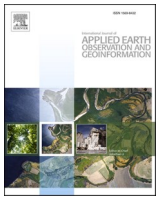
International Journal of Applied Earth Observations and Geoinformation 117 (2023) 103221

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



# Elucidation of spatial disparities of factors that affect air pollutant concentrations in industrial regions at a continental level

Zehua Zhang [a], Yongze Song [a,*], Peng Luo [b], Peng Wu [a], Xiaochi Liu [c,d], Mingshu Wang [e]

[a] School of Design and the Built Environment, Curtin University, Perth 6102, Australia
[b] Chair of Cartography, Department of Aerospace and Geodesy, Technical University of Munich, Munich, Germany
[c] School of Information Engineering, China University of Geosciences, Beijing 100083, China
[d] Earth and Environmental Sciences, Faculty of Science and Engineering, Macquarie University, Sydney 2109, New South Wales, Australia
[e] School of Geographical and Earth Sciences, University of Glasgow, Glasgow, United Kingdom

ARTICLE INFO

*Keywords:*
Spatial disparity
Air pollutant concentration
Industrial regions
Geographically weighted regression
Spatial statistics
Australia

ABSTRACT

Industrial regions and relevant infrastructures are known to contribute to air pollutant emissions; thus, a detailed investigation of the air pollutant concentrations of a region based on specific land uses, with spatial reasoning, can support smart regional planning. 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. Thus, in this study, we aimed to identify the factors that affect air-pollutant concentrations due to local spatial impacts in industrial regions across Australia. Considering the large spatial scale, the impact of a global factor can be overwhelmed by another factor due to local spatial impacts, and the phenomenon is a kind of spatial disparity. We developed a novel set of methods, including a point-of-interests-based spatial identification method and geographically weighted regression (with standardised coefficients), to: (i) identify the industrial regions in the study area, (ii) collect the remote sensing factors, and (iii) identify the factors that affect the spatial disparity of air-pollutant concentrations in industrial regions. The results indicated a significant spatial disparity in the air pollutant concentrations in the industrial region, at a continental scale. Anthropogenic factors significantly affected the spatial patterns of air pollutant concentrations in the industrial regions that were remote to cities, whereas meteorological and topographical factors had significant impacts on the air pollutant distributions in urban industrial regions. Furthermore, within the nationwide industrial lands, drives of the relatively high concentrations of ozone and sulphur dioxide, the drivers of the air pollutant concentrations were environmental factors; high concentrations of nitrogen dioxide were more associated with the topographical features of the region. The methods proposed in this study can serve as a reliable framework for analysing the air quality of industrial regions and can also, supplement future studies on emissions reduction in industrial parks.

## 1. Introduction

Air pollutants are detrimental to the natural environment (Wang et al., 2021); of note, in recent times, human health (Pope 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 ($N O_2$), and sulphur dioxide ($S O_2$) (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

* Corresponding author.
*E-mail address:* Yongze.song@curtin.edu.au (Y. Song).

https://doi.org/10.1016/j.jag.2023.103221
Received 27 October 2022; Received in revised form 25 January 2023; Accepted 1 February 2023

**Table 1**
Description of the spatial vector data used for industrial region identification.

| Type of vector data | Count | Rural | Urban | Source |
|---|---|---|---|---|
| Industrial land use polygon | 6237 | 2869 | 3368 | OSM |
| POI for utility and waste services | 1237 | 841 | 396 | NPI |
| POI for manufacturing | 1225 | 576 | 649 | NPI |
| POI for mining | 711 | 647 | 64 | NPI |
| POI supplement utility and waste services | 1076 | 1011 | 65 | OSM |
| POI converted from land-use polygons | 1221 | 708 | 513 | OSM |

investigations and is supported by remote sensing techniques. Several relevant previous studies analysed air pollutant concentration from the socioeconomic 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. Spatial association modelling is an effective approach to examine relationships between spatial variables and it has been widely implemented in geographical factor exploration and spatial prediction (Luo et al., 2022; Song et al., 2021; Song, 2022a, 2022b). 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 with standardised coefficients. In this study, we also explored the spatial impacts of various remote-sensing factors on air pollutant concentrations in detail.

## 2. Study area and data

### 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).

### 2.2. Data

#### 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 1 presents a general description of the spatial data used for industrial region identification.

#### 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, $O_3$, N $O_2$, and S $O_2$ , were considered as response variables in this study (Table 2). These remote sensing data were accessed from the Sentinel-5P mission carried out by the European Space Agency (ESA), and 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 2). For anthropogenic-activity data, 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

**Table 2**
Satellite measurements and raster data summary.

| Category | Data | Spatial resolution (m) | Temporal resolution | Statistics | Unit | Provider |
|---|---|---|---|---|---|---|
| Air pollutant concentrations | Column densities of CO, $O_3$, N $O_2$, and S $O_2$ | 1113 | Daily | Yearly average | mol/$m^2$ | Sentinel-5P |
| Socio-economic | NTL | 464 | Monthly | Yearly average | nanoWatts/ $cm^2$/sr | VIIRS |
| | Population | 100 | Yearly | – | count | WorldPop |
| Geography | DEM | 31 | – | – | m | Geoscience Australia |
| Vegetation | NDVI | 30 | 18 days | Yearly average | – | Landsat8 |
| | NDVI | 500 | 16 days | Yearly average | – | MODIS |
| Meteorology | Precipitation | 4638 | Monthly | Yearly sum | mm | TerraClimate |
| | Wind speed | 4638 | Monthly | Yearly average | m/s | TerraClimate |

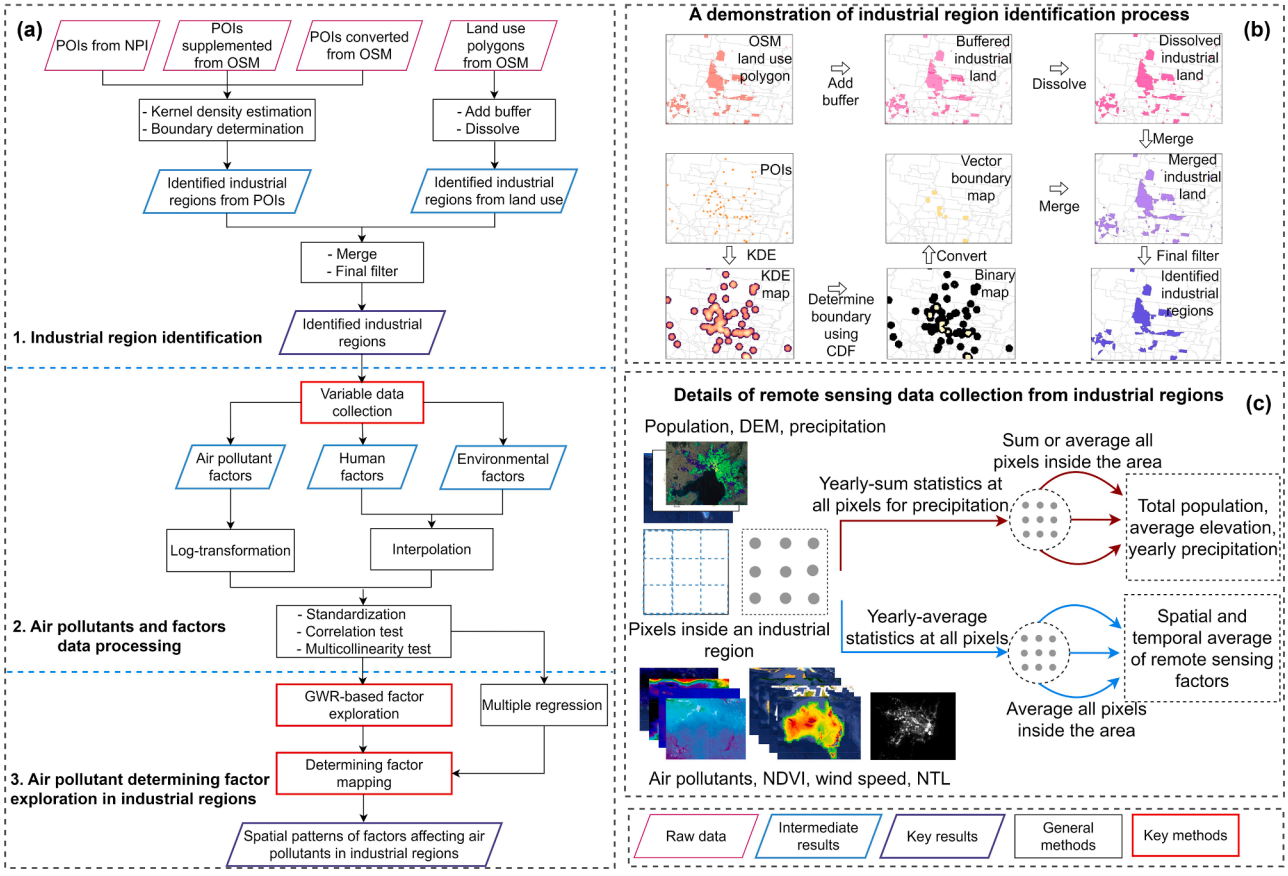



**Fig. 1.** Research workflow and detailed process. (a) Research workflow. (b) A demonstration of the industrial region identification process. (c) Details of remote sensing data collection.

datasets, which provide the monthly climate information on global terrestrial surfaces (Google Developers and University of California Merced, 2020).

## 3. Methods

Fig. 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.

### 3.1. Industrial region identification

#### 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).

#### 3.1.2. 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 Fig. 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., 2018).

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 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., 2018).

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$<$0.4642 km$^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.

### 3.2. 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 Fig. 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 spatio-temporal 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 spatio-temporal 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, $<$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 N $O_2$ 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. Multiple regression and GWR with standardized coefficients

Multiple regression and GWR were applied to quantitatively measure the factors that affected the air pollutant concentrations in the industrial regions, based on the factors selected in the previous process. A multiple regression model, as shown in Eq. (1), was utilised to quantify the relationship between the standardised air pollutant densities and the potential determining factors (Shmool et al., 2014).

$$Y_i = \beta_0 + \sum_{j=1}^{n} \beta_j X_{ij} + \varepsilon_i \tag{1}$$

where $Y_i$ is the air pollutant density at location $i$, and $\beta_j$ are the standardised coefficients of the selected influential factors computed using ordinary least squares (OLS). The absolute value of the standardised coefficient indicates the influence of air pollutant concentrations. $X_{ij}$ refers to the $j$-th influential factor value at location $i$, and $\varepsilon_i$ is the error term.

A GWR with an adaptive kernel was applied to quantify the relationship between the air pollutant concentrations and the selected influential factors, while considering spatial non-stationarity (Fotheringham et al., 2003). The GWR model is shown in Eq. (2).

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

where $Y_i$ refers to the four types of air pollutants, $X_{ij}$ is the selected anthropogenic or environmental influential factor, and $\hat{I}\mu_i$ is the error term. $\beta_j(u_i, v_i)$ denotes the local standardised coefficients for the influential factors; a higher absolute coefficient value indicates a stronger impact of the potential factors on the air pollutant concentration.

The values of $\beta_j(u_i, v_i)$ were computed using Eqs. (3) and (4). In our study, the industrial regions were sparsely distributed. Therefore, an adaptive kernel using the k-nearest neighbour (KNN) was chosen to compute the weight matrix. The bandwidth selection function was used to determine the final bandwidth value by minimising the model root mean square error value (Fotheringham et al., 2003).

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

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

The Gaussian weight kernel shown in Eq. (4) was applied to compute

**Table 3**
Summary of kernel density estimation (KDE) industrial boundary threshold selection based on points of interest (POI) density cumulative density function (CDF).

| POI 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 | 346,906 | | |

the spatial matrix. The adaptive bandwidth was determined using KNN. $d_{ij}$ refers to the distance between the two industrial regions. Notably, we used an R-language-based "spgwr" package, to determine the optimal bandwidth value, and analysed the 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 Fig. 1(a).

The input factors for both the global regression model [Eq. (1)] and the local spatial regression [Eq. (2)] 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).

## 4. Results

### 4.1. Identified industrial regions at a continental scale

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. 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.

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 to 51 km$^2$. The industrial region size in other areas were up to 107 km$^2$. Fig. 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).

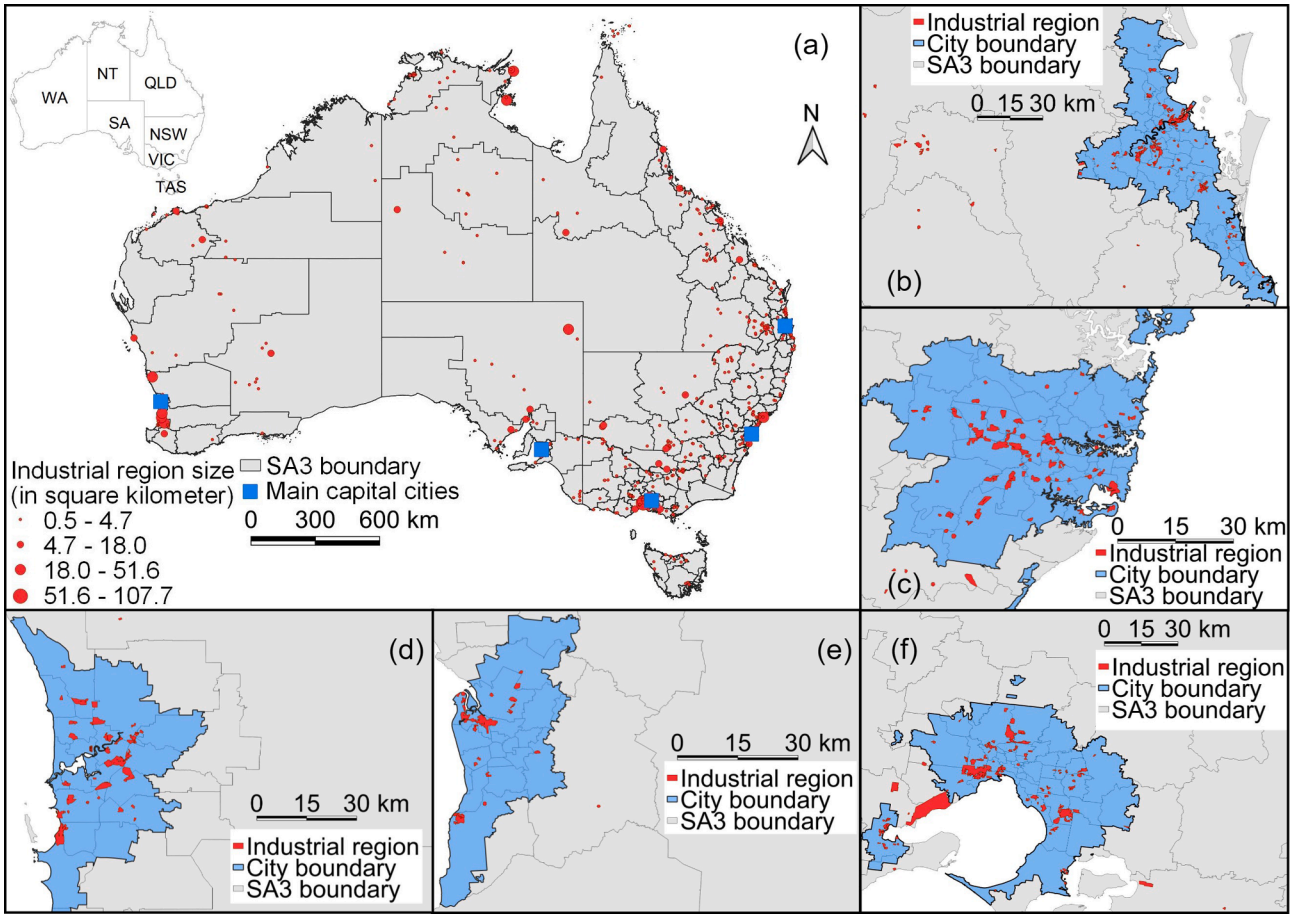



**Fig. 2.** Identified industrial regions in Australia and comparison with spatial distributions of major cities and Statistical Area Level 3 (SA3) boundaries. (a) Spatial distribution of industrial regions, main capital cities, and Australian state boundaries. Industrial regions in (b) Brisbane, (c) Sydney, (d) Perth, (e) Adelaide, and (f) Melbourne.

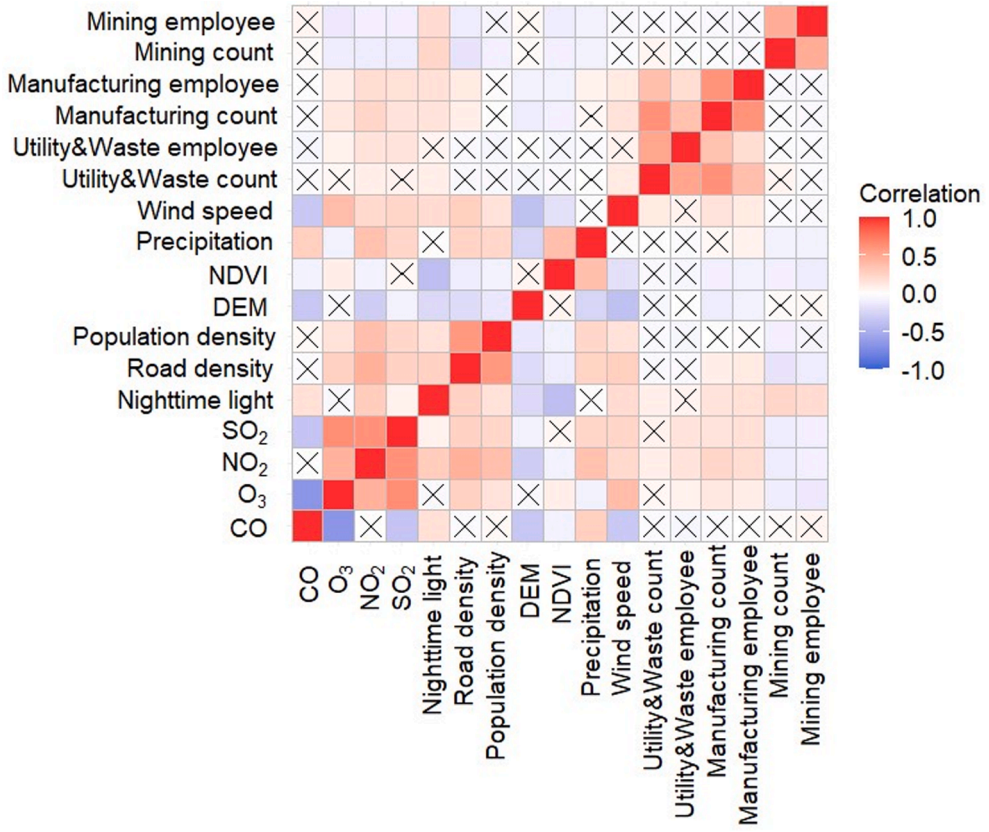



**Fig. 3.** Correlation test results.

**Table 4**
Multiple regression statistical results.

| | CO | $O_3$ | N $O_2$ | S $O_2$ |
|---|---|---|---|---|
| NTL | 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: (***) $p < 0.0001$, (**) $p < 0.001$, (*) $p < 0.01$, (.) for $p < 0.05$.

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.

### 4.2. Correlation test, influential factor selection, and multi-collinearity test

The correlation test results are shown in Fig. 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. N $O_2$ 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 S $O_2$ 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 4) were all less than 2.5, which is generally acceptable for the regression models.

### 4.3. Air pollutant determining factor exploration

#### 4.3.1. Analysis results from multiple regression and GWR with standardized coefficients

In this study, we determined the general air pollutant determining factors using OLS multiple regressions, as shown in Table 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 S $O_2$, while road

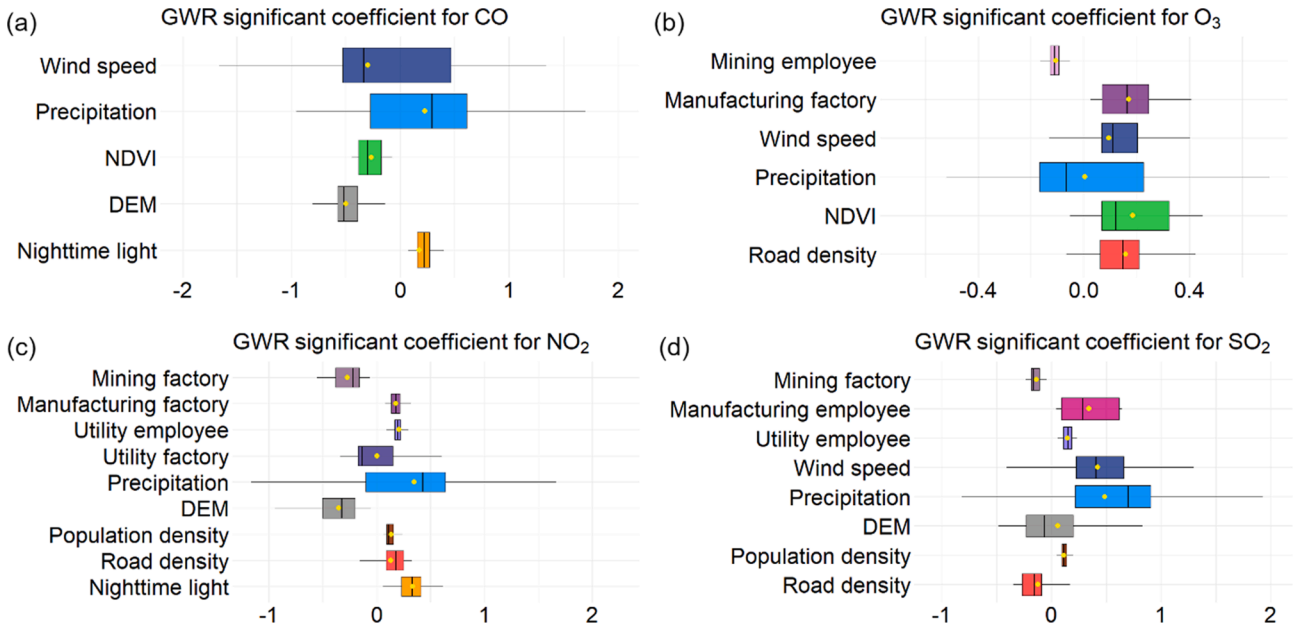



**Fig. 4.** Statistical summary of GWR significant coefficient for (a) CO model, (b) $O_3$ model, (c) N $O_2$ model, and (d) S $O_2$ model.

**Table 5**
Counts of geographically weighted regression (GWR) significant coefficient and determining factor.

| | Count of the industrial region where this factor is statistically significant and predominant / Count of industrial regions where the factor is statistically significant | | | |
|---|---|---|---|---|
| | CO | $O_3$ | N $O_2$ | S $O_2$ |
| NTL | 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: The top three determining factors for each model in terms of count are **bold fonts**.

**Table 6**
Model comparison.

| | CO | | $O_3$ | | N $O_2$ | | S $O_2$ | |
|---|---|---|---|---|---|---|---|---|
| | OLS | GWR | OLS | GWR | OLS | GWR | OLS | GWR |
| AIC value | 1768 | 660 | 1957 | −152 | 1841 | 850 | 2059 | 878 |
| RSS value | 436 | 77 | 555 | 32 | 474 | 111 | 630 | 118 |

density was the determining factor for N $O_2$. 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 N $O_2$ and S $O_2$. Additionally, anthropogenic activity indicators (i.e. NTL, road density, and population density) and manufacturing industry scales were positively related to air pollutant density.

The spatial disparities of the determining factors, on a continental level, were identified by GWR models, as shown in Fig. 4 and Table 5. Fig. 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 5 provides the supplementary information regarding the significant coefficient statistics and the determining factor statistics. Table 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 N $O_2$ density, precipitation, topography, and NTL were the most significant determining factors. For S $O_2$ density, precipitation was the primary determining factor, followed by wind speed and manufacturing employee scale. From Tables 4 and 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.

The multiple regression and GWR model performance are listed in Table 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 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

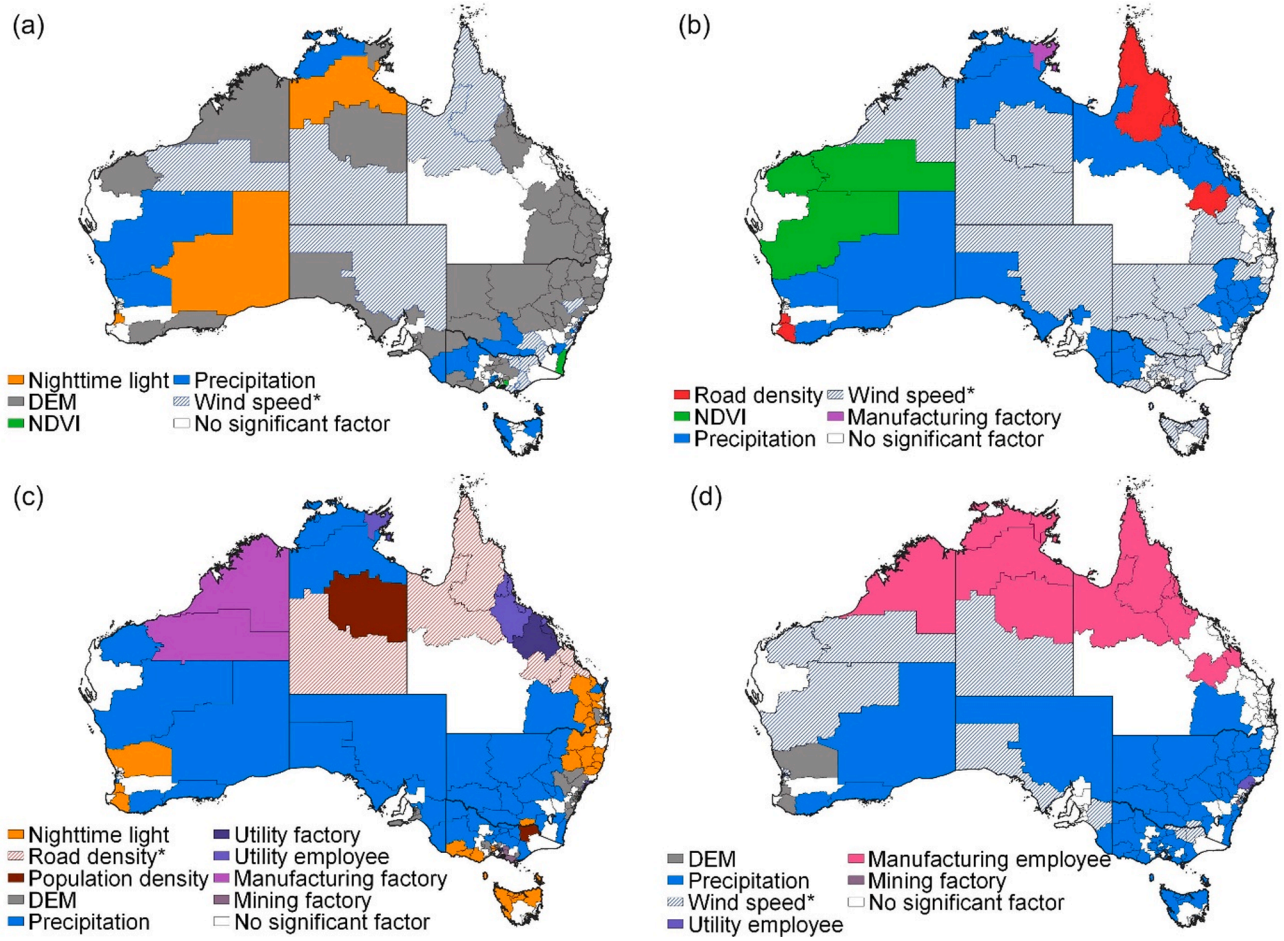



**Fig. 5.** Determining factors of air pollutant distributions in different SA3 areas: determining factor map for (a) CO, (b) $O_3$, (c) N $O_2$, (d) S $O_2$. "*" denoted a global factor not affected by local spatial impacts.

**Table 7**
Determining factors of air pollutant distributions in five major cities.

| | CO | N $O_2$ | S $O_2$ |
|---|---|---|---|
| Sydney | Precipitation, Wind speed | DEM | Precipitation, Wind speed, Mining factory |
| Melbourne | DEM, NDVI, Precipitation, Wind speed | NTL, DEM, Precipitation, Mining factory | Precipitation |
| Brisbane | DEM | NTL, DEM, Precipitation | – |
| Perth | DEM | Precipitation | Wind speed |
| Adelaide | DEM | DEM | – |

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

their spatial disparities.

*4.3.2. Air-pollutant-determining factor mapping: Which factor is more influential and at which location?*

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 Fig. 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 N $O_2$ 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 S $O_2$ 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 S $O_2$ concentrations in the northern part of the continent.

The factors affecting the air pollutants in the capital cities of Australia are summarised in Table 7. From a spatial perspective, meteorological factors and topography were the common attributes of air pollutants. Vegetation greenness also influenced the CO concentrations

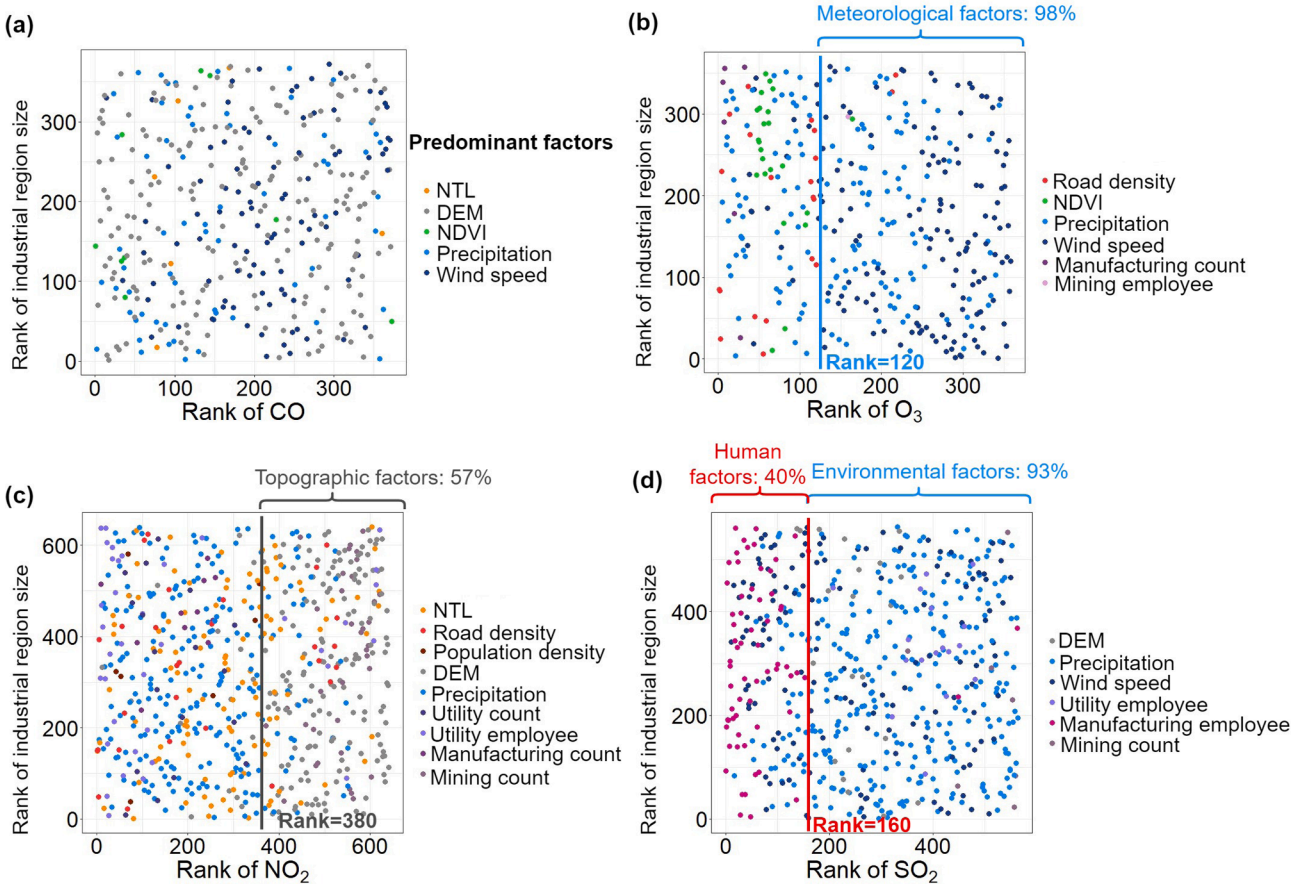



**Fig. 6.** Distributions of predominant factors under the rank of air pollutant concentrations and the rank of industrial region size. Distribution of factors affecting (a) CO, (b) $O_3$, (c) N $O_2$, (d) S..$O_2$

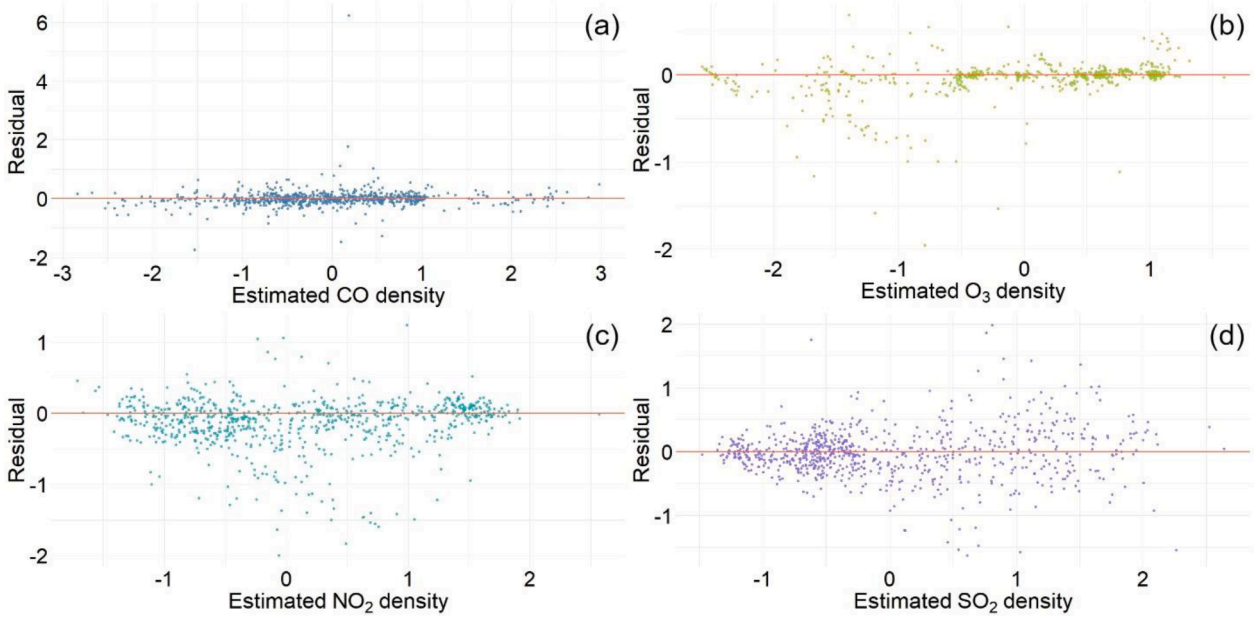



**Fig. A1.** Residuals vs. fitted plots for GWR models. (a) Residual plot of CO density. (b) Residual plot of $O_3$ density. (c) Residual plot of N $O_2$ density. (d) Residual plot of S $O_2$ density. Note: GWR made estimations on log-transformed CO, $O_3$, and N $O_2$ density.

in Melbourne, and the NTL and mining factory scale had an impact on the suburbs of Melbourne. NTL also influenced the N $O_2$ concentrations in the suburbs of Brisbane.

The distributions of the factors affecting the air pollutant concentrations in the study area are shown in Fig. 6. The scatter plots in Fig. 6 portray the distribution of various predominant factors in ascending

**Table A1**
List of abbreviations.

| Abbreviation | Word | Abbreviation | Word |
|---|---|---|---|
| AIC | Akaike information criterion | NT | Northern Territory |
| ASGS | Australian Statistical Geography Standard | NTL | nighttime light |
| CDF | cumulative distribution function | $O_3$ | ozone |
| CO | carbon monoxide | OLS | ordinary least square |
| DEM | digital elevation model | OSM | OpenStreetMap |
| EKC | Environmental Kuznets Curve | POI | point of interest |
| ESA | European Space Agency | QLD | Queensland |
| GEE | Google Earth Engine | RSS | residual sum of squares |
| GIS | geographic information system | SA | South Australia |
| GWR | geographically weighted regression | SA2 | statistical area level 2 |
| KDE | kernel density estimation | SA3 | statistical area level 3 |
| KNN | k-nearest neighbor | $S O_2$ | sulfur dioxide |
| MODIS | Moderate Resolution Imaging Spectroradiometer | TAS | Tasmania |
| $N O_2$ | nitrogen dioxide | VIC | Victoria |
| NDVI | normalized difference vegetation index | VIF | variance inflation factor |
| NPI | national pollutant inventory | VIIRS | Visible Infrared Imaging Radiometer Suite |
| NSW | New South Wales | WA | Western Australia |

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 $S O_2$ 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 $N O_2$, the topographic factors were more strongly associated with the air pollutant concentrations in the industrial regions.

## 5. Discussion

### 5.1. Spatial factors affecting air pollutant density at a continental scale

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 $N O_2$ density, but not to $S O_2$ density. The positive correlation with $N O_2$ 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 $S O_2$ may be due to the Environmental Kuznets Curve (EKC) effect. Road density was mainly negatively related to $S O_2$ 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-$S O_2$ relationship.

### 5.2. 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 $N O_2$ and $S O_2$ 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.

### 5.3. Limitations

This study has some limitations. The first limitation is the existence of heteroscedasticity in some measurement data, such as that of $S O_2$. 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.

## 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 unexplored. 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 on 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, 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 $S O_2$ were more associated

with meteorological factors of the area, while the higher concentrations of N $O_2$ 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.

## CRediT authorship contribution statement

**Zehua Zhang:** Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Writing – original draft. **Yongze Song:** Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Supervision, Writing – original draft, Writing – review & editing. **Peng Luo:** Formal analysis, Visualization, Writing – review & editing. **Peng Wu:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Xiaochi Liu:** Writing – review & editing. **Mingshu Wang:** Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgement

This work was partially supported by the University of Glasgow, Reinvigorating Research Funding (Grant No. 201644-20) and Research Enabling/Impact Generating Scheme (Grant No. 126438-01).

## Appendix

Fig. A1 and Table A1.

## References

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. J. Geovisualization Spatial Anal. 5 (1) https://doi.org/10.1007/s41651-021-00079-6.

Australian Bureau of Statistics, 2021a. National, state and territory population. Available from: <https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release>.

Australian Bureau of Statistics, 2021b. Australian Industry. Available from: <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. Available from: <>.

Borck, R., Schrauth, P., 2021. Population density and urban air quality. Reg. Sci. Urban Econ. 86 (103596), 103596 https://doi.org/10.1016/j.regsciurbeco.2020.103596.

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 Environ. (Oxford, England: 1994) 223 (117267), 117267. doi: 10.1016/j.atmosenv.2020.117267.

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

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

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

Department of the Environment and Energy, Australian Government, October 13, 2021. National Pollutant Inventory. Available from: <http://www.npi.gov.au/>.

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

Dons, E., Van Poppel, M., Kochan, B., Wets, G., & Int Panis, L., 2013. Modeling 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. doi: 10.1016/j.atmosenv.2013.03.050.

Erdogan, S., 2020. Analyzing the environmental Kuznets curve hypothesis: the role of disaggregated transport infrastructure investments. Sustain. Cities Soc. 61 (102338), 102338 https://doi.org/10.1016/j.scs.2020.102338.

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. https://doi.org/10.3390/su71115570.

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 Sens. (Basel) 12 (8), 1297.

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.

Geofabrik and OpenStreetMap contributors, 2020. Download OpenStreetMap for this region: Australia and Oceania [Data set]. Available from: <http://download.geofabrik.de/australia-oceania.html>.

Geoscience Australia, 2014. Area of Australia - states and territories. Available from: <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. Comput. Environ. Urban Syst. 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]. Available from: <https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C01_T1_8DAY_NDVI>.

Google Developers and Earth Observation Group, 2020. VIIRS Stray Light Corrected Nighttime Day/Night Band Composites Version 1 [Data set]. Available 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]. Available from: <https://developers.google.com/earth-engine/datasets/catalog/AU_GA_DEM_1SEC_v10_DEM-S>.

Google Developers and the European Space Agency, 2020. Sentinel-5P [Data set]. Available 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. Available 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]. Available from: <https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_TERRACLIMATE>.

Google Developers and Worldpop, 2020. WorldPop Global Project Population Data [Data set]. Available 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. Sci. Total Environ. 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. doi: 10.1007/s11869-020-00913-8.

Hadjisophocleous, G., Chen, Z., 2010. A survey of fire loads in elementary schools and high schools. J. Fire. Prot. Eng. 20, 55–71. https://doi.org/10.1177/1042391509360266.

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 Environ. (Oxford, England: 1994) 42(33), 7561–7578. doi: 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 Environ. (Oxford, England: 1994) 254 (118415), 118415. doi: 10.1016/j.atmosenv.2021.118415.

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. Front. Environ. Sci. Eng. 17 (1), 8. https://doi.org/10.1007/s11783-023-1608-1.

Li, J., Long, Y., Dang, A., 2018. Live-Work-Play Centers of Chinese cities: Identification and temporal evolution with emerging data. Comput. Environ. Urban Syst. 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. Comput. Environ. Urban Syst. 78 (101356), 101356 https://doi.org/10.1016/j.compenvurbsys.2019.101356.

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. Sci. Total Environ. 565, 607–615. https://doi.org/10.1016/j.scitotenv.2016.03.189.

Luo, P., Song, Y., Zhu, D., Cheng, J., Meng, L., 2022. A generalized heterogeneity model for spatial interpolation. Int. J. Geogr. Inf. Sci. 1–26 https://doi.org/10.1080/13658816.2022.2147530.

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. Environ. Res. 137, 308–315. https://doi.org/10.1016/j.envres.2015.01.003.

Pope, C.A., Thun, M.J., Namboodiri, M.M., Dockery, D.W., Evans, J.S., Speizer, F.E., Heath, C.W., 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am. J. Respir. Crit. Care Med. 151 (3), 669–674.

Roy, A., 2021. Atmospheric pollution retrieval using path radiance derived from remote sensing data. J. Geovisualization Spatial Anal. 5 (2) https://doi.org/10.1007/s41651-021-00093-8.

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.

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

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. Ecol. Ind. 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. Environ. Health: Global Access Sci. Source 13 (1), 91. https://doi.org/10.1186/1476-069X-13-91.

Song, Y., Long, Y., Wu, P., Wang, X., 2018. 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., Wu, P., 2021. An interactive detector for spatial associations. Int. J. Geogr. Inf. Sci. 35 (8), 1676–1701. https://doi.org/10.1080/13658816.2021.1882680.

Song, Y., 2022a. Geographically optimal similarity. Math. Geosci. 1–26 https://doi.org/10.1007/s11004-022-10036-8.

Song, Y., 2022b. The second dimension of spatial association. Int. J. Appl. Earth Obs. Geoinf. 111, 102834 https://doi.org/10.1016/j.jag.2022.102834.

Tian, Y., Yao, X., Chen, L., 2019. Analysis of spatial and seasonal distributions of air pollutants by incorporating urban morphological characteristics. Comput. Environ. Urban Syst. 75, 35–48. https://doi.org/10.1016/j.compenvurbsys.2019.01.003.

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. Comput. Environ. Urban Syst. 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 Environ. (Oxford, England: 1994) 265 (118721), 118721. doi: 10.1016/j.atmosenv.2021.118721.

VicRoads, 2021, August 18. Victoria's road network. Gov.Au. Available from: <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, 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. Ecol. Indicators 129 (107976), 107976. https://doi.org/10.1016/j.ecolind.2021.107976.

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 Sens. (Basel) 13 (11), 2152. https://doi.org/10.3390/rs13112152.

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 Int. J. Geo Inf. 6 (12), 389. https://doi.org/10.3390/ijgi6120389.

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). J. Environ. Engi. (Transactions of AIJ) 77 (681), 889–897.

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. Environ. Pollution (Barking, Essex: 1987) 254(Pt A), 112952. doi: 10.1016/j.envpol.2019.07.120.

York, R., Rosa, E.A., Dietz, T., 2003. STIRPAT, IPAT and ImPACT: analytic tools for unpacking the driving forces of environmental impacts. Ecol. Econ.: J. Int. Soc. Ecol. Econ. 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. Modeling the spatiotemporal dynamics of industrial sulfur dioxide emissions in China based on DMSP-OLS nighttime stable light data. PLoS One 15 (9), e0238696.

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 Environ. (Oxford, England: 1994), 181, 145–154. doi: 10.1016/j.atmosenv.2018.03.017.

Zhang, Z., Song, Y., Wu, P., 2022. Robust geographical detector. Int. J. Appl. Earth Observation Geoinformation: ITC J. 109 (102782), 102782 https://doi.org/10.1016/j.jag.2022.102782.

## Further reading

Dockery, D.W., 1995. An association between air pollution and mortality in six US cities. J. Occup. Environ. Med. 37 (2), 136. https://doi.org/10.1056/NEJM199312093292401.