features and air pollutants in industrial regions is limited. From the perspective of socio-economic sustainability assessment, the pace of development of industrial features is analysed from an urban scaling perspective. The development of industrial regions is tied to cities, and urban scaling is a theory that can identify the pace of development of urban features, including industrial features, population growth, and economic development. Furthermore, socio-economic sustainability for an industrial region means that the region's development may have a positive impact on economic equality in the local community.

Therefore, the scope of this thesis includes sustainability assessment of an industrial built environment via investigation of industrial regions' association with air pollutants from an environmental perspective and the scaling pace of industrial development from a socio-economic perspective. Furthermore, how industrial regions impact on economic inequality will be analysed using spatial methods. This thesis bridges the gaps between GIS innovation design and construction management by investigating a nationwide industrial sustainability assessment as a case study. We provide clear definitions of sustainability for three industries. First, Earth observation data were applied in an industrial sustainability assessment to measure various features and relevant factors. Second, Earth observation data were further explored using advanced spatial methods, and sustainability patterns from socio-economic and environmental perspectives were provided. Third, this study elucidated spatial methodologies with innovations, and the spatial features uncovered from new spatial methods were summarised and concluded in this sustainable construction case. This nationwide case study can further popularise the application of GIS and Earth observations in the construction industry.

The content of this manuscript is subject to the fields of construction and spatial science. These tasks, for sustainable construction and built environment design, will be completed using spatial analysis methods and Earth observation. This study involves strategies in urban theory, GIS, spatial science, environmental science, and construction management. Final recommendations and suggestions were provided to stakeholders.

1.3 Objectives

This Ph.D. research aims to assess sustainable development in industrial regions for smart built environment management using Earth observation big data, considering the gaps and needs of the current study. Sustainability assessment will be carried out from environmental and socio-economic perspectives. An advanced spatial analysis method will be developed to process spatial big data by exploring the advanced spatial relationships among industrial sustainable indicators and influential factors. Furthermore, the spatial impact of industrial development on economic inequality will be assessed. Upon completion of this study, theoretical and practical contributions were made to built environment management and the construction industry, owing to the methodology design used and results obtained. New research methodologies have been proposed and developed to solve the gaps in previous spatial methods. These new spatial methods can be further applied for spatial planning and management purposes at the design and maintenance stages of the construction management lifecycle, with reliable performance. The results of this study provide a systematic overview of industrial built environment sustainability assessment, which is beneficial to further urban and regional governance during the maintenance phase of construction management, as well as future industrial planning at the design phase. Therefore, to achieve the expected outcomes, the following four objectives were established.

1. To explore the environmental sustainability of industrial regions by studying the spatial disparities of factors affecting air pollutants nationwide. This objective focuses on measuring industrial sustainability from an environmental perspective by

analysing the spatial heterogeneity of the determining factors of air pollutants in Australian industrial regions. First, spatial methodology of industrial region identification is proposed, and industrial regions of interest are identified. The identification methodology and geographical boundaries of industrial regions are the foundations of the subsequent objectives. Second, various air pollutant data and influential factor Earth observation data were accessed and pre-processed. Then, the spatial heterogeneity of air pollutant determining factors was analysed using geographically weighted regression with standardised coefficients. Finally, spatial advice has been given based on the identified spatial patterns summarised at a higher level of spatial granularity.

2. To develop an advanced and innovative spatial association identification method for spatial analysis and planning. A robust geographic detector was developed and adapted to overcome previous spatial discretisation limitations in a general geographic detector using an optimisation algorithm. This innovative method was then applied to uncover the advanced associations between environmental sustainability indicators and influential factors. Both objectives 1 and 2 aim at investigating the spatial patterns of factors affecting air pollutant concentration in industrial regions, while the second research objective further explores spatial association between air pollutant density and remote sensing factors from the view of spatial stratified heterogeneity. Results delivered from the second objective extend the understanding of the environmental sustainability of industrial regions using remote sensing datasets. The innovative method, as a kind of geographical detector with a powerful factor analysis performance, can be further applied in other fields.

3. To assess the socio-economic sustainability of the properties of industrial features from an urban scaling perspective based on an innovative spatial method. Industrial development is tied to cities. The pace of development of industrial features (i.e. industrial region scale, industrial employees, and industrial company count) has been assessed based on urban scaling theory. Urban features may be correlated and associated. Thus, the spatial association between industrial regions and other urban indicators was analysed using a robust geographical detector.

4. To explain the spatial impact of industrial development on economic inequality using the concept of geocomplexity. Variance in industrial features across space may lead to the distribution of economic inequality. The concept of geocomplexity has been proposed to represent the spatial impact of factors. Impact of

development of industrial features on economic inequality was explored based on the concept of geocomplexity.

1.4 Significance

This study makes three main contributions to academics and stakeholders according to the developed methodology framework and implied results.

(1) **The development of an advanced and innovative spatial analysis method supporting spatial planning and built environment management.**

This study developed a new analysis method for spatially stratified heterogeneity identification with innovation. This new method, named the RGD, overcomes the limitations of spatial zone determination in the previous geographical detector (GD) model. By developing and integrating an optimisation algorithm, a RGD can indicate spatial associations between factors with greater accuracy, robustness, and reliability. In addition to the model design, an advanced algorithm is applied to assess the sustainability of industrial development from both environmental and socio-economic perspectives. This newly designed spatial analysis model can also be applied to different case studies involving spatial datasets in various fields, including but not limited to environmental engineering, urban planning, and construction management. These methods are particularly useful for the application of sustainable built environment design and smart urban planning with clear spatial data support.

(2) **The development of an analysis method indicating the complexity of spatial impact with an application to industrial features' association with economic inequality.**

This study also proposes and redefines the concept of complexity in spatial science. This concept is called 'geocomplexity' or 'spatial local complexity'. In spatial studies, there is a phenomenon that the spatial distribution of numerical variables may vary across space. In some cases, the distribution may not be consistent with spatial autocorrelation. Thus, geocomplexity describes whether the current spatial distribution is complex according to the spatial autocorrelation law. Furthermore, the concept of geocomplexity was quantified and applied to the association between industrial features and economic inequality. The consideration of geocomplexity as a factor can improve the performance of traditional models in further studies.

(3) **Nation-wide key industrial regions being identified and the sustainability of industrial development being assessed, from both environmental and socio-economic perspectives, for smart built environment management.**

At the application level, newly designed algorithms and traditional models are utilised to assess the sustainability of industrial development from both environmental and socio-economic perspectives. Nationwide industrial regions were identified based on a spatial framework and the Australian Statistical Geography Standards. From an environmental perspective, determining factors, along with their spatial distributions of air pollutant densities in industrial regions, are indicated from the computational results. From a socio-economic perspective, the pace of development of industrial features is assessed based on the urban scaling theory. The association between industrial features and other urban indicators was also discussed. These results at the application level demonstrate the sustainability of industrial regions from different perspectives. Scientific advice for policymakers and stakeholders was provided based on the research outcomes.

1.5 Thesis structure

The structure of the remaining parts is shown in Figure 1-1.

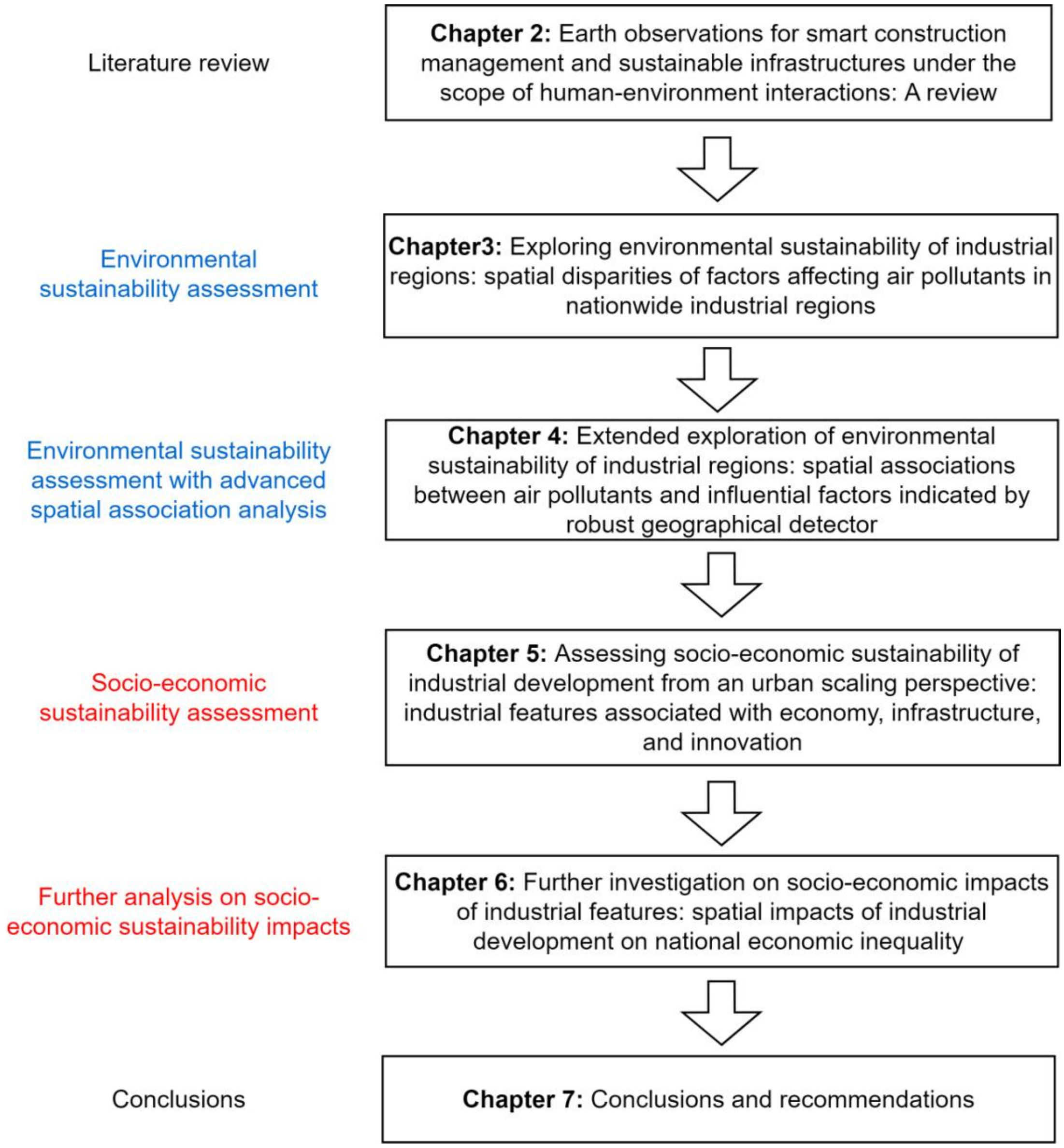


Figure 1-1. Thesis structure

Chapter 2 reviews the application of Earth observation in urban construction and sustainable infrastructure management under the scope of HEIs. This chapter provides the theoretical foundations for the four objectives.

Chapter 3 explores the environmental sustainability of the industrial built environment by analysing the spatial disparity of factors affecting air pollutants in nationwide industrial regions.

Chapter 4 further demonstrates the environmental sustainability of industrial regions by analysing the spatial associations between air pollutants and influential factors. A RGD was developed to overcome the limitations of previous models, to demonstrate the spatial patterns of environmental sustainability.

Chapter 5 assesses the socio-economic sustainability of industrial features from the perspective of the urban scaling theory. The power-law scaling pace of industrial development was estimated. The spatial association between industrial features and other urban indicators was analysed using a RGD.

Chapter 6 further investigates the socio-economic impacts of industrial features by explaining the spatial impacts of industrial development on economic inequality using the concept of geocomplexity. The complexity of the spatial impact was defined and applied to a case study of economic inequality and industrial features.

Finally, chapter 7 concludes the paper by listing important findings and provides suggestions to academics, policymakers, and stakeholders.

Chapter 2. Earth observations for smart construction management and sustainable infrastructures under the scope of human-environment interactions: A review

2.1 Introduction

The development of methodologies for smart urban construction and sustainable infrastructure is important. Given the case of Australian mining, manufacturing, utility supply, and waste services industries, industrial regions and relevant infrastructures are developing at the cost of environmental degradation. The development of these key industries has had a significant impact on the national economy. However, current knowledge regarding the sustainable development of key industries is limited. The sustainable development of industrial regions is a part of smart urban construction and sustainable infrastructure design. Sustainable industry assessment is also a part of human-environment interactions (HEIs). Thus, this chapter reviews the capability of Earth observations for HEIs, smart urban construction, and sustainable infrastructure. This chapter lays a solid theoretical foundation for the four objectives of this thesis.

HEIs are dynamic processes involving a wide range of research areas, and these complex processes have been investigated for decades (Kefalas et al., 2019). In this study, HEIs have been interpreted based on a compound relationship among human activities, land cover, and the ecological environment. In this article, the word ‘environment’ refers to the ecological environment, which includes abiotic elements (from soil and water quality to precipitation and temperature) and biotic components (e.g. vegetation cover, agricultural production, and biodiversity) (Chang et al., 2019). Human activities at a large spatial scale, such as urban construction, industrial development, and government-level investment programs, can impose forces leading to land cover change over both short-term and long-term periods (Desjeus et al., 2015; Gellrich and Zimmermann, 2007). In this chapter, we review urban construction as a kind of human activity and discuss the smart urban construction enabled by Earth observations. Land cover has intensive interactions with human activities and functions as a key intermediate variable in the HEI dynamic process (Wang et al., 2021). Changes in land cover from urban construction could affect cities facing the urban heat island effect (Song et al., 2014), in addition to other potential landscape ecological changes. In addition to human forces, land cover can be influenced by natural features, including wind speed, humidity, precipitation, and topography, through chronic geomorphological processes (Dai et al., 2014). Furthermore, there are dense interactions between human activities and the environment. For instance, the aggregation of the construction industry can lead to air and water pollution (Dong et al., 2019), and poor air quality would impose negative effects on an individual's health, leading to other social risks. Urban construction levels may vary across landscapes in different places. Therefore, it is difficult to investigate the potential relationships under human-environment dynamics in a systematic way using fixed models. The evolution of Earth observations and remote sensing (RS) techniques enables environmental variables and land cover dynamics to be monitored from a more comprehensive global or detailed local perspective (Ustin and Middleton, 2021). Moreover, spatial analysis approaches for Earth observation data reveal spatial patterns and create a spatial simulation for landscape management and decision making.

Earth observation techniques help determine physical and chemical properties of the Earth's surface in terms of spectral information and physical compositions. These advanced technologies have been applied from space to the ground, which includes various topics in HEI studies. This chapter summarises and reviews the contributions and significance of Earth observations toward HEI research, based on findings of previous studies. There have been numerous studies revealing impact of humans on the ecological environment (Jin et al., 2019), the ecological environment's feedback on human society and its influence on human decision-making (Zhai et al., 2020), and the role of Earth observations in environmental monitoring and human development monitoring (Phiri et al., 2020). Furthermore, theories and models have been developed to explain the impacts and consequences of human activities on the planet. Relevant topics include studies on the carbon and nitrogen cycles (Erisman et al., 2013), climate change (Brody et al., 2018), and other topics relevant to environmental change. Among these topics, numerous issues, including biology, environmental engineering, spatial engineering, urban planning, and social science, have been discussed based on datasets generated from Earth observations and spatial techniques. Generalised Earth observations include data generated from monitoring stations and mobile technologies, which could represent natural factors (Gellrich and Zimmermann, 2007), social behaviours (Ristea et al., 2020), and natural hazards (Bruneau et al., 2021). Apart from the academic benefits of Earth surface property disclosure, the value of Earth observation can be added with the help of analytical methods with innovations. In general, these methods help further explore the value of Earth observation data by providing valuable information about spatial relationship identification (She et al., 2017), which is used for spatial pattern and distribution analysis, spatial estimation to overcome the spatial limitations of field sampling, and spatial decision-making for governance policy. The remainder of this chapter is organised as follows. The second subsection shows the key topics of land cover, urban construction, and the environment using Earth observation data under the scope of HEI analysis. Earth observation data and analysis methods are explained in the third and fourth subsections. Smart urban construction plays a key role in sustainable management in HEI research, and Earth observation and spatial services enable sustainable construction. The relationship between Earth observation development and smart urban construction is explained in the fifth subsection. The last subsection concludes the chapter.

2.2 Implementation of Earth observation to land use, land cover, and HEIs

Land cover is a key component in HEI research, which has already been investigated using Earth observation techniques from multiple perspectives (Foley et al., 2005). We summarise Earth observation-based land cover studies from three perspectives: land cover change, land cover mapping, and land cover management and monitoring. There are three types of land cover studies: those that focus on human endeavours to identify reasons for generating land cover change in the past (Gellrich and Zimmermann, 2007), those that focus on human development to learn land cover compositions and contemporary patterns (Milenov et al., 2014), and those that focus on management of land cover to improve quality of life in the future. Additionally, investigations have been conducted to explore natural and social factors relevant to land cover change using Earth observation data (Kefalas et al., 2019). From a mapping perspective, the spatial patterns of various land cover types, including urban areas, green spaces, forests, farming-pastoral ecotones, cropland, terrain, shoreline, and other land covers for specific uses, were determined based on Earth observation data using spatial approaches. From a land management perspective, soil properties and landscape metrics have been investigated intensively. Numerous studies have been conducted on soil moisture, aboveground carbon, soil nitrogen, soil organic matter, mineral chemicals, and heavy metals. Furthermore, landscape metrics, as indicators of land patterns, are utilised to show landscape design and ecological risk in urban and rural regions (Sahraoui et al., 2021). In HEI studies, Earth observation has presented irreplaceable values for the aforementioned indicators. Generally, numerous land cover indicators and relevant spatial metrics are identified using satellites from Landsat, moderate resolution imaging spectroradiometer (MODIS), and Sentinel-2, as well as commercial products and airborne-based photos. Soil property information can be collected via field sampling and from ground in situ analysis.

Landscape composition and land cover change are represented by spatial and aspatial patterns of land cover composition and changes over time. Landscape composition and land cover change are critical concepts in Earth observation and land cover management (Van et al., 2013). Landscape composition and its changes are considered a key median process when investigating HEIs using Earth observation data. Changes in landscape composition are subject to short- or long-term human forces, including urbanisation and population change, governance policies for ecological restoration, economic development, and infrastructure development. Landscape changes are also influenced by long-term natural forces caused by precipitation, topography, temperature, humidity, and wind speed (Kefalas et al., 2019). The dominance of human or natural forces was determined by the development level of the study area. Human forces can influence HEI processes in highly urbanised regions (Chen et al., 2019). As a process of HEI, land cover changes caused by urbanisation or vegetation recovery may lead to the urban heat island effect or heat mitigation (Song et al., 2014). Typically, land covers are water bodies, urban and built-up areas, soil, cropland, and vegetation, which can be monitored by RS indices. The normalised difference vegetation index (NDVI), enhanced vegetation index (EVI), normalised difference soil index (NDSI), soil adjusted vegetation index (SAVI), difference vegetation index (DVI), and impervious surface fraction (ISF) are indicators generated from Earth observations to monitor the composition of land cover. Landscape patterns and spatial metrics are other indicators of morphology and general spatial patterns in urban studies (Herold et al., 2003). Typically, these spatial metrics were originally derived from land composition (type of land cover or land composition). Landscape patterns and spatial metrics include patch number, total urban area, mean urban patch size, patch density, Shannon's diversity index, interspersion juxtaposition index, landscape aggregation index, and eccentricity (McCarty and Kaza, 2015). Land cover is further processed and translated into a new term named 'patch', which refers to homogeneous areas for a specific landscape property of research interest, such as ‘industry region’, ‘residential region’, and ‘green space’. These patch-based indicators are measurements for modelling forms of urban sprawl and quantifying the shape and spatial patterns of vegetation in natural land cover.

The construction industry focuses on the design and construction of infrastructure that supports cities. This urban infrastructure includes roads, dams, utility supply systems, waste service systems, and green infrastructure. Such infrastructure supports both daily lives and future development. Construction activity, a representative of human activities, is an important process of urbanisation and industrialisation. Therefore, the construction industry can support population growth and urbanisation to some extent. Population change and urbanisation, which are causes of environmental pollution, ecological risk, and land cover change in HEI research, are interpreted as social change and urban development caused by human activities and construction industry aggregations (Dong et al., 2019). Population is generally indicated by population or population density in a certain area, and urbanisation is regarded as a phenomenon of human aggregation from rural to urban areas. Therefore, the urbanisation level can be transformed and measured by the proportion of the urban population or other human activity indicators. Population-based data in HEI studies can be accessed from census data published by governance authorities, and night-time light (NTL) Earth observation data can be utilised as ancillary data when measuring the urbanisation level.

With the help of Earth observation development, ecological risks and environmental pollution can be measured using RS and spatial techniques. The environmental sensitivity index (ESI), habitat quality index (HQ), and ecological risk index (ERI) are three indicators that are highly relevant to environmental patterns and spatial metrics, demonstrating the vulnerability of the ecological environment. As the ecological environment is a complex system composed of various spatial features with dynamic interactions, ESI, HQ, and ERI have been proposed, based on human needs, to evaluate the quality of the physical environment for further sustainable development decision-making. Currently, spatial studies have already verified strong relationships among landscape ecological risk, urban infrastructure development, governance policy, and urbanisation level (Lin et al., 2019).

The ecological and physical environments can also convey feedback to cities or urban areas when environmental degradation, owing to environmental pollution or the urban heat island effect, occurs. In this article, the term ‘environmental pollution’ refers to pollutants that are detrimental to an individual's health, including but not limited to wastewater, particulate matter, carbon dioxide, SO_2, NO_2, and aerosol optical depth. These environmental pollutants are relevant to governance policies, population and urbanisation, economic development, construction industry development, technology innovation, and landscape design. In most cases, these pollutants can be measured through air pollution monitoring stations and updated hourly through open-access platforms. In broad-area research, air pollutants can be monitored more efficiently by Earth observation data using satellites or unmanned aerial vehicles (UAVs). Current studies have demonstrated the feasibility of air pollutant monitoring using MODIS products (Zhang et al., 2018), and other forms of air pollutants can be measured using UAVs mounted with specific sensors (Lambey and Prasad, 2021).

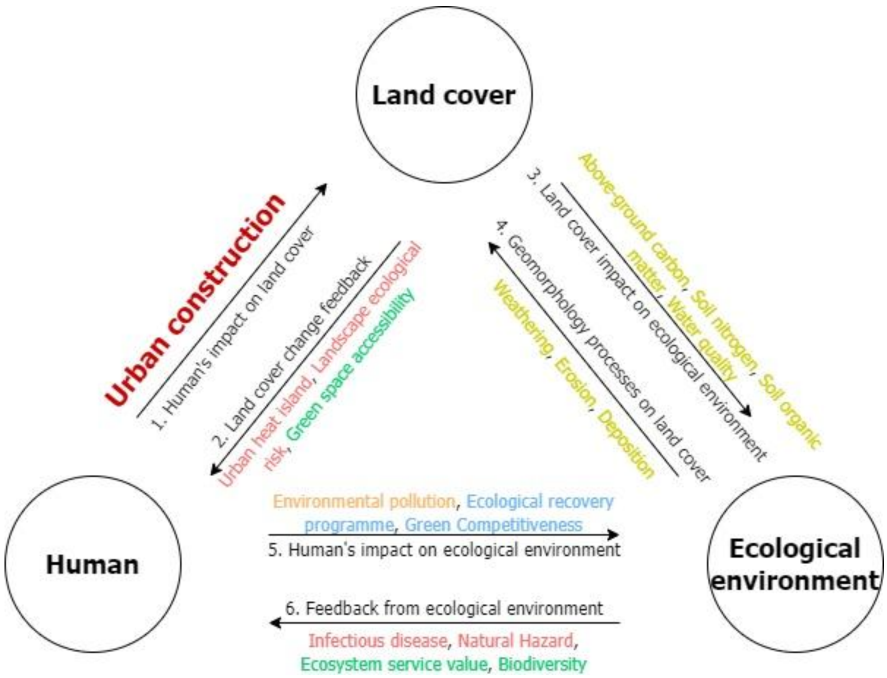


Figure 2-1. Key topics under the scope of human-environment interactions

Figure 2-1 summarises the key topics within the scope of HEI. Human activities, land cover, and ecological environment are mutually related through various processes. Urban construction is a type of human activity within the scope of HEI, which can lead to land cover changes and environmental changes. The rest of the chapter will show

how Earth observation data and analysis methods can be applied in HEI studies and how Earth observation development could lead to smart urban construction.

2.3 Earth observation data for land cover, infrastructure, and the environment

As free open-access data, Landsat and MODIS products have been widely and intensively applied in HEI research. Land information from land cover properties and landscape metrics to land surface temperature and ecological risk indicators can be generated from Landsat and MODIS products. The derived information and indicators are valuable for most HEI research relevant to land cover status and construction design. These data can be further processed using spatial or arbitrary methods. Landsat and MODIS products can generally be utilised to identify the spatial relationships among variables. These RS products can also be used to explore spatial patterns and distributions, spatial estimations, and spatial decision making in HEI studies. Advanced requirements in HEI can be fulfilled by commercial satellites or satellite missions for specific purposes. When mapping skinny land features, such as streams and roads, the spatial resolutions for Landsat or Sentinel-2 are too coarse to provide valid information. High-resolution commercial products can facilitate detailed mapping tasks (Biotto et al., 2009). Furthermore, the general commercial satellite image acquisition cost is lower than cost of obtaining data from airborne-based sensors or UAVs (Lambey and Prasad, 2021). Moreover, the use of Tropical Rainfall Measuring Mission (TRMM) for precipitation (Chen et al., 2018) and Shuttle Radar Topography Mission (SRTM) for topography information facilitates free access to global natural factors measured from satellites.

Airborne-based Earth observation data refer to RS images captured from airborne or UAVs. The variability of sensors mounted on airborne vehicles or UAVs determines the capability of airborne-based Earth observation data. Typically, airborne-based sensors support spectral detection from the visible band to VNIR and SWIR bands. Active RS, such as LiDAR, is also performed using airborne vehicles. In HEI research, a few land cover monitoring, NTL collection, and species statistics tasks are completed with the help of airborne RS, as aerial images provide higher spatial resolutions than satellite images (Kuechly et al., 2012).

Ground-based Earth observation data are mainly obtained from field sampling and monitoring stations. Field samplings, sometimes known as in situ data, typically refer to samples taken at a specific location; the physical or chemical properties of those samples are subsequently analysed in the laboratory. This data collection methodology has been intensively utilised for analysis of soil properties, such as heavy metal concentration, soil moisture, mineral chemistry, and aboveground carbon and nitrogen. Although field sampling has time and budget requirements and spatial limitations, more physical properties and chemical compositions, from above-ground to deep soil, could be comprehensively studied using such data. In addition to soil quality sampling, field sampling has been applied for water quality testing (Sun et al., 2016). Ground-based Earth observation data also include data collected from the monitoring stations. Ground monitoring stations have been established mainly for natural factor monitoring and air quality monitoring. Natural factors (e.g. temperature, humidity, and precipitation) and air quality indicators can be measured frequently from stations (McCarty and Kaza, 2015). Other stations are designed for specific monitoring tasks, such as forest soil carbon efflux (Crabbe et al., 2019) and water quality. The drawback of the spatial limitation for missing values from field sampling and monitoring stations can be mitigated by spatial interpolation or RS image complementarity (Dang et al., 2018).

Currently, mobile technology generates a new type of Earth observation data, which is published via Twitter, online mapping applications, and other web-based social media platforms. This type of Earth observation data generally contains geographic location, event occurrence, and occurrence time. The superiority of the immediate message-sharing function of mobile technology enables the real-time monitoring of rapid land cover changes (Ristea et al., 2020). This new data type has been applied for hazard monitoring purposes (Bruneau et al., 2021).

2.4 Spatial analysis methods for Earth observation applications

2.4.1 Spatial analysis methods and applications in sustainability

To identify relationships among spatial variables, spatial statistical methods, including spatial regression models, geographically weighted regression (GWR), and bi-variable Moran's I, can be used. Spatial regressions containing spatial lag, spatial error, or other advanced forms, along with GWR, are improved forms of analysis methods derived from standard regressions by introducing a spatial matrix. The bi-variable Moran's I method enables one-to-one spatial relationship identification, and one response variable can also be tested with a compound explanatory variable representing multiple raw explanatory variables added based on a specific criterion (Balducci and Ferrara, 2018). Relationships among spatial variables can also be identified using aspatial methods covering non-spatial regression, multivariable linear regression, and PCA (Chang et al., 2019).

Spatial autocorrelation of spatial variables and topology features of land cover are two issues in EO-HEI. Spatial autocorrelation can be visualised using Moran's I and Gi* (Chen et al., 2020), whereas land topology can be elucidated using digital image processing and spatial metrics. Spatial estimations in HEI mainly refer to spatial interpolation methods to overcome the spatial limitations of ground sampling. Numerous Kriging-based methods and deterministic interpolations have been utilised to estimate soil quality and precipitation. Previous studies have shown that kriging performs best when estimating soil organic matter (Long et al., 2020). Spatial decision-making in HEI includes identification of spatial factors that could influence policymaking, as well as identification and assessment of influential consequences of spatial variables (Jackson, 2003). Spatial decision-making can be considered a compound spatial issue at a different level. Spatial decision-making interprets the results generated from spatial relationships, spatial patterns, and spatial estimations and draws a further conclusion for smart policy-making purposes (Cheng et al., 2019).

Earth observations and relevant spatial analysis methods have been used to analyse the factors affecting air pollution. Air pollutants are detrimental to the natural environment (Wang et al., 2021) and human health (Dockery et al., 1995); monitoring of air pollution has become a critical environmental justice issue (Xie et al., 2017; Cai et al., 2020; Tong et al., 2021). Considering the impact of air pollutants, their monitoring has long been a key research issue. Relevant research efforts have been undertaken from socio-economic and environmental perspectives (Shmool et al., 2014; Fang et al., 2015; Ge et al., 2018). These studies have demonstrated that a few spatial factors can lead to high air pollutant density (Gómez-Losada et al., 2019; Xu et al., 2019). From an environmental interaction perspective, meteorological factors, including precipitation and wind speed (Hu et al., 2021), vegetation greenness (Wang et al., 2020), and topography (Sabrin et al., 2020), are closely related to air pollutant concentrations. From the perspective of human activity, road development (Dons et al., 2013), population growth (York et al., 2003; Cui et al., 2019), urbanisation (She et al., 2017), and industrialisation (Cheng, 2016) can increase air pollution.

2.4.2 Geographical detector (GD): application and future development

A GD is a method for measuring spatial stratified heterogeneity (SSH) using statistical variance (Wang et al., 2016). In GD models, the association between dependent and explanatory variables is quantified using the power of determinant (PD) value, which is a comparison between the variance within strata and across the entire study area (Wang et al., 2010). GD has been proposed and widely applied in geography for a decade with proven solid theories. Current GD is well-developed in diverse applications and methodology extensions. From an application perspective, GD is a powerful tool for examining spatial differences (Chen et al., 2019), identifying driving factors (He et al., 2019), and providing spatial advice (Dong et al., 2021). The spatial analysis advantage of GD has been shown in various studies, from human settlement management to HEI investigation, at different spatial scales (Raghavan et al., 2013; Qu et al., 2018; Maus et al., 2020; Song et al., 2021b). From a methodology extension perspective, optimal parameters regarding the break interval and spatial scale have been investigated to improve the GD performance (Cao et al., 2013). A spatial association interactive detector based on the GD theory was proposed to quantify spatial associations between spatial causes and effects (Song & Wu, 2021a). Moreover, a geographically optimal zone-based heterogeneity model was developed to improve the measure of SSH based on GD (Luo et al., 2022). These advanced GD methods have been applied to infrastructure management and soil moisture modelling.

However, the process of spatial data discretisation has been a sensitive stage for the exploration of spatial associations, computation of PD values, and identification of geographical variables. This means that changes in the spatial discretisation method and the number of spatial zones can typically affect the relative importance of the variables. In studies in which explanatory variables are continuous numerical data, spatial data discretisation is a necessary and essential step before using GD models (Wang et al., 2016). In natural and social environment studies, continuous numerical data of explanatory variables, such as population, economic conditions, wind speed, precipitation, air pollutant indicators, and vegetation coverage, are common. Therefore, developing an effective discretisation approach for continuous numerical data is important for the practical implementation of GD models. To address this issue, an optimal parameter-based geographic detector (OPGD) was developed to improve the GD factor detector by providing various discretisation strategies based on the statistical distribution of explanatory variables (Song et al., 2020). However, these discretisation strategies do not fully address the limitations of deriving reliable strata using robust discretisation approaches. PD values derived from OPGD fluctuate with the increase in interval breaks during the process of selecting optimal discretisation parameters (Song et al., 2020; Luo et al., 2021), implying that the stability and robustness of spatial data discretisation are limited. This is because most of the current spatial discretisation strategies, including the strategy developed in OPGD, are performed based on observations of samples, rather than in-depth data characteristics. Therefore, more effective and robust spatial data discretisation strategies are required to improve GD modelling.

2.5 Future development of Earth observation, smart urban construction, and sustainable infrastructure

2.5.1 Earth observation development

In the future, changes may occur in terms of the data sources, analytical techniques, and platforms used. From a data source and platform evolution perspective, sentinel series products can be potentially powerful competitors to MODIS and Landsat series products. UAV and radar images share a large proportion of this research area, owing to their irreplaceable advantages. Furthermore, as a geospatial processing platform that has already been developed and researched, Google Earth Engine (GGE) will also be a popular tool in HEI studies. From a data processing improvement perspective, image fusion has shown potential for Earth observation data pre-processing, and machine learning-based methods can be efficient approaches for spatial feature identification.

Radar products and UAV images also play a significant role in HEI based on their unique features. Owing to the advantages of data acquisition costs and high usability, UAVs mounted with a variety of sensors are utilised to monitor air quality and coastal regions (Adade et al., 2021). Furthermore, the feasibility of UAVs also enables more frequent and higher-spatial-resolution observations for air pollutant monitoring based on the needs of research (Lambey and Prasad, 2021). As active RS, radar techniques release and receive microwaves, which are not subjected to bad weather or other negative natural risks (Minh et al., 2020) Currently, radar has been applied for plateau and non-residential area mapping (Reinosh et al., 2020). Although there is low spectral variety and no current open-source platform for data sharing, UAVs and radar products have been applied to some HEI issues, including land cover monitoring and air quality monitoring, owing to their irreplaceable feasibility and utility.

GEE, which exists as a new cloud-based geospatial data platform, will play a key role in future HEI studies from a big data handling perspective. Earth observation data can be categorised as big data, as the nature of Earth observation coincides with the "3V" (volume, velocity, and variety) big data definition. The cloud-based GEE platform, in the form of software as a service (SaaS), is designed to handle spatial big data tasks at the petabyte level, from raw datasets to final valuable products. From a data storage perspective, Earth observation big data can be smartly and efficiently stored in such a large cloud system via distributed networks. Furthermore, GEE, as an integrated and well-managed platform, also increases the accessibility and availability of Earth observation data from multiple sources worldwide. From a processing algorithm perspective, GEE application programming interfaces (APIs) make better basic Earth observation data processing algorithm code sharing and accessing environment possible, which saves time for experts and non-experts. Moreover, the GEE computational infrastructure with high-speed parallel processing and distributed computing techniques is an efficient tool for manipulating advanced machine learning or image processing tasks (Tamiminia et al., 2020).

The coming years may witness the prosperity of image fusion and machine learning evolution in HEI research. Although not included in current HEI research, image fusion has the potential to be introduced in the future, as its functionality is beneficial. Image fusion for hyperspectral images (HSIs) and multispectral images (MSIs) has been commonly applied in Earth observation studies to improve the HSI spatial resolution and MSI spectral resolution (Dian et al., 2021). High-quality fused images, created with more detailed and precise geographical features, can be further utilised to monitor land cover change.

2.5.2 Earth observation and spatial tools support smart urban construction and sustainable infrastructure

The construction industry has served to change the world through infrastructure development. These infrastructures include roads, bridges, utility supply systems, buildings, green infrastructure, and other forms of built systems that support modern societies physically and spiritually (Bansal, 2011). Understanding and recognising real problems, establishing models, and providing feasible solutions are the key stages of construction management. These key stages can be further broken down into multiple tasks from various disciplines, including but not limited to spatial engineering, environmental engineering, civil engineering, and human research (Chiu and Russell, 2011). Each task requires specific tools to provide solutions. Geographic information systems (GISs) are integrated systems designed to store and analyse spatial information using strategies from statistics and spatial science perspectives. GIS software includes multiple computational functionalities, including database management, spatial data visualisation, spatial data computation, construction scheduling, environmental modelling, and safety planning. One of the goals of GIS applications is to provide scientific decision-making for stakeholders (Worboys and Duckham, 2021). Therefore, the functionality of GIS software meets the demands of the current construction industry, and GIS has been applied in construction management from the planning, bidding, and construction phases. Improvements and innovations in GIS technologies can stimulate construction management decision-making (Cheng and Chen, 2002).

Although GIS software has been utilised in construction management from multiple perspectives, three limitations must be addressed (Bansal, 2012). GIS software and relevant strategies are tools for sustainable construction. However, this application has not been fully developed because of the complexity of the interactions among multiple influential factors. Furthermore, the concept of sustainability is general and inclusive, and it can be redefined in different scenarios. Second, risk assessment is a critical part of the construction process, and the utility of GIS functionality in construction risk assessment has not been fully developed. Risk assessment from a sustainable development perspective can be further analysed using GIS with advanced functions. Third, the popularity of GIS software and spatial functions in construction management may be hampered by educational background, working experiences, and other social issues. Construction management is a multi-disciplinary field, and not all construction professionals are familiar with GIS and the latest changes.

Earth observation based on RS techniques can be an effective tool for measuring the influential factors relevant to sustainability issues. Sustainability in construction management includes, but is not limited to, air pollutant emissions, land cover change, and the urban heat island effect. The implementation of Earth observation applications could make sustainability measurements feasible. For air pollutant emission evaluation, Sentinel products and other commercial products can provide global measurements of air pollutant density with a high temporal resolution. For land cover change monitoring, satellite-based and UAV-based images can be used to monitor changes in land use and land cover with various images, uncovering different physical properties. For urban heat island effect assessment, RS products with thermal infrared have been utilised in several metropolitan areas as case studies.

Construction can be classified into agricultural, residential, commercial, industrial, and environmental categories. Industrial construction, including infrastructure for mining, manufacturing, waste services, and utility supplies, is critical to Australian society. The utility of such infrastructure contributes approximately 25% of the national GDP and approximately 99% of human-dominated air pollutant emissions (Department of Environment and Energy, Australian government, 2020). Considering the importance of the economic contribution and environmental risk, it is necessary to investigate and assess the sustainable construction of these three industries. Therefore, the development of Earth observation and spatial tools can help assess the sustainable construction of the three industries using Earth observation data and GIS techniques.

The development of Earth observation and spatial tools can help in the implementation of smart urban construction by overcoming these three gaps. First, in terms of the requirements of various sustainability features for urban areas, different sources of Earth observation could provide measures of urban sustainability from socio-economic and environmental perspectives. Second, in terms of risk assessment in sustainable construction, spatial tools, geographical measures, and big data analysis approaches can provide reliable sustainable construction analysis and deliver scientific decision-making. Third, in terms of the popularity of GIS and Earth observation in the construction and urban design industries, the demand for and development of GEE could popularise the application of spatial tools and Earth observation. Furthermore, an advanced spatial methodology framework can provide more feasible suggestions that are acceptable and understandable for experts from non-spatial fields.

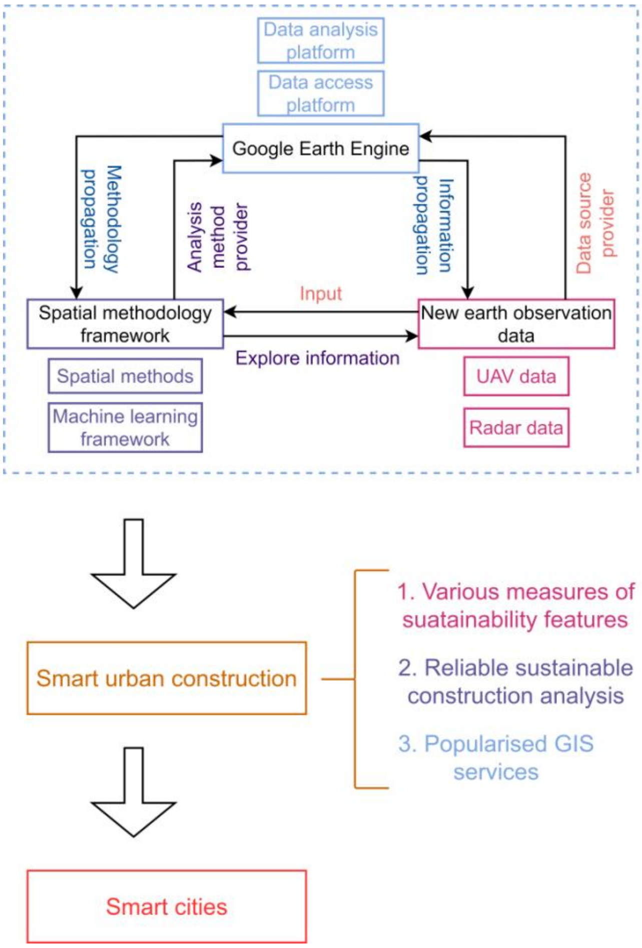


Figure 2-2. The development of Earth observation techniques for a better smart urban construction purpose.

Figure 2-2 summarises how Earth observation and GIS services can support smart urban construction and smart city development. The new spatial methodology framework GEE and Earth observation data form a mutually enhanced system from the perspective of Earth observation application. GEE works as a platform for sharing new RS datasets and advanced spatial methodologies, and GEE could popularise the application of spatial services and RS data among researchers who are experts and non-experts in spatial services. New forms of Earth observation datasets can provide input for an innovative spatial methodology framework, and new spatial methods can elucidate more values associated with novel Earth observation data. Within the scope of a smart city, the mutually enhanced Earth observation system enables sustainable urban construction. This system can support urban construction activities with measures of sustainability features from socio-economic and environmental perspectives, reliable sustainable construction analysis, and popularised spatial services. Under the scope of HEI, the construction industry, as a type of human activity, plays a key role in HEI; moreover, urban construction stimulates urbanisation, industrialisation, and population growth. Sustainable urban construction with Earth observation as a management tool could be beneficial for smart environment design and smart land cover management.

2.6 Conclusion

This chapter summarises the contribution, potential, and significance of Earth observation for urban construction and land cover under the scope of the HEI. This article presents the application of Earth observations in construction management, land cover monitoring, and environmental topics. This chapter also summarises important topics for applying Earth observation in land cover monitoring, urban construction, and environments. By utilising spatial methods, the academic values of Earth observations can be explored using numerous analysis methods. In general, Earth observation could provide further valuable information regarding spatial relationship identification, spatial pattern and distribution, spatial estimation to overcome the spatial limitation of field sampling, and spatial decision-making for governance policy via spatial or aspatial methods. Finally, there is potential for future HEI research to present technological and analytical evolution, and advancements can be dominated by three mutually enhanced technologies, namely, spatial algorithms, the GEE platform, and new Earth observation data sources. The development of Earth observation and spatial framework systems could help us achieve smart construction and smart city design goals.

By reviewing spatial methods and Earth observation for urban construction and industrial sustainability, several research gaps can be bridged at the current stage. Currently, the geographical detector, a spatial method applied in environmental assessment and urban construction analysis, needs to be improved by overcoming the sensitivity to the statistical distribution of variables. Furthermore, some of the spatial phenomena are not simple and may be related to the status of complex. Thus, it is also beneficial to extend the understanding of complexity from the view of spatial dependence. Despite the capacity of Earth observation to represent remote sensing properties of land uses and land covers has been demonstrated in previous studies, an extended research application on how Earth observation can be fully utilised to analyse industrial sustainability at a continental level remains to be shown. The following research components of the thesis from Chapter 3 to Chapter 6 overcomes the limitations of spatial methods in urban construction and built environment and Earth observation data for industrial sustainability analysis with specific case studies.

Chapter 3. Exploring environmental sustainability of industrial regions: spatial disparities of factors affecting air pollutants in nationwide industrial regions

3.1 Introduction

Air pollutants are detrimental to the natural environment (Wang et al., 2021); of note, in recent times, human health (Dockery et al., 1995) and air-pollution monitoring have become a critical environmental justice issue (Xie et al., 2017; Cai et al., 2020). It is a known fact that industrial regions and relevant infrastructures contribute to air pollutant emissions, and the related spatial studies imply that air pollutant investigations, with specific industrial land use as the geographical boundaries (rather than administrative boundaries), can be accurate (Satterthwaite, 2008). Thus, a detailed study that investigates the air pollutant concentrations of a region based on specific land uses, with spatial reasoning, can support smart regional planning with a clear focus. In case of the Australian industrial region, factories, and infrastructures, three key industries, i.e. manufacturing, mining, and utility supply and waste services, are highly relevant to air pollutants, including carbon monoxide (CO), ozone (O_3), nitrogen dioxide (NO_2), and sulphur dioxide (SO_2) (Department of the Environment and Energy, Australian Government, 2021).

The concentrations of air pollutants have long been monitored using remote sensing and earth observations (Akinwumiju et al., 2021; Roy, 2021). Factor analysis is one of the most important topics in air pollutant investigations and is supported by remote sensing techniques. Several relevant previous studies analysed air pollutant concentration from the socio-economic and environmental perspectives (Shmool et al., 2014; Fang et al., 2015). These studies demonstrated that a few spatial factors can lead to a high air-pollutant density (Gómez-Losada et al., 2019). From an environmental interaction perspective, meteorological factors, including precipitation and wind speed (Hu et al., 2021), vegetation greenness (Wang et al., 2020), and topography (Sabrin et al., 2020), are closely related to air pollutant concentrations. From an anthropological perspective, road development (Dons et al., 2013), population growth (York et al., 2003), urbanisation (She et al., 2017), and industrialisation (Cheng, 2016) can increase air pollution significantly.

The relationship between the air pollutant concentrations and the characteristics of industrial regions can be determined using traditional statistical methods (Yang et al., 2019). Furthermore, advanced machine-learning algorithms, including support vector machines and random forests, were developed to measure industrial air-pollutant concentrations (Ju et al., 2023). However, the current knowledge about the spatial patterns that indicate the relationship between the anthropological or environmental features and the air pollutant concentrations in industrial regions is limited. Currently, the spatial disparities of the factors that affect air pollutant concentrations, which indicate the internal characteristics of industrial regions, remain undiscovered. To explore such spatial relationships, geographically weighted regression (GWR) is a suitable method (Fotheringham, 2002). In previous studies, this method has been applied to identify the spatial heterogeneity in the relationship between air pollutants and their driving factors (Tian et al., 2019; Guo et al., 2021). The practical feasibility of applying the GWR method to such industrial land-use studies has been proven in previous works (Fotheringham et al., 2003; Tu et al., 2021).

This study was designed to identify the factors that affect the air pollutant concentrations, due to the local spatial impacts in industrial regions, at a continental level. In this study, we identified nationwide industrial regions, based on a spatial methodology framework, and demonstrated the spatial patterns of the factors that affect air pollutant concentrations in industrial regions, using the GWR method (with standardised coefficients). In this study, we also explored the spatial impacts of various remote-sensing factors on air pollutant concentrations in detail.

3.2 Study area and data

3.2.1 Study area

In this study, we focused on the industrial regions that support key industry activities across Australia. As of 2020, Australia comprised eight states or territories and had a population of 25 million. According to the remoteness structure defined by the Australian Statistical Geography Standard (ASGS), approximately 20,000 km^2 (accounting for 0.26% of the country) of the nations' total area is covered by major cities. Approximately 70% of the national population resides in major capital cities (Geoscience Australia, 2014; Australian Bureau of Statistics, 2021a).

3.2.2 Data

3.2.2.1 Data for industrial region identification

We identified the industrial regions in the country using land-use polygons and points of interest (POI). The industrial land use polygons were acquired from the OpenStreetMap (OSM) software (Geofabrik and OpenStreetMap contributors, 2020). The polygons, tagged as industrial areas, delineated the areas designed for relevant industrial activities. Note that the industrial polygons from OSM met the definition of areas for manufacturing, mining, utility supply, and waste services. The POI data, acquired from the National Pollutant Inventory (NPI) data, included all the officially registered locations of facilities built for manufacturing, mining, utility, and waste services listed in the NPI (Department of the Environment and Energy, Australian Government, 2020).

The raw OSM industrial polygons were coarse in size and needed to be processed prior to computation. According to the ASGS, polygons covering an area of 5000 m^2 should be the minimal resolution of a region having at least one functional facility, which also holds true for maintaining the same spatial granularity for an area having an infrastructure that supports the daily activities of a society (Hadjisophocleous and Chen, 2010; Yamaguchi et al., 2012). Therefore, land use polygons of areas less than 5000 m^2 were considered as points, rather than regions. Thus, extremely small land-use polygons were converted into points and treated as supplementary POI. Table 3-1 presents a general description of the spatial data used for industrial region identification.

Table 3-1. Description of spatial vector data for industrial region identification.

Type of vector data	Count	Rural	Urban	Source
Industrial land use polygon	6237	2869	3368	OSM
POIs for utility and waste services	1237	841	396	NPI
POIs for manufacturing	1225	576	649	NPI

POIs for mining	711	647	64	NPI
POIs supplement utility and waste services	1076	1011	65	OSM
POIs converted from land-use polygons	1221	708	513	OSM

3.2.2.2 Air pollutant and explanatory factor data

In this study, the air pollutants were studied by investigating the air pollutant concentrations measured by satellites. Air pollutant concentrations, including the column densities of CO, O_3, NO_2, and SO_2 , were considered as response variables in this study (Table 3-2). These remote sensing data were accessed from the Sentinel-5P mission carried out by the European Space Agency (ESA), acquired from the Google Earth Engine (GEE) (Google Developers and the European Space Agency, 2020).

The explanatory factors for air pollutant density can be categorised into anthropological and environmental factors (Table 3-2). For anthropogenic-activity data, we used the night-time light (NTL) data and estimated population were accessed from GEE, and road density was obtained from OSM. The NTL data used in this study were monthly radiance composite images from the visible infrared imaging radiometer suite (VIIRS) day/night band information provided by the Earth Observation Group (Google Developers and Earth Observation Group, 2020). The population data were sourced from the Real-WorldPop Global Population Project of GEE (Google Developers and Worldpop, 2020). The road data were obtained from OSM open-access Big Data (Geofabrik and OpenStreetMap contributors, 2020). The size of the industry was represented by the total number of factories and employees within an industrial region. The factory counts and employee numbers were acquired from the government's open-access database (Department of the Environment and Energy, Australian Government, 2020).

The environmental factor data consisted of remote sensing data accessed from the GEE, including the digital elevation model (DEM), normalised difference vegetation index (NDVI), precipitation, and wind speed. The Australian DEM provided by Geoscience Australia, with geomorphological information, was acquired from GEE (Google Developers and Geoscience Australia, 2010). Note that Landsat-8 remote sensing products (Google, 2020) were the primary source of NDVI information, owing to their high spatial resolution. The moderate resolution imaging spectroradiometer (MODIS) NDVI products (Google Developers and the United States Geological Survey, 2020) provided the information that was missing in the data acquired from Landsat-8. The precipitation and wind speed data were accessed from the TerraClimate datasets, which provide the monthly climate information on global terrestrial surfaces (Google Developers and University of California Merced, 2020).

Table 3-2. Satellite measurement and raster data summary.

Category	Data	Spatial resolution	Temporal resolution	Statistics	Unit
Air pollutants in industrial regions	Column density of CO, O_3, NO_2, and SO_2 (Sentinel-5P)	1113 meters	Daily	Yearly average	mol/m^2
Socio-economic	Nighttime light	464 meters	Monthly	Yearly average	nanoWatts/cm^2/sr
	Population	100 meters	Yearly	-	count
Geography	DEM	31 meters	-	-	meter
Vegetation	NDVI (Landsat8)	30 meters	18 days	Yearly average	-
	NDVI (MOD13A1)	500 meters	16 days	Yearly average	-
Meteorology	Precipitation	4638 meters	Monthly	Yearly sum	mm
	Wind speed	4638 meters	Monthly	Yearly average	m/s

3.3 Methods

Figure 3-1 presents a flowchart of the method applied in this study. We adopted three major steps: industrial region identification, air pollutant and factor data processing, and air pollutant determinant factor exploration. Details of the study methods are explained in the following sections.

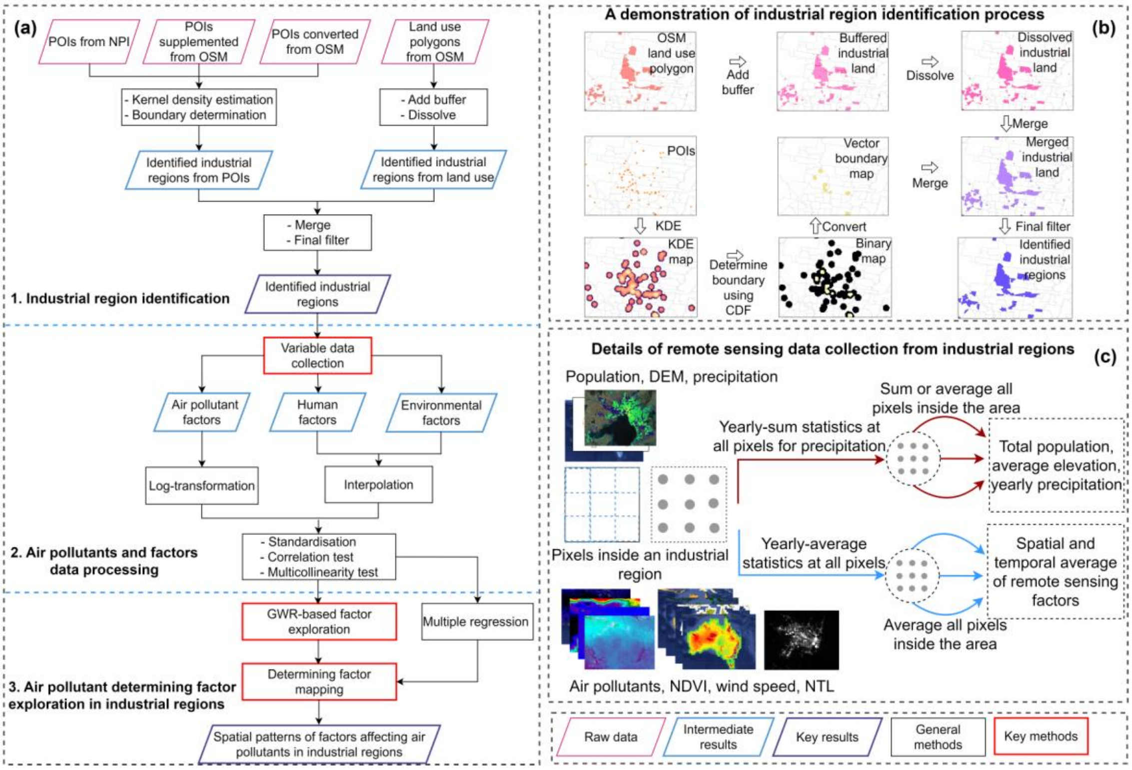


Figure 3-1. Research workflow and detailed process. (a) Research workflow. (b) A demonstration of industrial region identification process. (c) Details of remote sensing data collection.

3.3.1 Industrial region identification

3.3.1.1 Definition of industrial regions

We redefined the industrial regions serving three key industrial land use: mining, manufacturing, and utility supply and waste services (Australian Bureau of Statistics, 2021b). The definition and spatial boundary properties of the industrial regions applied in this study were consistent with the concept of functional areas adopted by the Australian Bureau of Statistics (Australian Bureau of Statistics, 2021c). Thus, industrial regions should have a dense infrastructure and must be large enough to preserve the industrial functions by providing utility and waste services and manufacturing or mining products. Note that a single industrial region should be equivalent to the industrial functional area of Statistical Area Level 2 (SA2).

3.3.1.2 Industrial region identification

In this study, the industrial regions were a combination of industrial land use and industrial areas with a high density of infrastructure. A demonstration of the industrial region-identification process is shown in Figure 3-1(b). The industrial regions were determined based on the OSM polygons and POI. Note that POI and OSM have been used previously for similar purposes in studies on urban planning (Li et al., 2019; Tu et al., 2020). A spatial methodology framework for region identification, based on POI and OSM, through kernel density estimation (KDE) and the use of geographic information systems (GIS) was adopted in previous studies (Li et al., 2018; Song et al., 2018a).

The industrial polygons acquired from OSM were a part of the industrial regions. Areas with high densities of industrial infrastructure were also considered to be industrial regions. The POI representing the facilities that supported the three key industries listed in Table 3-1 were used to identify he regions with high densities of industrial infrastructure, by KDE. In general, the KDE method is used to estimate the density of infrastructure within an area defined by a searching radius for a given kernel shape. The geographic boundary between the highly dense and non-dense areas, determined using KDE, can be identified when the cumulative density function (CDF) change is sufficiently small. An Epanechnikov kernel was used for the KDE function, while considering a theoretically lower mean square error than that of the Gaussian and uniform kernels (Chen, 2017). The KDE function was processed with a search radius of 1000 m, equivalent to the size of the SA2 functional area. The pixel size of the KDE was set at 194 m, equivalent to the finest spatial granularity of the ASGS products (Zhang et al., 2022). After processing the KDE function, the threshold value used to determine the boundaries of areas with high densities of infrastructure was 0.5%; the feasibility and effectiveness of this threshold value has been previously proven (Song et al., 2018b).

Finally, the industrial land use polygons and the areas having dense infrastructure were merged. Small-sized industrial regions were filtered after the merging process. Accordingly, the merged industrial regions of less than 0.4642 km^2—the size of which was equivalent to the smallest recorded SA2 functional areas—were not regarded as valid functional industrial regions and were, therefore, filtered. The final products of industrial regions were large enough to be functional areas and have dense industrial infrastructures.

3.3.2 Air pollutants and factors data processing

Remote sensing factors, including air pollutants, NTL, population densities, DEM, NDVI, precipitation, and wind speed, were collected from the GEE. Note that in this study, we aimed to investigate how remote sensing and spatial factors affected the air pollutant concentrations in the industrial regions of Australia in 2020. Spatial factors, including road density and industrial size, were computed using the GIS data with the OSM or NPI vector data. The detailed process of remote sensing factor generation is shown in Figure 3-1(c). Various remote-sensing factors for industrial regions were calculated using two different methods. The first category of factors, including air pollutant concentrations, NTL, NDVI, and wind speed, were represented by a spatiotemporal average for each industrial region. After accessing the remote sensing datasets, we computed the temporal average of the pixels at the same location for the entire year and obtained the yearly average of the remote sensing factor at that location. Then, we estimated the spatiotemporal average of that factor within the industrial region by computing the mean value of all the temporally averaged pixels inside the industrial region. The population, DEM, and precipitation factors were treated differently. The population was the spatial sum of all the pixels within the area, DEM was the spatial average of all the values inside the area, and yearly precipitation was the spatial average of the sum of monthly precipitation.

For the NDVI images acquired from Landsat 8, less than 0.1% of the pixels were not sampled, and these missing values were estimated and interpolated using the MODIS NDVI products at the same location. Thus, a data fusion method, based on cubist regression, was used to estimate the missing Landsat-8 NDVI values, using MODIS NDVI values with high accuracy, as proven by previous studies (Filgueiras et al., 2020). To maintain consistency with the homoscedasticity assumption in the regression, we calculated the logarithms for the CO, O_3, and NO_2 values, prior to presenting the final results.

When the factor data were processed, we standardised all the factors. The standardisation process enabled the variables to be unitless and comparable. Thus, the coefficients from the regression models were standardised coefficients, and the absolute value of the standardised coefficients implied the strength of the impacts of different factors on the air pollutant concentrations in the industrial regions (Wu et al., 2021). Furthermore, we used the Pearson's correlation coefficient to indicate the relationships between the standardised response and the explanatory variables. Then, a multi-collinearity test was performed to remove the variables containing multi-collinearity, using a variance inflation factor (VIF) threshold of 2.5.

3.3.3 Multiple regression and GWR with standardised coefficients

Multiple regression and geographically weighted regression are applied to quantitatively measure factors affecting air pollutants in industrial regions based on factors selected in the previous process. A multiple regression model, as shown in equation (3-1), is utilised to quantify the relationship between standardised air pollutant densities and potential determining factors.

$$Y_i = \beta_0 + \sum_{j=1}^{n} \beta_j X_{ij} + \varepsilon_i \qquad (3\text{-}1)$$

where Y_i is air pollutant density at location i, and β_j are standardised coefficients of selected influential factors, which is computed through the ordinary least squares (OLS). The absolute value of standardised coefficient indicates the influential strength to air pollutant density. X_{ij} refers to the j-th influential factor value at location i, and ε_i is the error term.

A geographically weighted regression with an adaptive kernel is applied to quantify the relationship between standardised air pollutant density and selected influential factors considering spatial non-stationarity. The GWR model is shown in equation (3-2).

$$Y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{n} \beta_j(u_i, v_i) X_{ij} + \varepsilon_i \qquad (3\text{-}2)$$

where Y_i refers to four types of air pollutants, X_{ij} are selected human or environmental influential factors, and ε_i is the error term. Values of $\beta_j(u_i, v_i)$ are local standardised coefficients for influential factors, and a higher absolute coefficient value indicates stronger impacts of potential factors on air pollutants.

Values of $\beta_j(u_i, v_i)$ are computed from equations (3-3) and (3-4). In our study, industrial regions are sparsely distributed across the nation. For this reason, an adaptive kernel using k nearest neighbour (KNN) is chosen to compute the weight matrix. To determine the final bandwidth value, the bandwidth selection function by minimizing the model root mean square error value is used (Fotheringham et al., 2003).

$$\beta(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \qquad (3\text{-}3)$$

$$W_{ij} = \begin{cases} e^{-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2}, if\ d_{ij} < b \\ 0, Otherwise \end{cases} \qquad (3\text{-}4)$$

The Gaussian weight kernel shown in equation (3-4) is applied to compute the spatial matrix. The adaptive bandwidth is determined by KNN. The d_{ij} refers to the distance between two industrial regions. This study uses an R-language-based “spgwr” package to determine optimal bandwidth value and analysis datasets using GWR.

In this study, spatial disparity refers to the phenomenon in which the strength of the global influential factor is overwhelmed by another factor owing to local spatial impacts. The key results, including the spatial pattern and statistical distribution of the factors affecting air pollutant concentrations, were mainly generated and interpreted, using the GWR method (with standardised coefficients), as shown in Figure 3-1(a). The input factors for both the global regression model and the local spatial regression were standardised. Thus, the strength of the factors could be compared using the absolute value of the standardised coefficients from both the models, as these coefficients were unitless and at the same scale. The global and local predominant factors had the greatest absolute value of standardised coefficients, as indicated by the standardised OLS and GWR results. Finally, the local spatial results were compared with the global outcomes to identify the places where the global factors were overwhelmed by other factor(s).

3.4 Results

3.4.1 Identified industrial regions on a continental level

The regions with POI density values greater than 1.95 were selected as the potential industrial regions. The details of the KDE industrial boundary selection are presented in Table 3-3. The top 2.5% of the regions with high POI density values, covering an area of

326 km^2 across the whole nation, was a part of the study area.

Table 3-3. Summary of KDE industrial boundary threshold selection based on POIs density CDF.

POIs density range	KDE pixel count	CDF value	Change of CDF
……	……	……	……
(1.45, 1.55]	2842	94.45%	0.82%
(1.55, 1.65]	2753	95.25%	0.79%
(1.65, 1.75]	2505	95.97%	0.72%
(1.75, 1.85]	2224	96.61%	0.64%
(1.85, 1.95]	**1938**	**97.17%**	**0.56%**
(1.95, 2.05]	1140	97.50%	0.32%
(2.05, 2.15]	778	97.72%	0.22%
(2.15, 2.25]	722	97.93%	0.21%
(2.25, 2.35]	667	98.12%	0.19%
……	……	……	……
(34.65, 34.75]	1	100%	-
Total	346906		

Notably, we identified 755 industrial regions across Australia. The industrial regions covered an area of 1827 km^2, which occupied 0.025% of the total Australian land area. The size of the industrial regions of interest in major cities ranged from 0.46–51 km^2. The industrial region size in other areas were up to 107 km^2. Figure 3-2 portrays the size and spatial distribution of the identified industrial regions. According to the statistical distribution by state, New South Wales (NSW) had the most industrial regions (in terms of counts; 226). Queensland (QLD) had 192 industrial regions (ranking second), followed by Victoria (VIC; 150). Western Australia (WA) contained 94 industrial regions and ranked fourth, almost equal to the total count in Tasmania (TAS), South Australia (SA), and Northern Territory (NT). According to the ASGS remoteness definition, 322 of these regions were clustered within major cities.

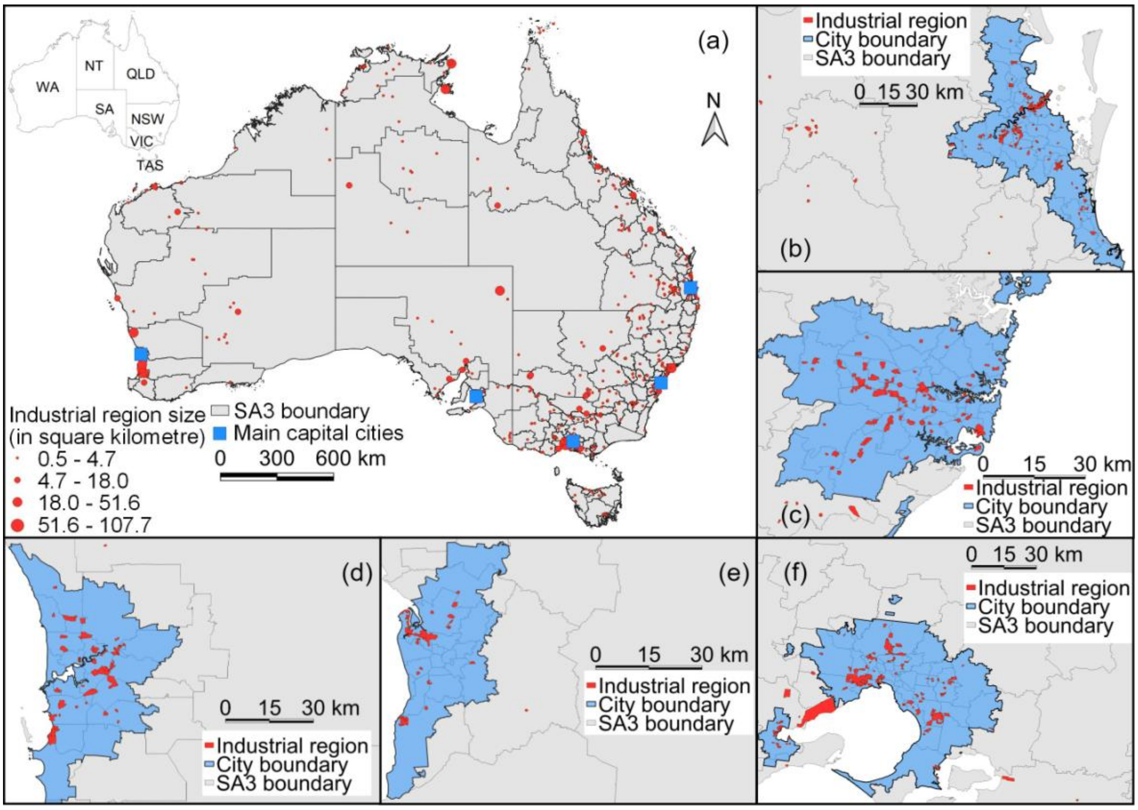


Figure 3-2. Identified industrial regions in Australia and comparison with spatial distributions of major cities and SA3 boundaries. (a) Spatial distribution of industrial regions and main capital cities. (b) Industrial regions in Brisbane. (c) Sydney. (d) Perth. (e) Adelaide. (f) Melbourne.

3.4.2 Correlation test, influential factor selection, and multicollinearity test

The correlation test results are shown in Figure 3-3. According to the factor selection process, NTL, DEM, NDVI, precipitation, and wind speed were the influential factors selected for the CO density analysis. Road density, NDVI, precipitation, wind speed, manufacturing factory count, and mining employee scale were considered as influential factors for O_3 density. NO_2 density had nine influential factors: NTL, population density, road density, DEM, precipitation, utility and waste service represented by the factory and employee scale, respectively, manufacturing factory count, and mining factory count. The SO_2 density was influenced by road density, population density, DEM, precipitation, wind speed, utility and waste service employee scale, manufacturing employee scale, and mining factory count. In terms of the multi-collinearity test, the VIF values for the selected variables (listed in Table 3-4) were all less than 2.5, which is generally acceptable for the regression models.

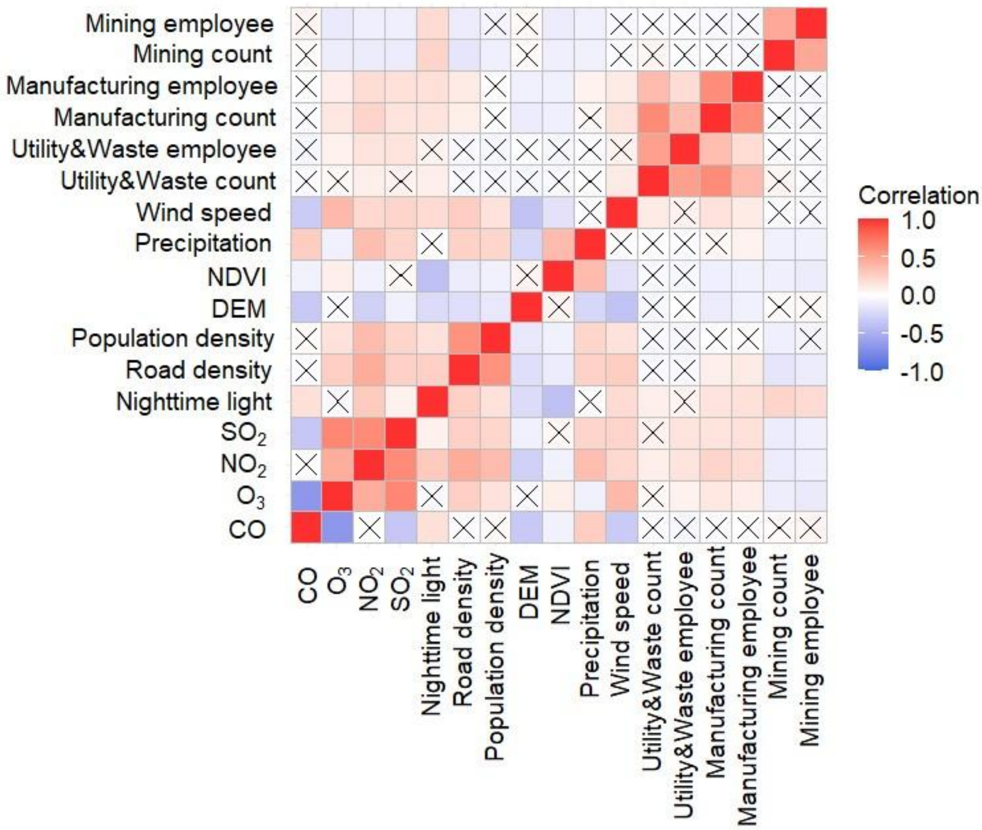


Figure 3-3. Correlation test results.

3.4.3 Air pollutant determining factor exploration

3.4.3.1 Analysis results from multiple regression and GWR with standardised coefficients

In this study, we determined the general air pollutant determining factors using OLS multiple regressions, as shown in Table 3-4. The determining factor for each air pollutant density was the factor with the greatest absolute value of the standardised coefficient in multiple regressions. In general, meteorological factors affected the air pollutant concentrations in the region more than anthropogenic activities. Wind speed was a global determining factor for CO, O_3, and SO_2, while road density was the determining factor for NO_2. Topography was the second determining factor for CO density, and NDVI was the second determining factor for O_3 concentration. Precipitation was the second most influential factor on NO_2 and SO_2. Additionally, anthropogenic activity indicators (i.e. NTL, road density, and population density) and manufacturing industry scales were positively related to air pollutant density.

Table 3-4. Multiple regression statistical results.

	CO	O_3	NO_2	SO_2
Nighttime light	0.107 (***)	-	0.182 (***)	-
Road density	-	0.264 (***)	**0.233** (***)	0.079 (*)
Population density	-	-	0.128 (***)	0.113 (*)
DEM	-0.458 (***)	-	-0.111 (***)	0.090 (*)
NDVI	-0.199 (***)	0.289 (***)	-	-
Precipitation	0.229 (***)	-0.274 (***)	0.226 (***)	0.209 (***)
Wind speed	**-0.556 (***)**	**0.351 (***)**	-	**0.218 (***)**
Utility and waste factory count	-	-	-0.109 (**)	-
Utility and waste employee	-	-	0.154 (***)	0.129 (***)
Manufacturing factory count	-	0.092 (**)	0.179 (***)	-
Manufacturing employee	-	-	-	0.099 (**)
Mining factory count	-	-	-0.083 (**)	-0.064 (.)
Mining employee	-	-0.070 (*)	-	-
p –value for F-statistic	<0.001	<0.001	<0.001	<0.001
R squared value	0.436	0.282	0.387	0.185

Note: Determining factors in global multiple regressions were **bold fonts**. Significance code: (***) for $p < 0.0001$, (**) for $p < 0.001$, (*) for $p < 0.01$, (.) for $p < 0.05$.

The spatial disparities of the determining factors, on a continental level, were identified by GWR models, as shown in Figure 3-4 and Table 3-5. Figure 3-4 portrays a statistical summary of the GWR-significant coefficients of the four air pollutant density models used in this study. Evidently, NTL, population density, and manufacturing industry scales were positively related to the air pollutant concentrations across the study area; these factors had disparities in their strength. However, road density and other environmental factors had spatial disparities in both their influential strengths and directions. Table 3-5 provides the supplementary information regarding the significant coefficient statistics and the determining factor statistics. Table 3-5 summarises the number of industrial regions, where the impact of each spatial factor was statistically significant, and the number of industrial regions, where the absolute value of the regression coefficient of each spatial factor, was statistically significant. A spatial factor was known to be predominant in an industrial region, when the absolute value of the standardised regression coefficient was larger than other factors; the standardised coefficient should be statistically significant at the 0.05 level. For CO density, topography was the most common influential factor in the Australian industrial regions, followed by precipitation and wind speed. For O_3 density, precipitation and wind speed were the two key determining factors, at a similar level (in terms of count). For NO_2 density, precipitation, topography, and NTL were the most significant determining factors. For SO_2 density, precipitation was the primary determining factor, followed by wind speed and manufacturing employee scale. From Tables 3-4 and 3-5, the spatial disparities of the air-pollutant-determining factors across the whole nation were apparent. Although wind speed and road density appeared to be the determining factors for air pollutant densities in global models, in our study, the predominant factors varied significantly.

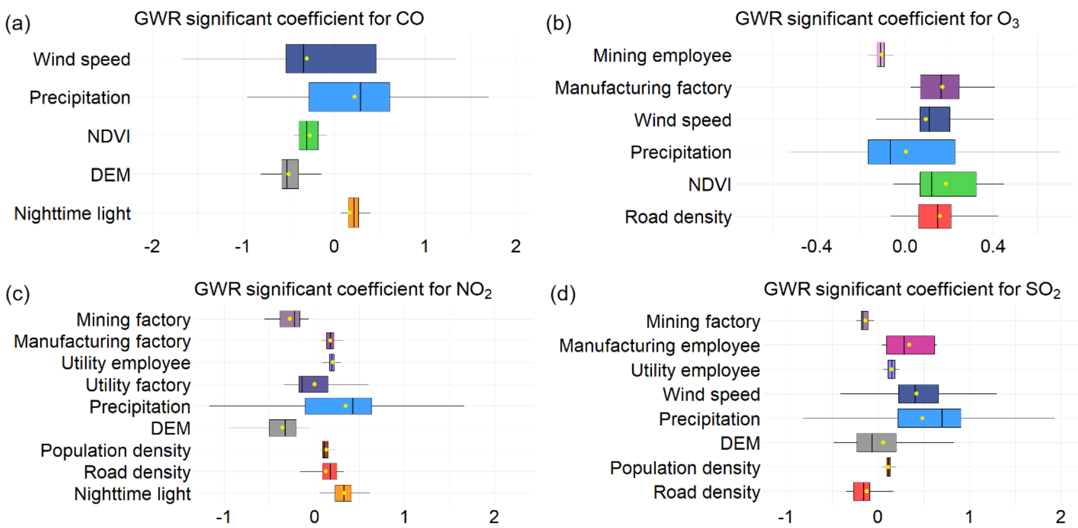


Figure 3-4. Statistical summary of significant coefficient. (a) GWR model for CO. (b) GWR model for O_3. (c) GWR model for NO_2. (d) GWR model for SO_2.

Table 3-5. Counts of GWR significant coefficient and determining factor.

	Count of industrial region where this factor is statistically significant and predominant / Count of industrial regions where the factor is statistically significant			
	CO	O_3	NO_2	SO_2
Nighttime light	6 / 43	-	**112** / 434	-
Road density	-	21 / 119	28 / 171	0 / 109
Population density	-	-	7 / 231	0 / 105
DEM	**198** / 272	-	**166** / 406	20 / 284
NDVI	9 / 66	**23** / 114	-	-
Precipitation	**64** / 158	**153** / 259	**226** / 355	**317** / 385
Wind speed	**95** / 185	**154** / 254	-	**136** / 372
Utility and waste factory count	-	-	18 / 100	-
Utility and waste employee	-	-	37 / 210	17 / 210
Manufacturing factory count	-	6 / 89	11 / 124	-
Manufacturing employee	-	-	-	**65** / 98
Mining factory count	-	-	35 / 132	7 / 72
Mining employee	-	1 / 43	-	-
Quasi-global R squared value	0.899	0.958	0.856	0.847

Note: Top three determining factors for each model in terms of count are **bold fonts**.

The multiple regression and GWR model performance are listed in Table 3-6. By comparing all air pollutant models, we concluded that the GWR had better Akaike information criterion (AIC) and residual sum of squares (RSS) values than the multiple regression. The figures in Table 3-6 indicate that the GWR model had a better goodness-of-fit and model quality. That is, the GWR method provided a better explanation of the air-pollutant-determining factors on a continental level by considering their spatial disparities.

Table 3-6. Model comparison.

	CO		O_3		NO_2		SO_2	
	OLS	GWR	OLS	GWR	OLS	GWR	OLS	GWR
AIC value	1768	660	1957	-152	1841	850	2059	878
Residual sum of squares	436	77	555	32	474	111	630	118

3.4.3.2 Air pollutant determining factor mapping: Which factor is more influential and where

Generally, the SA3 areas are higher-level spatial areas that enable a global view of regional planning (Australian Bureau of Statistics, 2021b). This study provides planning and management evidence for stakeholders based on the spatial patterns summarised at the SA3 level. In terms of count, the most frequent determining factor for industrial regions inside the same SA3 area was regarded as the determining factor of this SA3 region. The spatial patterns of factors affecting air pollutants are summarised at the SA3 level and are shown in Figure 3-5. The spatial disparities of factors affecting air pollutant concentrations across the nation are obvious and are summarised as follows.

For CO concentrations, wind speed remained a dominant factor in the inner SA, WA, NT, and northern QLD. However, CO density was dominated by topography in most parts of the continent, from the eastern coast to the inner NSW and from the southern coast to the northwest regions. The CO density was determined by precipitation and NTL in the inner WA and northern NT. For O_3 concentrations, wind speed dominated the northern WA and the inner parts of the other six states, or territories. Nevertheless, precipitation was more influential on the southern coast of WA and SA, northeast of NSW, and northern coast of NT and QLD. Vegetation greenness was dominant in central WA. Road density was dominant in a minor area of the QLD. For NO_2 concentrations, road density only dominated the inner NT and northern QLD, although it worked as a determining factor in the global regression. Precipitation was more influential in most parts of the continent. The manufacturing factory scale had a significant impact on northern WA. A variety of factors, including NTL and utility industry scales, dominated the coast of the QLD. For SO_2 concentrations, wind speed had an impact on central WA, inner NT, and a minor region of SA and VIC. However, precipitation was more influential in southern WA and most parts of SA, VIC, and NSW. Furthermore, the number of manufacturing employees was an important factor influencing the SO_2 concentrations in the northern part of the continent.

The factors affecting the air pollutants in the capital cities of Australia are summarised in Table 3-7. Meteorological factors and topography were the common attributes of air pollutants. Vegetation greenness also influenced the CO concentrations in Melbourne, and the NTL and mining factory scale had an impact on the suburbs of Melbourne. NTL also influenced the NO_2 concentrations in the suburbs of Brisbane.

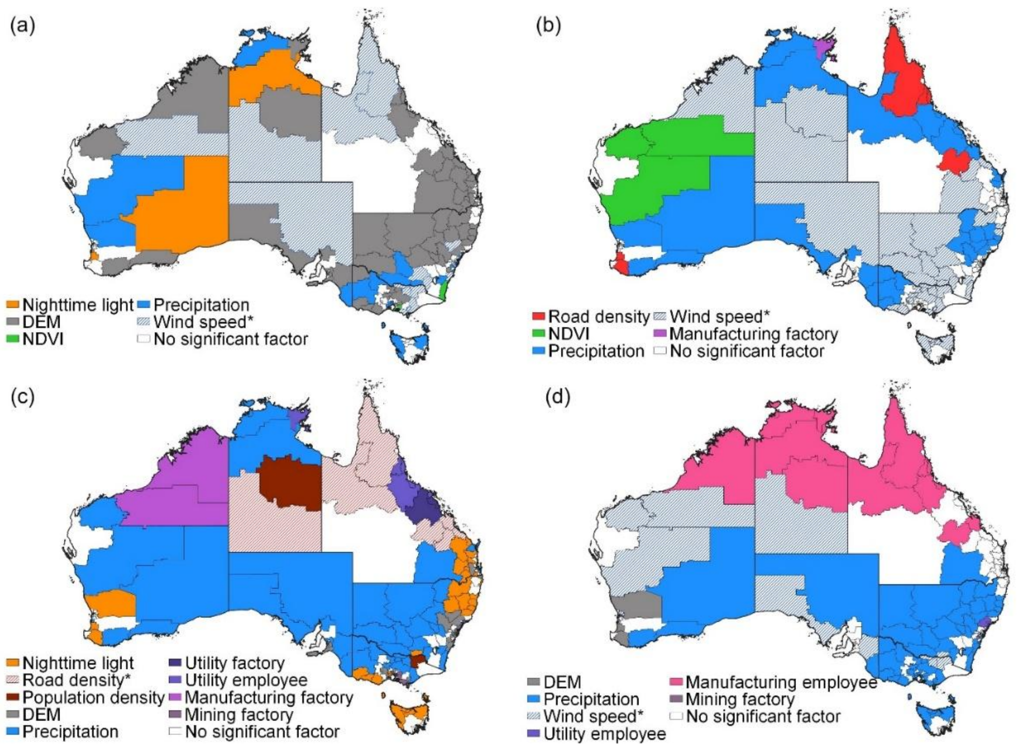


Figure 3-5. Air pollutant determining factors in different SA3 areas. (a) Determining factor map for CO. (b) Determining factor map for O_3. (c) Determining factor map for NO_2. (d) Determining factor map for SO_2. "*" means a global determining factor.

Table 3-7. Air pollutant determining factors in five major cities.

	CO	NO_2	SO_2
Sydney	Precipitation, Wind speed	DEM	Precipitation, Wind speed, Mining factory
Melbourne	DEM, NDVI, Precipitation, Wind speed	Nighttime light, DEM, Precipitation, Mining factory	Precipitation
Brisbane	DEM	Nighttime light, DEM, Precipitation	-
Perth	DEM	Precipitation	Wind speed
Adelaide	DEM	DEM	-

Note: No significant determining factors for O_3 density in all five major cities.

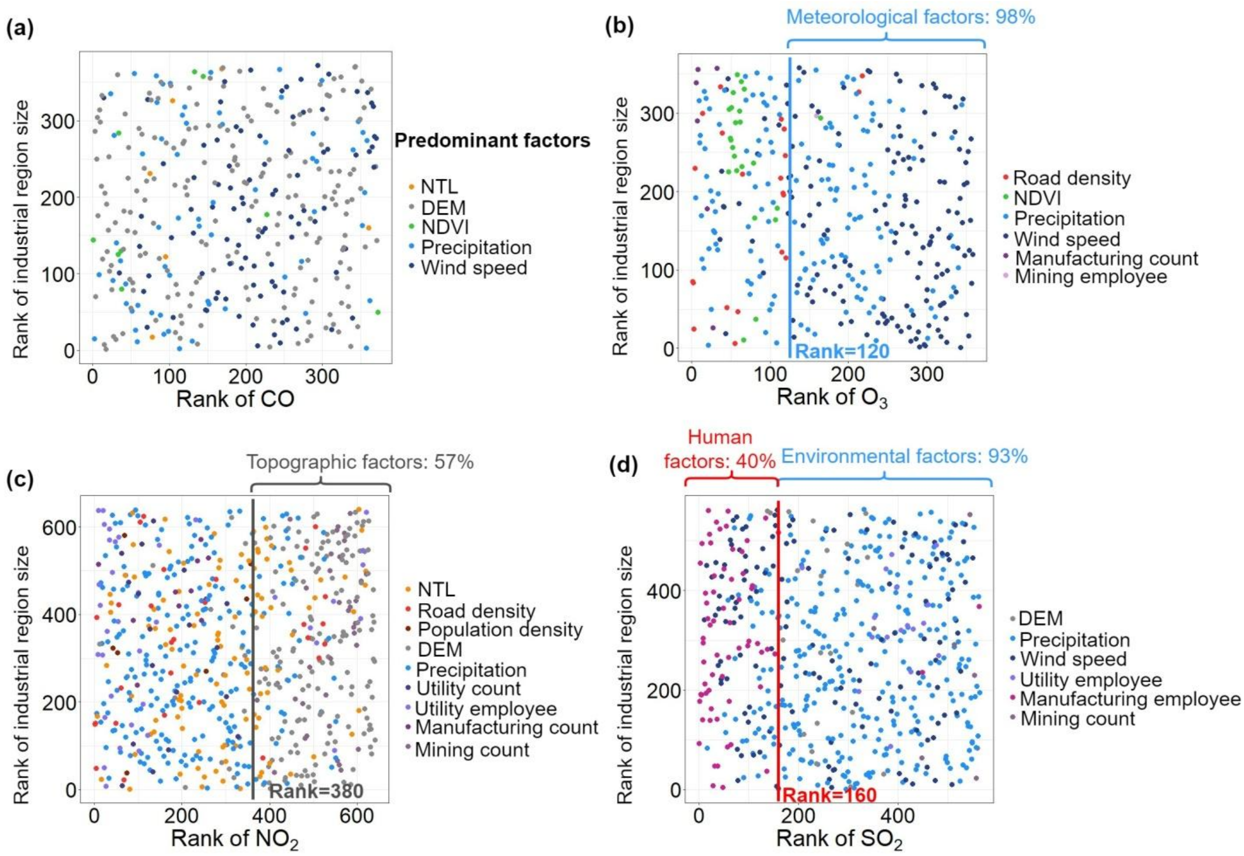


Figure 3-6. Distributions of predominant factors under the rank of air pollutant concentrations and the rank of industrial region size. Distribution of factors affecting (a) CO, (b) O_3, (c) NO_2, (d) SO_2

The distributions of the factors affecting the air pollutant concentrations in the study area are shown in Figure 3-6. The scatter plots in Figure 3-6 portray the distribution of various predominant factors in ascending ranks of their influence on air pollutant concentrations and the industrial region size. The levels of air pollutant concentrations predominantly effective in industrial regions are summarised in the scatter plots. Evidently, environmental factors, especially meteorological factors, influenced the higher concentrations of O_3 and SO_2 in industrial regions. More than 90% of the higher air-pollutant concentrations in the industrial regions were affected by non-anthropogenic factors. In terms of the concentrations of NO_2, the topographic factors were more strongly associated with the air pollutant concentrations in the industrial regions.

3.5 Discussion

3.5.1 Spatial factors affecting air pollutant density on a continental level

In this study, we investigated the spatial disparities in the factors that affected the air pollutant concentrations in the industrial regions of Australia. The case study demonstrated evident spatial disparities in the determining factors. Meteorological attributes and topography were the dominant factors that influenced the air pollutant densities in most industrial regions; however, the information on anthropogenic factors and their spatial patterns is non-negligible. Note that this study is the first study to determine the factors of air pollutant concentrations in industrial regions at a continental scale.

General findings include the unitary relationships between air pollutant densities and anthropogenic factors. Notably, the study indicates that anthropogenic activities, including NTL and population density, and the industry scales of manufacturing, utility supply, and waste services are positively related to the air pollutant concentrations. These positive relationships were statistically significant in both the nationwide and the local industrial areas, mostly coinciding with the outcomes of studies on urban industrial regions. During urban expansion, NTL (Yue et al., 2020), population density (Liu et al., 2016; Borck and Schrauth, 2021), and industrial land-use scales (Cai et al., 2020) are positively related to air pollutant densities. As part of human settlements, industrial regions follow a similar pattern.

Local results from GWR models, indicating non-unitary relationships between air pollutants and influential factors, were consistent with previous studies. Meteorological factors, including wind speed and precipitation, affected the air pollutants in various numerical directions at different locations. This local spatial variance was observed in a previous study that investigated the correlation between industrial air pollutants and their influential factors (Yang et al., 2019). Road density was positively related to NO_2 density, but not to SO_2 density. The positive correlation with NO_2 density was consistent with the air-pollutant monitoring models used in previous studies (Hoek et al., 2008; Meng et al., 2015; Zhai et al., 2018). The non-unitary relationship with SO_2 may be due to the Environmental Kuznets Curve (EKC) effect. Road density was mainly negatively related to SO_2 density in VIC, where the road infrastructure was well-developed and denser than that in other places (Geofabrik and OpenStreetMap contributors, 2020; Vicroads, 2021). Anthropogenic activities at the primary stage can lead to environmental degradation, while post-development anthropogenic activities, being highly invested, would have the opposite effect (Erdogan, 2020; Guo et al., 2021). Therefore, the EKC hypothesis could be a reason leading to the non-unitary road density-SO_2 relationship.

3.5.2 The necessity of studying air pollutants based on specific industrial land uses

Satterthwaite (2008) regarded industrial land use, rather than administrative boundaries, as an exact geographical feature. According to the 2020 NPI report, approximately 98% of NO_2 and SO_2 is emitted from industrial regions from three key industries. Previous air pollutant monitoring and relevant environmental justice study projects paid more attention to urban areas or administrative boundaries (Cooper et al., 2019; Haddad and Vizakos, 2020). Nevertheless, planning advice based on exact emission sources would be effective, and policy-makers, planners, and study teams are advised to pay more attention to the impact of human forces on industrial land use, when monitoring air pollutants, as industrial impacts in remote areas are sufficiently large and thus, should not be underestimated.

3.5.3 Limitations

This study has some limitations. The first limitation is the existence of heteroscedasticity in some measurement data shown in Figure 3-7, such as that of SO_2.

Additionally, the potential estimation residuals in the population count were difficult to calculate, due to dynamic changes. Considering the changes caused by migration, birth, death, and other reasons, the population at a fine spatial granularity would vary in different statistical years. Therefore, to estimate the population in industrial regions, future studies may use the Real-WorldPop products for the year 2020.

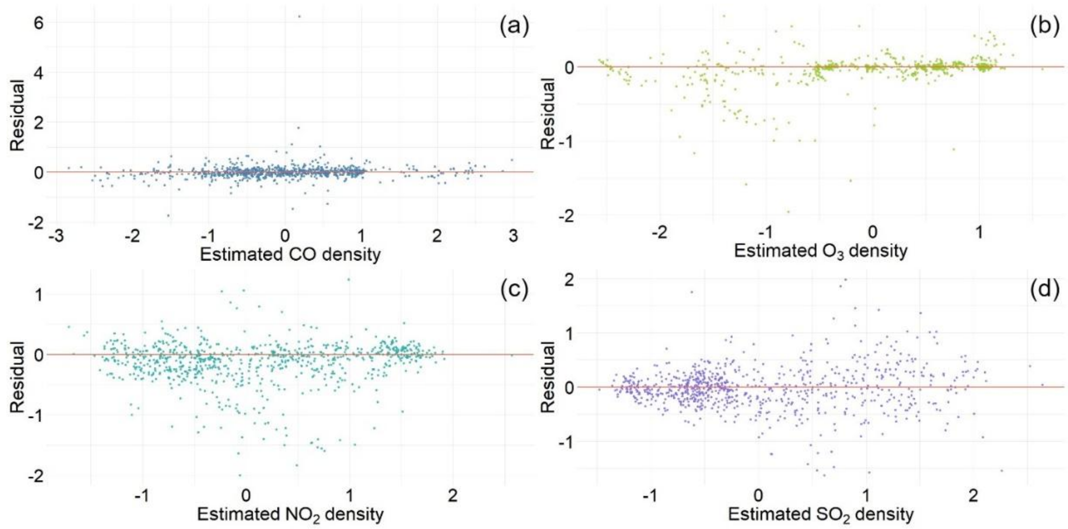


Figure 3-7. Residuals vs. fitted plots for GWR models. (a) Residual plot of CO density. (b) Residual plot of O_3 density. (c) Residual plot of NO_2 density. (d) Residual plot of SO_2 density. Note: GWR made estimations on log-transformed CO, O_3, and NO_2 density.

3.6 Conclusion

The association between air pollutant concentration and industrial development has been reported in several previous studies. However, the spatial patterns of the factors that indicate the internal properties of industrial regions, which can affect air pollutant concentration, remain un-explored. In this study, we developed a novel set of methods, wherein we included a specific land use identification method and GWR (with standardised coefficients), to identify the industrial regions in Australia as the exact study areas, collect the information of potential factors that may affect air pollutants (using remote sensing data), and assess the spatial disparity of the factors that affect the air pollutant concentrations in the industrial regions. Our results demonstrate evident local spatial impacts on the air pollutant concentrations in continent-level industrial regions.

Notably, anthropogenic factors influenced the air pollutant concentrations in the remote industrial lands, especially in north Australia, and meteorological and topographical factors significantly impacted the concentrations in urban industrial regions. Furthermore, within the nationwide industrial land use systems, higher concentrations of O_3 and SO_2 were more associated with meteorological factors of the area, while the higher concentrations of NO_2 were more related to the topographic features of the region. In this study, we explored the spatial features of the factors that affect the air pollutant concentrations in industrial regions, while providing the results for specific land uses. Notably, our study can serve as a reliable framework for future studies on the air quality of industrial regions, while providing viable suggestions for the environmental and spatial management of industrial lands.

Chapter 4. Extended exploration of environmental sustainability of industrial regions: spatial associations between air pollutants and influential factors indicated by robust geographical detector

4.1 Introduction

This chapter aims to demonstrate the advanced spatial associations between industrial sustainability indicators and influential factors based on spatial stratified heterogeneity. To address the limitation of the discretisation strategy for spatial stratified heterogeneity, this chapter presents a robust geographical detector (RGD) to effectively explore more reliable and robust spatial associations between dependent and explanatory variables from a spatial heterogeneity perspective. RGD determines discretisation interval breaks using an optimisation algorithm for variance-based change point detection (CPD). In this study, RGD was implemented in a nationwide case study to explore the potential factors of nitrogen dioxide (NO_2) density in industrial regions across Australia, wherein data on both NO_2 and potential factors were sourced from satellite images and remote sensing (RS) products using Google Earth Engine (GEE). A sensitivity analysis was performed to evaluate the effectiveness of RGD in exploring spatial associations.

4.2 Robust geographical detector (RGD)

RGD is an improvement of the geographical detector (GD), with optimal spatial zones determined using the optimisation of spatial data discretisation of explanatory variables. Figure 4-1 shows the process of the RGD model for spatial determinant exploration, which includes four steps. The first step is the equivalence transformation for RGD using a ranking approach, which guarantees the measure of spatial stratified heterogeneity (SSH) and creates opportunities for solving an optimisation problem. The second step is to redescribe the objective of spatial discretisation as an almost-solved optimisation problem. This means that the process of identifying breaks of spatial discretisation is transformed into a CPD problem, where change points of explanatory variable ranks are identified using a dynamic programming method, and within the sum of squares (SSW) is minimised using a least squared deviation cost function. The third step was to calculate the power of determinant (PD) values of explanatory variables. In RGD, a B-value is used to quantify the PD of variables, with the detected robust change points determined by spatial zones. Finally, the sensitivity of the RGD was evaluated by comparing it with previous GD models. In this study, RGD was implemented to explore the spatial determinants of air pollutants in industrial regions of Australia, as described in Section 4.3.

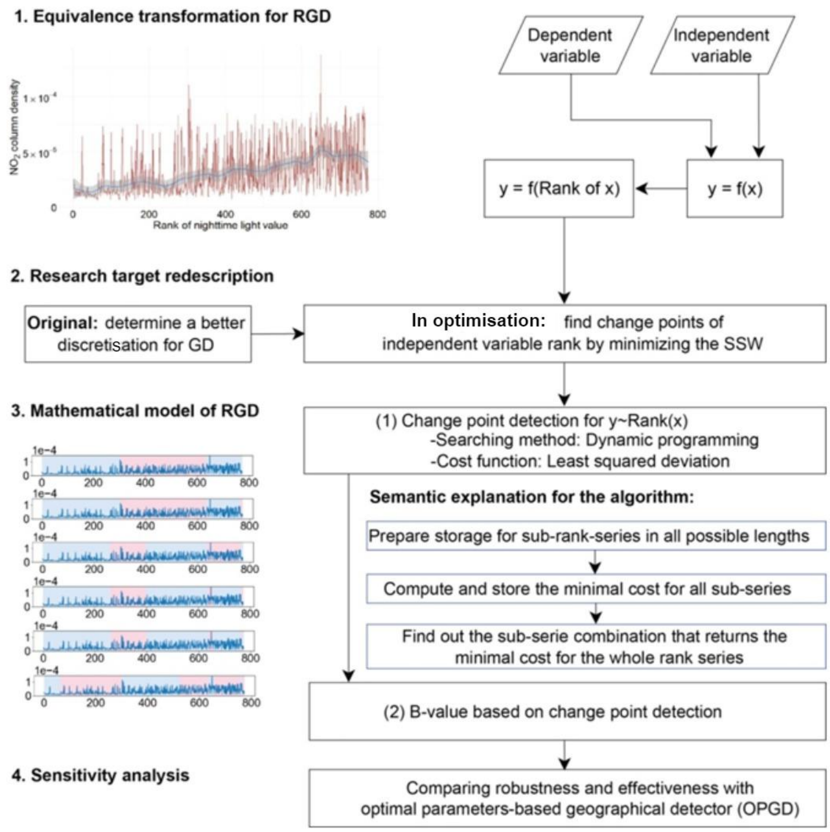


Figure 4-1. Process of using RGD model for determinant exploration.

4.2.1 Equivalence transformation for RGD

RGD is a variant of GD with a robust optimisation discretisation strategy for a more effective estimation of spatially stratified heterogeneity. The PD of the explanatory variables is computed as a B-value in RGD, as shown in equation (4-1).

$$B = 1-\frac{SSW^R}{SST}= 1-\frac{\sum_{z=1}^{k} N_z \sigma_z^2}{N\sigma^2} \qquad (4\text{-}1)$$

where SSW^R is the sum of squares within spatial zones identified using the robust optimisation strategy of spatial data discretisation for explanatory variables, SST is sum of squares total of observations in the whole study area, *z* is an RGD spatial zone, N_z and

σ_z^2 are the number and variance of observations within zone z by discretising an explanatory variable, respectively, and N and σ^2 are the number and variance of data in the whole study area, respectively. Similar to the Q-value in GD, the B-value measures the spatial association between dependent and explanatory variables and ranges from 0 to 1.

The basic idea for equation (4-1) in the previous optimal parameter-based geographic detector (OPGD) quantile method is how much of the spatial heterogeneity of the dependent variable can be explained by dividing the sorted explanatory factor value (Song et al., 2020), which is identical to dividing the rank of an explanatory factor value. RGD quantifies the extent to which the spatial heterogeneity of a dependent variable can be explained by the ranks of explanatory variables, instead of the values of explanatory variables in GD. Spatial zone z across the space in equation (4-1) is determined by discretising the numerical continuous explanatory variable, which means that a division of the study area to show the spatial heterogeneity of the dependent variable is based on the segmentation of sorted and ranked explanatory variable series from the observed minimum to maximum. If dependent and explanatory variables are spatially associated, there should be a consistent mathematical relationship between the sorted dependent variable value (A) and the sorted rank of this explanatory variable (B), meaning that the mapping from A to B is a bijection with a simultaneous increase. Discretisation of explanatory variable for RGD is equivalent to categorising the sorted rank of explanatory variables. Thus, the process of determining spatial zones using explanatory variables can be converted into that of determining spatial zones using the ranks of explanatory variables, which is a robust approach without the impact of outliers and extreme values, to calculate the PD values based on spatially stratified heterogeneity.

Therefore, a rank transformation has two advantages: it can both guarantee the measure of the SSH value based on equation (4-1) and work as an input for the optimisation algorithm, simultaneously. Considering these advantages, RGD accepts equivalence transformation and investigates how much of the spatial heterogeneity of the dependent variable can be explained by the rank of the explanatory variable. This transformation converts the original distribution of explanatory variables into a sequence of natural numbers, starting from a value of one to the number of total observations. Even with no direct computing advantage, transformed relationships (i.e. the relationship between the dependent variable and rank of the explanatory variable) can be treated as a simulated signal recorded within a time series, which can be further transformed into an optimisation problem.

4.2.2 Research target redescription

In previous studies, great efforts have been made to improve GD through various methods of spatial discretisation; however, deriving an explicit mathematical approach for reliable and robust modelling is still a challenge. The RGD provides a robust solution to address this issue. In RGD, the research target can be stated more clearly after equivalence transformation because the explanatory variables are continuous sequences of natural numbers. The relationship between the dependent variable and the rank of the explanatory variable is equivalent to a simulated offline signal series, where the dependent variable acts as a signal pulse and the explanatory variable rank is the time series. As previous GD discretisation strategies do not fully explore the relationship between variables when determining the segmentation point and generating spatial zones z, RGD treats minimising SSW as an optimal target to segment the transformed signal series. Determining a better discretisation for GD can be rephrased by a clear optimisation problem. Given a simulated signal series, it is possible to find a specified number of segmentation points that can have the least SSW for a dependent variable. The answer to this question is ‘yes’, and the solution is CPD (Page, 1955; Truong et al., 2020), which is introduced in the subsequent section.

4.2.3 Mathematical model of RGD

The mathematical model of RGD is composed of CPD for simulated signal series generated from equivalence transformation for variables and B-value derived from change-point segmentation. CPD is a method used to detect the time points of a signal series where significant specified types of changes occur. CPD is composed of the cost function, searching method, and constraints. The cost function defines the types of change to be detected, and this function is also the optimisation target. To minimise the SSW, the least squared deviation cost function was selected for RGD. The search method is a computing strategy for finding the required change points, and RGD selects a suitable search method to overcome past limitations. The previous GD discretisation method generated a fluctuating spatially stratified heterogeneity value in equation (4-1) with an increase in the interval number. From a computer science perspective, there is a lack of learning experience from the relationship between response and explanatory variables and previous segmentation outcomes that lead to unexpected phenomena. Dynamic programming is based on the view that an optimal solution to the original problem comprises several optimal solutions toward overlapping sub-problems. In CPD, dynamic programming divides the task of finding a specified number of change points into multiple tasks to find fewer required points. To determine the *K* number of break intervals, the dynamic programming-based CPD algorithm executes (*K*-1) times and finds a change point that meets the optimisation goal once a time. Each subprocess of the *K* interval determination task starts from the last executed optimised results. By minimising SSW as the primary optimisation target in each searching sub-task, this bottom-up computing strategy guarantees the increment of the spatially stratified heterogeneity value with the increase in the specified interval number. Therefore, a dynamic programming search method was selected to determine the required segmentation points. There is no CPD constraint because the number of intervals can be specified by users based on their needs. The key concepts of RGD have been translated into the following pseudocodes:

Algorithm: Change point detection for RGD – dynamic programming and least squared deviation

---

Input: simulated signal series, denoted by $\{y_x\}_{x=1}^{N}$; cost function $c(y_I) = \sum_{x \in I} ||y_x - \bar{y}||_2^2$; specified number of spatial zone *K* (no less than 2); lists to record costs within each interval given interval number from 1 to *K*-1, denoted by C_1, C_2, …, C_{k-1}; and an empty list denoted by *L*, with a length of *K* [the top (*K*-1) elements are to store segmentation points for generating spatial zones, and the last element is the number of observations].

for all rank series pairs (p, q), $1 \le p < q \le$ N (number of observations) **do**

    $C_1(p, q) \leftarrow c(\{y_x\}_{x=p}^{q})$

end for

for all j from 2 to K-1 **do**

    for all rank series pairs (p, q), $1 \le p < q \le$ N, $q - p \ge j$ **do**

        $C_j(p, q) \leftarrow \min_{p+j-1 \le x < q} (C_{j-1}(p, x) + C_1(x+1, q))$

    end for

end for

$L[K] \leftarrow$ N; $j \leftarrow K$

10 **while** $j > 1$ **do**

11 $\quad m \leftarrow L(j)$

12 $\quad x^* \leftarrow \mathrm{argmin}_{j-1 \le x < m} (C_{j-1}(1, x) + C_1(x+1, m))$

13 $\quad L[j - 1] \leftarrow x^*$

14 $\quad j \leftarrow j - 1$

15 **end while**

16 **return** list L

Note: The transformed rank observation series and simulated signal series as the algorithm inputs refer to the same processed sequence of data.

The algorithm provided above indicates how RGD intervals are determined using CPD, while minimising SSW as the optimisation target. The algorithm from lines 1 to 8 is a preparation for dynamic programming searching, which is composed of two key steps. The algorithm from lines 1 to 3 is used to prepare the storage memory for all possible lengths of the sub-series. With the least squared deviation as the cost function, the algorithm from lines 4 to 8 computes the cost for all subseries. The remainder of the algorithm demonstrates the dynamic programming search process. This algorithm returns a vector of segmentation points using a dynamic programming search. K-1 change points divide the explanatory variable value range into K groups. It is worth mentioning that algorithm line 12 is the process to find the optimal combination of sub-series, which minimises the cost function by searching all possible lengths of the sub-series. When there is an outlier, to minimise the cost function, CPD will detect the extreme value or outliers and categorise outliers into a new group if it can minimise the cost. Then, the explanatory variables are discretised and categorised based on the segment value range. Finally, the spatially stratified heterogeneity value for RGD is calculated using equation (4-1) and is denoted as the B-value. To distinguish between the terms for OPGD, we regard break intervals determined by RGD as robust geographic zones.

4.3 Advanced spatial associations between air pollutants and influential factors in industrial regions

In this chapter, NO_2 and potential variables that affect NO_2 distributions in the identified industrial regions were collected from satellite images and RS products using the GEE platform (Google Developers and the European Space Agency, 2020). Table 4-1 shows a summary of the data used in this case study. The air pollutant NO_2 density was the dependent variable in this case study. Sentinel-5P products from the European Space Agency (ESA) provide satellite measurements of the NO_2 column density. The NO_2column density data in this case study were collected and processed using the GEE.

Table 4-1. Summary of remote sensing datasets and factors

Data	Variable	Spatial resolution	Temporal resolution	Unit
Sentinel-5P Nitrogen Dioxide	NO_2 density	1113 meters	Daily	mol/m^2
GCOM-C/SGLI L3 product	LAI	4638 meters	8-day	m^2/m^2
MODIS Combined EVI	EVI	463 meters	16-day	-
Landsat8	NDVI	30 meters	18-day	-
TerraClimate climate data	Wind speed	4638 meters	Monthly	m/s
VIIRS Day/Night Band	Night-time light	464 meters	Monthly	nanoWatts/cm^2/sr

In addition, a series of potential variable data were collected to explain the spatial pattern of NO_2 in industrial regions. Vegetation was represented by the normalised difference vegetation index (NDVI), leaf area index (LAI), and enhanced vegetation index (EVI). High spatial resolution NDVI information was accessed from the Landsat8 collection provided by Google (Google, 2020). LAI information was derived from GCOM-C/SGLI Level 3 spatially and temporally averaged products from Global Change Observation Mission-Climate and provided by Google (Google Developers and Global Change Observation Mission, 2020). Moderate resolution imaging spectroradiometer (MODIS) combined 16-day EVI information was accessed from the GEE (Google, 2020). Wind speed is a meteorological factor that influences NO_2 density. Wind speed information was generated from the TerraClimate datasets accessed from the GEE platform (Google Developers and University of California Merced, 2020). TerraClimate has been used in spatial research since 1958 to monitor global climate with high spatial resolution. Night-time light (NTL) is an explanatory variable that represents human activity in industrial regions. NTL data were remotely sensed by the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band and provided by the Earth Observation Group and Colorado School of Mines (Google Developers and Earth observation Group, 2020). VIIRS provides monthly updated NTL data and can be accessed from the GEE platform.

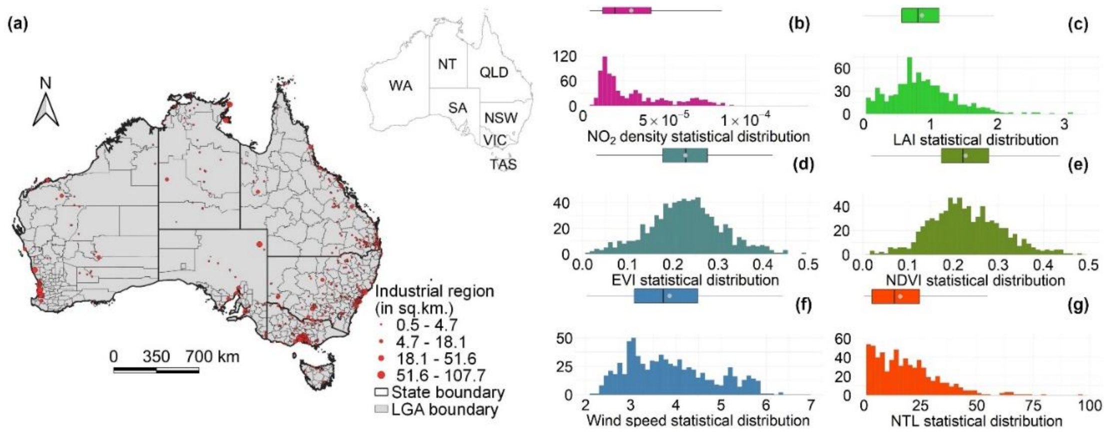


Figure 4-2. Study area and data summary. Size and spatial distribution of Australian industrial regions (a), the statistical distribution of dependent variable NO_2 column density in industrial regions (b), and statistical distributions of explanatory variables leaf area index (LAI) (c), enhanced vegetation index (EVI) (d), normalised difference vegetation index (NDVI) (e), wind speed (f), and night-time light (NTL) (g).

4.4 Results

4.4.1 B-value for air pollutant and sensitivity analysis of RGD

In RGD, the B-values of variables are used to quantify the spatial associations between the dependent and explanatory variables. Figure 4-3 shows the B values of the variables affecting NO_2 density in the industrial regions of Australia. The B-values of the variables generally increase with an increase in the number of intervals for spatial data discretisation and determining spatial zones. However, the increase rates of the B-values gradually decrease with an increase in the number of intervals. Thus, the optimal number of intervals for discretisation was selected when the change rates were lower than 0.05, which has been used in a series of previous studies on spatial discretisation (Song et al., 2020, 2021a; Luo et al., 2021, 2022). The results showed that the optimal discretisation intervals for LAI, EVI, NDVI, wind speed, and NTL were 9, 9, 10, 9, and 5, respectively. Figure 4-3 (f) shows the B-values of the variables with optimal discretisation. The NTL had the highest contribution to the spatial patterns of NO_2 density in the industrial regions, with a contribution of 0.260 ($p < 0.01$). The contributions of wind speed, LAI, EVI, and NDVI were 0.161 ($p < 0.01$), 0.161 ($p < 0.01$), 0.149 ($p < 0.01$), and 0.097 ($p < 0.01$), respectively. This indicated that in industrial regions, industrial and human activities contribute more to the spatial pattern of air pollutants than climate and vegetation variables.

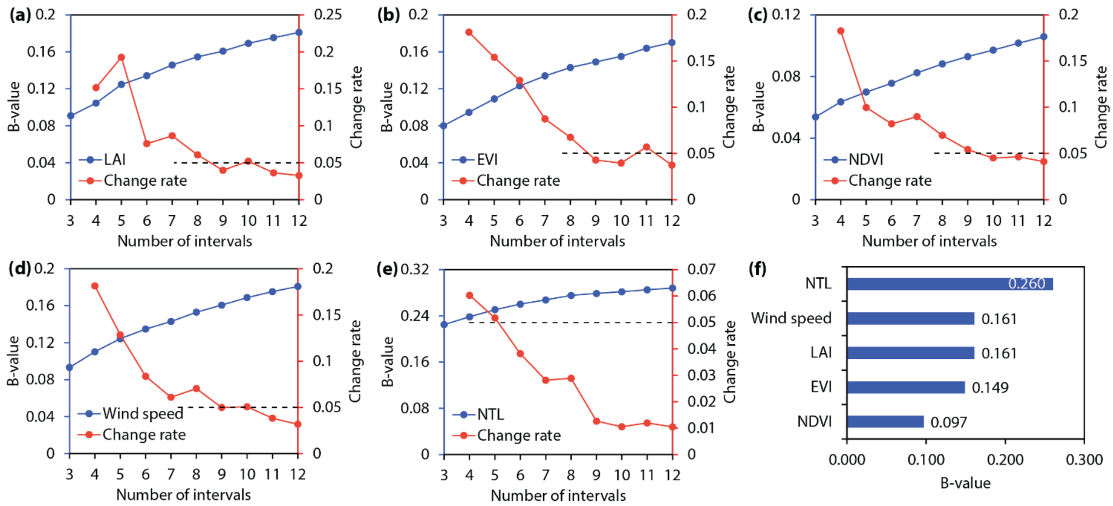


Figure 4-3. Process of selecting optimal numbers of intervals for the robust spatial discretisation for variables LAI (a), EVI (b), NDVI (c), wind speed (d), and NTL (e) in RGD, and B-values of variables in affecting NO_2 density in industrial regions (f).

The robustness and reliability of RGD for exploring spatial associations and potential variables were evaluated by comparing it with OPGD, which is an improved, effective, and commonly used GD model. The sensitivity of RGD and OPGD was assessed with different numbers of intervals ranging from 3 to 12, which were used to determine the spatial zones. Figure 4-4 shows the comparison of PD values of the five explanatory variables explored by RGD and OPGD, which have been computed as B-values and Q-values, respectively. The results show that RGD is more effective and reliable than OPGD models in exploring spatial determinants. The advantages of RGD include the following aspects.

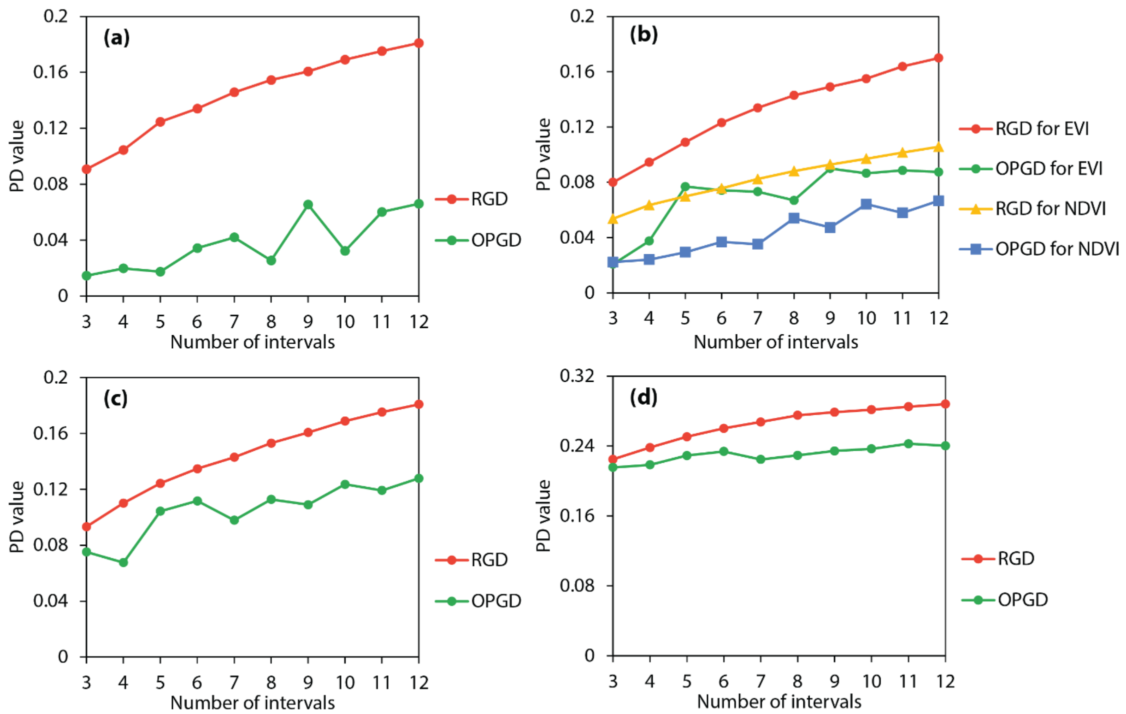


Figure 4-4. Comparisons of power of determinant (PD) values, including B-values from RGD and Q-values from OPGD, of variables LAI (a), EVI and NDVI (b), wind speed (c), and NTL (d). Note: All PD values shown in this figure have a statistical significance level with p-value lower than 0.01.

First, the comparison between B-values and the corresponding Q-values with identical numbers of intervals shows that RGD can determine better spatial zones that can enable stronger spatial associations between the dependent variable and explanatory variables. Second, RGD guarantees the increment of the B-value with the increase in interval break for explanatory factors, whereas the Q-value from OPGD fluctuates. This phenomenon also confirms the robustness of RGD in assessing spatial associations. B-values based on a higher number of intervals would have more dynamic programming searching processes to find the required breakpoints. The performance of RGD was consistent with the model expectation. In detail, OPGD quantified the PD value for wind speed from 0.07 with 3 intervals to 0.12 with 12 intervals, and RGD calculated the PD value from 0.09 to 0.18 for 10 intervals. The RGD-based B-value of the NTL increased from 0.22 to 0.28 with an increase in the interval number, which was generally higher than the corresponding Q-values from OPGD. It is worth mentioning that the PD value of LAI was at least 200% improved by RGD, when comparing the B-values and corresponding Q-values. This finding is discussed in detail in the next section. For EVI and NDVI, RGD improved by no less than 100% of the PD value compared with the OPGD results. In summary, NTL had the highest spatial association with NO_2 density in the Australian industrial region, followed by wind speed and LAI.

4.4.2 Analysis of RGD-based robust spatial zones

In addition to the robustness of RGD shown in the previous section, RGD provided robust spatial zones in terms of the comparison of PD ranks of variables between RGD and OPGD. A discretisation method is essential for numerical variables prior to presenting PD values from RGD or OPGD, and how the explanatory factor is discretised has a significant impact on PD values and result interpretation. We give an example of the significant impact issue in Figure 4-5. When assessing OPGD-derived Q-values only, wind speed seemed to have the second strongest PD with the dependent variable, while LAI's PD ranked the lowest. In both Figure 4-3 and Figure 4-4, OPGD regards LAI as a factor with a relatively low association with NO_2 density in industrial regions. However, the RGD analysis results indicate that the spatial association between LAI and NO_2 density is notably underestimated using previous methods. According to Figure 4-3 and Figure 4-4, the spatial association of NO_2 density with LAI is as strong as that of wind speed when robust geographic zones determined by RGD are used. The PD values of LAI and wind speed were 0.12 with 5 intervals and 0.18 with 12 intervals, respectively. The RGD method does not make a special treatment for LAI but manages to find suitable intervals with minimised SSW for driving factors to be tested.

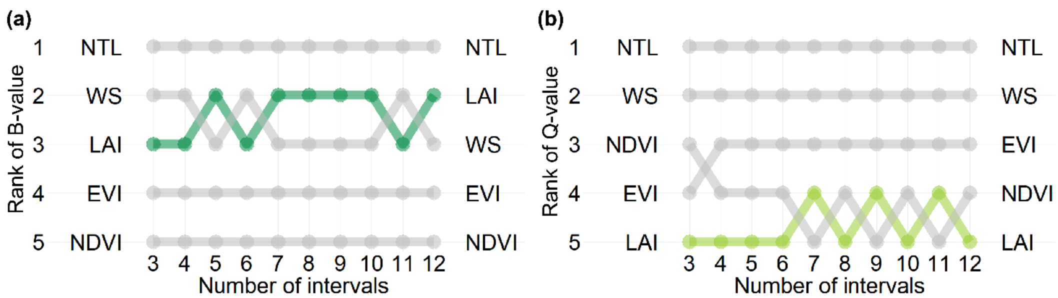


Figure 4-5. Ranks of PD values, including RGD-based B-values (a) and OPGD-based Q-values (b).

Furthermore, a map demonstrating robust spatial zones for LAI determined by RGD was generated, as shown in Figure 4-6. We present a case with an interval of five.

Five spatial zones were noted from ‘A’ to ‘E’, corresponding to the LAI intervals from the lowest to the highest. According to the Australian Remoteness Structure, group ‘A’ is distributed in nonurban regions. Group ‘B’ has the least group elements located in Sydney, Melbourne, and Adelaide. ‘C’ category industrial regions are sparsely distributed across the nation. ‘D’ and ‘E’ groups refer to clustering in urban and inner regional areas, respectively. As shown in Figure 4-6(f), statistical summaries indicate that urban industrial regions have NO_2 density at multiple levels, and rural industrial regions categorised in group ‘A’ maintain a low NO_2 density level.

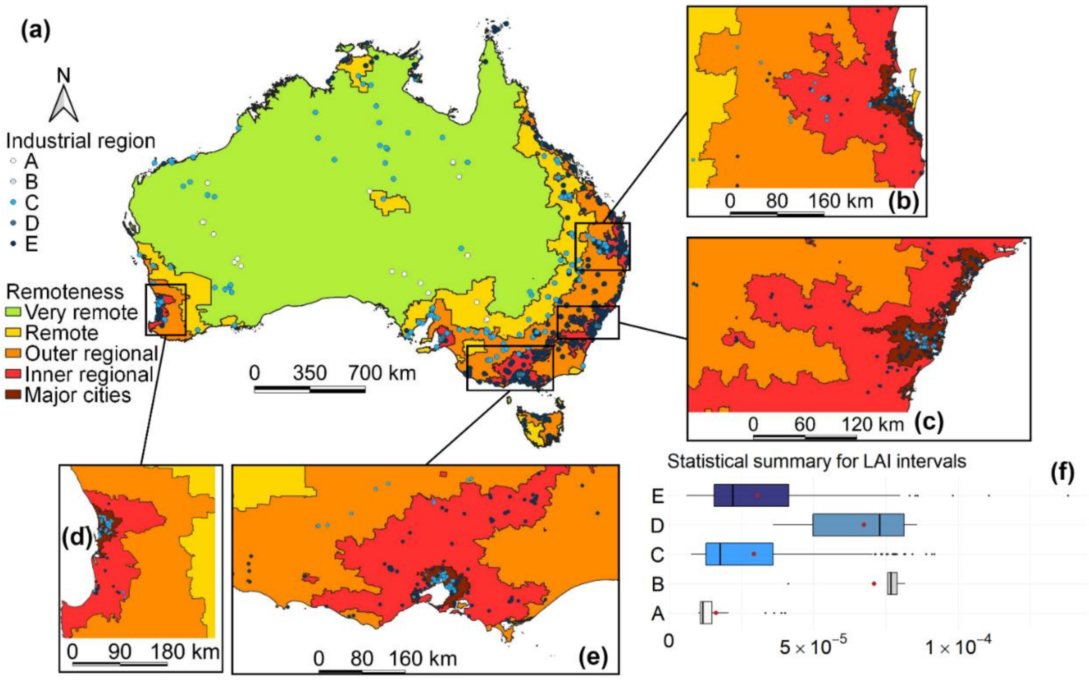


Figure 4-6. Robust spatial zones for LAI determined by RGD. (a) Spatial distribution of LAI robust geographical zones with 5 intervals. (b) Brisbane. (c) Sydney. (d) Perth. (e) Melbourne. (f) Statistical summary of association of NO_2 column density with LAI of spatial zones.

4.5 Discussion

4.5.1 Contribution

In this study, we proposed an RGD model for identifying spatial determinants through the optimisation of spatial data discretisation, deriving robust spatial zones, and exploring robust spatial associations. GD and its improvements, such as OPGD, are widely used approaches for measuring PD values of explanatory variables in terms of spatially stratified heterogeneity. The selection of a spatial data discretisation strategy can critically affect the measurement of PD values and interpretation of results. However, discretising continuous numeric variables in an effective, robust, and reliable manner remains a challenge. This study demonstrates that the developed RGD model can provide a robust solution for estimating spatial associations between dependent and explanatory variables. RGD has the following advantages for exploring spatial associations. First, by using the robust optimisation algorithm, RGD can explore the maximum spatial associations, which are much higher than the PD values explored by the OPGD models. Second, RGD guarantees the increment of PD values with the increase of interval numbers, as the optimisation processes can also be extended with the interval increase, which is hardly ensured in previous GD models. Third, RGD is robust for explanatory variables with different statistical distributions. In most previous spatial heterogeneity models, assumptions of statistical distributions of data are required, such as a normal distribution in geographically weighted regression, and modelling is affected by outliers. Compared with the OPGD method, which uses only sorted information, RGD further utilises and explores the functionality of rank information. Owing to the advantages provided by the rank function and CPD algorithm, RGD can effectively overcome the impacts of outliers and extreme values in explanatory variables, and assumptions of statistical distributions of data are not required. Finally, RGD can provide robust spatial zones for a more reliable and practical interpretation of the results.

4.5.2 Future recommendations

This study demonstrated the advantages of RGD in the robust estimation of PD measurement compared with OPGD because of the optimal interval determination using CPD. The original CPD method allowed researchers to make adjustments to the minimal segmentation length and control the size of the intervals. In this article, we present RGD results with a minimal interval length. Future GD-based spatial heterogeneity research could set the minimal segment length parameter based on case requirements or specific research targets. The setting of the minimal segment length parameter was related to the scale of the spatial analysis (Song et al., 2020). A small minimal segment length can enlarge SSW values, and a large minimal segment length can provide better spatial visualisation and interpretation for large-scale spatial analysis. In addition, the associations between RGD and other GD-based models can be compared in terms of PD values and spatial zone characteristics. For instance, it is interesting that the quantile OPGD method is a special case of the RGD. When the minimal segment length is equivalent to the number of elements in an equal-sized interval determined by a given number of breakpoints, the RGD becomes a quantile OPGD.

4.6 Conclusion

This chapter proposes a RGD model to explore robust spatial associations between dependent and explanatory variables. The RGD-based analysis of the case study indicates that RGD can effectively identify the robust PD values of explanatory variables using a rank function and CPD-based optimisation approach for robust spatial data discretisation. The analysis and visualisation of the results and sensitivity analysis for model evaluation demonstrate that RGD can explore the maximum spatial associations and guarantee a stable increase in PD values with an increase in the number of intervals. RGD is robust in dealing with variables with different statistical distributions, outliers, and extreme values and provides robust spatial zones for spatial analysis. In summary, the RGD provides a solution for an in-depth understanding of spatially stratified heterogeneity and spatial associations. RGD can be implemented in diverse fields for robust and optimal spatial zone identification, spatial determination or factor exploration, and assessment of spatial disparities.

Chapter 5. Assessing social-economic sustainability of industrial development from an urban scaling perspective: industrial features associated with economy, infrastructure, and innovation

5.1 Introduction

Power-law scaling is one of the urban theories that state the rule of development in cities during the expansion processes (Lobo et al., 2013b; Keuschnigg et al., 2019; Lei et al., 2022b). The urban power-law scaling theory proposes that the pace of urban development follows the rule of power-law scaling in general, rather than a linear developing rhythm (Bettencourt, 2013; Riascos, 2017). Cross-sectional scaling is the most frequently used analysis approach in urban scaling analysis, and this method analyses the scaling relationship between urban indicators and the population at a specific time (Bettencourt et al., 2020). Relevant cross-sectional scaling analysis methods have been intensively applied in research on cities to simulate development during city expansions from multiple perspectives, including urban morphology (Ovando-Montejo et al., 2021), population size (Khan & Pinter, 2016), economy (Xu et al., 2020), infrastructures (Lämmer et al., 2006; Kwon, 2018; Ma et al., 2018) and innovation (Lobo et al., 2013a).

Cross-sectional scaling estimates the average-aggregated scaling development of cities, while the characters or properties of each city are also valuable in research and practice. The residuals from urban scaling regression models, also known as scale-adjusted metropolitan indicators (SAMIs), provide supplementary scaling information regarding the characters and disparities of cities within the system (Bettencourt et al., 2010; Xiao & Gong, 2022). SAMIs indicate whether local urban scaling features are above or below the average level according to the sign of positive or negative. How a city is performed within the entire system can be implied by analyzing the rank of SAMIs. Studies have validated that the SAMIs have no bias on the urban population size (Bettencourt et al., 2010), while spatial effects of SAMIs are evident (Xiao & Gong, 2022). The spatial distribution of SAMIs in general and the spatial autocorrelation effect of SAMIs in specific have been explored and applied to urban policy-making (Yang & Zhao, 2023; Lei et al., 2022a). However, cities are developing from multiple perspectives and urban features may be spatially associated. Current knowledge on how urban scaling features are associated and how to explore spatial associations among urban features from a power-law scaling perspective is limited.

Industrial development is important and tied to cities. Given the case in Australia, three types of industries, including mining, manufacturing, utility supply and waste services play a key role in Australian people's daily lives. These industries contribute around a quarter of the national GDP in 2016 (Australian Bureau of Statistics, 2017c). These three industries are closely linked to cities in Australia. According to the national census in 2016, 84% of employees and companies, from three key industries, are resided and operated within cities (Australian Bureau of Statistics, 2020). Previous studies suggest that industrial features in cities, such as industrial land use, industrial employee, and industrial company, follow the rule of power-law scaling (Bettencourt et al., 2007; Lei et al., 2022a), and the development of industrial activities may relate to the local economy, infrastructure design, and innovation development (Rietveld et al., 1994; Kozlov et al., 2017; Wang et al., 2020). Some of the urban scaling properties of Australian cities, including personal income and economic inequality, have been investigated by Australian research groups (Sarkar et al., 2018; Sarkar, 2019). However, further scaling properties of Australian cities regarding industrial performance, infrastructure design, and innovation growth remain to be explored. Given the importance of industrial development, spatial associations between industrial scaling features and urban indicators from the economy, infrastructure, and innovation categories, deserve more research effort. Understanding the association among urban features is beneficial to scientific decision-making and smart urban design.

Considering research gaps in exploring spatial associations among urban scaling features and limited knowledge on the power-law scaling development of Australian cities especially the industrial performance, this study aims to identify spatial disparity of urban performance from a power-law scaling perspective with Australian city system as a case study, and further explore spatial associations between industrial features and urban indicators from multiple perspectives. This study develops a set of methods to analyse urban scaling indicators and spatial associations relevant to industrial features. Firstly, we apply power-law regression models to assess power-law scaling rules in cities. Secondly, SAMIs, i.e., residuals from power-law regression, are used to show properties and spatial disparities of urban performance from the view of power-law scaling. Lastly, spatial associations between industrial features and other urban indicators are assessed by a robust geographical detector (RGD) with optimisation algorithms. In this research, eight urban indicators under the categories of industrial features, economy, infrastructure, and innovation are analysed by processing open-access datasets.

5.2 Study area and datasets

5.2.1 Study area

This study mainly investigates cities in Australia. These Australian cities are significant urban areas (SUAs) defined by Australian Bureau of Statistics (ABS). An SUA refers to a geographical boundary that contains at least one urban center with a resident population of at least 10,000 (Australian Bureau of Statistics, 2017b). There are altogether 101 SUAs in Australia by the year 2016, including Sydney, Melbourne, Brisbane, Perth, Adelaide, and other non-capital cities. The area of SUAs ranges from 6189 square kilometers in Melbourne to 38 square kilometers in Emerald. The urban population of SUAs ranges from 4.4 million in Sydney to 10.2 thousand in Kingaroy. This study investigates the scaling properties of 101 all SUAs from the perspectives of industrial features, economy, infrastructure, and innovation. The spatial association between industrial features and other urban indicators is further explored.

5.2.2 Urban indicators and datasets

This study investigates eight urban indicators and demonstrates the scaling properties of SUAs from the perspectives of industrial development, economy, infrastructure, and innovation. Industrial features include industrial scale, industrial employee, and industrial company. The industrial scale indicator refers to the total area of industrial regions within each city that support industrial activities of mining, manufacturing, utility supply and waste services. Industrial regions are defined to be large enough to be functional areas. To study economic development in cities, our research investigates total estimated GDP value and total income as urban economic indicators. Our study further analyses road length and dwelling count as infrastructure indicators. The innovation indicator is measured by the total number of companies providing professional, scientific and technical services under the scope of the Australian industry. These companies contribute research outputs and development to society, and they are thereafter noted as ‘R&D company’. The study area and statistical distribution of eight urban indicators are visualised in Figure 5-1.

The industrial scale is represented by the area of industrial regions. As there is no prepared data representing industrial regions, our study identifies these areas from raw data provided by OpenStreetMap (OSM) and National Pollutant Inventory (NPI). This study defines industrial regions as industrial areas supporting mining, manufacturing, utility supply and waste services. Our research pre-processed OSM land use data and identified dense infrastructure areas based on POIs. The identified industrial regions are composed of large industrial land use and dense industrial infrastructure areas. The industrial region identification process is based on the Australian Statistical Geography Standards (ASGS) and a POI-based interested spatial region identification workflow (Song et al., 2018b). Details of the industrial region identification process are explained as follows. Industrial land use data from OSM are coarse and extremely small areas should be filtered. According to the ASGS, 5000 square meters is the minimal scale of a region with at least one functional infrastructure, and therefore industrial land uses smaller than 5000 square meters are filtered and then considered industrial POIs. Then, 3977 industrial POIs, including converted POIs from land use and spatial points representing infrastructures, from OSM and NPI are processed by the kernel density estimation (KDE) to generate high-density industrial infrastructure areas. The KDE method is executed with 1000 meters as a searching radius, 194m x 194m (i.e., the median size of the mesh block, which is the finest spatial granularity of ASGS products) as the pixel size for the Epanechnikov kernel (Zhang et al., 2022). The 1000-meter searching radius refers to the minimal level of the square root of the Statistical Area Level 2 (SA2) functional area. Next, the high-density industrial infrastructure areas are determined by the change of cumulative distribution function (CDF) from KDE with a threshold value of 0.5%. Finally, large industrial land use and dense industrial infrastructure areas are merged, and aggregated industrial areas smaller than 0.49 square kilometers (i.e., the minimal scale of SA2 functional areas in 2016) are filtered.

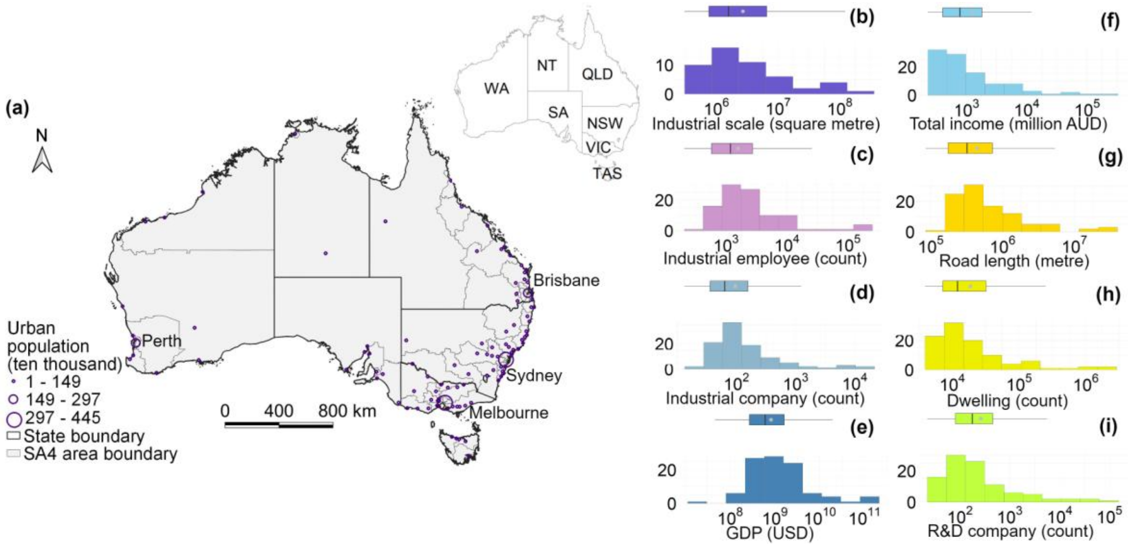


Figure 5-1. Study area and statistical distribution of indicators. Significant urban areas in Australia by population (a), and statistical distribution of urban factors, including industrial scale (b), industrial employee (c), industrial company (d), GDP (e), total income (f), road length (g), dwelling (h), and R&D company (i). Note: WA – Western Australia, SA – South Australia, NT – Northern Territory, TAS – Tasmania, VIC – Victoria, NSW – New South Wales, QLD – Queensland, and SA4 – Statistical Area Level 4 (sub-state boundaries).

Urban indicators, data descriptions, and sources are shown in Table 5-1. All datasets are accessed from OpenStreetMap (OSM), National Pollutant Inventory (NPI), ABS, and Dryad. OSM provides raw datasets including industrial land use and point of interests (POIs) for industrial region identification, and spatial data of nationwide Australian roads (Geofabrik and OpenStreetMap contributors, 2021). NPI provides supplementary spatial information on industrial region identification POIs, and these POIs are recorded facilities supporting industrial activities from the Australian government (Department of the Environment and Energy, Australian Government, 2016). The ABS census data provide a collection of urban indicators, covering industrial employees and companies, total income, dwelling, and R&D companies (Australian Bureau of Statistics, 2020). Last, the GDP estimation, with USD as a unit, is available from the Dryad platform provided by a research team from Aalto University (Matti et al., 2020). As an explanatory variable of urban scaling, the urban population is available from the ABS national census (Australian Bureau of Statistics, 2017d). This research uses all datasets representing urban scaling status during the financial year 2015-2016.

Table 5-1. A summary of urban scaling indicators

Urban indicator	Unit	Data description	Source
Industrial scale	m^2	Total key industrial area within cities	OSM, NPI
Industrial employee	Count	Total employees working for key industries within cities	ABS
Industrial company	Count	Total companies registered for key industries within cities	ABS
GDP	USD	Total estimated GDP of cities	Dryad, Aalto University
Total income	AUD (million)	Total reported income of cities	ABS
Road length	Meter	Total road length of cities	OSM
Dwelling	Count	Total dwelling within cities	ABS
R&D company	Count	Total companies providing professional, scientific and technical services within cities	ABS

5.3 Methods

Figure 5-2 shows the research method workflow, including three main steps. The first step is to compute the power-law scaling of eight urban indicators after industrial region scale identification and urban factor data computation. The second step is to demonstrate the spatial disparity of urban performance in the scaling development, indicated by SAMIs. The last step is to identify the spatial association between industrial features and other urban indicators by analyzing SAMIs using a robust geographical detector.

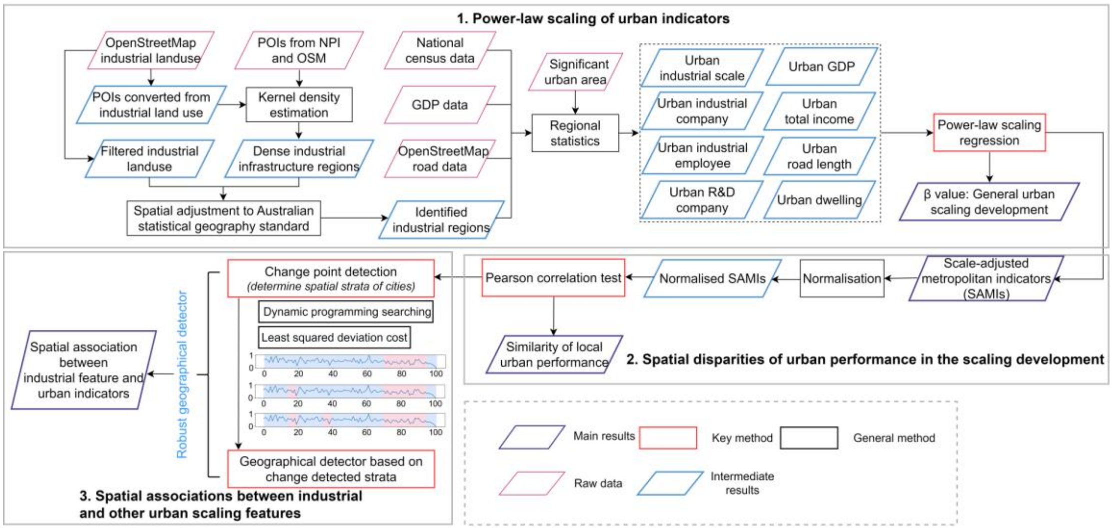


Figure 5-2. Flowchart of research methods, including power-law scaling of urban indicators, spatial disparities of urban performance from a scaling perspective, and spatial associations between industrial and other scaling features.

5.3.1 Power-law scaling of urban indicators

5.3.1.1 Urban indicator data statistics

Eight urban indicators are computed by functions in geographic information systems

(GIS). The industrial scale indicator, referring to the total industrial area with cities, is calculated by a ‘Statistics by categories’ method in GIS after all industrial regions are categorised by their belonging cities. ABS provides five urban indicators, including industrial employee and company, total income, dwelling, and R&D company, at the spatial granularity of SA2 level. As cities defined by SUAs are aggregations of SA2 areas (Australian Bureau of Statistics, 2017a), these five urban indicators, measured by ABS, are the sum of values recorded by all SA2 areas within cities. The road length indicator is the total length of road lines with cities computed by a ‘sum line length’ function in GIS. The GDP of cities is estimated by the sum of all GDP pixel values within city boundaries using the ‘Raster layer zonal statistics’ method in GIS.

5.3.1.2 Urban power-law scaling

Cities are developing at the pace of power-law scaling. With the city size represented by urban population, the relationship between the logarithm-transformed urban indicator and logarithm-transformed urban population is linear. The urban power-law scaling relationship is shown in equation (5-1).

$$Y(t) = Y_0 \times P(t)^{\beta} \qquad (5\text{-}1)$$

where $Y(t)$ refers to the urban indicator; Y_0 is a normalization constant; $P(t)$ is the urban size represented by population; β is the scaling exponent indicating urban indicators’ scaling development at a specific time; t, which is the year 2016, is the time when these power-law relationships are measured. When urban indicators and urban population are logarithm-transformed, the scaling coefficient values can be calculated via linear regression models.

5.3.2 Spatial disparities of urban performance in the scaling development

5.3.2.1 Scale-adjusted metropolitan indicators

Spatial disparities in urban performance, demonstrating characters of each city, are represented by SAMIs (Bettencourt et al., 2010). SAMIs, which are dimensionless and not influenced by city size, are quantified by residuals from urban scaling regressions, shown in equation (5-2).

$$\varepsilon_i = log(Y_i) - log(Y_0 P_i^{\beta}) = log\frac{Y_i}{Y_0 P_i^{\beta}} \qquad (5\text{-}2)$$

where ε_i is the SAMI at city '*i*'; Y_i is the observed value of the urban indicator at city '*i*'; β is the scaling exponent; Y_0 is a normalization constant, and $Y_0 P_i^{\beta}$ is the estimated value of the urban indicator at city '*i*'.

In this research, the SAMIs values are interpreted by residual-rank plots to demonstrate the unique properties of each city in urban scaling. Normalised SAMI values can be further analysed using quantitative models for the relationship between scaling properties. Thus, we use normalised SAMIs in the following spatial analysis processes.

5.3.2.2 Similarity among cities based on scaling disparities

The similarity among cities is measured by the Pearson correlation coefficient from normalised SAMIs, as shown in equation (5-3). Similar cities may have common scaling features from industrial development, economy, infrastructure, and innovation perspectives, and the features of each city are represented by normalised SAMIs. In this research, we test the similarity among top-populated cities.

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \qquad (5\text{-}3)$$

where X_i is the *i*-th normalised SAMI value, representing a kind of scaling feature, of the city *X*; $\bar{X}$ is the mean value of all normalised SAMI scaling features of the city *X*; Y_i is the *i*-th normalised SAMI value of another city *Y*; $\bar{Y}$ is the mean value of normalised SAMI values of the city *Y*; *n* is the number of scaling features, which corresponds to eight urban indicators.

5.3.3 Spatial analysis on urban scaling disparities – Robust geographical detector

Spatial stratified heterogeneity (SSH) is a representation of spatial association based on the fact that the within-strata variance is less than the inter-strata variance (Wang et al., 2016; Song et al., 2020). The spatial association is determined by spatial strata from the division of the sorted rank of the independent variable. Thus, the SSH value represents how the dependent variable is associated with spatial strata by dividing the rank of the independent variable. The robust geographical detector (RGD) is a measure of SSH value using an optimised discretisation algorithm to improve the geographical detector by minimizing the inter-strata variance and increasing the within-strata variance (Zhang et al., 2022). The optimisation strategy for the RGD model is a change point detection algorithm with a least squared deviation as a cost function and dynamic programming as a searching method. The pseudo code demonstrates how the spatial strata of cities are determined.

Pseudo code: Determining robust spatial strata of cities

Input: a list of dependant variable values of 101 SUAs sorted by the rank of an independent variable $\{y_x\}_{x=1}^{N}$, cost function $c(y_I) = \sum_{x \in I} ||y_x - \bar{y}||_2^2$ (least squared deviation), the number of spatial strata *K* determined by (*K*-1) change points, number of observations *N*.

for all possible lengths ($1 < r < N$) of sub-list *(p, q)* **do**

  store the cost value of each sub-list;

end for

for all the sub-list *(p, q)* with a length from 2 to *K*-1

  find and store the division point that divides sub-list *(p, q)* into 2 smaller lists with a minimum sum of the cost;

end for

set the last observation as the first change point; $j \leftarrow K$;

while $j > 1$ **do**

start with the latest change point and find a division point that divides the sub-list into 2 smaller lists with a minimum sum of the cost;

store the division point as the latest change point;

$j \leftarrow j - 1$;

end while

return the change point list

The SSH value representing the spatial association between factors can be calculated from the geographical detector after the determination of robust spatial strata of cities. The formula for SSH computation is shown in equation (5-4) (Wang et al., 2016; Song et al., 2020; Zhang et al., 2022). In this research, our RGD model sets the minimum number of observations in each group as no less than five in the optimisation algorithm processing. To be consistent with the SAMI-rank analysis, the RGD model uses the descending order rank to determine spatial zones:

$$B = 1 - \frac{\sum_{h=1}^{k} N_h \delta_h^2}{N\delta^2} = 1 - \frac{SSW}{SST} \qquad (5\text{-}4)$$

where N_h is the number of observations within the sub-stratum h by discretizing the explanatory factor; N is the total number of observations; δ_h^2 is the variance within a sub-stratum h by discretizing the explanatory factor; δ^2 is the variance of the whole study area. *SST* refers to the Sum of Squares Total and *SSW* refers to the Sum of Squares Within spatial strata. The *B* value ranges from 0 to 1. A higher *B* value indicates a stronger spatial association between two variables. The *B* value is a measure of spatial stratified heterogeneity, which is also an indicator of the power of determinants. One of the primary research purposes of this study is to identify the spatial relationship between industrial features and other urban indicators based on scaling properties. The response variable is the normalised SAMI value from one of the industrial feature indicators, and the explanatory variable is the normalised SAMI value from one of the urban indicators under the category of economy, infrastructure, and innovation. All 101 SUAs are divided into multiple strata based on the rank of an explanatory variable using RGD.

5.4 Results

5.4.1 Power-law scaling of urban indicators

The average scaling development of eight urban indicators from four categories is measured based on urban scaling regressions, and statistical results are demonstrated in Figure 5-3 and Table 5-2. According to urban scaling coefficients and their 95% CI value range, scaling paces of industrial scale and infrastructure indicators are sub-linear, scaling paces of industrial employee development and industrial company development are almost linear, and power-law developments of economy and innovation indicators are super-linear. All urban scaling regression models are robust with adjusted R-squared values no less than 0.7 and all p-values for F-statistics are significant at the 0.01 level. The general results, which show the sigmoidal development of infrastructure scaling indicators and booming development of wealth and innovation, coincide with general findings and assumptions of the urban scaling theory. It is worth noting that this study differentiates the computation of industrial scale and other urban indicators. As the industrial scale considers regions large enough to be functional areas only, not every city has one or more industrial regions inside. There are 50 large cities that have industrial regions inside. Therefore, when analyzing the industrial scale indicator in urban scaling regression and further spatial or aspatial relationships involving this factor, computation results are based on these 50 observations with industrial regions.

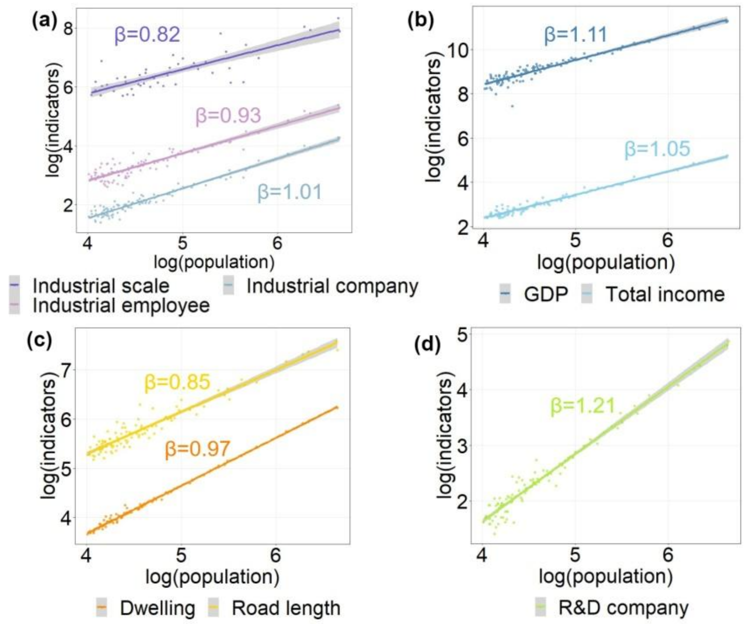


Figure 5-3. Urban scaling regression results. (a) Urban scaling of industrial features. (b) Urban scaling of economic indicators. (c) Urban scaling of infrastructure indicators. (d) Urban scaling of the innovation indicator.

Table 5-2. Urban scaling of industrial features, economy, infrastructure, and innovation

Urban indicator	β value	95% CI	Adj-R^2	Category
Industrial scale	0.82	[0.67, 0.97]	0.70	Industrial feature
Industrial employee	0.93	[0.86, 1.00]	0.88	Industrial feature
Industrial company	1.01	[0.97, 1.06]	0.95	Industrial feature
GDP	1.11	[1.03, 1.18]	0.90	Economy
Total income	1.05	[1.01, 1.09]	0.97	Economy
Road length	0.85	[0.80 , 0.90]	0.93	Infrastructure
Dwelling	0.97	[0.96, 0.98]	0.99	Infrastructure
R&D company	1.21	[1.16, 1.26]	0.96	Innovation

5.4.2 Spatial disparities of urban performance from a scaling perspective

The urban scaling theory suggests a sub-linear scaling development for infrastructure indicators and super-linear development for wealth and innovation indicators. The sub-linear infrastructure scaling development is due to the optimised design and efficiency of the urban system, and this value also implies the share ratio of public facilities for cities. Local infrastructures are publicly shared by residents where the scaling coefficient is low. This fact further indicates that cities with negative SAMI values for infrastructure indicators have higher efficiency in infrastructure design than the average level. On the contrary, wealth and innovation are developing at an accelerating speed and relatively fast developments are indicated by high scaling coefficients. Therefore, cities with positive SAMI values for wealth and innovation indicators have greater development potential than the average level of the whole system.

Thus, apart from general scaling development indicated by scaling coefficients, the character of each city indicated by SAMIs can also provide valuable information. Statistical distributions of SAMIs of eight urban indicators are further analysed by their rank as shown in Figure 5-4. The top-ranked city of each scaling relationship and five Australian major capital cities are highlighted in these figures. According to SAMI-rank plots, Sydney surpasses the other four major cities in terms of scaling performance in infrastructure efficiency, economy, and innovation. Figure 5-4(a), Figure 5-4(f), and Figure 5-4(g) demonstrate that Sydney has negative SAMI values for industrial region scale, road, and dwelling. These facts indicate that the infrastructure design in Sydney is relatively efficient from the view of urban scaling. Sydney has higher SAMI values in both total income and R&D company as shown in Figure 5-4(e) and Figure 5-4(h), which further implies Sydney's higher accelerating achievement in economic and innovation than the average level. Two economic SAMIs from Figure 5-4(d) and Figure 5-4(e) also suggest that Perth has a good economic performance in urban scaling. Top-ranked cities according to SAMIs performance, including industrial design, infrastructure and economic development, and innovation, have no bias on the population size. This fact is consistent with the scale-independent assumption in the scaling theory (Bettencourt et al., 2010).

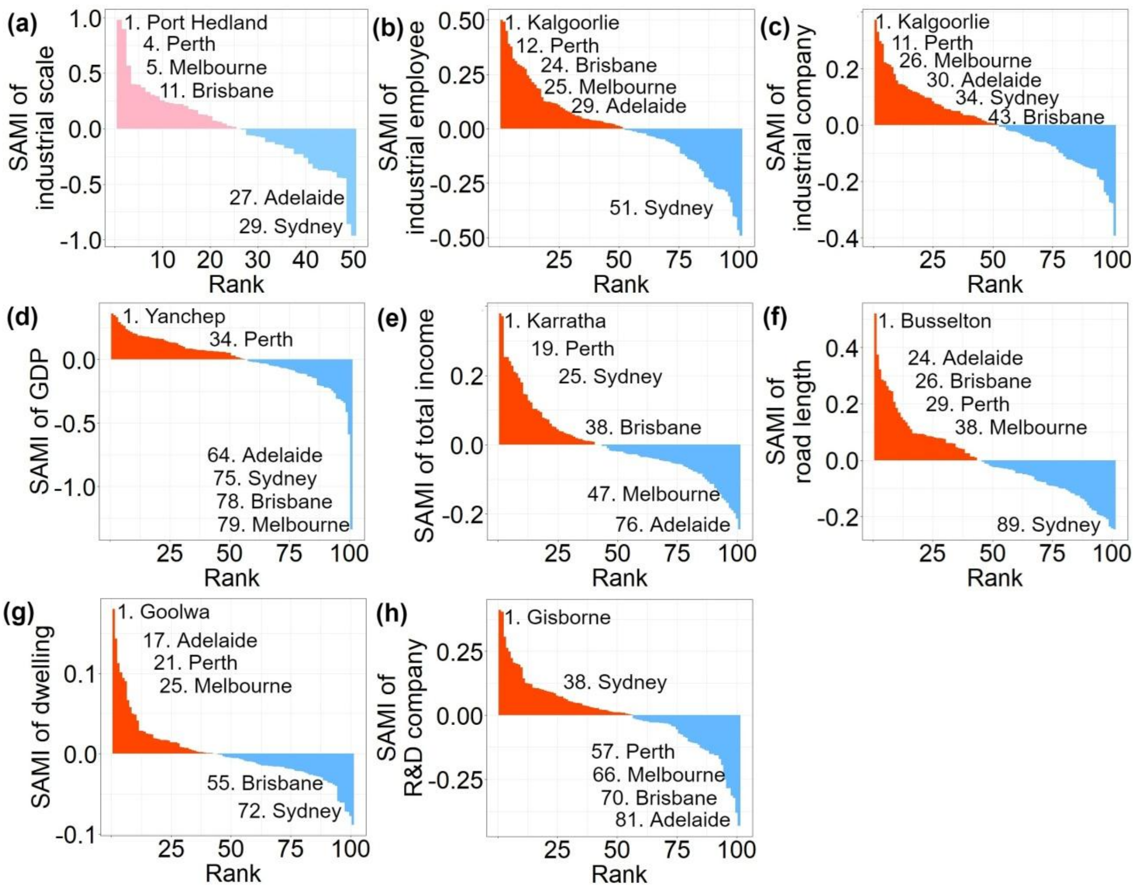


Figure 5-4. SAMI-rank plots. SAMIs rank of urban indicators, including (a) industrial scale, (b) industrial employee, (c) industrial company, (d) GDP, (e) total income, (f) road length, (g) dwelling, and (h) R&D company.

The similarity among large cities measured by eight urban indicators is summarised in Figure 5-5. The top four populated cities (i.e., Sydney, Melbourne, Brisbane, and Perth) are similar in eight urban scaling features. Scaling characters represented by SAMI values of Melbourne, Brisbane, and Perth are more correlated.

The scaling development of Sydney is similar to a majority of top-populated cities to different extents. It is worth noting that the city of Toowoomba is also similar to cities with top-ranked residents from the view of scaling development.

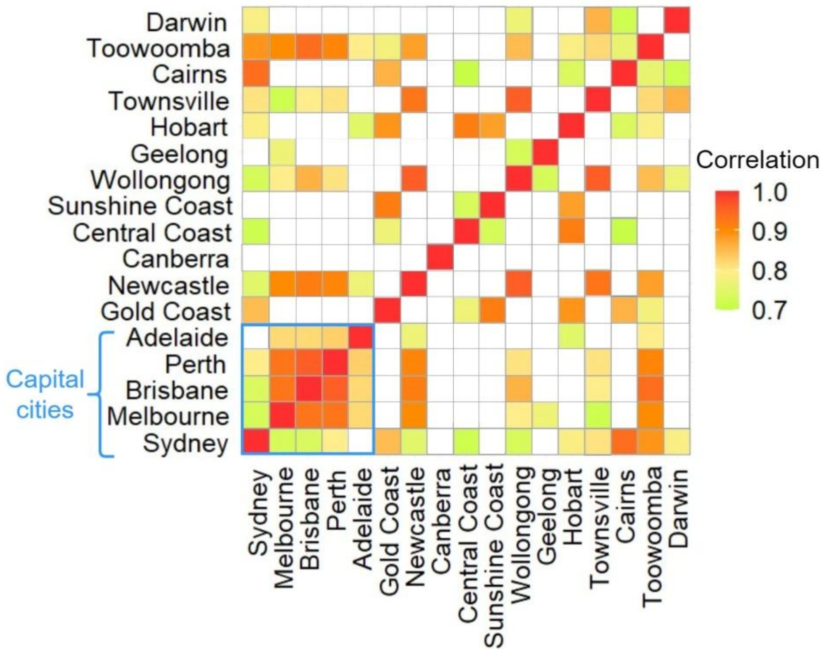


Figure 5-5. Similarity among Australian major cities assessed by normalised SAMIs. Note: Cities are sorted by the resident population. The white pixel in the correlation matrix indicates no statistically significant results.

5.4.3 Spatial association among urban scaling disparities

The spatial associations between industrial features and urban indicators from other categories are analysed by the RGD model, shown in Figure 5-6. According to a pilot study of the robust geographical detector, it is accurate to test the predominance of associated factors with different numbers of spatial strata prior to a conclusion (Zhang et al., 2022). Thus, in this study, the spatial association between industrial features and other urban indicators is demonstrated from multiple tests with an increasing number of spatial strata. Infrastructure indicators and the innovation indicator can partially explain the industrial scale of cities from 25% to 29% with the increase of spatial strata as shown in Figure 5-6(a). However, the difference in the power of determination, for the industrial scale feature, among indicators is not significant. As shown in Figure 5-6(b), the urban indicator of total income can significantly explain features of scaling in industrial employees, and the *B* value increases from 42% to 51% with the increase of spatial strata from three to seven. In other words, the feature of total income can explain around 50% of industrial employee scaling features. The power of explanation of income level is far stronger than the other urban indicators. The feature of scaling in industrial companies is more associated with the innovation feature compared with indicators from other categories, as shown in Figure 5-6(c). The innovation feature can explain 27% to 34% of the industrial company feature with the increase of the number of spatial strata, and the power of determination of innovation is much greater than that of road infrastructure, ranking second, from 11% to 21%. The geographical-detector-based model can imply spatial associations among factors according to the power of determinants (i.e., SSH values), and how pre-dominant variables are associated with industrial scaling features can be further inferred from the spatial and statistical distribution of response variable observations by strata determined by explanatory variables.

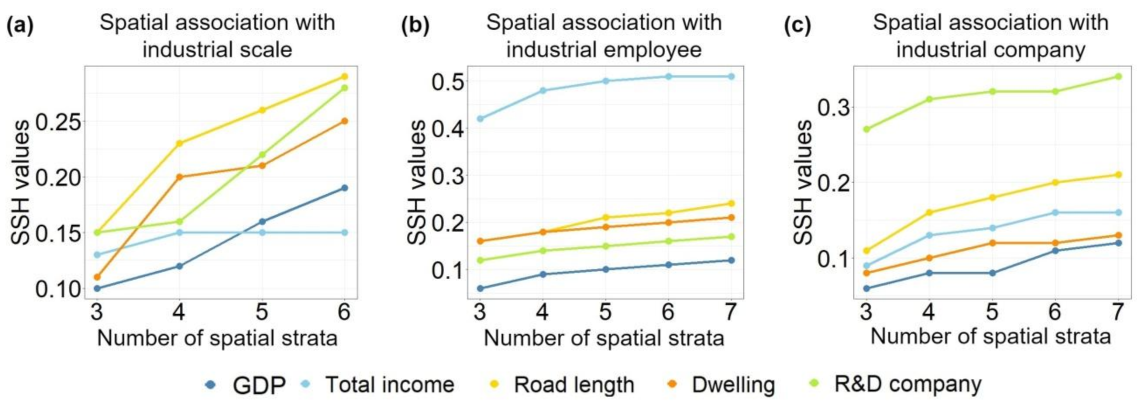


Figure 5-6. The value of industrial features spatially associated with economy, infrastructure, and innovation factors, from the view of urban power-law scaling

rule, with the increase of the number of spatial strata. (a) Industrial scale's association with other indicators. (b) Industrial employee's association with other indicators. (c) Industrial company's association with other indicators.

The industrial features associated with the predominant urban factors are assessed by the robust geographical detector, while how these features are spatially associated remains to be explored. Figure 5-7 provides a supplementary spatial explanation of the results shown in Figure 5-6. Figure 5-7 further demonstrates how the industrial employee is associated with the total income feature, and how the industrial company scaling feature is associated with the R&D company feature in cities. The spatial pattern of strata from industrial employees and industrial companies determined by their pre-dominant factors at the strata number of three is shown in Figure 5-7 to draw the conclusion, as the predominance of associated factors does not change with the increase of strata. The associations between industrial employees and total income, and industrial company and R&D company are both statistically significant at the strata of three. This study uses the descending order rank to determine spatial zones from RGD and be in line with the SAMI-rank analysis. Thus, take the association between industrial employees and total income as an example, 'stratum1' is the group determined by observations with the top-ranked normalised total income SAMI, and 'stratum3' is composed of the bottom-ranked normalised total income SAMI values. As shown in Figure 5-7(a), the top income SAMI group observations are distributed in remote areas of Australia, and the median group contains a majority of cities including Sydney, Melbourne, and Perth. Figure 5-7(b) indicates a general positive association between industrial employees and total income, as the average level of industrial employees SAMI drops with the decrease of total income SAMI values. How scaling feature of industrial company is associated with R&D company is demonstrated in Figure 5-7(c) and Figure 5-7(d). As shown in Figure 5-7(c), the top R&D company SAMI group includes Sydney, Melbourne, Brisbane, and Perth. This also implies that the synergy between industrial companies growth and innovation development in major capital cities is evident. The median group is located close to the coast of the Australia continent, and the bottom group is distributed in South Australia and the inner continent. Figure 5-7(d) also indicates a relative positive association between the industrial company feature and the R&D company feature. The average level of industrial company SAMI value decreases from the top-ranked group to the bottom-ranked group.

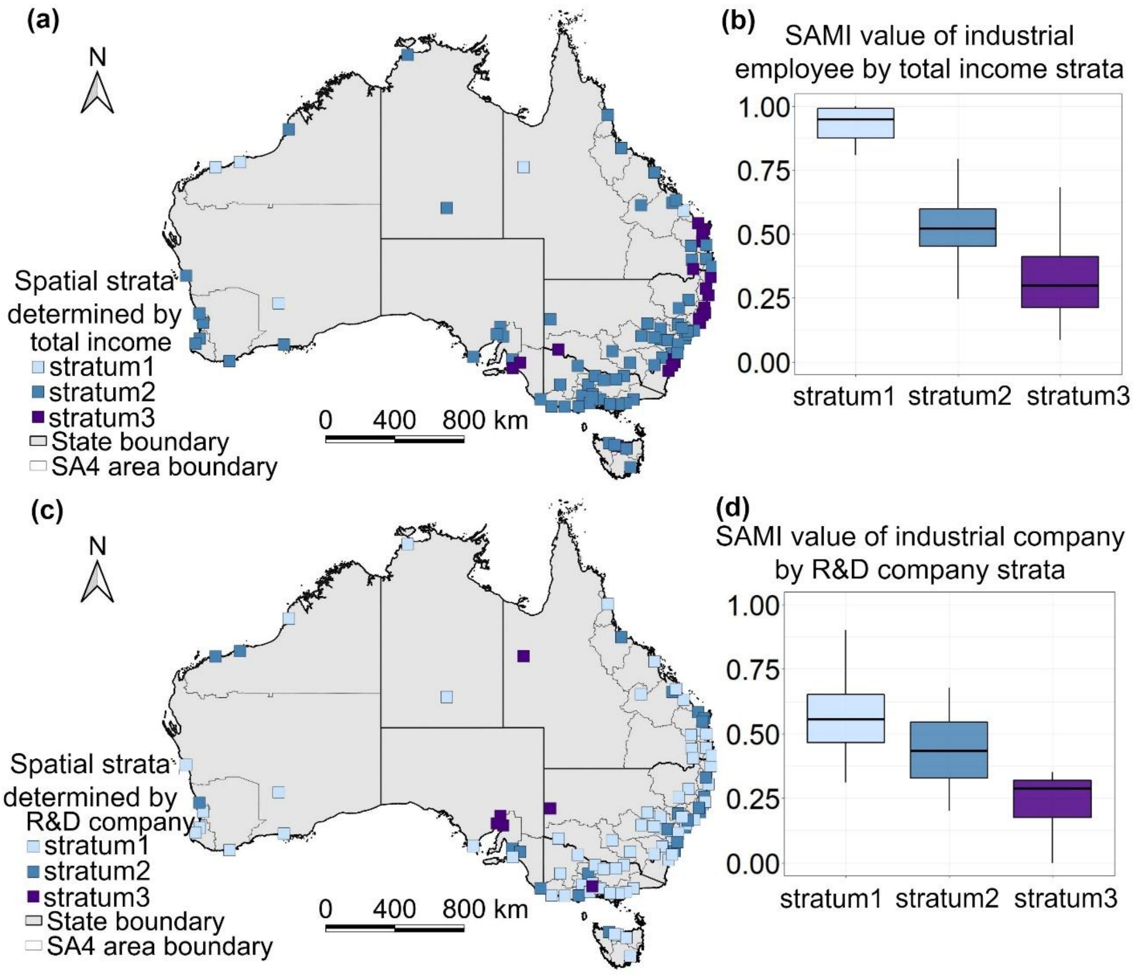


Figure 5-7. Spatial strata of industrial features by predominant influential factors. (a). Spatial strata of industrial employee by total income. (b). Statistical distribution of normalised SAMI value of industrial employee by spatial strata of total income. (c). Spatial strata of industrial company by R&D company. (d). Statistical distribution of normalised SAMI value of industrial company by

spatial strata of R&D company.

5.5 Discussion

This research identified the scaling development of eight urban features from industrial development, economy, infrastructure, and innovation perspectives. Properties and disparities of each city in scaling pace were indicated by SAMIs. Considering the importance and contribution of mining, manufacturing, utility supply and waste services, we further analysed the spatial association between industrial features and urban indicators from other categories. Through a series of computation processes, this research identified the paces of scaling achievement for Australian cities using urban scaling models, and the scaling coefficients coincide with scaling theory assumptions. This research also validates the scale independence of power-law scaling characters. Top-performance cities according to SAMIs are not related to the urban size, and this is consistent with the scaling assumptions (Bettencourt et al., 2010). Despite population size independence in scaling disparity, the urban performance of major capital cities is similar from an urban scaling perspective. Furthermore, among these top populated cities, Sydney has a good scaling development in economy, infrastructure, and innovation. Finally, results from RGD shows that industrial employee is highly and positively associated with income level, and the industrial company is also positively associated with R&D company. The synergy between industrial companies growth and innovation development in major capital cities is evident.

In terms of the methodological design, this research analyses scaling properties in Australian cities by fully investigating open-access socio-economic data from OSM and Australian census data. Apart from traditional statistical analysis on scaling coefficients and SAMIs, we further applied an advanced spatial association analysis method to figure out the pre-dominant indicator of industrial features. Spatial features of urban scaling have long been a key topic in urban studies, and this research explores spatial relationships among scaling indicators using an improved geographical detector algorithm.

This is a cross-sectional research in urban scaling, and research efforts in the future can make the work towards a temporal study. The Australian government is conducting an inclusive national census survey every five years, and both spatial and statistical datasets representing the same urban indicators are updated twice a decade. Therefore, in the future, Australian cities' scaling development in the post-pandemic era can be further analysed when national census data are processed and fully open to the public.

This research is a pilot investigation on spatial associations among urban scaling features, and future research can be implemented and enriched by various spatial methods and data. From a methodology design perspective, our study demonstrates the spatial associations between industrial scaling features and other urban characters, while interactive influences of multiple urban factors on industrial features remain to be explored. The development of urban systems is dynamic and synergistic, and thus economy, infrastructure, and innovation factors may impose an interactive impact on industrial size change. Therefore, the interactive detector of 'Geodetector' and its advanced models can be applied into urban scaling studies in the future work. From a data enrichment perspective, this research is supported by free open-access data sources including OpenStreetMap and ABS, and more urban factors for Australian cities can be accessed from cloud-based public spatial data platform, including Google Earth Engine and Australian Urban Research Infrastructure Network, in the future.

5.6 Conclusion

Urban features are developing at the pace of power-law scaling. Spatial effects in the scaling characters of cities are evident, while spatial associations among multiple urban features remain to be explored. Industrial features are one of the most important characteristics of urban development, and these features deserve more research efforts. Considering current gaps in investigating associations among urban scaling features and limited knowledge on the power-law scaling of Australian cities especially the industrial performance, this study aims to identify spatial disparity of urban performance from a power-law scaling perspective in Australia and explore spatial associations between industrial features and other urban indicators. We use power-law models to assess the scaling development of urban indicators and disparities of urban performance, and RGD models to explore spatial associations between industrial features and other urban indicators. This study validates the general consistency of the scaling development among Australian cities with the power-law theory and the similarity of scaling disparity features among top-populated cities. Spatial analysis results show that the urban innovation indicator and the income level are predominantly and positively associated factors with industrial companies and employees, indicating that the innovation growth and economic development in Australian cities would stimulate the performance of industrial companies and the employment scale. The robust geographical detector categorises major capital cities into a stratum with high industrial company performance corresponding to high innovation growth, indicating that the synergy between urban innovation and industrial company performance is especially significant in large capital cities. This study explores further spatial properties of urban performances from a power-law scaling perspective and reveals the pace of development in Australian cities, especially industrial features. Future research efforts on census data released at different periods can enrich current cross-sectional results toward temporal outcomes.

Chapter 6. Further investigation on socio-economic impacts of industrial features: spatial impacts of industrial development on national economic inequality

6.1 Introduction

Income inequality, which refers to the unequal distribution of income across different scales, is a form of economic disparity. This phenomenon has been on the rise in developed countries, including Australia, over the past decades (Athanasopoulos and Vahid, 2003). Understanding economic disparity and the associated factors can benefit Australian national policy-making. Previous studies, dating back to 1960s, indicate that economic inequality is especially associated with socio-economic factors, including industry activities, gender, education level, and employment (Fleming and Measham, 2015; Murray, 1978; Reeson et al., 2012). However, as circumstances have changed over time, the current association between economic inequality and socio-economic variables, including industrial features, requires further investigation. The Gini coefficient is currently the most widely used inequality indicator, with acceptable properties of scale and size independence (Athanasopoulos and Vahid, 2003). Detailed and precise information on Gini coefficient at different spatial scales is accessible from the Australian Bureau of Statistics (ABS) source (Australian Bureau of Statistics, 2020). Statistical Area Level 3 (SA3) is a spatial scale representing entire regions serviced by regional cities. Investigations conducted at this regional level can provide guidance for more detailed regional decision-making (Australian Bureau of Statistics, 2016). To address gaps and provide potential value to policy-making, we conduct a case study on nationwide economic inequality and its associated factors, including industrial features and other socio-economic indicators, at SA3 level in Australia. In this study, we introduce spatial local complexity as a spatial index to explain unknown residuals from traditional models.

6.2 The concept of spatial local complexity

6.2.1 The background of error explanation and spatial complexity

Geospatial modelling has long faced the challenge of explaining errors that vary across space (Henebry, 1995; Pringle & Lark 2006). Spatial autoregressive methods improve linear model performance by accounting for spatial impacts and explaining unknown errors through consideration of the spatial dependence of variables or residuals (Chi and Zhu, 2008). The effectiveness of spatial dependence concept in error illumination is demonstrated by the boom of autoregressive models in spatial analysis. In this study, we propose a new spatial complexity index by extending the spatial dependence expression to partially explain more spatial errors. Complexity, a concept opposite to simplicity, refers to being uncertain, unpredictable, hard to describe or explain, or difficult to solve (Suh, 1999), and this feature has been investigated in spatial research.

Describing and predicting distributions and explaining the cause of patterns in spatial science can be challenging when impacts of factors and interactions among factors change across space (Weisent et al., 2012), leading to the emergence of spatial complexity. Previous studies have investigated the complexity of geographical data using various perspectives, including geomorphology indicators, information theory, and structural complexity. First, geomorphology indicators, such as stand density and surface fraction, have been used to measure spatial complexity by quantifying spatial distribution. This complexity measure has been applied in investigations of geospatial patterns of various land covers (Owers et al., 2016; Rufino et al., 2020). Second, spatial information can be measured using Shannon entropy, which is based on information theory. This complexity measure has been applied in urban studies and water resource management (Batty et al., 2014; Ilunga, 2019). Third, structural complexity, which is based on a fractal dimension, can be used in ecological studies to understand spatial information (Yanovski et al., 2017). In addition to quantitative measures, spatial complexity can also be understood in terms of the existence of spatial scales. This is because patterns and relationships may differ at different spatial granularities due to the scale dependence effect (Cola, 1994).

Our study extends the understanding of complexity from the concept of spatial dependence and proposes a new spatial index. Spatial dependence, defining an ideal spatial pattern under geographical impacts, advises that spatial features at a location are more correlated with nearby features than distant ones (Epperson and Li, 1996), and similar spatial features tend to cluster. Spatial dependence, demonstrated by a spatial autocorrelation phenomenon, can be measured by Moran's I and Geary's C globally, and local spatial autocorrelation can be indicated by Local Indicators of Spatial Association (LISA) (Anselin, 2019; Poudyal et al, 2019). By extending the recognition of spatial dependence, this study proposes a geocomplexity measure, i.e., spatial local complexity indicator, to characterize the complexity among local spatial patterns and the spatial dependence for spatial neighbours. The spatial local complexity indicator indicates spatial distributions that exceptionally contradict the concept of spatial dependence as a complexity. We then developed a series of traditional models (i.e., linear regression, support vector regression, and geographically weighted regression) to model the economic inequality in Australia, and spatial local complexity is further applied to explain spatial errors from these traditional models. Finally, the explanation power of geocomplexity as a geospatial impact may vary across space, as indicated by spatial heterogeneity (Luo et al, 2022) and spatial association (Song and Wu, 2021a; Song, 2022a). This feature of spatial variation can be captured geographically weighted regression (GWR) (Fotheringham, 2002). Therefore, in this study, we employed GWR to model the power of error explanations using geocomplexity.

6.2.2 Spatial local complexity

The spatial local complexity quantifies the relationship between an area of interest and its surroundings, as well as the relationships among the spatial neighbours of the target area. In previous studies, indicators of spatial dependence, such as Moran's measures and Geary's measures, have been commonly used to measure the difference between the selected area and surrounding neighbours. According to the concept of spatial dependence, the spatial dependence phenomenon may exist in the surroundings of this selected area, where two surroundings are spatial neighbours or not spatially isolated. Thus, the spatial dependence between two spatial neighbours may exist under the geographical impact of the selected area. This study proposes the concept of spatial local complexity as a measure of geocomplexity, which extends previous measures and considers spatial dependence in this circumstance. The spatial local complexity uses a Moran measure to quantify local spatial patterns and the spatial dependence for its neighbours.

The process of measuring spatial local complexity is illustrated in Figure 6-1, and its formulas are shown in equation (6-1) and equation (6-2). Equation (6-1) is identical to equation (6-2), and equation (6-1) is the form with a spatial adjacency matrix. In this study, we further normalise the spatial local complexity value by transferring P_i value in equation (6-2) into G_i value in equation (6-3). The spatial local complexity is composed of two features: the local feature and the surrounding feature. The local feature is a measure of local spatial autocorrelation. The surrounding feature measures the spatial dependence of the surrounding environment under the spatial impacts of the measured central area. Both the local and surrounding features are quantified using a Moran-based measure. The Moran-based measure quantifies the spatial dependence between locations by summing the multiplication of Z-score values, which can represent both magnitude and direction. The spatial local complexity is explained as follows.

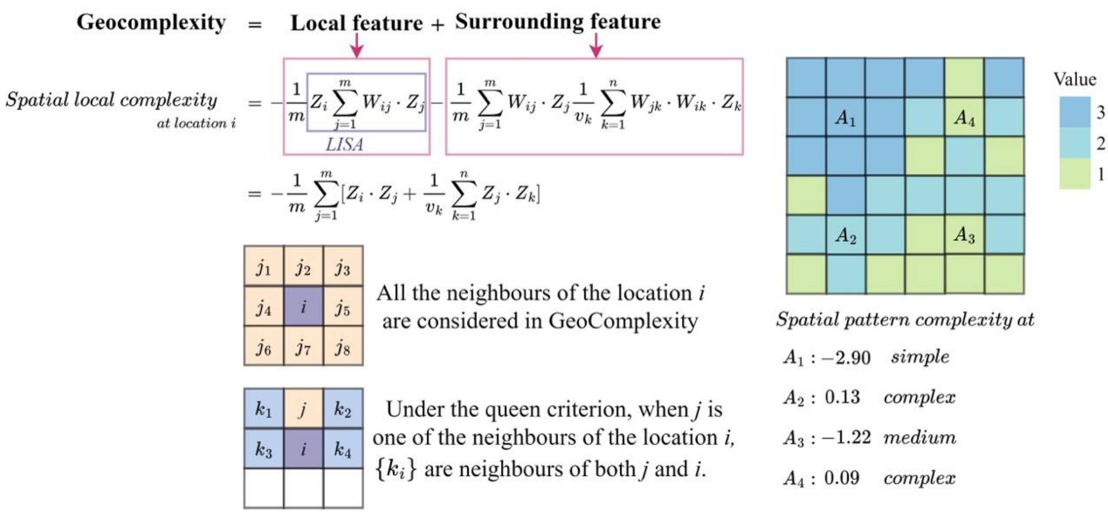


Figure 6-1. A measure of spatial local complexity.

$$P_i = -\frac{1}{m} Z_i \sum_{j=1}^{m} W_{ij} \bullet Z_j - \frac{1}{m} \sum_{j=1}^{m} W_{ij} \bullet Z_j \frac{1}{v_k} \sum_{k=1}^{n} W_{jk} \bullet W_{ik} \bullet Z_k \quad (6\text{-}1)$$

$$P_i = -\frac{1}{m} \sum_{j=1}^{m} [Z_i \bullet Z_j + \frac{1}{v_k} \sum_{k=1}^{n} Z_j \bullet Z_k] \quad (6\text{-}2)$$

$$G_i = \frac{P_i - min(P_i)}{max(P_i) - min(P_i)} \quad (6\text{-}3)$$

where P_i is the spatial local complexity for a location 'i' and G_i is the normalised indicator, i.e., the developed geocomplexity indicator in this study; W is a spatial adjacency matrix indicating the spatial relationship between observations. The adjacency matrix is represented by '0' and '1'. When location 'i' and location 'j' are spatial neighbours, the value W_{ij} is '1'. Z_i is the standardised value (Z-score) of the selected factor at location 'i'; Z_j is the Z-score of the selected factor at location 'j', and 'j' is the spatial neighbour of the location 'i'; Z_k is the Z-score of the selected factor at location 'k', and 'k' is the spatial neighbour of both location 'j' and 'i', and $\{k\}$ is a subset of $\{j\}$; m is the total number of spatial neighbours of location 'i'; v_k is the number of spatial neighbours of location 'j' while these neighbours for location 'j' should be spatial neighbours of location 'i' at the same time.

Equation (6-1) illustrates the computation of spatial local complexity using a spatial adjacency matrix, with the term '$Z_i \sum_{j=1}^{m} W_{ij} \bullet Z_j$' being based on a Moran measure for local spatial autocorrelation. This research makes further exploration by introducing the term '$\sum_{j=1}^{m} W_{ij} \bullet Z_j \frac{1}{v_k} \sum_{k=1}^{n} W_{jk} \bullet W_{ik} \bullet Z_k$' to capture the surrounding local spatial dependence from the view of location 'i'. This term calculates the product of Z-score in location 'j' (the spatial neighbour of location 'i') and Z-score in location 'k' (the spatial neighbour of both location 'i' and location 'j', and also the spatial neighbour of location 'j' restricted by spatial impact from location 'i'). In other words, this term partially computes the local Moran value of location 'j'. Only when the neighbour of 'j' is the neighbour of 'i' at the same time will location 'k' be considered in the computation. This computing strategy ensures that all computation results are related to the target location 'i'. The value of '$Z_i \sum_{j=1}^{m} W_{ij} \bullet Z_j$' is positive when a majority of surroundings have the same sign (positive or negative) as the location 'i', indicating strong local autocorrelation. Similarly, when a majority of Z-scores in location 'k' have the same sign (positive or negative) as those in each of the location 'j', the term '$\sum_{j=1}^{m} W_{ij} \bullet Z_j \frac{1}{v_k} \sum_{k=1}^{n} W_{jk} \bullet W_{ik} \bullet Z_k$' has a high positive value, indicating strong autocorrelation among the surroundings. To balance the impact of surrounding autocorrelation with the main local spatial autocorrelation, a weighted average strategy is employed by dividing v_k and m. The value of $\frac{1}{v_k}$ ensures that the impact of each surrounding autocorrelation does not exceed that the impact of main local spatial autocorrelation, while the value of $\frac{1}{m}$ ensures that the measure of spatial local complexity is not influenced by the number of neighbouring observations.

Equation (6-1) can be simplified to equation (6-2) under clear spatial relationships among 'i', 'j', and 'k'. Figure 6-1 illustrates the relationships among three types of locations, with an example under the queen criterion. When the location 'i' is at the center of the square, the set of $\{j\}$ refers to the eight surrounding squares.

For each element in $\{j\}$, the corresponding sets of $\{k\}$ and v_k are different. For instance, when an element 'j' is located at the top of location 'i', the set of $\{k\}$ consists of four squares marked in blue, and the value of v_k is 4.

Equation (6-1) and equation (6-2) employ the negative summation of local and surrounding features to measure spatial local complexity, where negative values represent more spatial complexity and lower values indicate less disorder. The negative measure of spatial features is consistent with the understanding of complexity, as a higher value indicates more spatial complexity. Areas with strong spatial dependence should exhibit similar Z-scores between the area and its surroundings, while regions with low geocomplexity are expected to have negative spatial local complexity values. For example, in Figure 6-1, A_1 exhibits a relatively low spatial local complexity value and follows a pattern of spatial autocorrelation, while the surrounding patterns are simple. However, both a local feature and surrounding features from A_2 and A_4 are complex, as demonstrated in disorder against spatial dependence, resulting in corresponding spatial local complexity values that are positive or close to zero according to equation (6-1).

This research further normalises the spatial local complexity in equation (6-3) to explain spatial error from traditional estimation models. As a region may have multiple spatial variables representing different attributes, the value range of the spatial local complexity of each variable may differ. The normalization process ensures that the value ranges of all relevant spatial local complexity from variables are the same and range from 0 to 1. This study uses the normalised complexity to examine to the extent to which spatial local complexity can explain unknown errors in linear models, machine learning models, and spatial heterogeneity models. Additionally, this study assesses to what extent the consideration of spatial local complexity could improve the performance of the GWR model.

6.3 Spatial impacts of industrial development on economic inequality

6.3.1 Economic inequality data

Gini coefficient is a critical economic indicator measuring the inequality of income distribution (Cowell, 1995; Mukhopadhyay and Sengupta, 2021). Its value ranges from 0 to 1, where 0 represents ideal economic equality and 1 indicates extreme inequality, with a coefficient over 0.5 indicating severe gaps in income distribution (Shorrocks, 1978). Understanding the determinants and spatial patterns of economic inequality can assist regional planners in designing regions with more effective policies. In Australia, nationwide economic inequality data are available from the ABS based on census data. Gini coefficient at the SA3 level, obtained from the government-maintained website (Australian Bureau of Statistics, 2020), is used as a dependent variable in this study to indicate economic inequality. The 2016 census report indicates that 16% of SA3 regions have Gini coefficients above 0.5, and all regions with Gini coefficients above 0.55 are located in major cities (Australian Bureau of Statistics, 2020). Figure 6-2 shows the distribution of economic inequality across Australia, with all five major cities containing regions with Gini coefficients greater than 0.5.

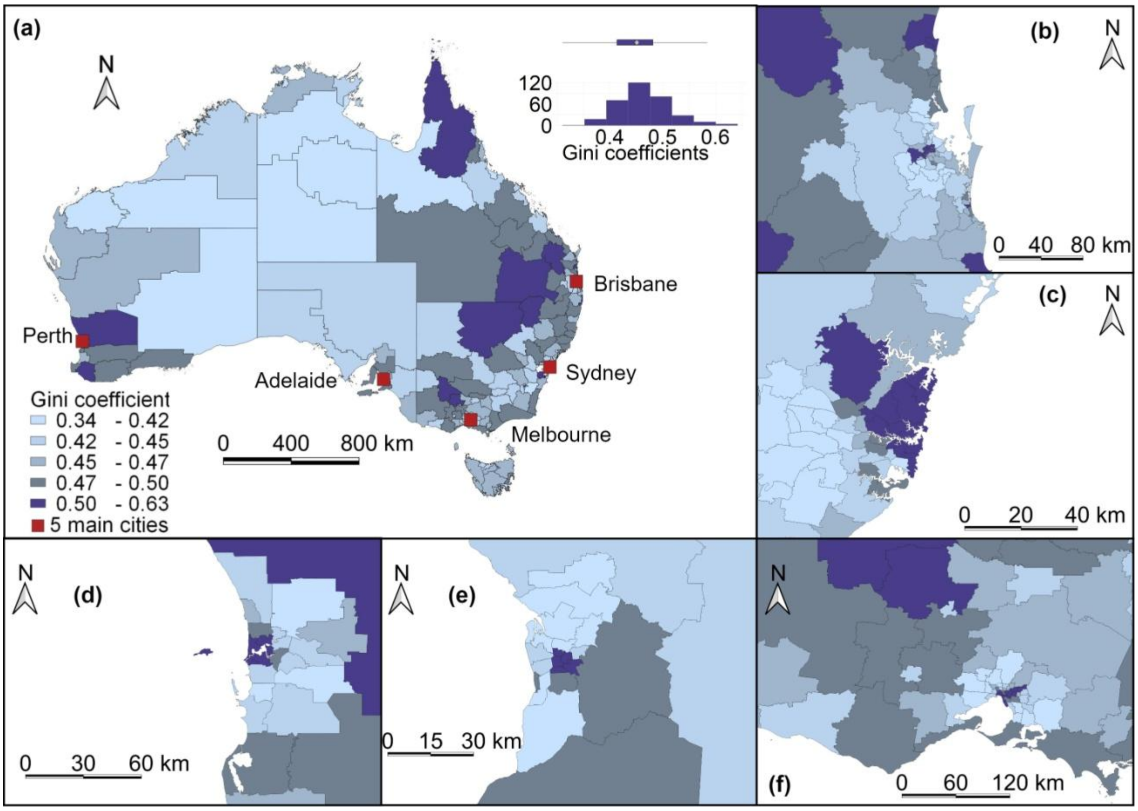


Figure 6-2. Spatial distributions of Gini coefficients in Australia (a) and major cities, including Brisbane (b), Sydney (c), Perth (d), Adelaide (e) and Melbourne (f).

Several factors may cause the unbalanced economic distribution, as indicated by the Gini coefficient. These factors include education, local income levels, gender, infrastructure development, and financial status of the community (Rodríguez-Pose and Tselios, 2010; Solga, 2014; Liczbińska and Sobkowiak, 2020; Klenert et al., 2018). In addition to these factors, the industrial sector also plays a key role in economic development. Three key industries in Australia, including manufacturing, mining, and utility supply and waste service, contributed 25% of the national GDP in 2016 (Australian Bureau of Statistics, 2018). A study has shown that industrialization also have impacts on wage inequality (Sbardella et al., 2017). Thus, industrial company, industrial employee scale, and industrial area scale may have an impact on economic inequality distribution. This study will explore the relationship between economic inequality and these influential variables. This study will also test the ability of geocomplexity to explain unknown residuals from traditional models by using different spatial matrices, which will be compared with another spatial index.

6.3.2 Explanatory variable data

The datasets containing potential explanatory variables in this study were sourced from ABS, OpenStreetMap (OSM), and National Pollutant Inventory (NPI), as demonstrated in Table 6-1. The study considers eight social and infrastructure variables that may be associated with economic inequality. These variables include sex ratio, internet accessibility rate, higher education level, median income level, house ownership rate, number of key industrial companies, and number of industrial employees. The data for these variables were sourced from ABS at the SA3 level in 2016 (Australian Bureau of Statistics, 2020), under categories of population and people, income, education and employment, and family and community. The corresponding SA3 spatial boundaries were also obtained from the ABS data archives (Australian Bureau of Statistics, 2016). Furthermore, this study also incorporates the industrial area scale as an independent variable. The industrial scale is calculated as the ratio of the total industrial area within the SA3 region to the corresponding SA3 area size. In this study, the industrial area encompasses regions and lands that are designed to support industrial activities such as manufacturing, mining and utility supply, and waste service. These industrial regions contain both land use planned for industrial activities and areas with a high density of industrial infrastructures. Industrial land uses are represented by OSM land use polygons tagged with 'industrial' in this study because the definition of OSM industry coincides with the three key industries under the scope of Australian industry (Australian Bureau of Statistics, 2018). High-density industrial infrastructure areas are regions that have a number of relevant infrastructures recorded by the Australian government at NPI or outlined by OSM, and the infrastructure density is measured by a kernel density estimation approach. OSM-based land use polygon and industrial infrastructure data for the year 2016 is available at the OSM website (Geofabrik and OpenStreetMap contributors, 2020). Industrial infrastructures recorded by the Australian government from NPI are available at the government-maintained website (Department of the Environment and Energy, Australian Government, 2020). The identification of final industrial areas follows a validated spatial methodology based on Australian Statistical Geography Standards (ASGS) (Song et al., 2018b; Zhang et al., 2022).

Table 6-1. A summary of explanatory variables

Variable name	Unit	Data description	Source
Sex ratio	%	Sex ratio	ABS
Internet coverage	%	Internet accessed from dwelling	ABS
Higher education ratio	%	Higher education level	ABS

Income	AUD	Median income level	ABS
House ownership	%	Percentage of people who owned a house	ABS
Industrial company	count	Number of industrial companies	ABS
Industrial employee	count	Number of industrial employees	ABS
		Industrial land use polygon	OSM
Industrial scale	%	Key industrial infrastructure	OSM
		Key industrial factory	NPI

6.3.3 Experiment design

This research contains five key stages as demonstrated in Figure 6-3. First, basic regression assumptions should be satisfied, which is indicated by correlation test and multicollinearity test. Variables that satisfy regression assumptions, given the case study of economic inequality, are selected. Second, geocomplexities for selected determinants are computed based on proposed definitions. By applying further tests of regression assumptions, geocomplexities that are relevant to economic inequality are chosen. Next, three traditional models (i.e., multiple regression, SVR and GWR) are applied with optimised parameters to estimate nationwide economic inequality. These three selected regression models have been applied in various applications, given their properties. Multiple regression (i.e., linear regression) is the simplest and most commonly used model for estimation and prediction. SVR is the regression model with machine learning techniques, which has more power of explanation and has also been widely used (Nieto et al., 2021). GWR is one of the most commonly used regression techniques modelling geospatial information (Fotheringham, 2002). These methods are selected as classic representatives of regression model. Then, selected geocomplexities are utilised to estimate residuals from three regression models using GWR to measure to what extent different types of model errors can be explained by the spatial index.

Finally, the power of error explanation of geocomplexity using binary spatial matrix is compared with that of local Geary's C and geocomplexity using row standardised matrix.

In this study, correlation test and multicollinearity test are applied to help select influential variables and geocomplexities that satisfy regression assumptions from a number of candidate variables. The correlation test is the test of Pearson correlation coefficient. A p-value greater than 0.05 for the Pearson correlation coefficient with Gini coefficient indicates that the corresponding variable or geocomplexity is not statistically relevant to economic inequality. This study applied two multicollinearity tests indicated by variance inflation factor (VIF) value to diagnose the inter-relation among variables and validate the basic regression assumption. A threshold of 2.5 is set for VIF value, indicating multicollinearity, and factors with VIF value higher than 2.5, indicating considerable collinearity, are filtered (Johnston et al., 2018). The multicollinearity test is applied twice in this study. The first multicollinearity test during the economic inequality factor selection process is applied to validate the regression assumption for selected influential factors. The second multicollinearity test for geocomplexity selection further diagnoses the inter-association among selected influential factors and their geocomplexities.

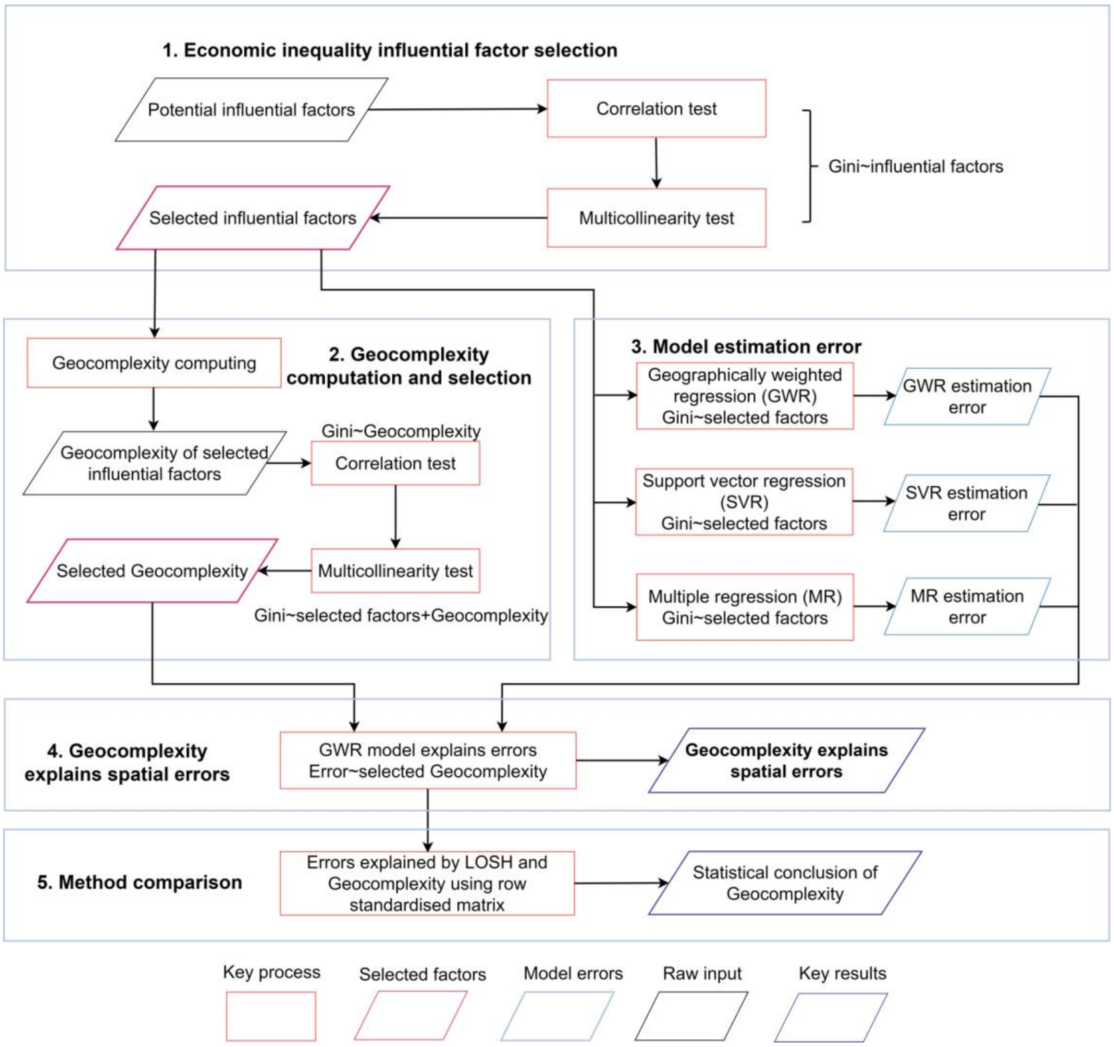


Figure 6-3. The workflow for assessing geocomplexity of economic inequality and its contribution to explaining spatial errors.

6.3.4 Models and errors

This study employs three models (i.e., multiple regression, support vector regression, and GWR) to investigate the relationship between economic inequality and determinants. For the multiple regression, model parameters are calculated based on the Least Squares using R packages. For the SVR model, optimal machine learning parameters, including the cost value and the gamma value, are determined by ten-fold cross-validation, and SVR estimation results are computed by the 'e1071' R language package. In this study, the cost of SVR is selected from all values from 0.1 to 100 with

0.1 as a step. For the GWR model, the optimal bandwidth is determined by cross-validation operated by the ‘spgwr’ R language package. Starting with an initial value from 0 to 1, the performance of bandwidth is assessed using leave-one-out cross validation. The optimal bandwidth should have the best estimation performance, and the change of performance is small enough till the next iteration for searching. The GWR bandwidth in this study is adaptive based on k-nearest neighbours. The GWR model is applied to explain the cause of economic inequality with spatial concerns. At the final stage, selected geocomplexities will be applied to explain errors from these three models. The GWR model is a quantitative analysis approach representing the second law of geography, and this spatial method is further used to show to what extent spatial disparity of geocomplexity can explain errors from three traditional models.

6.4 Results

6.4.1 Determinants of economic inequality

Based on the correlation test, six variables, including higher education ratio, sex ratio, income, house ownership, industrial employee, and industrial scale, were initially identified as potential determinants of economic inequality. The multicollinearity test revealed that the higher education ratio had a high degree of collinearity with other factors, with a VIF value over 8, which is higher than the threshold value 2.5. To satisfy regression assumptions, higher education ratio was removed, leaving the remaining five variables with VIF values lower than 2.5, including sex ratio, income, house ownership, industrial employee, and industrial scale, are selected. Table 6-2 shows the global coefficients and performance of the multiple regression model for the five selected determinants. The p-values for income, house ownership and industrial employee were less than 0.001, indicating their significant contributions in explaining economic inequality. The results suggest that a higher median income level, higher percentage of house ownership, and lower number of industrial employees could increase economic inequality. The general model with five selected factors is acceptable as indicated by the p-value for F-statistic. Spatial local complexities were computed for the five selected determinants, and relevant spatial local complexities were applied to explain unknown errors from three traditional models.

Table 6-2. Multiple regression for modelling Gini coefficients at the SA3 level

Variable	Coefficient (significance level)
Intercept	2.65e-01 (***)
Sex ratio	-9.01e-03
Income	2.12e-06 (***)
House ownership	3.04e-03 (***)
Industrial employee	-2.54e-06 (***)
Industrial scale	-3.27e-02
p-value for model	<0.001
R-squared value	0.47
AIC value	-1288

Note: (***) indicates significance level indicated by p-value less than 0.001.

6.4.2 Geocomplexity of explanatory variables

Spatial local complexities of income and industrial employee are correlated to economic inequality as indicated by the correlation matrix shown in Figure 6-4. By further introducing these two spatial local complexities, the VIF values for all previous selected factors and two correlated spatial local complexities are no greater than 2.5. Therefore, these two spatial local complexities are then selected and considered as influential geography impacts to economic inequality.

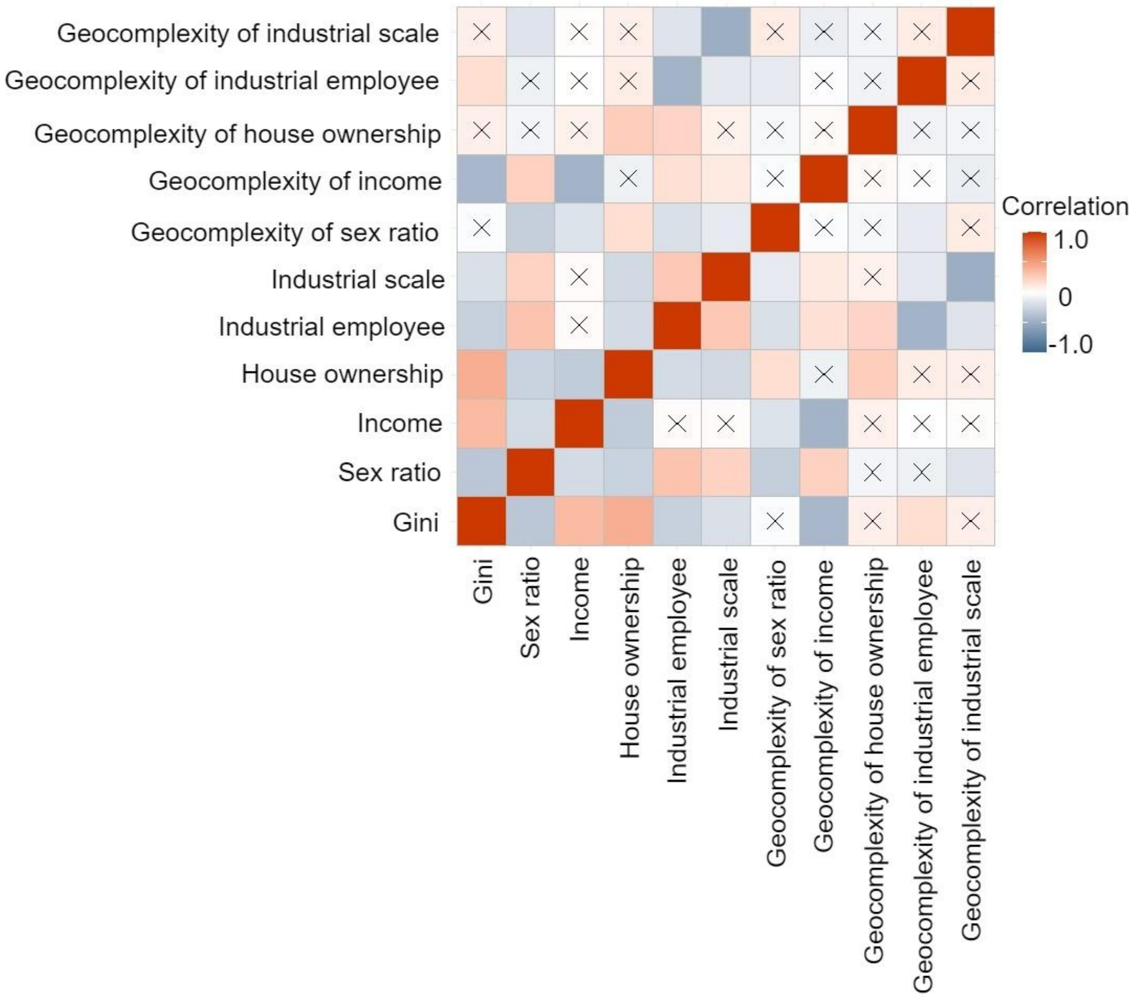


Figure 6-4. The correlation matrix among Gini coefficients, selected explanatory variables and geocomplexities of explanatory variables. Note: ‘x’ indicates that the correlation is not significant.

Figure 6-5 shows the statistical distributions of two selected spatial local complexities for factors compared with the corresponding factor values. The scatter plots show the distributions of income level and industrial employee against their local spatial complexity patterns, with each point representing an SA3 region. According to Figure 6-5(a), a higher spatial complexity of local income tends to locate in SA3 regions with low-medium income levels, and a similar trend is shown in the industrial employee count in Figure 6-5(b). The findings suggest that regions with either very high or very low values of income or industrial employee tend to exhibit simple local spatial patterns, with those having the highest values displaying the most simple local spatial patterns. In contract, regions with moderate values show the highest levels of geocomplexity.

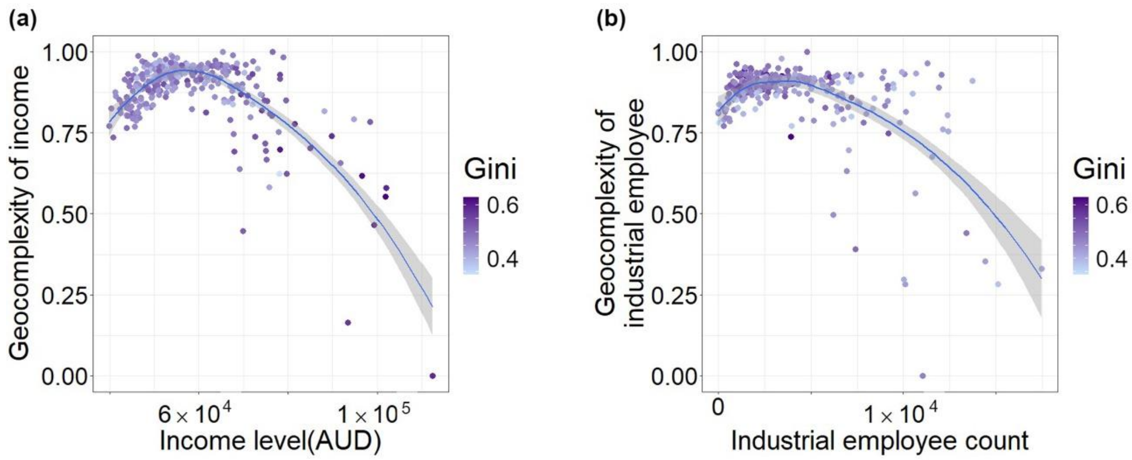


Figure 5. Geocomplexity of selected variables in SA3 regions colored by Gini coefficient. (a) Income level v.s. Geocomplexity of income. (b) Industrial employee count v.s. Geocomplexity of industrial employee.

6.4.3 Errors from three traditional models

Three models, including multiple regression, SVR, and GWR, are applied to estimate nationwide economic inequality. The process of determining model parameters is explained in section 6.3.4. Table 6-3 demonstrates the performance of the three models, while Figure 6-6 shows statistical distributions of estimation errors. Using the Least Squares method, the multiple regression can explain 47% of the economic inequality. Using ten-fold cross-validation, the SVR model selects a cost value of 1.4 and a gamma value of 0.2 for the radial kernel as optimal parameters. With these optimal parameters, the SVR model can explain 65% of the economic inequality. By applying cross-validation for the adaptive bandwidth selection, the GWR estimation performs the best among the three models with an R^2 of 76%.

Table 6-3. Performance of three models

	MR model	SVR model	GWR model
R squared value	0.472	0.646	0.761
RSS value	0.396	0.266	0.179
Parameters determined by	Least Squares	Cross-validation	Cross-validation

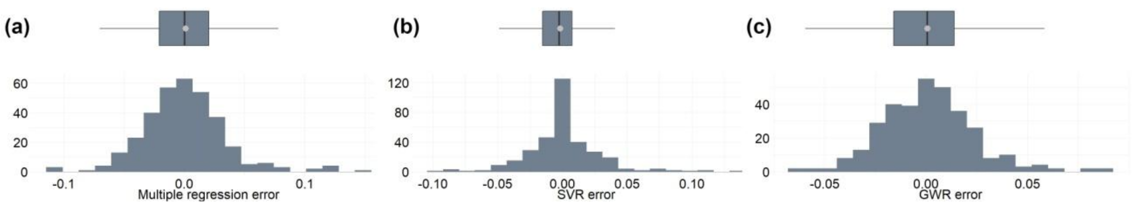


Figure 6-6. Statistical distribution of model errors. (a) Errors of multiple regression estimations. (b) Errors of SVR estimations. (c) Errors of GWR estimations.

6.4.4 Geocomplexity explains spatial errors

6.4.4.1 Overview of explain spatial errors using geocomplexity

This research aims to further investigate the extent to which spatial local complexities can explain estimation errors of traditional models, and how selected geocomplexities can explain spatial errors. Considering the spatial variation of geography impacts, we used GWR to quantify the explanation of selected spatial local complexities on estimation errors. Table 6-4 shows the statistical summaries of significant results of three error explanation models based on GWR. The spatial disparity of selected complexities can explain 47% of linear regression errors, 17% of SVR regression errors, and 14% of GWR errors. The residual-fitted plots of error explanation models are shown in Figure 6-7(a), and residuals are randomly distributed around the horizontal line of 0 without significant linear or non-linear patterns. Thus, there is no significant non-linear or heteroscedasticity problem for the three explanation models. Q-Q plots from Figure 6-7(b) to Figure 6-7(d) demonstrate probability distributions of residuals from three error explanations compared with randomly generated data with normal distribution. For residuals from multiple regression and GWR error explanation models, a majority of points fall on the reference line as shown in Figure 6-7(b) and Figure 6-7(d).

Table 6-4. A summary of the contribution of geocomplexity to explain spatial errors

Count of significant results and explanation model summaries	MR error explanation	SVR error explanation	GWR error explanation
Geocomplexity of income	237	105	53
Geocomplexity of industrial employee	95	70	3
Global R^2	0.467	0.166	0.136
AIC value	-1459	-1477	-1575
RSS value	0.211	0.221	0.155

Note: The statistical significance level is 0.05. The total observation count is 333.

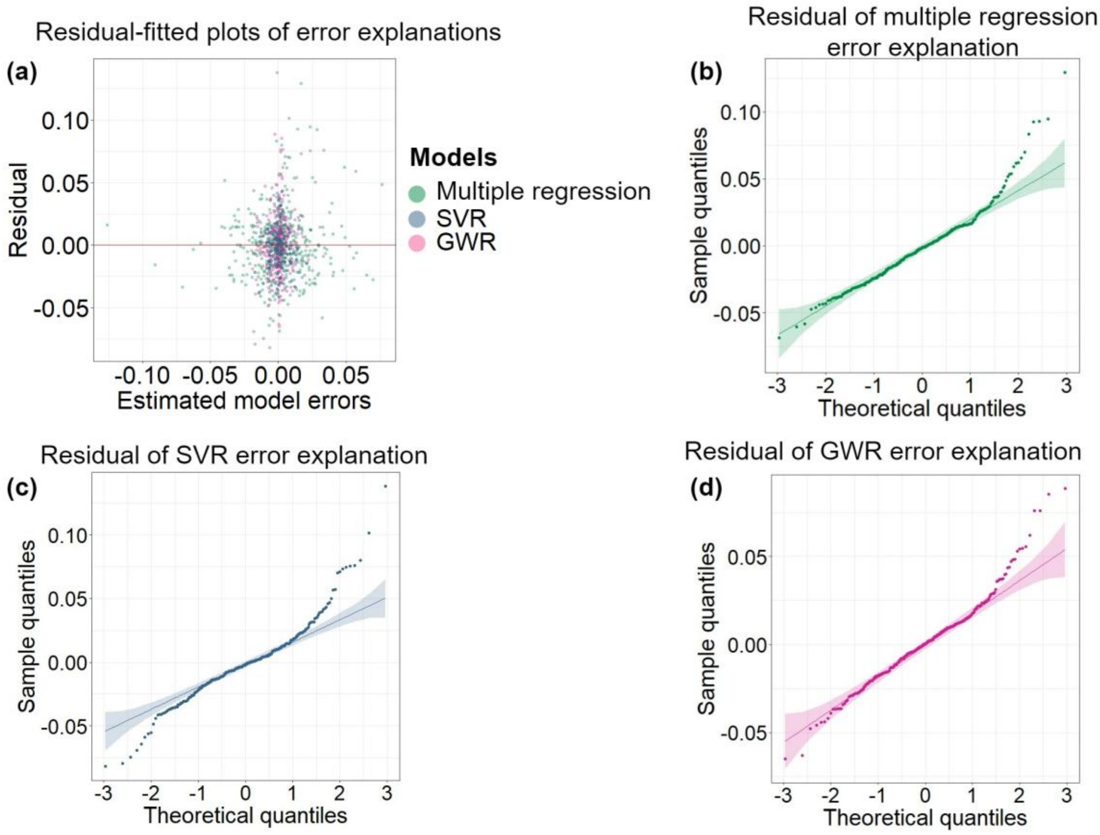


Figure 6-7. Residuals of error explanations. (a) Residual-fitted plot of error explanation from three models. (b) Q-Q plot showing residuals of multiple regression error explanation. (c) Q-Q plot showing residuals of SVR error explanation. (d) Q-Q plot showing residuals of GWR error explanation.

6.4.4.2 Spatial explanation of multiple regression errors by geocomplexities

Figure 6-8 demonstrates how estimation errors from multiple regression can be spatially explained by selected spatial local complexities. Spatial local complexity of income and industrial employees can significantly explain linear errors in a many regions in the Australian continent. The spatial local complexity of income is positively significant in Northern Territory and remote areas of New South Wales, while the geocomplexity coefficient is negative significant in a majority of regions, including Melbourne, Brisbane, Perth as well as other inner regional and remote areas. Figure

6-8(b) demonstrates the explanation of geocomplexity in Sydney's surrounding regions, and Figure 6-8(c) shows that the spatial explanation for linear errors varies in Melbourne. The spatial local complexity of industrial employees is positively significant in Melbourne and the inner region of New South Wales and Brisbane, but negative significant in Adelaide and surrounding regions.

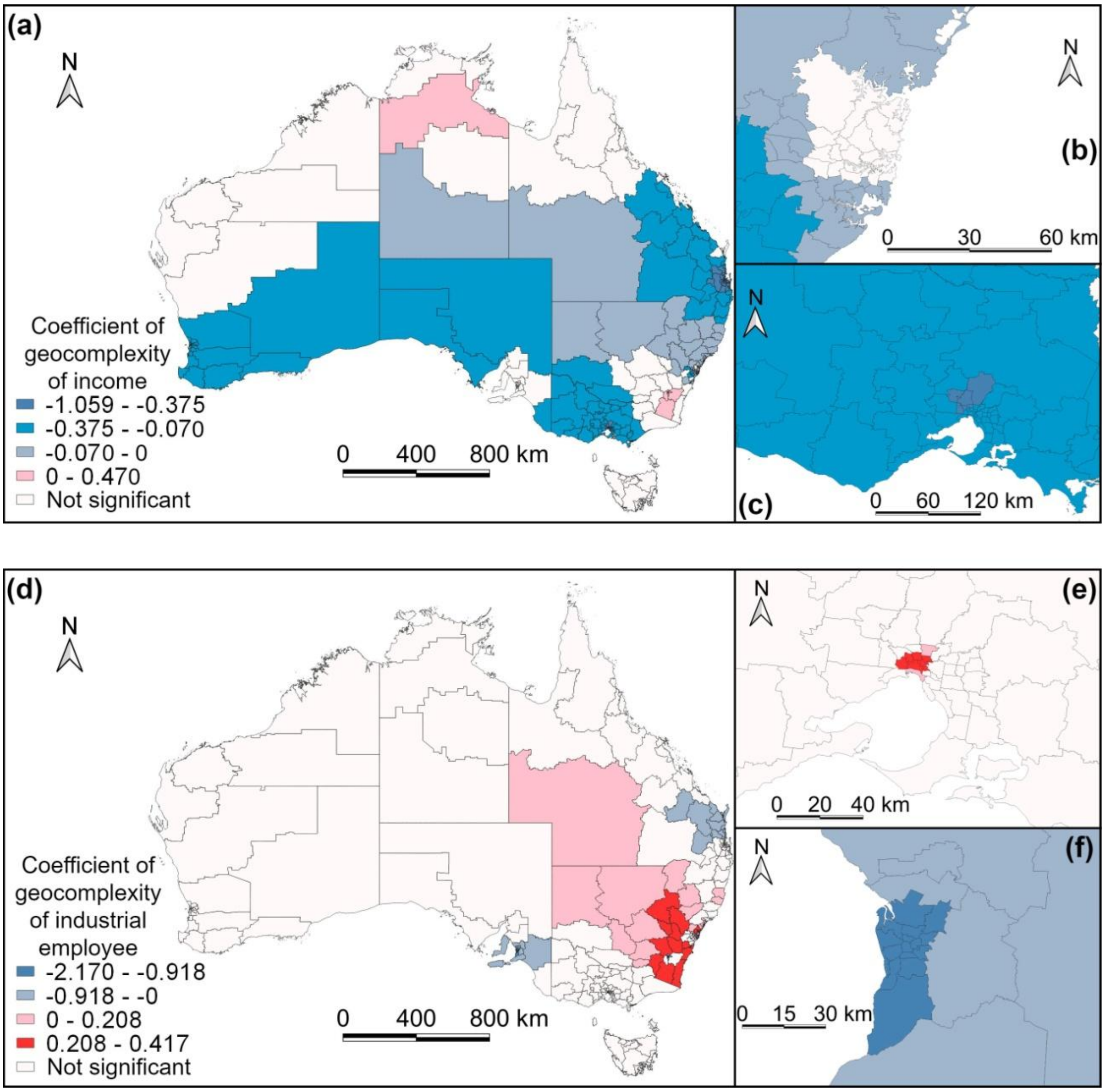


Figure 6-8. Coefficients of geocomplexity of income and industrial employee in multiple regression error explanation. Distribution of geocomplexity of income in Australia (a) and major cities including, Sydney (b) and Melbourne (c). Distribution of geocomplexity of industrial employees in Australia (d) and major cities including, Melbourne (e) and Adelaide (f).

6.4.4.3 Spatial explanation of support vector regression errors by geocomplexities

Figure 6-9 illustrates how spatial local complexity can explain errors from SVR estimations. As a result, significant results are distributed in the East part of the continent. Spatial local complexity of income is negatively associated with SVR errors, and significant results of income complexity are clustering in Melbourne and Brisbane as shown in Figure 6-9(a). Furthermore, the higher absolute value of the coefficient in Brisbane indicates that SVR errors are more sensitive to the change of spatial impact in Brisbane and surrounding areas. In terms of spatial local complexity of industrial employees, a majority of significant results demonstrate a positive relation, and a small number of negative significant results are distributed in the Northern Greater Brisbane Area as shown in Figure 6-9(d) and Figure 6-9(e). The positive significant coefficient of spatial local complexity of industrial employees is higher in the Southern of the continent.

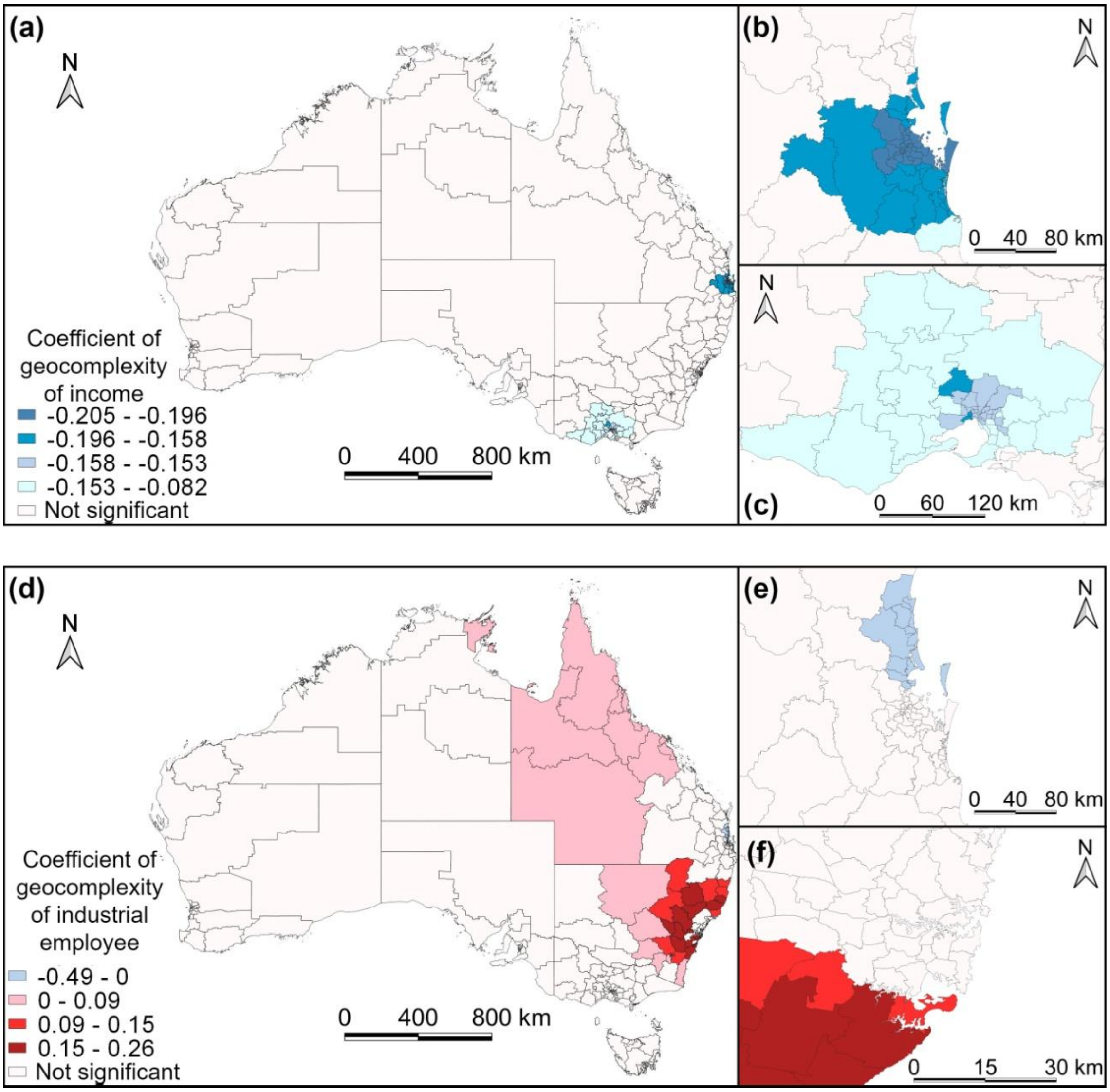


Figure 6-9. Coefficients of geocomplexity of income and industrial employee in SVR error explanation. Distribution of geocomplexity of income in Australia (a) and major cities including, Brisbane (b) and Melbourne (c). Distribution of geocomplexity of industrial employee in Australia (d) and major cities including, Brisbane (e) and Sydney (f).

6.4.4.4 Spatial explanation of geographically weighted regression errors by geocomplexities

The spatial disparity of two selected complexities can also further explain the remaining errors from the GWR model only concerning five influential factors. As shown in Figure 6-10(a), spatial local complexity of income is negatively associated with GWR errors in 53 SA3 areas and spatial local complexity of industrial employees is positively associated with GWR errors in three SA3 areas in New South Wales. Significant results of spatial local complexity of income are distributed in the surrounding inner regions of the Greater Brisbane Area as shown in Figure 6-10(b). The absolute value of the income complexity coefficient is higher in regions close to Brisbane than in other regions in Victoria and South Australia.

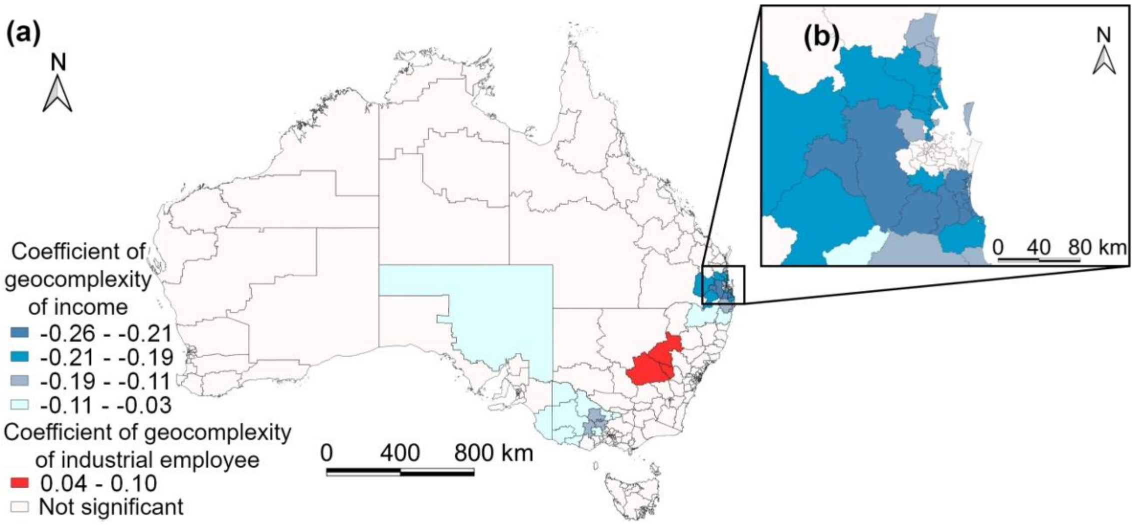


Figure 6-10. Coefficients of geocomplexity of income and industrial employee in GWR error explanation (a). Distribution of geocomplexity of income in Brisbane (b).

6.4.5 Method comparison

Our study proposes a new spatial index named 'geocomplexity', which indicates the pattern of spatial local complexity based on a Moran's measure. This index is composed of LISA and all spatial associations between connecting neighbours of the target area with Moran's measure. This spatial index is derived from and naturally connected to the local spatial heterogeneity index (LOSH). In this section, we add further tests, given the same case study of the association between Gini coefficient and independent variables, using LOSH represented by local Geary's C and geocomplexity with row standardised spatial weight matrix to explain errors from three traditional models. Further tests follow the same workflow and criteria of factor selection as shown in Figure 6-3. As a result, two spatial indexes for both income and industrial employee, identical to factor selection results shown in the workflow, are selected to explain errors from three models.

The performances in terms of the power of error explanation using three spatial indexes are shown in Table 6-5. Errors from three traditional models are explained by spatial indexes of local income level and industrial employee count. Geocomplexity with binary spatial weight matrix, with detailed results demonstrated before, has the best power of explanation for errors among all three methods. By replacing with row standardised spatial weight matrix for the computation of geocomplexity, the power of error explanation decreases. Geocomplexity with row standardised matrix can explain more linear regression errors than local Geary's C, but the power of explanation for the other two model errors is no different from that of LOSH.

Table 6-5. The comparison between geocomplexity and local spatial heterogeneity index (LOSH) on error explanation.

Performance (Global R^2, RSS, AIC)	Geocomplexity with binary spatial weight matrix	Geocomplexity with row standardised spatial weight matrix	local spatial heterogeneity index (LOSH)
Multiple regression error explanation model	**47%, 0.21, -1459**	42%, 0.23, -1430	38%, 0.25, -1418

SVR error explanation model	**17%, 0.22, -1477**	6%, 0.25, -1444	7%, 0.25, -1448
GWR error explanation model	**14%, 0.16, -1575**	12%, 0.16, -1567	12%, 0.16, -1569

6.5 Discussion

This study proposes a spatial index to measure the concept of complexity regarding spatial dependence. As geospatial impacts are not constant across space, the spatial complexity may vary in different places. Therefore, this research applies GWR models to quantify the spatial association between economic inequality and selected variables and explain unknown errors from traditional models by capturing the spatial heterogeneity of relevant complexities. The aim of a series of experiments is to examine whether the consideration of spatial local complexity could explain more unknown errors from traditional models and improve model performance. As a result, this assumption is validated.

The power of error illumination from spatial local complexity is also compared with spatial autoregressive models, given the case study of economic inequality estimation. The Lagrange multiplier test of spatial dependence indicates that the spatial error model (SER) is the most suitable among spatial autoregressive models. With p-value in F-statistics is less than 0.001, the SER has an R-squared value of 0.64, indicating the fact that spatial autocorrelation of the error term can explain 32% of the original error. Thus, spatial heterogeneity of geocomplexity with either binary or row standardised spatial matrix can explain more errors. It is worth noting that the concept of geocomplexity in our study is more compatible with a symmetric spatial matrix, given our definition. As shown in Figure 6-1 and equation (6-1) - (6-2), we consider the spatial dependence between neighbours of the target area as a mutual relationship. In other words, the connection status from location '*j*' to ‘*k*’ (or ‘*i*’ to ‘*k*’) and ‘*k*’ to ‘*j*’ (or ‘*k*’ to ‘*i*’) is preferred to be identical. The main purpose of geocomplexity is to further include associations among neighbours when considering spatial dependence and additional information from the row standardised process cannot be reflected.

This research aims to provide an innovative understanding of complexity by extending the understanding of spatial dependence. The geocomplexity is a Moran-based measure (i.e., the multiplication of the Z-score of observations). We also analysed different measures representing differences between observations including Geary’s C measure (i.e., the square of the difference between two values). As a result, the Moran-based measure performs better. Despite stronger power of error illumination, the spatial local complexity may not be a final or the only answer to the question ‘what is complexity in spatial science’. We redefine the concept of ‘spatial complexity’ considering spatial dependence between neighbours, and this could not be the only definition. For instance, the ‘spatial complexity’ can also be redefined considering the third law of geography, which defines a concept of ‘spatial similarity’ that the target variable between two locations is similar if geographic configurations at these two places are identical (Zhu et al., 2018; Song, 2022b). That means a region can be spatially complex where geographic configurations are not closely associated. In future work, the content of ‘complexity’ under the spatial scope can be filled by other meaningful spatial concepts. In terms of further application, the geocomplexity can also be a complexity measure in future remote sensing studies, especially hyperspectral image unmixing (Jia and Qian, 2017). Last but not the least, in future studies, spatial filtering tools can be a method supporting geospatial information modelling along with spatial local complexity. The applications of spatial filtering for selecting influential variables together with geocomplexity for error explanation have the potential to explore the limits of spatial data modelling (Paez, 2019). To sum up, we proposed a geocomplexity index, which is an extension from a spatial autocorrelation indicator. A short comparison between geocomplexity and spatial dependence is shown in Table 6-6.

Table 6-6. The similarity and difference between geocomplexity and spatial dependence from various aspects

	Geocomplexity	Spatial dependence
Concept	To describe a status of complex from the view of spatial dependence	Tobler's first law of geography can best describe spatial dependence and spatial autocorrelation.
Representation form	The extension of LISA by further considering the relationship among neighbours	Global: Moran’s I, Geary’s C Local: LISA, LOSH
Application	Provides error explanation for various models	Spatial autocorrelation indicators; Spatial autoregressive models
Capacity with local models	Capable with local models	Capable with local models

6.6 Conclusion

This study proposes a spatial index named ‘geocomplexity’ to measure the concept of complexity regarding spatial dependence. The index demonstrates the complexity of spatial dependence by considering both the relationship between an interested area and surroundings, and the associations among spatial neighbours of the target area. Given a case study of association between nationwide economic inequality and influential variables, results show that spatial local complexity can explain unknown errors and improve the traditional model performance. The proposed spatial complexity index is suited well to symmetric spatial weight matrix and geocomplexity with binary spatial weight matrix has better power of error explanation than geocomplexity with asymmetric matrix and local Geary’s C. The proposed geocomplexity indicator has the potential for spatial data analysis and relevant applications in both implying complexity in spatial science and illuminating unknown errors from traditional models.

Chapter 7. Conclusion and recommendation

7.1 Introduction

This Ph.D. thesis assesses the sustainability of industrial regions in Australia from both environmental and socio-economic perspectives. For completing the four objectives of this thesis, first the following aspects were systematically reviewed in chapter 2: capability of applying Earth observation data to smart urban construction, sustainable built environment and infrastructure design, and human-environment interactions (HEIs). The review article in chapter 2 lays a theoretical foundation for the assessment of industrial sustainability issues using scientific methodologies and reliable tools. Nationwide industrial regions were identified using a point of interest (POI)-based spatial method, and industrial sustainability was assessed from environmental and socio-economic perspectives. To investigate the spatial association between sustainable indicators and influential factors comprehensively, a robust geographical detector (RGD) was developed. The RGD overcame the gap of the previous method by introducing an optimisation algorithm. To further analyse the impact of industrial features on economic inequality, a geocomplexity was proposed and quantified. The consideration of geocomplexity can improve the performance of the traditional spatial and aspatial models. Finally, scientific recommendations were provided based on the computational results. This thesis provides academic contributions, in terms of both innovative methodology design and computational outputs from the models.

From the perspective of methodology innovation, a RGD was developed to satisfy the research objectives and bridge previous gaps. The RGD model overcame the limitations of previous spatial stratified heterogeneity (SSH) models and was applied in this thesis to detect spatial associations among factors. This method can be applied for spatial planning and management purposes at the design and maintenance stages of the construction management lifecycle, with reliable performance. The design of this novel geographical detector (GD) can also be applied to studies in various research fields, including environmental engineering, urban planning, social data analysis, and spatial science. This thesis also proposes the concept of geocomplexity, which has been quantified and utilised to improve model performance. The concept of geocomplexity is also suitable for application to research topics in spatial analysis.

From the perspective of research results, industrial built environment sustainability has been comprehensively assessed from both environmental and socio-economic perspectives by exploring spatial big data. Factors determining air pollutants and spatial disparities in industrial regions were identified and analysed. The pace of industrial development was also assessed based on the urban scaling theory. These results demonstrate the sustainable development of industrial regions from multiple perspectives. Further scientific decision-making is supported by these reliable outcomes.

In this chapter, the thesis has been concluded from the following three perspectives. First, the four research objectives listed in chapter 1 are revisited. The key results and findings are summarised. Scientific advice is then given to policymakers and stakeholders, followed by a summary of the research contributions. Finally, recommendations for future research are presented in the final section.

7.2 Revisiting research objectives

7.2.1 Exploring environmental sustainability of industrial regions: spatial disparities of factors affecting air pollutants in industrial regions

The first objective was to investigate the environmentally sustainable development of industrial regions nationwide in Australia. Environmental sustainability was assessed by investigating the factors affecting air pollutants and the relevant spatial disparities in industrial regions. The first objective involved assessment of environmental sustainability in industrial regions and laid a solid foundation for subsequent research.

This objective has been achieved by addressing the following sub-objectives. First, industrial regions were identified based on the POI and OpenStreetMap (OSM) datasets by adjusting the spatial framework. Then, a variety of remote sensing (RS) and spatial datasets representing air pollutant densities, environmental forces, industrial scales, and human activities were collected from the Google Earth Engine (GEE) and open-access platforms. Factors affecting air pollutant densities and relevant spatial patterns were calculated using geographically weighted regression with standardised coefficients. Finally, spatial patterns of factors affecting air pollutants in industrial regions were summarised at a higher level of spatial granularity.

The final results from the models revealed precipitation, wind speed, and road density to be factors affecting air pollutant densities in industrial regions. The disparity in determining the factors affecting air pollutants was also evident. In northern Australia, industrial scale and human activities are the predominant factors affecting air pollutant densities. However, in major cities, vegetation, elevation, and meteorological factors are the dominant factors affecting air pollutant levels.

The sub-objectives of the first objective are the foundations of the following research topics. The geographical boundaries of the industrial regions were identified using POI-based spatial methods. These spatial methods and results were consistent with the Australian Statistical Geography Standard (ASGS) and were further utilised in addressing the subsequent objectives. Furthermore, RS datasets collected from open-access platforms represent the properties of industrial regions, and these data have been applied in subsequent analyses.

7.2.2 Extended exploration of environmental sustainability of industrial regions: spatial associations between air pollutants and influential factors indicated by RGD

The second objective was to explore environmental sustainability in industrial regions nationwide. Despite the results generated from the first objective, the spatial association between air pollutants and influential factors required exploration. The second objective was to deliver spatial outcomes using an innovative methodology design. This research objective paid more attention to the spatial association between selected air pollutant density and remote sensing factors, including night-time light, vegetation indexes, and wind speed. The completion of the second objective yielded information supplementary to the findings from the previous objective.

This objective was satisfied by developing a RGD model that overcame the limitations of previous SSH models. By introducing a change detection optimisation algorithm into a GD model, a RGD was constructed, which could determine the optimal spatial zones for different numbers of intervals. As a result, the RGD exhibited greater accuracy, stability, and robustness of results than the previous model. This innovative method was applied for industrial sustainability assessment. The spatial association between nitrogen dioxide (NO_2) density and RS factors was analysed using a RGD. The results showed that NO_2density (followed by wind speed and vegetation greenness) has the greatest association with the night-time light (NTL) factor.

7.2.3 Assessing socio-economic sustainability of industrial development from an urban scaling perspective: industrial features associated with economy, infrastructure, and innovation

The third objective was to assess the socio-economic sustainability of nationwide industrial features. Current knowledge regarding the pace of development of industrial features (i.e. industrial region scale, industrial companies, and industrial employees) is limited. Furthermore, urban factors associated with industrial growth deserve further exploration.

This objective was satisfied by analysing the scaling pace of industrial features and other urban indicators in 101 Australian cities. This objective mainly focused on the development of industrial and other urban indicators in Australian cities. Further datasets representing economic, infrastructure, and innovation indicators were collected from the census data provided by the Australian Bureau of Statistics. The pace of development of multiple urban indicators was identified by urban scaling models, and the association between industrial features and urban indicators was assessed using the RGD.

The development of industrial regions is tied to cities, and this phenomenon indicates that the pace of industrial expansion can be measured using the urban scaling theory. Scaling models indicate that industrial features develop at a sublinear pace. The super-linear pace of economic development and sublinear pace of infrastructure development are also indicated by the scaling models. This study provides both general results on Australian urban scaling development and specific spatial association results on industrial scaling features. In general, this study validates the consistency of scaling development among Australian cities using power-law theory and the similarity of scaling disparity features among top-populated cities. Specifically, the urban innovation indicator and income level were predominantly and positively associated with industrial companies and employees, indicating that innovation growth and economic development in Australian cities would stimulate the performance of industrial companies and the employment scale. The synergy between urban innovation and industrial company performance is particularly significant in major capital cities. The developed spatial models hold broad potential to address the spatial and scaling characteristics of industrial features.

7.2.4 Further investigation on socio-economic impacts of industrial features: spatial impacts of industrial development on national economic inequality

The final objective was to explore the impact of industrial features on economic inequality. The sustainability of industrial development has been assessed from both the environmental and socio-economic perspectives. However, the impact of industrial development remains to be investigated. The manner in which industrial features are related to economic inequality was demonstrated through the completion of this objective. The impact of industrial features on the distribution of economic inequality was assessed spatially and aspatially.

The final objective was satisfied by modelling the nationwide economic inequality status with consideration of spatial local complexity at a higher level of spatial granularity. First, relevant socio-economic datasets were collected from the national census. Second, the impact of industrial features on economic inequality was assessed using traditional spatial and aspatial models. Then, the concept of geocomplexity was proposed and quantified using the indicator of spatial local complexity. Finally, spatial local complexity was applied to explain the errors in traditional models.

The results from the traditional model indicate that the nationwide economic inequality phenomenon is related to income distribution, house ownership, and the local industrial employee scale. The spatial distribution of economic inequality is also subject to the complexity of the spatial impacts of the local income distribution and local industrial employee scale. The disparity in spatial local complexity can explain the errors in linear regression, machine learning models, and spatial models.

7.3 Advice for smart industrial built environment in Australia

Advice for achieving and maintaining a smart Australian industrial built environment can be derived based on the results yielded upon completion of the four research objectives. According to the results of the first two research objectives on environmental sustainability assessment, although road density and meteorological factors are generally influential, spatial disparities of factors affecting air pollutants in nationwide industrial regions are evident. In urban areas, air pollutants in industrial regions are mainly influenced by natural factors. However, in northern Australia, human activity factors and the industrial scale are key factors affecting air pollutant levels in industrial regions. From the perspective of spatially stratified heterogeneity, NTL is more strongly associated with air pollutants in industrial regions than meteorological factors and vegetation greenness. Given the environmental sustainability results, policymakers and stakeholders of industrial regions are advised to pay more attention to environmental issues and human activities in northern areas. Environmental sustainability monitoring can be based on the analysis of Earth observation data. Considering the high maintenance cost of air pollutant monitoring stations in northern areas, researchers and scientific teams are advised to monitor air pollutants in these regions based on open-access satellite data that is updated daily.

According to the results of the socio-economic sustainability assessment, industrial features are tied to cities and urban industrial features (i.e. industrial region scale, industrial employees, and industrial companies) are developing at linear and sub-linear speeds. The development of industrial features is strongly associated with the pace of income increase and innovation development. Thus, policymakers should pay more attention to professional services from industries with innovation outputs, while planning smart industrial regions for mining, manufacturing, utility supply, and waste services. Furthermore, the concept of geocomplexity indicates that the spatial impact of industrial features on economic inequality is significant. Therefore, the industrial development of surrounding areas is tied to the local economic distribution, and accordingly, policymakers should consider the spatial effect.

7.4 Contributions of this study

7.4.1 Theoretical contributions to spatial science

The design of a RGD and concept of geocomplexity make theoretical contributions to spatial science. First, the RGD design proved the feasibility of overcoming the sensitivity of spatial zone determination for traditional GDs. GDs have been proposed and applied in spatial data analysis for a decade, and the optimal parameter-based GD is an advanced model used to improve GDs. Despite improvements in model performance and result interpretation, SSH values are sensitive to spatial zone segmentation based on independent variables. By tracking the trend of spatial stratified heterogeneity values from these two models, the value fluctuates with an increase in the number of spatial zones, while spatial associations between variables are supposed to become clearer with the increase in number of spatial zones. The design of the RGD proves that it is possible to further improve the performance of GD models by clearly indicating the association between factors. By transforming the spatial zone segmentation problem into a pure optimisation problem and applying a change point detection (CPD) algorithm, the RGD guarantees an increasing association between factors with additional spatial zones, and the model results are consistent with intuitiveness.

Second, the expression of spatial local complexity defines the concept of complexity from spatial impact. The concept of complexity is ubiquitous in both science and engineering. Local spatial complexity defines the concept of complexity in spatial science. Furthermore, this research proves that the complexity of the spatial distribution of factors can explain unknown errors in traditional models. In other words, it is necessary to consider the complexity of spatial science.

7.4.2 Practical contributions to smart built environment management

The application of innovative spatial methods and analysis results on industrial built environment sustainability makes practical contributions to smart built environment management. The application of new spatial methods adds value to smart built environment management. To propagate usage of geographic information system (GIS) in built environments and construction fields, new spatial methods—consistent with intuitiveness—with reliable and accurate results are acceptable to experts from non-spatial fields. Thus, these innovative methods can popularise the application of GIS functions, particularly spatial planning functions, in the built environment and construction management. Furthermore, new spatial methods support scientific decision-making and smart built environment management, with improved model performance.

The results of industrial built environment sustainability assessment contribute to smart built environment and construction management during the design and maintenance phases. This study uncovered the properties of industrial built environment development from an urban scaling perspective and identified the synergy between industrial features and other urban indicators. These socio-economic results, demonstrating the pace of industrial development in Australian cities, are helpful in understanding the industrial expansion process and achieving efficient resource allocation. These outcomes are beneficial to the smart built environment and construction management during the design and planning phases. Furthermore, this study assessed environmental sustainability by investigating air pollutants and their influential factors, as well as spatial impacts of industrial features on economic inequality. The results demonstrate the impacts of current industrial built environment development and provide guidelines and advice for urban and regional governance during the maintenance phase of the construction lifecycle.

7.5 Recommendations for future research

This Ph.D. thesis provides academic contributions, in the form of innovative methodology design and computational results. Future research can be undertaken through spatial method improvement, software development, systematic sustainability analysis, and smart urban construction and built environment management practices.

Despite innovation and breakthrough in geospatial analysis methods design, the theories and models from this thesis is not perfect. Future research can be undertaken to overcome relevant limitations. First, a robust geographical detector has been proposed to overcome the geographical detector’s sensitivity to variables’ statistical distributions. This method explores the maximum limit of spatial associations between two factors. However, in some cases, the target variable may be subjected to the interactive forces of multiple variables. Thus, a further optimisation algorithm is required to determine optimal spatial zones indicating the interaction of multiple variables. Second, we have proposed the concept of ‘geocomplexity’, which is quantified by spatial local complexity, from the view of spatial dependence. The effectiveness of geocomplexity in error explanations has been demonstrated, while this might not be the only understanding of complexity in spatial science. The meaning of complexity can be further extended from the view of spatial similarity.

In terms of software implementation, programming for the design of a RGD and assessing geocomplexity can be developed further. The RGD was accomplished using the public Python scientific library ‘ruptures’ for signal processing and the R language-based library ‘geodetector’. Currently, the Python library ‘ruptures’ computes the optimal spatial zones for the original GD, and the final value of SSH is computed by ‘geodetector’. This version of the RGD is supported by functions from two programming environments, and its design is complicated for experts from non-spatial fields. In future studies, an integrated and public R library can be developed for the RGD, to explore spatial heterogeneity. The implementation of this method can improve the efficiency of spatial analysis for experts from spatial fields and can propagate the concept of spatial science to non-spatial experts. To compute geocomplexity, this spatial concept was programmed with self-design in the R language. In future studies, a full R package can be developed to compute the spatial local complexity, explain spatial errors, and generate better results in an integrated function.

In terms of sustainability analysis, this thesis measures sustainable development in Australian industrial regions from environmental and socio-economic perspectives. Environmental sustainability is assessed by understanding the spatial patterns of influential factors of air pollutants, and socio-economic sustainability is measured from the scaling pace of development and economic inequality impacts.

These topics assess the development of industrial regions nationwide from multiple perspectives. However, sustainability covers various issues, and this study solves a minority of them, which can be assessed using Earth observations and spatial techniques. Thus, in future research, sustainable development of industrial regions from environmental and socio-economic perspectives can be explored through issues regarding climate change, ecosystems, energy consumption, vulnerable populations, and social safety. Furthermore, the sustainability analysis framework for industrial regions can be applied to other countries or states where mining, manufacturing, or utility supply industries play a key role.

Smart urban construction and built environment management can be achieved by the propagation of new geospatial functions and systematic sustainability analyses of industrial regions and relevant infrastructures. This thesis proposes new spatial analysis methods and measures sustainability by analysing Earth observation big data. Scientific decision-making and smart urban construction can be achieved with the help of popularised GIS methods and scientific analysis results for sustainable development. Thus, the value of smart urban construction can be increased by the implementation of spatial software and systematic sustainable analysis.

References

Adade, R., Aibinu, A. M., Ekumah, B., & Asaana, J. (2021). Unmanned Aerial Vehicle (UAV) applications in coastal zone management-a review. *Environmental Monitoring and Assessment*, 193(3), 154. https://doi.org/10.1007/s10661-021-08949-8

Akinwumiju, A. S., Ajisafe, T., & Adelodun, A. A. (2021). Airborne Particulate Matter Pollution in Akure Metro City, Southwestern Nigeria, West Africa: Attribution and Meteorological Influence. *Journal of Geovisualization and Spatial Analysis*, 5(1). http://doi:10.1007/s41651-021-00079-6

Anselin, L. (2019). A local indicator of multivariate spatial association: Extending geary's c. *Geographical Analysis*, 51(2), 133–150. https://doi.org/10.1111/gean.12164

Athanasopoulos, G., & Vahid, F. (2003). Statistical inference and changes in income inequality in Australia. *The Economic Record*, *79*(247), 412–424. https://doi.org/10.1111/j.1475-4932.2003.00141.x

Australian Bureau of Statistics. (2016). Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2016. [Data set]. Retrieved from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July%202016?OpenDocument

Australian Bureau of Statistics. (2017a). Australian Statistical Geography Standard (ASGS): Volume 1- Main Structure and Greater Capital City Statistical Areas, July 2016. [Data set]. Retrieved from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July%202016?OpenDocument

Australian Bureau of Statistics. (2017b). Australian Statistical Geography Standard (ASGS): Volume 4 - Significant Urban Areas, Urban Centres and Localities, Section of State. [Data set]. Retrieved from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.004July%202016?OpenDocument

Australian Bureau of Statistics. (2017c). Australian Industry. [Data set]. Retrieved from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/8155.02015-16?OpenDocument

Australian Bureau of Statistics. (2017d). Census of Population and Housing: Mesh Block Counts, Australia. [Data set]. Retrieved from https://www.abs.gov.au/ausstats/abs@.nsf/mf/2074.0

Australian Bureau of Statistics. (2018). Australian Industry. [Data set]. Retrieved from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/8155.02016-17?OpenDocument

Australian Bureau of Statistics. (2020). Data by region, 2014-19. [Data set]. Retrieved from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1410.02014-19?OpenDocument

Australian Bureau of Statistics. (2021a). *National, state and territory population.* https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release

Australian Bureau of Statistics. (2021b). *Australian Industry*. https://www.abs.gov.au/statistics/industry/industry-overview/australian-industry/latest-release

Australian Bureau of Statistics. (2021c). *Australian statistical geography standard (ASGS) edition 3*. https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/latest-release

Balducci, F., & Ferrara, A. (2018). Using urban environmental policy data to understand the domains of smartness: An analysis of spatial autocorrelation for all the Italian chief towns. Ecological Indicators, 89, 386–396.

Bansal, V. K. (2012). Application areas of GIS in construction projects and future research directions. *International Journal of Construction Management*, 12(4), 17–36. https://doi.org/10.1080/15623599.2012.10773198

Bansal, V. K., & Pal, M. (2011). Construction projects scheduling using GIS tools. *International Journal of Construction Management*, *11*(1), 1–18. https://doi.org/10.1080/15623599.2011.10773158

Batty, M., Morphet, R., Masucci, P., & Stanilov, K. (2014). Entropy, complexity, and spatial information. *Journal of Geographical Systems*, 16(4), 363–385. https://doi.org/10.1007/s10109-014-0202-2

Bertay, A. C., Dordevic, L., & Sever, C. (2020). Gender inequality and economic growth: Evidence from industry-level data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3658594

Bettencourt, L. M. A. (2013). The origins of scaling in cities. *Science (New York, N.Y.)*, *340*(6139), 1438–1441. https://doi.org/10.1126/science.1235823

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(17), 7301–7306. https://doi.org/10.1073/pnas.0610172104

Bettencourt, L. M. A., Lobo, J., Strumsky, D., & West, G. B. (2010). Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities. *PloS One*, *5*(11), e13541. https://doi.org/10.1371/journal.pone.0013541

Bettencourt, L. M. A., Yang, V. C., Lobo, J., Kempes, C. P., Rybski, D., & Hamilton, M. J. (2020). The interpretation of urban scaling analysis in time. *Journal of the Royal Society, Interface*, *17*(163), 20190846. https://doi.org/10.1098/rsif.2019.0846

Biotto, G., Silvestri, S., Gobbo, L., Furlan, E., Valenti, S., & Rosselli, R. (2009). GIS, multi-criteria and multi-factor spatial analysis for the probability assessment of the existence of illegal landfills. *Geographical Information Systems*, *23*(10), 1233–1244. http://doi.org/10.1080/13658810802112128

Borck, R., & Schrauth, P. (2021). Population density and urban air quality. *Regional Science and Urban Economics*, *86*(103596), 103596. https://doi.org/10.1016/j.regsciurbeco.2020.103596

Brody, S. D., Zahran, S., Grover, H., & Vedlitz, A. (2008). A spatial analysis of local climate change policy in the United States: Risk, stress, and opportunity. *Landscape and Urban Planning*, *87*(1), 33–41. http://doi.org/10.1016/j.landurbplan.2008.04.003

Bruneau, P., Brangbour, E., Marchand-Maillet, S., Hostache, R., Chini, M., Pelich, R.-M., Matgen, P., & Tamisier, T. (2021). Measuring the impact of natural hazards with citizen science: The case of flooded area estimation using Twitter. *Remote Sensing*, *13*(6), 1153. http://doi.org/10.3390/rs13061153

Cai, J., Ge, Y., Li, H., Yang, C., Liu, C., Meng, X., Wang, W., Niu, C., Kan, L., Schikowski, T., Yan, B., Chillrud, S. N., Kan, H., & Jin, L. (2020). Application of land use regression to assess exposure and identify potential sources in PM2.5, BC, NO2 concentrations. *Atmospheric Environment (Oxford, England: 1994)*, *223*(117267), 117267. https://doi.org/10.1016/j.atmosenv.2020.117267

Cao, F., Ge, Y., & Wang, J.-F. (2013). Optimal discretization for geographical detectors-based risk assessment. *GIScience & Remote Sensing*, *50*(1), 78–92. https://doi.org/10.1080/15481603.2013.778562

Chang, Y., Hou, K., Wu, Y., Li, X., & Zhang, J. (2019). A conceptual framework for establishing the index system of ecological environment evaluation–A case study of the upper Hanjiang River, China. *Ecological Indicators*, *107*(105568), 105568. http://doi.org/10.1016/j.ecolind.2019.105568

Chen, D., Lu, X., Liu, X., & Wang, X. (2019). Measurement of the eco-environmental effects of urban sprawl: Theoretical mechanism and spatiotemporal differentiation. *Ecological Indicators*, *105*, 6–15. http://doi.org/10.1016/j.ecolind.2019.05.059

Chen, M., Chen, Y., Wang, X., Tan, H., & Luo, F. (2019). Spatial difference of transit-based accessibility to hospitals by regions using spatially adjusted ANOVA. *International Journal of Environmental Research and Public Health*, *16*(11), 1923. https://doi.org/10.3390/ijerph16111923

Chen, Y. (2017). *Lecture 7: Density Estimation*. Washington.Edu. Retrieved November 18, 2021, from http://faculty.washington.edu/yenchic/17Sp_403/Lec7-density.pdf

Chen, Y, Huang, J., Sheng, S., Mansaray, L. R., Liu, Z., Wu, H., & Wang, X. (2018). A new downscaling-integration framework for high-resolution monthly precipitation estimates: Combining rain gauge observations, satellite-derived precipitation data and geographical ancillary data. *Remote Sensing of Environment*, *214*, 154–172. http://doi.org/10.1016/j.rse.2018.05.021

Chen, Y., Yue, W., & La Rosa, D. (2020). Which communities have better accessibility to green space? An investigation into environmental inequality using big data. *Landscape and Urban Planning*, *204*(103919), 103919. http://doi.org/10.1016/j.landurbplan.2020.103919

Cheng, M.-Y., & Chen, J.-C. (2002). Integrating barcode and GIS for monitoring construction progress. *Automation in Construction*, *11*(1), 23–33. http://doi.org/10.1016/s0926-5805(01)00043-7

Cheng, X., Long, R., Chen, H., & Li, Q. (2019). Coupling coordination degree and spatial dynamic evolution of a regional green competitiveness system – A case study from China. *Ecological Indicators*, *104*, 489–500. http://doi.org/10.1016/j.ecolind.2019.04.003

Cheng, Z. (2016). The spatial correlation and interaction between manufacturing agglomeration and environmental pollution. *Ecological Indicators*, *61*, 1024–1032. https://doi.org/10.1016/j.ecolind.2015.10.060

Chi, G., & Zhu, J. (2008). Spatial regression models for demographic analysis. *Population Research and Policy Review*, 27(1), 17–42. https://doi.org/10.1007/s11113-007-9051-8

Chiu, C.-Y., & Russell, A. D. (2011). Design of a construction management data visualization environment: A top–down approach. *Automation in Construction*, 20(4), 399–417. http://doi.org/10.1016/j.autcon.2010.11.010

Cola, L. D. (1994). Simulating and mapping spatial complexity using multi-scale techniques. *International Journal of Geographical Information Systems*, 8(5), 411–427. https://doi.org/10.1080/02693799408902011

Cooper, N., Green, D., & Knibbs, L. D. (2019). Inequalities in exposure to the air pollutants PM2.5 and NO2 in Australia. *Environmental Research Letters*, *14*(11), 115005. https://doi.org/10.1088/1748-9326/ab486a

Cowell, F. A. (1995). Measuring Inequality (2nd ed.). Prentice-Hall.

Crabbe, R. A., Janouš, D., Dařenová, E., & Pavelka, M. (2019). Exploring the potential of LANDSAT-8 for estimation of forest soil CO2 efflux. *International Journal of Applied Earth Observation and Geoinformation: ITC Journal*, *77*, 42–52. http://doi.org/10.1016/j.jag.2018.12.007

Dai, F., Zhou, Q., Lv, Z., Wang, X., & Liu, G. (2014). Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. *Ecological Indicators*, 45, 184–194. http://doi.org/10.1016/j.ecolind.2014.04.003

Dang, T. D., Cochrane, T. A., & Arias, M. E. (2018). Quantifying suspended sediment dynamics in mega deltas using remote sensing data: A case study of the Mekong floodplains. *International Journal of Applied Earth Observation and Geoinformation: ITC Journal*, *68*, 105–115. http://doi.org/10.1016/j.jag.2018.02.008

Dasgupta, S., Wheeler, D., Khaliquzzaman, M., & Huq, M. (2021). Siting priorities for congestion-reducing projects in Dhaka: a spatiotemporal analysis of traffic congestion, travel times, air pollution, and exposure vulnerability. *International Journal of Sustainable Transportation*, 1–19. https://doi.org/10.1080/15568318.2021.1969707

Davis, Z. Y. W., Baray, S., McLinden, C. A., Khanbabakhani, A., Fujs, W., Csukat, C., Debosz, J., & McLaren, R. (2019). Estimation of NO_x and SO_2 emissions from Sarnia, Ontario, using a mobile MAX-DOAS (Multi-AXis Differential Optical Absorption Spectroscopy) and a NO_x analyser. *Atmospheric Chemistry and Physics*, *19*(22), 13871–13889. https://doi.org/10.5194/acp-19-13871-2019

Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy*, *43*(10), 1785–1799. https://doi.org/10.1016/j.respol.2014.05.007

Department of the Environment and Energy, Australian Government. (October 13, 2021). *National Pollutant Inventory*. http://www.npi.gov.au/

Department of the Environment and Energy, Australian Government. (2016). National Pollutant Inventory [Data set]. Retrieved from http://www.npi.gov.au/npidata/action/load/browse-search/criteria/browse-type/Industry/year/2016

Department of the Environment and Energy, Australian Government. (2020). National Pollutant Inventory [Data set]. Retrieved from http://www.npi.gov.au/npidata/action/load/browse-search/criteria/browse-type/Industry/year/2020

Desjeux, Y., Dupraz, P., Kuhlman, T., Paracchini, M. L., Michels, R., Maigné, E., & Reinhard, S. (2015). Evaluating the impact of rural development measures on nature value indicators at different spatial levels: Application to France and The Netherlands. *Ecological Indicators*, *59*, 41–61. http://doi.org/10.1016/j.ecolind.2014.12.014

Dian, R., Li, S., Sun, B., & Guo, A. (2021). Recent advances and new guidelines on hyperspectral and multispectral image fusion. *An International Journal on Information Fusion*, *69*, 40–51. http://doi.org/10.1016/j.inffus.2020.11.001

Dockery, DW (1995). An association between air pollution and mortality in six US cities. *Journal of Occupational and Environmental Medicine*, *37*(2), 136. https://doi.org/10.1056/NEJM199312093292401

Dong, K., Hochman, G., Kong, X., Sun, R., & Wang, Z. (2019). Spatial econometric analysis of China's PM10 pollution and its influential factors: Evidence from the provincial level. *Ecological Indicators*, *96*, 317–328. http://doi.org/10.1016/j.ecolind.2018.09.014

Dong, F., Zhang, X., Liu, Y., Pan, Y., Zhang, X., Long, R., & Sun, Z. (2021). Economic policy choice of governing haze pollution: evidence from global 74 countries. *Environmental Science and Pollution Research International*, *28*(8), 9430–9447. https://doi.org/10.1007/s11356-020-11350-6

Dons, E., Van Poppel, M., Kochan, B., Wets, G., & Int Panis, L. (2013). Modelling temporal and spatial variability of traffic-related air pollution: Hourly land use regression models for black carbon. *Atmospheric Environment (Oxford, England: 1994)*, *74*, 237–246. https://doi.org/10.1016/j.atmosenv.2013.03.050

Epperson, B. K., & Li, T. (1996). Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. *Proceedings of the National Academy of Sciences of the United States of America*, 93(19), 10528–10532. https://doi.org/10.1073/pnas.93.19.10528

Erdogan, S. (2020). Analysing the environmental Kuznets curve hypothesis: The role of disaggregated transport infrastructure investments. *Sustainable Cities and Society*, *61*(102338), 102338. https://doi.org/10.1016/j.scs.2020.102338

Erisman, J. W., Galloway, J. N., Seitzinger, S., Bleeker, A., Dise, N. B., Petrescu, A. M. R., Leach, A. M., & de Vries, W. (2013). Consequences of human modification of the global nitrogen cycle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(1621), 20130116. http://doi.org/10.1098/rstb.2013.0116

Fang, C., Liu, H., Li, G., Sun, D., & Miao, Z. (2015). Estimating the impact of urbanization on air quality in China using spatial regression models. *Sustainability*, *7*(11), 15570–15592. http://dx.doi.org/10.3390/su71115570

Fang, Y., Wang, L., Ren, Z., Yang, Y., Mou, C., & Qu, Q. (2017). Spatial heterogeneity of energy-related CO2 emission growth rates around the world and their determinants during 1990–2014. *Energies*, *10*(3), 367. https://doi.org/10.3390/en10030367

Feng, R., Wang, F., Wang, K., Wang, H., & Li, L. (2021). Urban ecological land and natural-anthropogenic environment interactively drive surface urban heat island: An urban agglomeration-level study in China. *Environment International*, *157*(106857), 106857. https://doi.org/10.1016/j.envint.2021.106857

Filgueiras, R., Mantovani, E. C., Fernandes-Filho, E. I., Cunha, F. F. D., Althoff, D., & Dias, S. H. B. (2020). Fusion of MODIS and Landsat-Like images for daily high spatial resolution NDVI. *Remote Sensing*, *12*(8), 1297. https://doi.org/10.3390/rs12081297

Fleming, D. A., & Measham, T. G. (2015). Income Inequality across Australian Regions during the Mining Boom: 2001–11. *The Australian Geographer*, *46*(2), 203–216. https://doi.org/10.1080/00049182.2015.1020596

Foley, J. A., Defries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N., & Snyder, P. K. (2005). Global consequences of land use. *Science (New York, N.Y.)*, *309*(5734), 570–574. http://doi.org/10.1126/science.1111772

Fotheringham, A. S. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. John Wiley & Sons.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically Weighted Regression*. Wiley.

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment & Planning A*, *30*(11), 1905–1927. https://doi.org/10.1068/a301905

Gellrich, M., & Zimmermann, N. E. (2007). Investigating the regional-scale pattern of agricultural land abandonment in the Swiss mountains: A spatial statistical modelling approach. *Landscape and Urban Planning*, 79(1), 65–76. http://doi.org/10.1016/j.landurbplan.2006.03.004

Geofabrik and OpenStreetMap contributors. (2020). Download OpenStreetMap for this region: Australia and Oceania [Data set]. Retrieved from http://download.geofabrik.de/australia-oceania.html

Geofabrik and OpenStreetMap contributors. (2021). Download OpenStreetMap for this region: Australia and Oceania [Data set]. Retrieved from http://download.geofabrik.de/australia-oceania.html

Geoscience Australia. (2014). *Area of Australia - states and territories*. https://www.ga.gov.au/scientific-topics/national-location-information/dimensions/area-of-australia-states-and-territories

Gómez-Losada, Á., Santos, F. M., Gibert, K., & Pires, J. C. M. (2019). A data science approach for spatiotemporal modelling of low and resident air pollution in Madrid (Spain): Implications for epidemiological studies. *Computers, Environment and Urban Systems*, *75*, 1–11. https://doi.org/10.1016/j.compenvurbsys.2018.12.005

Google. (2020). Landsat 8 Collection 1 Tier 1 8-Day NDVI Composite [Data set]. Retrieved from https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C01_T1_8DAY_NDVI

Google. (2020). MODIS Combined 16-Day EVI [Data set]. Retrieved from https://developers.google.com/earth-engine/datasets/catalog/MODIS_MCD43A4_006_EVI

Google Developers and Earth Observation Group. (2020). VIIRS Stray Light Corrected Nighttime Day/Night Band Composites Version 1 [Data set]. Retrieved from https://developers.google.com/earth-engine/datasets/catalog/NOAA_VIIRS_DNB_MONTHLY_V1_VCMSLCFG

Google Developers and Geoscience Australia. (2010). DEM-S: Australian Smoothed Digital Elevation Model [Data set]. Retrieved from https://developers.google.com/earth-engine/datasets/catalog/AU_GA_DEM_1SEC_v10_DEM-S

Google Developers and Global Change Observation Mission. (2020). GCOM-C/SGLI L3 Leaf Area Index (V2) [Data set]. Retrieved from https://developers.google.com/earth-engine/datasets/catalog/JAXA_GCOM-C_L3_LAND_LAI_V2

Google Developers and the European Space Agency. (2020). Sentinel-5P [Data set]. Retrieved from https://developers.google.com/earth-engine/datasets/catalog/sentinel-5p

Google Developers and United States Geological Survey. (2020). MOD13A1.006 Terra Vegetation Indices 16-Day Global 500m [Data set]. Retrieved from https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD13A1

Google Developers and University of California Merced. (2020). TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces [Data set]. Retrieved from https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_TERRACLIMATE

Google Developers and Worldpop. (2020). WorldPop Global Project Population Data [Data set]. Retrieved from https://developers.google.com/earth-engine/datasets/catalog/WorldPop_GP_100m_pop_age_sex_cons_unadj?hl=en#bands

Guo, B., Wang, X., Pei, L., Su, Y., Zhang, D., & Wang, Y. (2021). Identifying the spatiotemporal dynamic of PM2.5 concentrations at multiple scales using geographically and temporally weighted regression model across China during 2015-2018. *The Science of the Total Environment*, *751*(141765), 141765. https://doi.org/10.1016/j.scitotenv.2020.141765

Haddad, K., & Vizakos, N. (2020). Air quality pollutants and their relationship with meteorological variables in four suburbs of Greater Sydney, Australia. *Air Quality, Atmosphere, & Health*. https://doi.org/10.1007/s11869-020-00913-8

Hadjisophocleous, G., &Chen, Z. (2010). A survey of fire loads in elementary schools and high schools. *Journal of Fire Protection Engineering*, *20*, 55-71. https://doi.org/10.1177/1042391509360266

He, J., Pan, Z., Liu, D., & Guo, X. (2019). Exploring the regional differences of ecosystem health and its driving factors in China. *The Science of the Total Environment*, *673*, 553–564. https://doi.org/10.1016/j.scitotenv.2019.03.465

Henebry, G. M. (1995). Spatial model error analysis using autocorrelation indices. *Ecological Modelling*, *82*(1), 75–91. https://doi.org/10.1016/0304-3800(94)00074-r

Herold, M., Goldstein, N. C., & Clarke, K. C. (2003). The spatiotemporal form of urban growth: measurement, analysis and modelling. *Remote Sensing of Environment*, *86*(3), 286–302. http://doi.org/10.1016/s0034-4257(03)00075-0

Ho T.M., D., Hanssen, R., & Rocca, F. (2020). Radar interferometry: 20 years of development in time series techniques and future perspectives. *Remote Sensing*, *12*(9), 1364. http://doi.org/10.3390/rs12091364

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment (Oxford, England: 1994)*, *42*(33), 7561–7578. http://doi.org/10.1016/j.atmosenv.2008.05.057

Hu, M., Wang, Y., Wang, S., Jiao, M., Huang, G., & Xia, B. (2021). Spatial-temporal heterogeneity of air pollution and its relationship with meteorological factors in the

Pearl River Delta, China. *Atmospheric Environment (Oxford, England: 1994)*, *254*(118415), 118415. https://doi.org/10.1016/j.atmosenv.2021.118415

Ilunga, M. (2019). Shannon entropy for measuring spatial complexity associated with mean annual runoff of tertiary catchments of the Middle Vaal basin in South Africa. *Entropy (Basel, Switzerland)*, 21(4), 366. https://doi.org/10.3390/e21040366

Jackson, L. E. (2003). The relationship of urban design to human health and condition. Landscape and Urban Planning, 64(4), 191–200. http://doi.org/10.1016/s0169-2046(02)00230-x

Jia, S., & Qian, Y. (2007). Spectral and Spatial Complexity-Based Hyperspectral Unmixing. *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society*, 45(12), 3867–3879. https://doi.org/10.1109/tgrs.2007.898443

Jiang, X.-T., Wang, Q., & Li, R. (2018). Investigating factors affecting carbon emission in China and the USA: A perspective of stratified heterogeneity. *Journal of Cleaner Production*, *199*, 85–92. https://doi.org/10.1016/j.jclepro.2018.07.160

Jin, X., Jin, Y., & Mao, X. (2019). Ecological risk assessment of cities on the Tibetan Plateau based on land use/land cover changes – Case study of Delingha City. *Ecological Indicators*, 101, 185–191. http://doi.org/10.1016/j.ecolind.2018.12.050

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity*, *52*(4), 1957–1976. https://doi.org/10.1007/s11135-017-0584-6

Ju, T., Lei, M., Guo, G., Xi, J., Zhang, Y., Xu, Y., & Lou, Q. (2023). A new prediction method of industrial atmospheric pollutant emission intensity based on pollutant emission standard quantification. *Frontiers of Environmental Science & Engineering*, *17*(1), 8. https://doi.org/10.1007/s11783-023-1608-1

Kefalas, G., Kalogirou, S., Poirazidis, K., & Lorilla, R. S. (2019). Landscape transition in Mediterranean islands: The case of Ionian islands, Greece 1985–2015. *Landscape and Urban Planning*, *191*(103641), 103641. http://doi.org/10.1016/j.landurbplan.2019.103641

Keuschnigg, M., Mutgan, S., & Hedström, P. (2019). Urban scaling and the regional divide. *Science Advances*, *5*(1), eaav0042. https://doi.org/10.1126/sciadv.aav0042

Khan, F., & Pinter, L. (2016). Scaling indicator and planning plane: An indicator and a visual tool for exploring the relationship between urban form, energy efficiency and carbon emissions. *Ecological Indicators*, *67*, 183–192. https://doi.org/10.1016/j.ecolind.2016.02.046

Klenert, D., Mattauch, L., Edenhofer, O., & Lessmann, K. (2018). Infrastructure and inequality: Insights from incorporating key economic facts about household heterogeneity. *Macroeconomic Dynamics*, 22(4), 864–895. https://doi.org/10.1017/s1365100516000432

Kong, H., Lin, J., Zhang, R., Liu, M., Weng, H., Ni, R., Chen, L., Wang, J., Yan, Y., & Zhang, Q. (2019). High-resolution (0.05 ° × 0.05 °) NOx emissions in the Yangtze River Delta inferred from OMI. *Atmospheric Chemistry and Physics*, 19(20), 12835–12856. https://doi.org/10.5194/acp-19-12835-2019

Kozlov, A., Gutman, S., Zaychenko, I., & Rytova, E. (2017). The formation of regional strategy of innovation-industrial development. In *Information Systems Architecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology – ISAT 2016 – Part IV* (pp. 115–126). Springer International Publishing.

Kuechly, H. U., Kyba, C. C. M., Ruhtz, T., Lindemann, C., Wolter, C., Fischer, J., & Hölker, F. (2012). Aerial survey and spatial analysis of sources of light pollution in Berlin, Germany. *Remote Sensing of Environment*, *126*, 39–50. http://doi.org/10.1016/j.rse.2012.08.008

Kwon, O. (2018). Scaling laws between population and a public transportation system of urban buses. *Physica A*, *503*, 209–214. https://doi.org/10.1016/j.physa.2018.02.193

Lambey, V., & Prasad, A. D. (2021). A review on air quality measurement using an unmanned aerial vehicle. *Water, Air, and Soil Pollution*, *232*(3). http://doi.org/10.1007/s11270-020-04973-5

Lämmer, S., Gehlsen, B., & Helbing, D. (2006). Scaling laws in the spatial structure of urban road networks. *Physica A*, *363*(1), 89–95. https://doi.org/10.1016/j.physa.2006.01.051

Lei, W., Jiao, L., & Xu, G. (2022a). Understanding the urban scaling of urban land with an internal structure view to characterize China's urbanization. *Land Use Policy*, *112*(105781), 105781. https://doi.org/10.1016/j.landusepol.2021.105781

Lei, W., Jiao, L., Xu, G., & Zhou, Z. (2022b). Urban scaling in rapidly urbanising China. *Urban Studies (Edinburgh, Scotland)*, *59*(9), 1889–1908. https://doi.org/10.1177/00420980211017817

Li, H., Jia, P., & Fei, T. (2021). Associations between taste preferences and chronic diseases: a population-based exploratory study in China. *Public Health Nutrition*, *24*(8), 2021–2032. https://doi.org/10.1017/S136898002000035X

Li, J., Long, Y., & Dang, A. (2018). Live-Work-Play Centers of Chinese cities: Identification and temporal evolution with emerging data. *Computers, Environment and Urban Systems*, *71*, 58–66. https://doi.org/10.1016/j.compenvurbsys.2018.04.002

Li, Q., Zhou, S., & Wen, P. (2019). The relationship between centrality and land use patterns: Empirical evidence from five Chinese metropolises. *Computers, Environment and Urban Systems*, *78*(101356), 101356. https://doi.org/10.1016/j.compenvurbsys.2019.101356

Liczbińska, G., & Sobkowiak, A. (2020). Did the sex ratio at birth reflect social and economic inequalities? The pilot study from the Poznań province, 1875–1913. *Przeszlosc Demograficzna Polski*, 42, 95–121. https://doi.org/10.18276/pdp.2020.42-04

Lin, Y., Hu, X., Zheng, X., Hou, X., Zhang, Z., Zhou, X., Qiu, R., & Lin, J. (2019). Spatial variations in the relationships between road network and landscape ecological risks in the highest forest coverage region of China. *Ecological Indicators*, *96*, 392–403. http://doi.org/10.1016/j.ecolind.2018.09.016

Liu, C., Henderson, B. H., Wang, D., Yang, X., & Peng, Z.-R. (2016). A land use regression application into assessing spatial variation of intra-urban fine particulate matter (PM2.5) and nitrogen dioxide (NO2) concentrations in City of Shanghai, China. *The Science of the Total Environment*, *565*, 607–615. http://dx.doi.org/10.1016/j.scitotenv.2016.03.189

Lobo, J., Strumsky, D., & Rothwell, J. (2013a). Scaling of patenting with urban population size: evidence from global metropolitan areas. *Scientometrics*, *96*(3), 819–828. https://doi.org/10.1007/s11192-013-0970-3

Lobo, José, Bettencourt, L. M. A., Strumsky, D., & West, G. B. (2013b). Urban scaling and the production function for cities. *PloS One*, *8*(3), e58407. https://doi.org/10.1371/journal.pone.0058407

Long, J., Liu, Y., Xing, S., Zhang, L., Qu, M., Qiu, L., Huang, Q., Zhou, B., & Shen, J. (2020). Optimal interpolation methods for farmland soil organic matter in various landforms of a complex topography. *Ecological Indicators*, *110*(105926), 105926. http://doi.org/10.1016/j.ecolind.2019.105926

Luo, P., Song, Y., Huang, X., Ma, H., Liu, J., Yao, Y., & Meng, L. (2022). Identifying determinants of spatio-temporal disparities in soil moisture of the Northern Hemisphere using a geographically optimal zones-based heterogeneity model. *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, *185*, 111–128. https://doi.org/10.1016/j.isprsjprs.2022.01.009

Luo, P., Song, Y., & Wu, P. (2021). Spatial disparities in trade-offs: economic and environmental impacts of road infrastructure on continental level. *GIScience & Remote Sensing*, *58*(5), 756-775. https://doi.org/10.1080/15481603.2021.1947624

Luo, P., Song, Y., Zhu, D., Cheng, J., & Meng, L. (2022). A generalized heterogeneity model for spatial interpolation. *International Journal of Geographical Information Science*, 1-26. https://doi.org/10.1080/13658816.2022.2147530

Ma, Q., Wu, J., He, C., & Hu, G. (2018). Spatial scaling of urban impervious surfaces across evolving landscapes: From cities to urban regions. *Landscape and Urban Planning*, *175*, 50–61. https://doi.org/10.1016/j.landurbplan.2018.03.010

Matti,K., Maija, T., Joseph, G. (2020) Data from: Gridded global datasets for Gross Domestic Product and Human Development Index over 1990-2015. [Data set]. Retrieved from https://datadryad.org/stash/dataset/doi:10.5061/dryad.dk1j0

Maus, V., Giljum, S., Gutschlhofer, J., da Silva, D. M., Probst, M., Gass, S. L. B., Luckeneder, S., Lieber, M., & McCallum, I. (2020). A global-scale data set of mining areas. *Scientific Data*, *7*(1), 289. https://doi.org/10.1038/s41597-020-00624-w

McCarty, J., & Kaza, N. (2015). Urban form and air quality in the United States. *Landscape and Urban Planning*, *139*, 168–179. http://doi.org/10.1016/j.landurbplan.2015.03.008

Meng, X., Chen, L., Cai, J., Zou, B., Wu, C.-F., Fu, Q., Zhang, Y., Liu, Y., & Kan, H. (2015). A land use regression model for estimating the NO2 concentration in Shanghai, China. *Environmental Research*, *137*, 308–315. http://dx.doi.org/10.1016/j.envres.2015.01.003

Milenov, P., Vassilev, V., Vassileva, A., Radkov, R., Samoungi, V., Dimitrov, Z., & Vichev, N. (2014). Monitoring of the risk of farmland abandonment as an efficient tool to assess the environmental and socio-economic impact of the Common Agriculture Policy. *International Journal of Applied Earth Observation and Geoinformation: ITC Journal*, *32*, 218–227. http://doi.org/10.1016/j.jag.2014.03.013

Mirchooli, F., Kiani-Harchegani, M., Khaledi Darvishan, A., Falahatkar, S., & Sadeghi, S. H. (2020). Spatial distribution dependency of soil organic carbon content to important environmental variables. *Ecological Indicators*, *116*(106473), 106473. http://doi.org/10.1016/j.ecolind.2020.106473

Mukhopadhyay, N., & Sengupta, P. P. (2021). Gini Inequality Index: Methods and Applications (N. Mukhopadhyay & P. P. Sengupta, Eds.). CRC Press.

Murray, D. (1978). Sources of income inequality in Australia, 1968-69. *The Economic Record*, 54(2), 159–169. https://doi.org/10.1111/j.1475-4932.1978.tb00327.x

Nieto, D. M. C., Quiroz, E. A. P., & Lengua, M. A. C. (2021). A systematic literature review on support vector machines applied to regression. 2021 *IEEE Sciences and Humanities International Research Conference (SHIRCON)*.

Ottaviano, G. I. P., & Puga, D. (1998). Agglomeration in the global economy: A survey of the 'new economic geography.' *World Economy*, *21(6)*, 707–731. https://doi.org/10.1111/1467-9701.00160

Ovando-Montejo, G. A., Kedron, P., & Frazier, A. E. (2021). Relationship between urban size and configuration: Scaling evidence from a hierarchical system in Mexico. *Applied Geography (Sevenoaks, England)*, *132*(102462), 102462. https://doi.org/10.1016/j.apgeog.2021.102462

Owers, C. J., Rogers, K., & Woodroffe, C. D. (2016). Identifying spatial variability and complexity in wetland vegetation using an object-based approach. *International Journal of Remote Sensing*, 37(18), 4296–4316. https://doi.org/10.1080/01431161.2016.1211349

Paez, A. (2019). Using spatial filters and exploratory data analysis to enhance regression models of spatial data: Using spatial filters and exploratory data analysis. *Geographical Analysis*, 51(3), 314–338. https://doi.org/10.1111/gean.12180

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, *42*(3–4), 523–527. https://doi.org/10.1093/biomet/42.3-4.523

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing*, *12*(14), 2291. http://doi.org/10.3390/rs12142291

Poudyal, N. C., Butler, B. J., & Hodges, D. G. (2019). Spatial analysis of family forest landownership in the southern United States. *Landscape and Urban Planning*, 188, 163–170. https://doi.org/10.1016/j.landurbplan.2018.10.018

Pringle, M. J., & Lark, R. M. (2006). Spatial analysis of model error, illustrated by soil carbon dioxide emissions. *Vadose Zone Journal: VZJ*, *5*(1), 168–183. https://doi.org/10.2136/vzj2005.0015

Qu Y, Jiang G, Yang Y, Zheng Q, Li Y, Ma W. (2018). Multi-scale analysis on spatial morphology differentiation and formation mechanism of rural residential land: A case study in Shandong Province, China. *Habitat International*, *71*, 135–146. https://doi.org/10.1016/j.habitatint.2017.11.011

Raghavan, R. K., Brenner, K. M., Harrington, J. A., Jr, Higgins, J. J., & Harkin, K. R. (2013). Spatial scale effects in environmental risk-factor modelling for diseases. *Geospatial Health*, *7*(2), 169–182. https://doi.org/10.4081/gh.2013.78

Reeson, A. F., Measham, T. G., & Hosking, K. (2012). Mining activity, income inequality and gender in regional Australia: Mining activity, income inequality and gender. *The Australian Journal of Agricultural and Resource Economics*, 56(2), 302–313. https://doi.org/10.1111/j.1467-8489.2012.00578.x

Reinosch, E., Buckel, J., Dong, J., Gerke, M., Baade, J., & Riedel, B. (2020). InSAR time series analysis of seasonal surface displacement dynamics on the Tibetan Plateau. *The Cryosphere*, *14*(5), 1633–1650. http://doi.org/10.5194/tc-14-1633-2020

Riascos, A. P. (2017). Universal scaling of the distribution of land in urban areas. *Physical Review. E*, *96*(3–1), 032302. https://doi.org/10.1103/PhysRevE.96.032302

Rietveld, R., Vlaanderen, N., Kame, D., & Schipper, Y. (1994). Infrastructure and industrial development: The case of central java. *Bulletin of Indonesian Economic Studies*, *30*(2), 119–132. https://doi.org/10.1080/00074919412331336617

Ristea, A., Al Boni, M., Resch, B., Gerber, M. S., & Leitner, M. (2020). Spatial crime distribution and prediction for sporting events using social media. *Geographical Information Systems*, *34*(9), 1708–1739. http://doi.org/10.1080/13658816.2020.1719495

Rodríguez-Pose, A., & Tselios, V. (2010). Inequalities in income and education and regional economic growth in western Europe. *The Annals of Regional Science*, 44(2), 349–375. https://doi.org/10.1007/s00168-008-0267-2

Roy, A. (2021). Atmospheric Pollution Retrieval Using Path Radiance Derived from Remote Sensing Data. *Journal of Geovisualization and Spatial Analysis*, 5(2). http://doi:10.1007/s41651-021-00093-8

Rufino, M. M., Bez, N., & Brind'Amour, A. (2020). Ability of spatial indicators to detect geographic changes (shift, shrink and split) across biomass levels and sample sizes. *Ecological Indicators*, 115(106393), 106393. https://doi.org/10.1016/j.ecolind.2020.106393

Sabrin, S., Karimi, M., Fahad, M. G. R., & Nazari, R. (2020). Quantifying environmental and social vulnerability: Role of urban Heat Island and air quality, a case study of Camden, NJ. *Urban Climate*, *34*(100699), 100699. https://doi.org/10.1016/j.uclim.2020.100699

Sahraoui, Y., Clauzel, C., & Foltête, J.-C. (2021). A metrics-based approach for modelling covariation of visual and ecological landscape qualities. *Ecological Indicators*, *123*(107331), 107331. http://doi.org/10.1016/j.ecolind.2020.107331

Sarkar, S. (2019). Urban scaling and the geographic concentration of inequalities by city size. *Environment and Planning. B, Urban Analytics and City Science*, *46*(9), 1627–1644. https://doi.org/10.1177/2399808318766070

Sarkar, S., Phibbs, P., Simpson, R., & Wasnik, S. (2018). The scaling of income distribution in Australia: Possible relationships between urban allometry, city size, and economic inequality. *Environment and Planning. B, Urban Analytics and City Science*, *45*(4), 603–622. https://doi.org/10.1177/0265813516676488

Satterthwaite, D. (2008). Cities' contribution to global warming: notes on the allocation of greenhouse gas emissions. *Environment and Urbanization*, *20*(2), 539–549. https://doi.org/10.1177/0956247808096127

Sbardella, A., Pugliese, E., & Pietronero, L. (2017). Economic development and wage inequality: A complex system analysis. *PloS One*, *12*(9), e0182774. https://doi.org/10.1371/journal.pone.0182774

She, Q., Peng, X., Xu, Q., Long, L., Wei, N., Liu, M., Jia, W., Zhou, T., Han, J., & Xiang, W. (2017). Air quality and its response to satellite-derived urban form in the Yangtze River Delta, China. *Ecological Indicators*, *75*, 297–306. https://doi.org/10.1016/j.ecolind.2016.12.045

Shmool, J. L. C., Kubzansky, L. D., Newman, O. D., Spengler, J., Shepard, P., & Clougherty, J. E. (2014). Social stressors and air pollution across New York City communities: a spatial approach for assessing correlations among multiple exposures. *Environmental Health: A Global Access Science Source*, *13*(1), 91. http://dx.doi.org/10.1186/1476-069X-13-91

Shorrocks, A. (1978). Income inequality and income mobility. Journal of Economic Theory, 19(2), 376–393. https://doi.org/10.1016/0022-0531(78)90101-1

Solga, H. (2014). Education, economic inequality and the promises of the social investment state. Socio-Economic Review, 12(2), 269–297. https://doi.org/10.1093/ser/mwu014

Song, J., Du, S., Feng, X., & Guo, L. (2014). The relationships between landscape compositions and land surface temperature: Quantifying their resolution sensitivity with spatial regression models. *Landscape and Urban Planning*, *123*, 145–157. http://doi.org/10.1016/j.landurbplan.2013.11.014

Song, Y., Wright, G., Wu, P., Thatcher, D., McHugh, T., Li, Q., ... & Wang, X. (2018a). Segment-based spatial analysis for assessing road infrastructure performance using monitoring observations and remote sensing data. *Remote Sensing*, *10*(11), 1696. https://doi.org/10.3390/rs10111696

Song, Y., Long, Y., Wu, P., & Wang, X. (2018b). Are all cities with similar urban form or not? Redefining cities with ubiquitous points of interest and evaluating them with indicators at city and block levels in China. *Geographical Information Systems*, *32*(12), 2447–2476. https://doi.org/10.1080/13658816.2018.1511793

Song, Y., Wang, J., Ge, Y., & Xu, C. (2020). An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: cases with different types of spatial data. *GIScience & Remote Sensing*, *57*(5), 593–610. https://doi.org/10.1080/15481603.2020.1760434

Song, Y., & Wu, P. (2021a). An interactive detector for spatial associations. *Geographical Information Systems*, *35*(8), 1676–1701. https://doi.org/10.1080/13658816.2021.1882680

Song, Y., Wu, P., Gilmore, D., & Li, Q. (2021b). A spatial heterogeneity-based segmentation model for analysing road deterioration network data in multi-scale infrastructure systems. *IEEE Transactions on Intelligent Transportation Systems: A Publication of the IEEE Intelligent Transportation Systems Council*, *22*(11), 7073–7083. https://doi.org/10.1109/tits.2020.3001193

Song, Y. (2022a). The second dimension of spatial association. *International Journal of Applied Earth Observation and Geoinformation*, 111, 102834. https://doi.org/10.1016/j.jag.2022.102834

Song, Y. (2022b). Geographically Optimal Similarity. *Mathematical Geosciences*, 1-26. https://doi.org/10.1007/s11004-022-10036-8

Suh, N. P. (1999). A theory of complexity, periodicity and the design axioms. *Research in Engineering Design*, 11(2), 116–132. https://doi.org/10.1007/pl00003883

Sun, W., Xia, C., Xu, M., Guo, J., & Sun, G. (2016). Application of modified water quality indices as indicators to assess the spatial and temporal trends of water quality in the Dongjiang River. *Ecological Indicators*, *66*, 306–312. http://doi.org/10.1016/j.ecolind.2016.01.054

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of* *the International Society for Photogrammetry and Remote Sensing (ISPRS)*, *164*, 152–170. http://doi.org/10.1016/j.isprsjprs.2020.04.001

Tian, Y., Yao, X., & Chen, L. (2019). Analysis of spatial and seasonal distributions of air pollutants by incorporating urban morphological characteristics. *Computers, Environment and Urban Systems*, *75*, 35–48. https://doi.org/10.1016/j.compenvurbsys.2019.01.003

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, *167*(107299), 107299. https://doi.org/10.1016/j.sigpro.2019.107299

Tu, W., Zhu, T., Xia, J., Zhou, Y., Lai, Y., Jiang, J., & Li, Q. (2020). Portraying the spatial dynamics of urban vibrancy using multisource urban big data. *Computers, Environment and Urban Systems*, *80*(101428), 101428. https://doi.org/10.1016/j.compenvurbsys.2019.101428

Tu, Y., Xu, C., Wang, W., Wang, Y., & Jin, K. (2021). Investigating the impacts of driving restriction on NO2 concentration by integrating citywide scale cellular data and traffic simulation. *Atmospheric Environment (Oxford, England: 1994)*, *265*(118721), 118721. https://doi.org/10.1016/j.atmosenv.2021.118721

Ustin, S. L., & Middleton, E. M. (2021). Current and near-term advances in Earth observation for ecological applications. *Ecological Processes*, *10*(1), 1. http://doi.org/10.1186/s13717-020-00255-4

Van der Zanden, E. H., Verburg, P. H., & Mücher, C. A. (2013). Modelling the spatial distribution of linear landscape elements in Europe. *Ecological Indicators*, *27*, 125–136. http://doi.org/10.1016/j.ecolind.2012.12.002

VicRoads. (2021, August 18). *Victoria's road network*. Gov.Au. https://www.vicroads.vic.gov.au/traffic-and-road-use/road-network-and-performance/victorias-road-network

Wang, H., Li, J., Gao, M., Chan, T.-C., Gao, Z., Zhang, M., Li, Y., Gu, Y., Chen, A., Yang, Y., & Ho, H. C. (2020). Spatiotemporal variability in long-term population exposure to PM2.5 and lung cancer mortality attributable to PM2.5 across the Yangtze River Delta (YRD) region over 2010-2016: A multistage approach. *Chemosphere*, *257*(127153), 127153. https://doi.org/10.1016/j.chemosphere.2020.127153

Wang, H., Liu, X., Zhao, C., Chang, Y., Liu, Y., & Zang, F. (2021). Spatial-temporal pattern analysis of landscape ecological risk assessment based on land use/land cover change in Baishuijiang National nature reserve in Gansu Province, China. *Ecological Indicators*, *124*(107454), 107454. http://doi.org/10.1016/j.ecolind.2021.107454

Wang, J.-F., Li, X.-H., Christakos, G., Liao, Y.-L., Zhang, T., Gu, X., & Zheng, X.-Y. (2010). Geographical detectors-based health risk assessment and its application in the neural tube defects study of the heshun region, China. *Geographical Information Systems*, *24*(1), 107–127. https://doi.org/10.1080/13658810802443457

Wang, J.-F., Zhang, T.-L., & Fu, B.-J. (2016). A measure of spatial stratified heterogeneity. *Ecological Indicators*, *67*, 250–256. https://doi.org/10.1016/j.ecolind.2016.02.052

Wang, Y., Ding, Y., & Yan, L. (2020). Analysis on the association of financial development, industrial structure optimisation and economic growth via panel VAR. *Proceedings of the 2020 3rd International Conference on E-Business, Information Management and Computer Science*. https://doi.org/10.1145/3453187.3453352

Wang, Y., Guo, Z., & Han, J. (2021). The relationship between urban heat island and air pollutants and them with influencing factors in the Yangtze River Delta, China. *Ecological Indicators*, *129*(107976), 107976. https://doi.org/10.1016/j.ecolind.2021.107976

Weisent, J., Rohrbach, B., Dunn, J. R., & Odoi, A. (2012). Socioeconomic determinants of geographic disparities in campylobacteriosis risk: a comparison of global and local modelling approaches. *International Journal of Health Geographics*, *11*(1), 45. https://doi.org/10.1186/1476-072X-11-45

Worboys, M. F., & Duckham, M. (2021). *GIS: A Computing Perspective, Third Edition* (3rd ed.). CRC Press.

Wu, T., Zhou, L., Jiang, G., Meadows, M. E., Zhang, J., Pu, L., Wu, C., & Xie, X. (2021). Modelling spatial heterogeneity in the effects of natural and socioeconomic factors, and their interactions, on atmospheric PM2.5 concentrations in China from 2000–2015. *Remote Sensing*, *13*(11), 2152. https://doi.org/10.3390/rs13112152

Xiao, Y., & Gong, P. (2022). Removing spatial autocorrelation in urban scaling analysis. *Cities (London, England)*, *124*(103600), 103600. https://doi.org/10.1016/j.cities.2022.103600

Xie, X., Semanjski, I., Gautama, S., Tsiligianni, E., Deligiannis, N., Rajan, R., Pasveer, F., & Philips, W. (2017). A review of urban air pollution monitoring and exposure assessment methods. *ISPRS International Journal of Geo-Information*, *6*(12), 389. http://doi.org/10.3390/ijgi6120389

Xu, G., Xu, Z., Gu, Y., Lei, W., Pan, Y., Liu, J., & Jiao, L. (2020). Scaling laws in intra-urban systems and over time at the district level in Shanghai, China. *Physica A*, *560*(125162), 125162. https://doi.org/10.1016/j.physa.2020.125162

Yamaguchi, Y., Suzuki, Y., Yamazaki, M., Shimoda, Y., Murakami, S., Bogaki, K., Matsunawa, K., Kametani, S., Takaguchi, H., Hanzawa, H., Yoshino, H., Asano, Y., Okumiya, M., Murakawa, S., & Yoda, H. (2012). Comparison of energy consumption per unit floor area among retail categories based on the database of energy consumption for commercial buildings (decc). *Journal of Environmental Engineering (Transactions of AIJ)*, *77*(681), 889–897. http://doi.org/10.3130/aije.77.889

Yang, C., & Zhao, S. (2023). Scaling of Chinese urban CO2 emissions and multiple dimensions of city size. *The Science of the Total Environment*, *857*(Pt 2), 159502. https://doi.org/10.1016/j.scitotenv.2022.159502

Yang, J., Ji, Z., Kang, S., Zhang, Q., Chen, X., & Lee, S.-Y. (2019). Spatiotemporal variations of air pollutants in western China and their relationship to meteorological factors and emission sources. *Environmental Pollution (Barking, Essex: 1987)*, *254*(Pt A), 112952. https://doi.org/10.1016/j.envpol.2019.07.120

Yanovski, R., Nelson, P. A., & Abelson, A. (2017). Structural complexity in coral reefs: Examination of a novel evaluation tool on different spatial scales. *Frontiers in Ecology and Evolution*, 5. https://doi.org/10.3389/fevo.2017.00027

York, R., Rosa, E. A., & Dietz, T. (2003). STIRPAT, IPAT and ImPACT: analytic tools for unpacking the driving forces of environmental impacts. *Ecological Economics: The Journal of the International Society for Ecological Economics*, *46*(3), 351–365. https://doi.org/10.1016/S0921-8009(03)00188-5

Yue, Y., Wang, Z., Tian, L., Zhao, J., Lai, Z., Ji, G., & Xia, H. (2020). Modelling the spatiotemporal dynamics of industrial sulphur dioxide emissions in China based on DMSP-OLS nighttime stable light data. *PloS One*, *15*(9), e0238696. https://doi.org/10.1371/journal.pone.0238696

Zhai, L., Li, S., Zou, B., Sang, H., Fang, X., & Xu, S. (2018). An improved geographically weighted regression model for PM2.5 concentration estimation in large areas. *Atmospheric Environment (Oxford, England: 1994)*, *181*, 145–154. https://doi.org/10.1016/j.atmosenv.2018.03.017

Zhai, T., Wang, J., Jin, Z., Qi, Y., Fang, Y., & Liu, J. (2020). Did improvements of ecosystem services supply-demand imbalance change environmental spatial injustices? *Ecological Indicators*, *111*(106068), 106068. http://doi.org/10.1016/j.ecolind.2020.106068

Zhang, Z., Song, Y., & Wu, P. (2022). Robust geographical detector. *International Journal of Applied Earth Observation and Geoinformation: ITC Journal*, 109(102782), 102782. https://doi.org/10.1016/j.jag.2022.102782

Zhu, A.-X., Lu, G., Liu, J., Qin, C.-Z., & Zhou, C. (2018). Spatial prediction based on Third Law of Geography. Annals of GIS, 24(4), 225–240. https://doi.org/10.1080/19475683.2018.1534890

Zuo, L., Gao, J., & Du, F. (2021). The pairwise interaction of environmental factors for ecosystem services relationships in karst ecological priority protection and key restoration areas. *Ecological Indicators*, *131(108125)*, 108125. https://doi.org/10.1016/j.ecolind.2021.108125