



Geostatistical predictive modeling for asthma and chronic obstructive pulmonary disease using socioeconomic and environmental determinants

R. M. K. Kumarihamy · N. K. Tripathi

Received: 3 February 2017 / Accepted: 20 March 2019
© Springer Nature Switzerland AG 2019

Abstract The spatial distribution of the prevalence of asthma and chronic obstructive pulmonary disease (COPD) remains under the influence of a wide array of environmental, climatic, and socioeconomic determinants. However, a large proportion of these influences remain unexplained. In completion, this study examined the spatial associations between asthma/COPD morbidity and their determinants using ordinary least squares (OLS) and geographically weighted regressions (GWR). Inpatient records collected from the secondary and tertiary care hospitals in Kandy from 2010 to 2014 were considered as the dependent variable. Potential risk factors (explanatory variables) were identified in four distinguished classes: 1) meteorological factors, (2) direct and indirect factors of air pollution, (3) socioeconomic factors, and (4) characteristics of the physical environment. All possible combinations of candidate explanatory variables were evaluated through an exploratory regression. A comparison between the regression models was also explored. The best OLS regression models revealed about 55% of asthma variation and 62% of

COPD variation while GWR models yielded 78% and 74% of the variation of asthma and COPD occurrences respectively. Relative humidity, proximity to roads (0–200 m), road density, use of firewood as a source of fuel, and elevation play a vital role in predicting morbidity from asthma and COPD. Both local and global regression models are important in assessing spatial relationships of asthma and COPD. However, the local models exhibit a better prediction capability for assessing non-stationary relationships of asthma and COPD than global models. The geostatistical aspects used in this study may also provide insights for evaluating heterogeneous environmental risk factors in other epidemiological studies across different spatial settings.

Keywords Asthma · COPD · Spatial determinants · Ordinary least square regression · Geographically weighted regression

Introduction

Chronic respiratory diseases (CRD), essentially asthma and chronic obstructive pulmonary disease (COPD), are influenced by a wide array of environmental, climatic, and socioeconomic factors (Abramson et al. 2015; Burney 2017). Epidemiological studies have convincingly explained the association between CRD morbidity and different individual levels risk factors such as particulate matter (PM₁₀), temperature, and poverty. For example, an elevated rate of CRD symptoms, medication use, and hospitalization was found among

This article is part of the Topical Collection on Geospatial Technology in Environmental Health Applications

R. M. K. Kumarihamy (✉) · N. K. Tripathi
Remote Sensing and Geographic Information System AoS, School of Engineering and Technology, Asian Institute of Technology, P.O. Box 4, Klong Luang, Pathumthani 12120, Thailand
e-mail: kumudunik@gmail.com

R. M. K. Kumarihamy
Department of Geography, University of Peradeniya, Peradeniya, Sri Lanka

communities with higher concentrations of PM₁₀ and other ambient air pollutants (Schwartz et al. 1993; Burnett et al. 1994; Van Vliet et al. 1997; Xiong et al. 2015). Moreover, socioeconomic status (SES) measured with lower levels of education and income has shown a positive association with morbidity and mortality of respiratory diseases (Prescott and Vestbo 1999; Bacon et al. 2009; Burney 2017; Keet et al. 2017). Further, epidemiological studies have linked bioclimatic factors such as humidity and temperature to respiratory health (Donaldson et al. 1999; Villeneuve et al. 2005; Mireku et al. 2009; Ding et al. 2017). Most of these studies have examined the association between CRD and individual risk factors using bivariate regression. However, these risk factors are inextricably linked together to affect the respiratory health of communities. Therefore, considerable inconsistencies are yet to be explained in terms of CRD surveillance at the local level.

A growing body of literature confirms the interconnected association of CRDs and their risk factors. For example, Wheeler and Ben-Shlomo examined the linked relationship between air quality, SES, and respiratory health using a non-spatial ordinary least squares regression model (Wheeler and Ben-Shlomo 2005). Similarly, the impact of air pollutants, meteorological variables, and thermal comfort indices on respiratory morbidity has quantified using the Poisson regression model (de Sousa Zanotti Stagliorio Coelho et al. 2010). Although the techniques utilized in such studies have some capability to assess the relationship of multifactorial determinants in a population, they are not able to assess the relationship in a spatial context. Thus, these techniques are adequate to assess spatial aspects such as dependency and non-stationarity, and multi-collinearity of the variables is questionable. According to Wong et al., multi-collinearity is an issue of two or more different risk factors being highly correlated with one another, especially in a multi-pollutant model (Wong et al. 1999). Moreover, socioeconomic variables often exhibit properties of spatial dependency when similar variables are measured at locations within a close proximity (Park and Kim 2014). Further, spatial and aspatial data have unique properties; thus, different statistical methods developed for aspatial data are not effective to assess spatially related phenomena (Fotheringham et al. 2003).

The association between multifactorial determinants and respiratory health is a spatial problem by its very nature (Scherber et al. 2014; So and Lai 2007). Moreover, the burden of CRD varies by space and time in

response to spatiotemporal variation of their determinants (Yin et al. 2016). Thus, CRD surveillance should consider all these aspects.

The term “disease surveillance” generally refers to the process of systematic collecting, organizing, analyzing, interpreting, and presenting of health-related data and risk factors for health planning and evaluation (Berkelman et al. 2009; Choi 2012). Disease surveillance provides a platform for policy makers to understand the prevalence of diseases and risk factors better. In order to develop a proper surveillance strategy, there should have a broad approach that integrates multifactorial determinants in terms of space and time. However, spatially and temporally integrated CRD risk factor assessments are often limited in the literature due to methodological issues and lack of data (Corvalán et al. 1997; Cruz et al. 2007).

The scarcity and rudimentary nature of communal spatial data on disease exposure and their determinants are one of the major barriers to such research in most developing countries (Cruz et al. 2007; Ait-Khaled et al. 2001; Mortimer et al. 2015). Data on CRD and risk factors are reported for less than 25% of the world's population and are mostly from high income countries (Cruz et al. 2007). Data on respiratory disease risk factors and burden are often incomplete, fragmented in many countries, and only available as country-wide aggregated data (national level data) (Cruz et al. 2007). Therefore, the resolution of the data is too coarse to do an accurate analysis (Tonne et al. 2017; Comber et al. 2011). In the case of Sri Lanka, mortality and morbidity data are aggregated into second-level administrative units known as districts. Local variations in health statistics, therefore, remain unknown. This type of statistical bias due to spatial aggregation is generally known as a modifiable areal unit problem (MAUP) (Openshaw 1984). Geocoding capabilities of geographic information systems (GIS) allow derivation of more detailed, precise, and flexible spatial data from aspatial health information systems (Rushton et al. 2007). Such detailed finer resolution spatial data on disease empowers different spatial analysis such as recording disease incidence in relation to pollution sources (Rushton et al. 2007). Combining spatially incompatible disease data received from a variety of sources and scales is a challenge in CRD surveillance. However, GIS provides capabilities to combine spatially disparate data (Crawford and Young 2004). Ad hoc methods such as union, intersection, and zonal-averaging are extensively

used data integration methods in GIS (Devries et al. 2016). Although such GIS capabilities facilitate addressing such problems, spatial epidemiological studies on CRD are still limited in some countries due to the lack of expertise (Cruz et al. 2007).

Rapidly growing geostatistical analytical capabilities of GIS are widely used in health and other social sciences to overcome aforesaid methodological issues (Robertson and Nelson 2010). Further, GIS facilitates examining of the spatial association among risk factors and disease (Jerrett et al. 2010). The available literature on CRDs shows how GIS capabilities assist in understanding spatial patterns underlying the relationship between CRDs and their risk factors. Many studies have used GIS techniques such as inverse distance weighting (IDW), or Kriging for analyzing and mapping the association between exposures to air pollution (Gorai et al. 2014; Peled et al. 2006). A study done in China explores the utility of geocoding capabilities, Kriging interpolation technique, and Global Moran's I to investigate the spatial association between air pollution and COPD hospitalization (Wang et al. 2015).

Ordinary least squares (OLS) is one of the most widely recognized regression analysis methods that can be used in epidemiological studies (Ahmad et al. 2015). The underlying assumption of OLS method is that the relationship between the predictors and the dependent variable is stationary across space (Bascuñán and Quezada 2016). Thus, OLS computes a single regression statistic for all spatial entities in a global model. Moreover, it is considered as a good starting point for spatial regression analysis (Scott and Janikas 2010). Miller in 2012 endorsed that OLS method can be used to assess spatially non heterogeneous relationships (Kala et al. 2017). However, in the real world, the relationship between variables might show spatial heterogeneity. In such cases, global regression methods may provide inaccurate predictions (Nkeki and Osirike 2013).

In contrast, geographically weighted regression (GWR) is a local regression method that explores spatial heterogeneity of the relationship (Fotheringham et al. 2003; Matthews and Yang 2012; Comber et al. 2011). GWR provides a platform to assess spatial relationship across spatial entities and yields multiple regression statistics for each spatial entity in the analysis. GIS provides an effective platform to accomplish both global and local spatial analysis; thus, the use of these methods in public health and epidemiological studies has

increased dramatically. Further, empirical literature has shown that both global and local regression methods have been applied in risk factor analysis for diseases including malaria (Kala et al. 2017), cholera (Nkeki and Osirike 2013), leishmaniasis (Ali-Akbarpour et al. 2012), tuberculosis (Wubuli et al. 2015), and diabetes (Wubuli et al. 2015). Most of these studies emphasize that the GWR method exhibits a better fit compared with the OLS method in assessing the spatially varied relationship of diseases.

In the light of the above facts, this study investigates the spatial associations that exist between morbidity due to CRD and its socioeconomic and environmental determinants in Kandy, Sri Lanka. A range of geospatial analytical capabilities of GIS were utilized to bring spatial and aspatial data into a common platform. Both local and global regression models were employed in the relationship assessment and compared with the modeling results in order to find the best fit.

Materials and methods

Research design

This is a retrospective ecological study design, which is primarily based on the inpatient records of public health care institutions in Kandy district in Sri Lanka. Aggregated data was provided at Grama Niladari Division level (GND)¹ which enabled a detailed examination of factors influencing disease incidence rates (Fig. 1). The respiratory disease incidence data in the study area covering the period between 1 January 2010 and 31 December 2014 were considered as the dependent variable. Based on the literature review, potential explanatory variables of CRDs were identified in four distinguished classes: (1) meteorological factors, including rainfall, relative humidity, and temperature; (2) direct and indirect factors of air pollution, including particulate matter with mean aerodynamic diameter 10 μm or less (PM_{10}), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), road density, proximity to road, and firewood as a source of fuel; (3) socioeconomic

¹ Sri Lanka is divided into nine provinces, which are further subdivided into 25 districts. Districts are further subdivided into divisional secretariat divisions (DSD). Each DSD is divided into Grama Niladhari divisions. A Grama Niladhari division is a fourth-level subdivision in the country, so they are the lowest-level administrative units in the country.

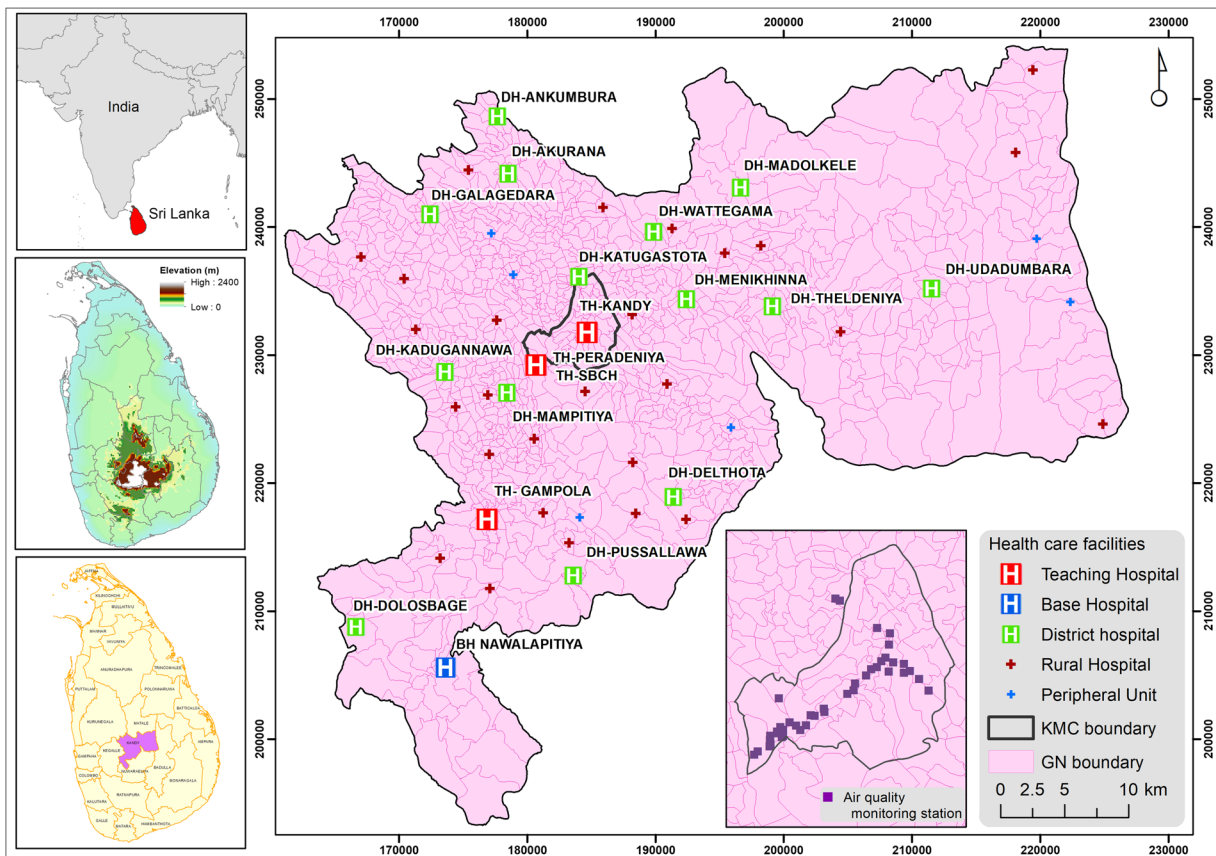


Fig. 1 Location map of the study area, Kandy district, Sri Lanka. Legend: The map shows the location of Kandy district in the regional and national context, Grama Niladiri division boundaries, locations of hospitals, and air quality sampling sites. Kandy district

is one of the 25 different second-level administrative divisions of Sri Lanka. The extent of the district is about 1940.3 km². Accordingly to the administrative structure of the country, the district has divided to 1188 lowest-level units called Grama Niladiri divisions.

factors, including level of education and income; and finally (4) characteristics of the physical environment, including topography and type of land cover. Analytical capabilities of ArcGIS 10.3.1 (Desktop 2011), GeoDa 1.6.7 (Anselin and Rey 2005), and SPSS 16.0 (Released 2007) were utilized in data preparation and analysis. The overall methodology of the study is summarized in Fig. 2.

Data preparation

Respiratory incidence data The indoor morbidity and mortality record (IMMR) system is the sole source of morbidity data in Sri Lanka (Gamage 2004; Kariyawasam et al. 2011). The IMMR keeps daily inpatient records of public sector health care institutions (Gamage 2004). Kandy district has 52 different public health care institutions including four teaching

hospitals, one base hospital, 14 district hospitals, 27 rural hospitals, and six peripheral units as shown in Fig. 1. The rural hospitals and peripheral units are primary health care institutions; therefore, they were not considered in the study. Four district level hospitals, namely, Dolosbage, Ankumbura, Modolkele, and Ududumbara (Fig. 1), do not have completed IMMR for the study period. Therefore, daily inpatient records of asthma (International Classification of Disease J45) and COPD (International Classification of Disease J44) were collected from the remaining hospitals (in total 15) in Kandy district for 5 consecutive years from 2010. Based on the patients' addresses, CRD inpatient records were geocoded to a relatively good matching rate (76%) and positional accuracy (90% of the addresses geocoded to within 300 m of their true location). Subsequently, the geocoded CRD records were aggregated into GNDs

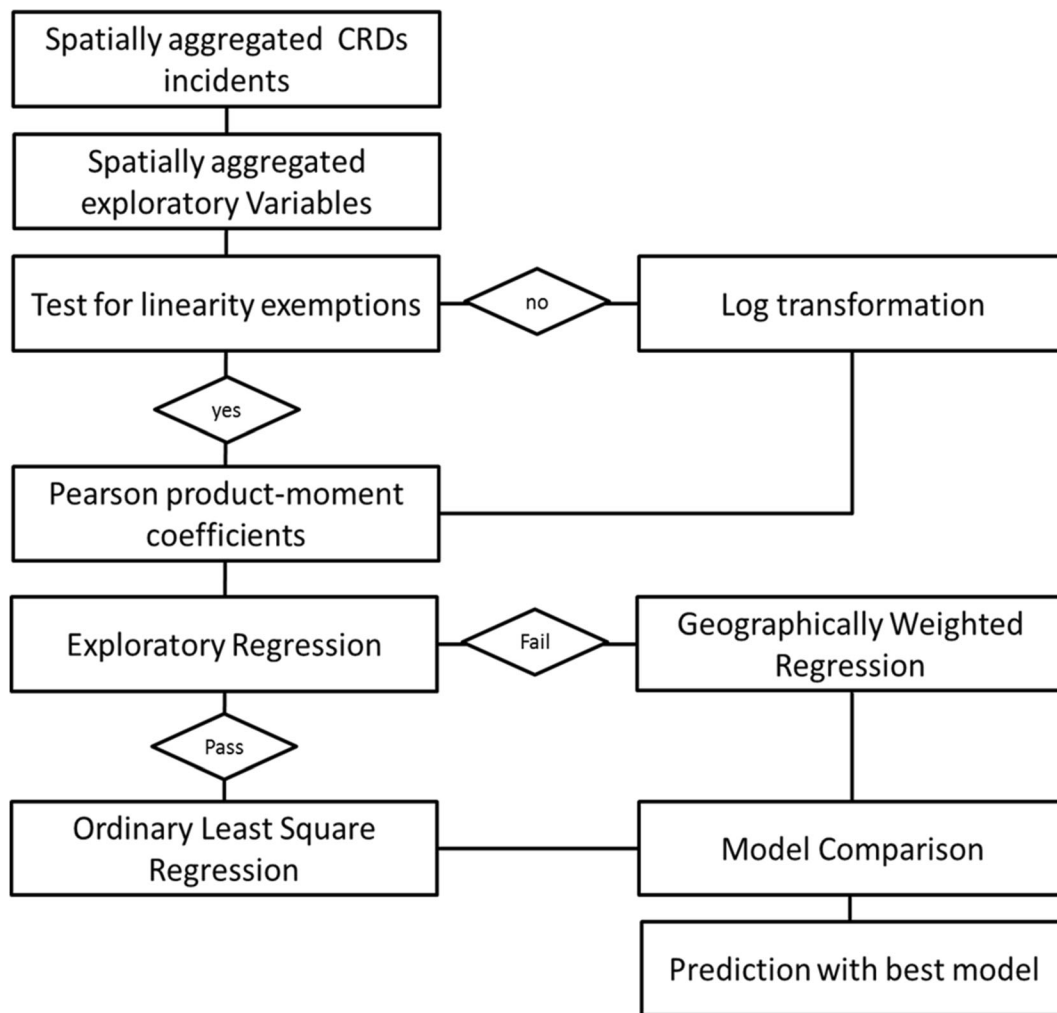


Fig. 2 The stochastic predictive modeling methodology

in the study area and average annual disease incidence rates were calculated.

Meteorological factors Rainfall, relative humidity, and temperature are well documented CRD predictors (Hervás et al. 2015; Chavarria 2001; Tseng et al. 2013; Stevanović et al. 2016). Therefore, ground-observed monthly records of rainfall (Rf) (36 different rain gauge stations), temperature (Te), and relative humidity (RH) (seven different weather stations for each) for the period of 2010 to 2014 were obtained from the Department of Meteorology, Sri Lanka. Average annual rainfall (mm), relative humidity (%), and temperature (°C) surfaces for the study area were derived using the ordinary Kriging interpolation technique (Johnston et al. 2001) in ArcGIS 10.3.1. The maximum, minimum, and mean amount of rainfall, temperature, and

relative humidity for each GN division in the study area were obtained using the zonal statistics extension of ArcGIS 10.3.1 (McCoy et al. 2001; Soysal et al. 2012).

Air pollution data Several studies have emphasized the association between ambient air pollution and respiratory disease (Ding et al. 2017; Moore et al. 2016; Jassal 2015). Ambient air pollutants such as particulate matter, nitrogen dioxide, sulfur dioxide, and carbon monoxide can be considered significant predictors of CRD. In order to explore the reported levels of air pollution in Kandy, a comprehensive literature survey was carried out; however, the result revealed that there are no fixed ground monitoring stations available for Kandy district (Wickramasinghe et al. 2011). Reported data are available only for Kandy Municipal Council (KMC) area.

Therefore, the most recent air pollution data (NWS&DB 2014) of average monthly particulate matter (PM₁₀), nitrogen dioxide (NO₂), sulfur dioxide (SO₂), and carbon monoxide (CO) for the year 2014 were interpolated using IDW method (Johnston et al. 2001) in order to derive the amount of pollution over space (only for the KMC area as shown in Fig. 1). Cell statistics for each GN division of the KMC area were obtained using zonal statistics. In addition, road density, the ratio of the length of the GND's total road network to the GND's land area (Rd), and proximity to roads from the patients' locations were considered as proxy determinants of air pollution because both factors are known to exacerbate CRD (Hansell et al. 2014; Li et al. 2011; Muralidharan et al. 2014). A road layer was derived from the 1:10,000 topographic maps and both road density (km of road per square km of land area) and proximity to roads were calculated. Three different proximities were identified: 0–200 m (B1), 200–400 m (B2), and more than 400 m (B3). Exposure to indoor air pollution such as biomass combustion is reliably associated with CRD (Kurmi et al. 2010; Van Miert et al. 2012). Therefore, GN level census data on the use of firewood (Fw) as a source of energy in lighting and cooking at the household level (%) were considered as a proxy to quantify the indoor air pollution in the study area.

Land use/land cover data Land use and land cover data are an integral part of respiratory health studies (Rao 2016). Generally, they provide an indication of biomass combustion, the level of air pollution, and exposure (Jagger and Shively 2014). Therefore, land use and land cover data for the study area at the scale of 1:10,000 were examined through an exploratory factor analysis in SPSS using principle factor extraction and verimax rotation. Land cover data were classified into three broad classes: agricultural crops (Ag), homestead (Hs), and other (Or) type of land cover. The land use class Ag includes paddy, tea, rubber, and chena cultivations; the class Hs includes homesteads, settlement areas, and built-up areas, whereas the class Or contains all other land use types such as forest, water bodies, and barren land. The extents of each land use class by GN divisions (km²) were calculated accordingly.

Topographic data High altitude and cold air exposure are also a known trigger of CRD (Seys et al. 2013; Cote et al. 1993). Further, distribution of human settlements and human activities such as mobility are often

manifested in topography (Zhang et al. 2014). Altitude for example indirectly affects respiratory health. Mean elevation was found as a good determinant which can play an important role in detecting respiratory disease exacerbation (Hales et al. 1998) and human settlement distribution (Zhang et al. 2014). Therefore, to assess the elevation levels in the area, a digital elevation model (DEM) was created using contours and spot heights at the scale of 1:50000. Based on the DEM, average elevation (El) of each GN division (m) was obtained using the zonal statistics.

Socioeconomic data There is a direct relationship between SES and CRD. Lower SES typically shows a positive association with CRD (Uphoff et al. 2015; Curtis et al. 2012; Ore and Ireland 2015). Therefore, census data on educational attainment and economic activity measured at the GN level were considered as proxy parameters for SES of the study. The census of educational attainment is presented in six different levels: the population who never attended a school, only primary school, up to secondary, up to G.C.E O/L or equivalent, G.C.E A/L or equivalent, and degree and above. Using principle factor extraction and verimax rotation-based exploratory factor analysis, the six variables were reduced to two. The percentage of the population who never attended a school and those who only completed primary school were categorized as having a low level of education (Le) while those who completed secondary school or higher were considered as having a higher level of education (He). The percentage of the population that was employed was identified as economically active (Ea), while the unemployed population was considered economically inactive (Ei).

Data analysis

Twenty explanatory variables pertaining to the four categories of respiratory determinants were first examined to test the linearity and normality assumptions. The exploratory spatial data analysis (ESDA) tools of ArcGIS 10.3.1 were used to examine the distribution of data (Johnston et al. 2001). Histograms and normal QQ plots were created for each variable and it was graphically determined if highly skewed results or kurtosis occurred. In a prior analysis, scatter plot matrices were visually inspected in order to assess the assumption of linearity for a continuously regressed variable, later the assessment was replaced with residual plots.

Outstanding variables were log transformed (natural log) to yield approximate normality in the variables. The robustness of the expected relationship between respiratory incidence and each of the individual potential explanatory variables was investigated by using Pearson product-moment correlation coefficients (Mukaka 2012). The goodness of fit of the variables was quantified through the data mining tool called “exploratory regression” in ESRI ArcGIS 10.1 (Kauhl et al. 2016; Rosenshein et al. 2011). Accordingly, the best fitting models with the best combinations of the explanatory variables within and out of the broad variable classes were assessed by the threshold criteria fitted with adjusted R -squared value ($R^2 > 0.5$), coefficient p values ($p < 0.1$), variance inflation factor ($VIF < 7.5$) values, and Jarque–Bera p values ($p > 0.1$) (Jarque 2011). Multi-collinearity was assessed with the variance inflation factor (VIF) where a larger VIF value(s) greater than 7.5 indicates the existence of redundancy among the explanatory variables. The Jarque–Bera test was utilized to assess bias and normality of the model’s variables. When the Jarque–Bera test is statistically significant, variables are not normally distributed which indicates that the model is biased or that one or more key variables are missing from the model (Rosenstein et al. 2011; Haque et al. 2012). The Koenker–BP statistic (Koenker’s studentized Bruesch–Pagan statistic) was utilized to determine whether the explanatory variables in the model had a consistent relationship to the dependent variable both in geographic space and in data space (spatial stationary). When the probability associated with the Koenker–BP statistic is statistically significant, regional variation is indicated (spatial non-stationary) and the model is a worthy candidate for geographically weighted regression (Nkeki and Osirike 2013). The spatial independency and normality of the residuals (the residuals are spatially random) were assessed using the global spatial autocorrelation coefficient Moran’s I test (Lee and Wong 2001). Statistically, significant clustering of the model residuals (spatial dependence in the model residuals) indicates a violation of the model assumption. In practice, this assumption may not always be met; in such cases, spatial lag and spatial error models were found to be acceptable alternatives (Ward and Gleditsch 2007). Spatial regression analysis tools in GeoDa can facilitate running these two alternative models as discussed by Anselin (Anselin 1992; Anselin et al. 1996; Anselin et al. 2006). Theoretically, both models yield unbiased estimations; however, based

on the model performance parameters, the R -squared and Log likelihood values of a good model can be adopted (Anselin et al. 2006). Finally, OLS (Charlton et al. 2009) and GWR (Brunsdon et al. 1996) models were fitted to predict respiratory incidence over space with significant geo-collated explanatory variables as presented in Eqs. (1) and (2) respectively.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (1)$$

where Y is the dependent variable, the betas (β_0 to β_n) represent the corresponding coefficients of explanatory variables, X_1 to X_n represents the corresponding number of explanatory variables, and ϵ is the random error term of the residuals.

$$Y = \beta_0 + \sum_k \beta_k X_{ik} + \epsilon_i \quad (2)$$

where Y is the dependent variable at location i , β_0 represents the corresponding coefficients of explanatory variables, β_k measures the relationship between the independent variable X_k at location i , and ϵ is the random error term of the residual at location i . The corrected Akaike information criterion (AICc) was used for model comparison. The smaller the values of AICc were, the better the model performed (Li et al. 2017). Random patterns of over and/or under predictions of the model residuals indicate the goodness of fit of the models (Fotheringham et al. 2003). Therefore, standard deviation maps of the model residuals and the global Moran I test were used to assess the distribution of the residuals and validity of the models.

Results

Demographic profile of CRD incidence

Table 1 presents the summary statistics of the respiratory incidence in the entire study area. A total of 33,384 respiratory disease incidences were recorded for the years 2010 to 2014, representing an average annual rate of 49/10,000 population. The total asthma incidence is higher than the total of COPD as 19,909 (29/10,000) and 13,475 (20/10,000) respectively. Moreover, asthma incidence is higher among females (10,462) than males (9447) while COPD incidence is significantly more frequent in males (11,293) than females (2182). Both asthma and COPD incidences are higher among the elderly population (age > 60), followed by the age

Table 1 Summary statistics of CRDs incidence in Kandy district—2010 to 2014

	Asthma		COPD		Total	
	No. of incidents	Average annual rate ^a (per 10,000)	No. of incidents	Average annual rate (per 10,000)	No. of incidents	Average annual rate ^a (per 10,000)
Male	9535	14	11,205	16	20,740	30
Female	10,505	15	2139	3	12,644	18
Total	20,040	29	13,344	20	33,384	49

^a The ratio of the number of asthma/COPD cases occurred in a specific sex group to the 10,000 persons of the total population in a given year in Kandy district

groups 46–60, 31–45, 16–30, 6–15, and 0–5 years respectively (Table 2).

Robustness of the explanatory variables

Pearson product-moment correlation coefficient (r) values of the 20 different explanatory variables revealed a distinct level of association between the respiratory disease incidences as shown in Table 3. Apparently, the correlation coefficients exhibit a moderate association between the exploratory variables and asthma/COPD incidence. Further, all variables show a positive relationship between the explanatory variables and asthma/COPD except Te, Ea, and He.

Among meteorological variables, rainfall (asthma $r = 0.38$, COPD $r = 0.49$) and relative humidity (asthma $r = 0.54$, COPD $r = 0.46$) showed a moderately strong positive association while temperature (asthma $r = -0.48$, COPD $r = -0.41$) showed a negative association for both asthma and COPD hospitalizations. Moreover, rainfall and temperature showed a relatively stronger relationship with COPD than asthma hospitalizations. Particle matter (PM₁₀) showed a low to moderate, but still statistically significant positive association with asthma ($r = 0.34$) and COPD ($r = 0.25$) while followed by CO, NO₂, and SO₂ respectively. However, among other proxy parameters of ambient air pollution, road

density and proximity to roads with 0–200 m showed a relatively strong relationship with CRD. Moreover, the use of firewood (Fw) as a source of cooking fuel also showed reasonably positive correlation coefficients for both asthma and COPD. Therefore, the study was continued with the proxy parameters of ambient air pollution. The highest correlation coefficient for land cover type was recorded with the homestead. The elevation also showed a positive relationship with both asthma and COPD; however, it is stronger with COPD than asthma hospitalizations. Accordingly, people who lived in high altitude are more vulnerable for COPD. For the socioeconomic variables, the population with economically active and higher level of education indicated a negative association while economically inactive and low level of education categories showed a positive association.

Goodness of fit of the candidate explanatory variables

The goodness of fit of the candidate explanatory variables was first evaluated through the exploratory regression. The result of the exploratory regression reveals the best combinations (the passing models) of explanatory variables that satisfy all the fitted criteria in asthma and COPD prediction (Table 4). Accordingly, all 10 possible passing models returned with lower values of VIF (less

Table 2 CRDs incidence by age in Kandy district—2010 to 2014

Age (years)	0–5	AAR*	6–15	AAR*	16–30	AAR*	31–45	AAR*	46–60	AAR*	> 60	AAR*
Asthma	1570	28	1705	15	1745	11	2107	15	4518	35	8395	151
COPD	234	4	104	1	344	3	474	3	3147	25	9041	162
Total	1804	32	1809	16	2089	14	2581	18	7665	60	17,436	313

*AAR—average annual rate: the ratio of the number of asthma/COPD cases occurred in a specific age group to the 10,000 population of that age group in a given year in Kandy district)

Table 3 Pearson product-moment correlation coefficients (*r*) values of the individual explanatory variables

Exploratory variables		Correlation coefficient (<i>r</i>) ^a	
		Asthma	COPD
Meteorological factors	Rainfall (Rf)	0.38	0.49
	Relative humidity (RH)	0.54	0.46
	Temperature (Te)	− 0.43	− 0.47
Land use	Agricultural crops (Ag)	0.22	0.24
	Homestead (Hs)	0.43	0.43
	Other (Or)	0.18	0.20
Outdoor air pollutant (only for Kandy municipal area)	Carbon monoxide (CO)	0.17	0.19
	Nitrogen dioxide (NO ₂)	0.26	0.19
	Particle matter 10 (PM ₁₀)	0.34	0.25
	Sulfur dioxide (SO ₂)	0.24	0.12
Indoor air pollutant: use of firewood as a source of energy (Fw)		0.32	0.38
Road density (Rd)		0.47	0.47
Proximity to roads	0–200 m (B1)	0.46	0.38
	200–400 m (B2)	0.18	0.10
	> 400 m (B3)	0.10	0.02
Elevation (El)		0.29	0.46
Poverty	Economically active (Ea)	− 0.35	− 0.35
	Economically inactive (Ei)	0.34	0.35
Education level	Low level of education (Le)	0.33	0.28
	Higher level of education (He)	− 0.32	− 0.27

^a The correlation coefficients are significant at the 0.001 level (*p* value > 0.001)

than 7.5) indicating that none of the variables were redundant. Statistically, non-significant *p* values (larger than 0.1) of the Jarque–Bera statistics indicated that the residuals are normally distributed; thus, the model's predictions were free from bias and misspecification of

the model variables. The predictive power of the best asthma and COPD models fluctuated between 50 to 56% and 60 to 63% respectively. The lowest value of AICc and the highest value of adjusted *R*-squared indicated the most outstanding models (model 3 for asthma

Table 4 The candidate models of the exploratory regression for asthma and COPD prediction

Model no.		AdjR ²	AICc	JB	VIF	Model variables ^a
Asthma	Model 1	0.554	3111	0.11	1.34	RH + Hs + Fw − He + El
	Model 2	0.549	3125	0.17	1.16	RH + Hs + Fw − Ea + El
	Model 3	0.557	3104	0.11	1.73	RH + Hs + Fw − He − Ea + El
	Model 4	0.554	3113	0.13	1.43	RH + Hs + Fw + Le − Ea + El
	Model 5	0.502	3245	0.23	1.41	Rf + RH + Rd + Fw − He + El
COPD	Model 6	0.604	2812	0.22	1.14	RH + B1 + Fw − Ea + El
	Model 7	0.619	2766	0.12	2.15	RH − Te + B1 + Fw − He + El
	Model 8	0.614	2784	0.31	2.09	RH − Te + B1 + Fw − Ea + El
	Model 9	0.613	2787	0.12	1.42	RH + B1 + Fw + Le − Ea + El
	Model 10	0.626	2746	0.15	2.31	RH − Te + B1 + Fw − He + Ei + El

^a The “+” and “−” signs indicate positive and negative relationships respectively

and model 10 for COPD) among the candidate models. Five different explanatory variables (out of 16) (relative humidity (RH), rainfall (Rf), use of firewood (Fw), elevation (El), a higher level of education (He), and economically active (Ea)) play a key role in both asthma and COPD prediction. Moreover, relative humidity (RH), rainfall (Rf), use of firewood (Fw), and elevation (El) revealed a positive relationship while a higher level of education (He) and economically active (Ea) showed a significant negative relationship with asthma and COPD. Further, homestead (Hs), low level of education (Le), and road density acted as the significant supplementary predictors in asthma models, whereas proximity to road (B1), temperature (Te), low level of education (Le), and economically inactive (Ei) were the significant supplementary predictors in COPD models.

Predictive model summaries

Summary statistics for the most outstanding OLS models (model 3 for asthma and model 10 for COPD in Table 4) proved the overall significance of the predictive capabilities of the explanatory variables (Tables 5 and 6). There was no evidence of heteroscedasticity or multi-collinearity among variables since VIF values were less than the set threshold value of 7.5. The ANOVA returned a statistically significant joint *F*-value and joint Wald statistic returned statistically significant chi-squared values. With this, both OLS models proved to be statistically significant. However, statistically significant Moran's *I* statistics indicated that the models did not satisfactorily account for spatial dependency in both asthma and COPD. The two alternative models (spatial lag model and spatial error model) were applied in order to incorporate the violated OLS assumptions of spatial dependency. The spatial lag model determined slightly better performance with no significant error of spatial dependency (Moran's *I* index for asthma is 0.0016 and for COPD is -0.0029). The effect of the explanatory variables and other statistics remained virtually the same as the standard OLS models; thus, the detail statistics are not reported. However, the models still resumed statistically significant chi-squared value for the Koenker-BP statistics. It is an indication of spatial mobility. Thus, the relationship between some (or all) explanatory variables and COPD/asthma incidence varied across the study area.

The results of the tested combinations of GWR models revealed significantly better prediction ability

with elevated adjusted R^2 values 0.78 and 0.74 for asthma and COPD respectively. The difference between adjusted R^2 values from OLS to GWR implied that local model fitness was higher in predicting CRD over the study area. Moreover, a comparison of AICc values between OLS and GWR models indicated that GWR models were an improvement over OLS models (Table 7). The associated condition numbers for each model returned less than 30, indicating that the models were free from local multi-collinearity. Standard deviation maps of the model residuals clearly show that over- and under-predicted asthma and COPD incidences are randomly distributed over study area. Statistically, non-significant Moran's *I* statistics (Table 7) are further evidence for randomly distributed model residuals. This was good evidence for a properly specified and bias-free predictive model.

Spatial variation of CRDs prediction

The OLS regression models conclude that there is a firm but non-stationary predictive power of asthma/COPD incidence for all GN divisions. The results of the GWR models confirmed it with locally varied regression values as shown in Figs. 3a and 4a. At the local level, the predictive strength (R^2) of asthma and COPD fluctuated between 0.62–0.82 (Fig. 3a) and 0.58–0.78 (Fig. 4a) respectively. Generally, the West half of the study area shows a higher level of predictive capability than the East in both cases. This can be warranted by two reasons: first, the occurrence of CRDs is quite high in the western half so prediction capability may also high in the region. Second, it could be a MAUP raised issue because spatial extent of a GN division in the western half is very small compared with the extent of a GN division in the Eastern half (the average size of a GND in the western half is about 1 km² whereas the average size of a GND in the Eastern half is approximately 4 km²). Thus, the aggregation process yields more generalized and homogeneous data for the Eastern half which may result in a low prediction over space. Further, the resultant regression coefficients revealed strong local differences of association within each variable of the study area (Figs. 3c–h and 4c–h). In these figures, the areas with dark-red shade represent where that particular variable exhibited a strong influence on the corresponding CRD occurrence, while the dark-blue shade represents the areas which have a relatively low influence on the corresponding CRD occurrence. The

Table 5 OLS summary statistics for the best asthma model

Variables (x)	Coefficients value	Std. error	t statistic	p value	VIF
Intercept	− 54.57	3.01	− 18.11	< 0.001	–
Relative humidity (RH) (%)	11.59	0.65	17.92	< 0.001	0.38
Homestead (Hs) (km ²)	0.43	0.03	17.92	< 0.001	0.30
Use of firewood as a source of energy (Fw) (%)	0.53	0.05	10.91	< 0.001	0.24
Higher level of education (He) (%)	− 0.59	0.12	− 4.76	< 0.001	− 0.12
Economically active (Ea) (%)	− 0.39	0.13	− 3.04	0.002	− 0.07
Elevation (El) (m)	0.95	0.06	15.11	< 0.001	0.29
OLS diagnostics parameters	value			p value	
Joint F-statistic	250.25			< 0.001	
Joint Wald statistic	1487.90			< 0.001	
Koenker-BP statistic	97.39			< 0.001	
Jarque–Bera statistic	4.35			0.114	
Moran's I statistics for model residuals	0.406			< 0.001	

Each coefficient value represents the change in asthma incidence relative to a 1-unit change in *X* variables. For example, if relative humidity differs by 1 unit, it would expect asthma incidence to differ by approximately 11.59

variable “road proximity” showed a positive association with asthma and COPD in the entire study area (Figs. 3d and 4d respectively); however, the influence is stronger in the southern half than the northern half of the study area for asthma, while the western half of the study area displays a more significant influence of COPD. This result further confirmed that the risk of asthma and COPD was increased among those who are living in close proximity to roads. However, the level of impact

was inconsistent with the influence of other risk factors over space. Conversely, the local coefficients of the remaining variables revealed a combination of positive and negative associations over the study area. The positive association means that adding one increment of a particular variable leads to a proportional increase in the corresponding CRD. On the other hand, negative association implies that inverse association with the corresponding CRD. Accordingly, a higher level of education

Table 6 OLS summary statistics for the best COPD model

Variables (x)	Coefficients value	Std. error	t statistic	p value	VIF
Intercept	− 19.40	4.77	− 4.07	< 0.001	–
Relative humidity (RH) (%)	4.87	0.77	6.36	< 0.001	0.17
Temperature (T) (°C)	− 3.12	0.54	− 5.77	< 0.001	− 0.15
Proximity to roads 0–200 m (B1) (m)	0.29	0.02	11.77	< 0.001	0.23
Use of firewood as a source of energy (Fw) (%)	0.57	0.04	12.83	< 0.001	0.27
Higher level of education (He) (%)	− 1.02	0.10	− 10.56	< 0.001	− 0.22
Economically inactive (Ei) (%)	0.49	0.10	4.69	< 0.001	0.10
Elevation (El) (m)	1.34	0.05	24.66	< 0.001	0.44
OLS diagnostics parameters	value			p value	
Joint F-statistic	285.68			< 0.001	
Joint Wald statistic	1795.37			< 0.001	
Koenker-BP statistic	101.57			< 0.001	
Jarque–Bera statistic	3.78			0.151	
Moran's I statistics for model residuals	0.383			< 0.001	

Each coefficient value represents the change in COPD incidence relative to a 1-unit change in *X* variables. For example, if relative humidity differs by 1 unit, it would expect COPD incidence to differ by approximately 4.87

Table 7 Model fitness comparison

Fitness parameters	Asthma		COPD	
	OLS	GWR	OLS	GWR
AICc	3262	3159	2776	2671
R^2	0.58	0.79	0.63	0.75
Adjusted R^2	0.57	0.78	0.62	0.74
Moran's I statistics for model residuals	0.337	0.005	0.274	0.009
Moran's I statistics p value	< 0.001	0.467	< 0.001	0.226

Both OLS and GWR models run for entire GN divisions (1188) in Kandy district

showed a positive relationship in the southwest region while other areas showed an inverse relationship for both asthma and CRD (Figs. 3c and 4c). In the case of road density, the southwest region showed an inverse relationship, where other areas showed a positive association with CRD. At the same time, rainfall and elevation showed positive association in the southwest region, indicating that the relative importance and predictive power of an explanatory variable can vary with location. According to these results, the Southwest region revealed a distinct relationship with the examined risk factors. The risk factors elevation and rainfall may be the most influencing factors in predicting CRD in the Southwest region. Ultimately, these results revealed that CRD risk factors are not globally significant. Instead, they vary, showing a positive association with risk factors in some areas, and negative relationships in other areas. As shown in the standard residual maps, asthma incidences are over-predicted around Delthota, Paradeka, Medagoda, Mampitiya, and Nawalapitiya areas where as COPD incidences are over-predicted around Gampola, Nawalapitiya, Delthota, Peradeniya Mahanuwara, and Mahadoraliyadda areas. Both these over- and under-predicted disease incidences locations further imply that the local variation of prediction ability and areas need to be further investigations.

Discussion

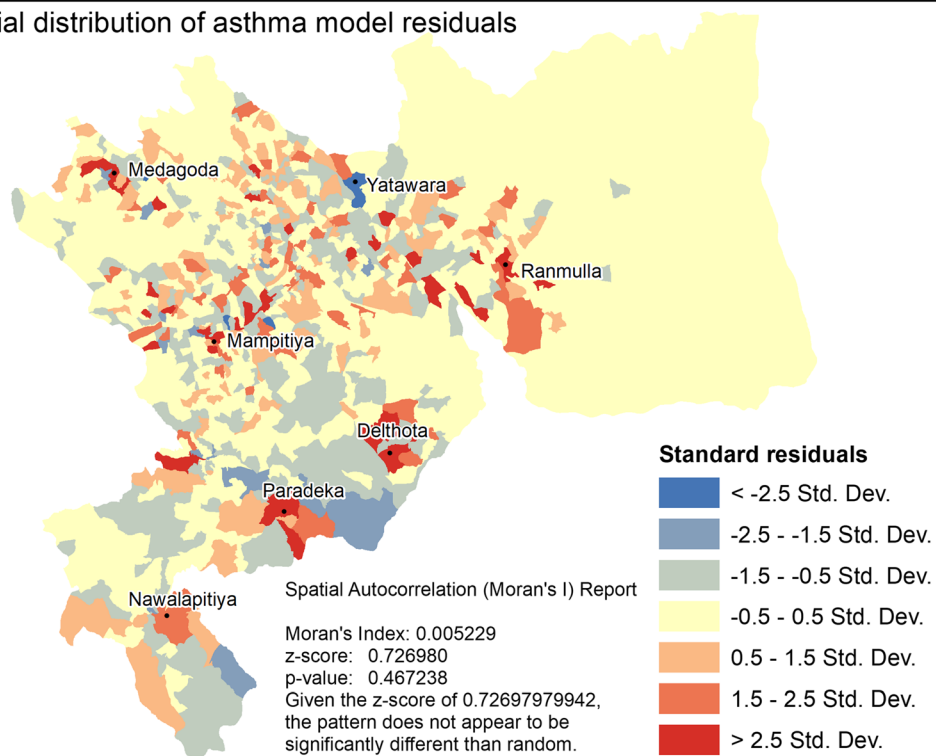
The result of Pearson product-moment correlation coefficients for each individual explanatory variable generally revealed a low strength of association with CRD morbidity. There are two primary explanations. For example, outdoor air pollution showed very weak association with CRD yet several comparative case studies were

done in Sri Lanka (Holgate et al. 1999; Nandasena et al. 2010) have shown a strong positive association. Basically, this would be a consequence of the lack of appropriate air pollution data in Kandy to gauge the actual magnitude of the problem. Air pollution studies in Kandy are mostly limited to few sample locations in the KMC area. For example, a study done in 2002 revealed that air pollution levels in Kandy exceeded the national standards for Sri Lanka; however, the sample size of the study was limited to two locations in KMC area (Abeyratne and Ileperuma 2002). In 2008, suspended particle matter pollution was monitored in 25 locations in KMC and surrounding areas (Elangasinghe and Shanthini 2009). Eight-hour exposure levels to PM₁₀ at 20 locations in KMC over a period of 9 months measured by Wickramasinghe et al. (2011) revealed that the concentration levels are high along roadways. Spatial variation of air pollution and level of exposure were assessed in 2011; however, this study also revealed that air pollution studies in Kandy are limited to a few sampling locations within KMC area (Premasiri et al. 2012). Although these preliminary studies stressed the need for long-term monitoring of air pollution using fixed monitoring stations, there is currently no mechanism to monitor air pollution in Kandy (Abeyratne and Ileperuma 2002; Premasiri et al. 2012).

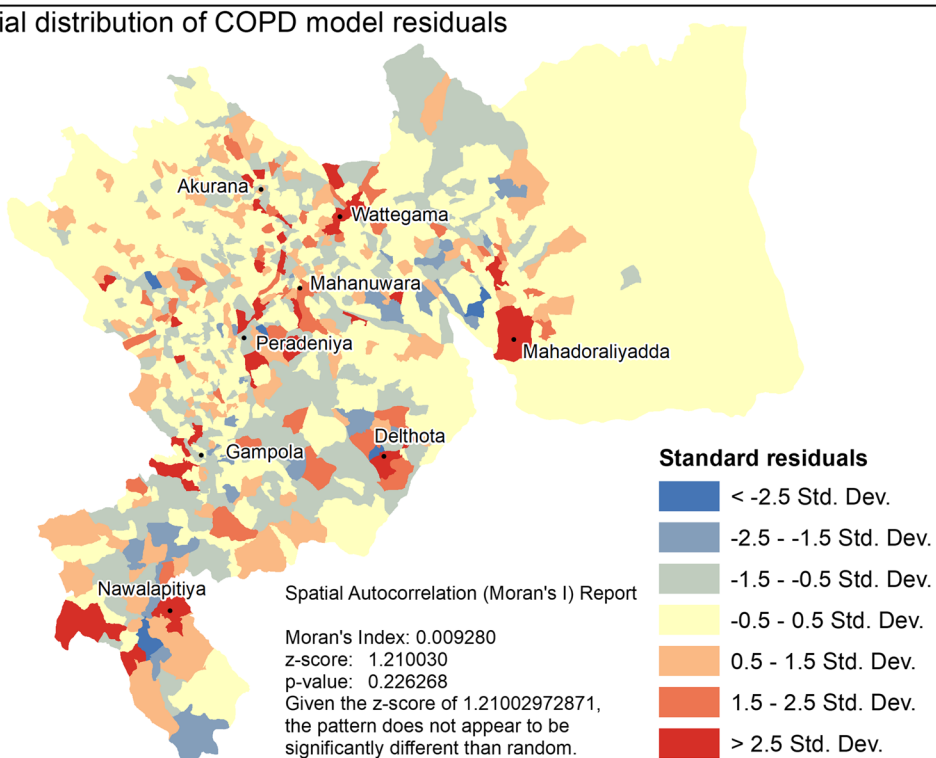
On the other hand, the individual level variable analysis is insufficient for CRD morbidity prediction. Cohen et al. stressed the same that a range of low to moderate correlation coefficients values is quite normal in any field that attempts to predict human processes due to

Fig. 3 Standard deviation maps of the model residuals. Legend: The hot (red)/cold (blue) rendering color scheme represents the patterns of over and under predictions of the model residuals respectively. The map shows the randomly distributed asthma model residuals (a) and depicts the randomly distributed COPD model residuals (b)

(a) Spatial distribution of asthma model residuals



(b) Spatial distribution of COPD model residuals



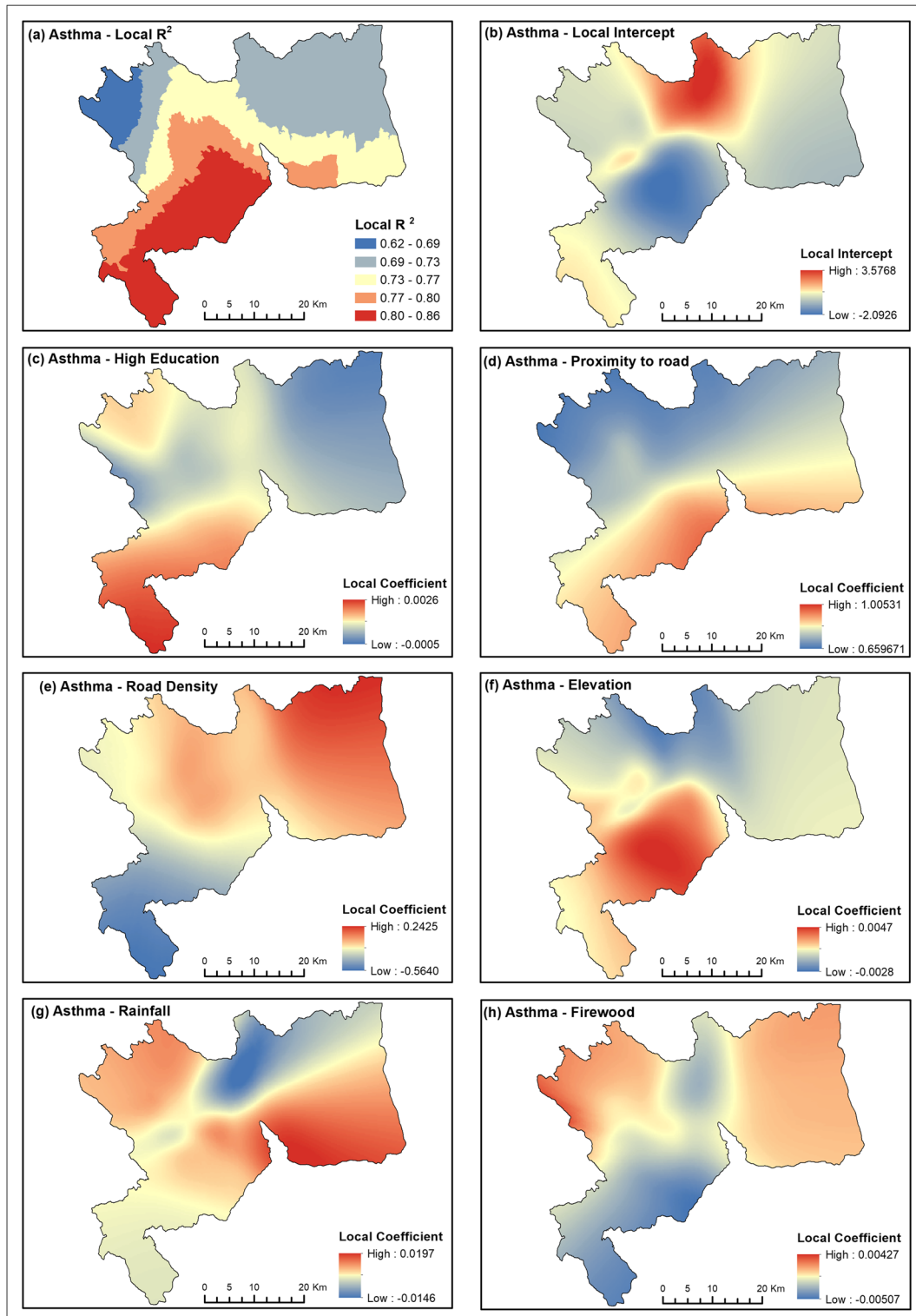


Fig. 4 Local r^2 values and coefficient raster surfaces for asthma model. Legend: The map shows the spatial variation of local R -squared values of Asthma model (a) and indicates the local intercept estimated by GWR (b). The maps c–h indicate how asthma

incidence would change for each unit change of the variables high education, proximity to the road, road density, elevation rainfall, and firewood respectively

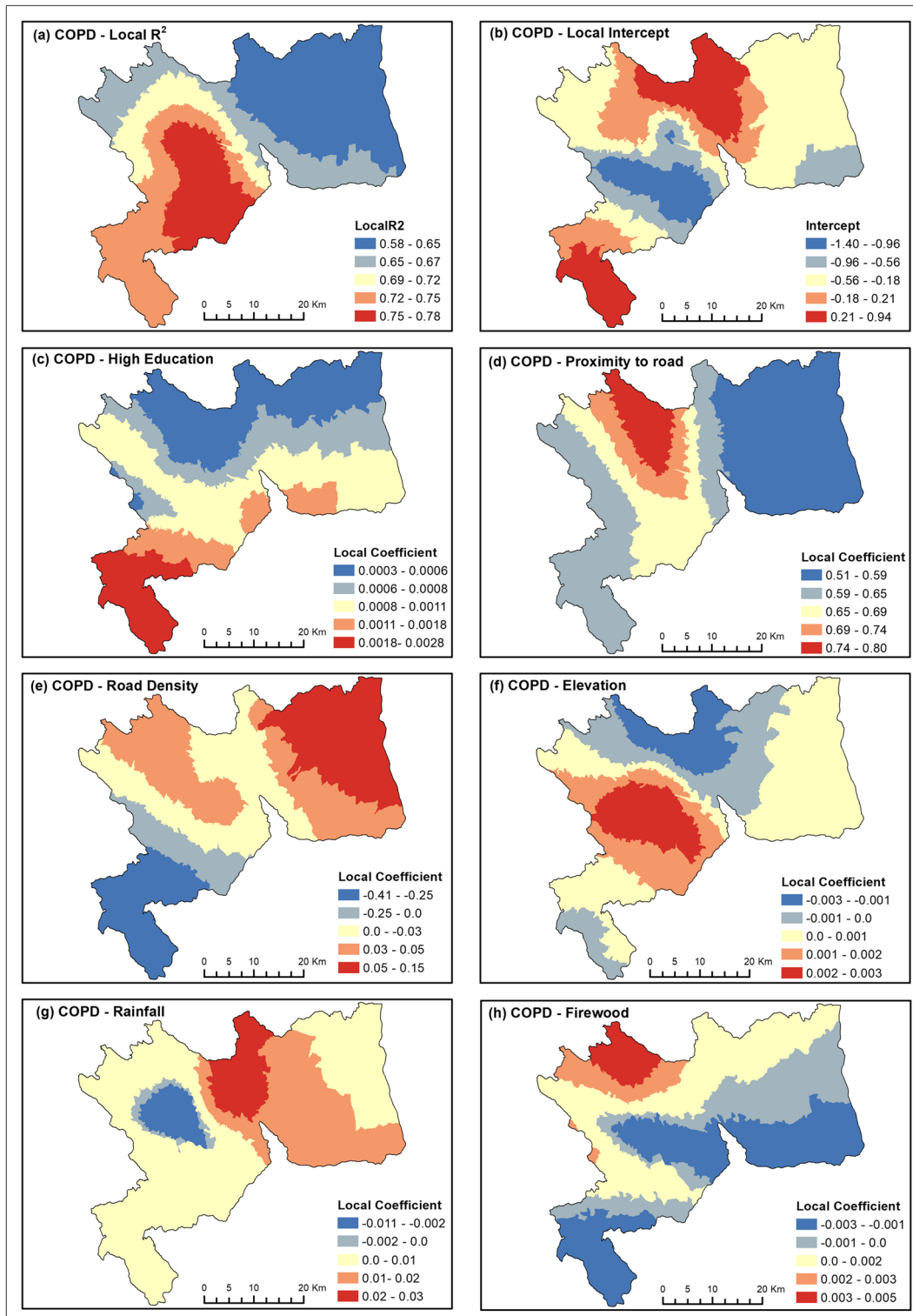


Fig. 5 Local r^2 values and coefficient raster surfaces for COPD model. Legend: The map shows the spatial variation of local R -squared values of COPD model (a) and indicates the local intercept estimated by GWR (b). The maps c-h indicate how asthma

incidence would change for each unit change of the variables high education, proximity to the road, road density, elevation rainfall, and firewood respectively

the complexity and multiplicity of influencing variables; thus, individual variables will not be able to represent their separate influences accurately (Cohen et al. 2013). Moreover, spatially varied coefficients of the local models (Figs. 4 and 5) further revealed that there is a locally varying association between CRDs and each variable; thus, a strong global correlation coefficient cannot be expected. However, the quantified directions of the association between the explanatory variables and CRDs incidence are compatible with existing knowledge. For example, a recent study done in Serbia revealed that there is a significant negative association between temperature and asthma patients (Stevanović et al. 2016). The perception is further confirmed by numerous studies (Seys et al. 2013; Qiu et al. 2015; Guo et al. 2012; Lim et al. 1991). Similarly, there is evidence that variables such as ambient air pollutants (Delamater et al. 2012; Guo et al. 2012), rainfall (Xirasagar et al. 2006; Silva et al. 2011), relative humidity (Guo et al. 2012), road density and proximity (Hansell et al. 2014; Gasana et al. 2012; Lindgren et al. 2009), poverty (Burney et al. 2014), and exposure to firewood combustion (Pan et al. 2016) are positively associated with CRD. Although such etiological studies of CRD have not been comprehensively elucidated in Sri Lanka, a few cohort level studies have confirmed a positive association between ambient air pollutants (Uragoda 1980; Thishan Dharshana and Coowanitwong 2008), exposure to firewood combustion (Nandasena et al. 2012; Nandasena et al. 2013; Phillips et al. 2016), and CRD. Moreover, there is a general perception that Kandy is vulnerable for asthma/COPD due to the high road traffic and the results of this study confirmed it further. However, more detailed epidemiological studies are needed to assess the relationship between aetiological factors and CRD exacerbation in Kandy.

Consistent with the spatial relationship analysis, the results of the exploratory regression and the OLS regression models helped to identify possible risk factors with a better prediction capability. Generally, OLS regression is considered as a good starting point for spatial relationship assessment (Ahmad et al. 2015). Accordingly, OLS regression models revealed that relative humidity, rainfall, proximity to the road (0–200 m), road density, use of firewood as a source of fuel, and elevation are the significant positive predictors of asthma and COPD in Kandy district. Temperature, a higher level of education, and economically active variables are

significant negative predictors of CRDs. Relative humidity is the most influencing risk factor for CRD prevalence in Kandy. This is consistent with the findings of the reported epidemiological studies in the literature which suggest higher respiratory infections in relatively high humidity (Arundel et al. 1986; Delamater et al. 2012; Ayres-Sampaio et al. 2014). The influence of RH on mounting the availability of dust mites and their allergens in the environment (Arlan et al. 2001; Arundel et al. 1986) and air pollutants such as PM₁₀ (Devries et al. 2016) could be the reason; however, this should be further verified in detail through proper studies. Consistent with the existing knowledge, CRD prevalence in Kandy district was inversely related to temperature as the coldest areas reported high CRD incidence. Thus, the contribution of meteorological factors in CRDs prediction is significant in Kandy district. Moreover, the contribution of three proxy parameters of air pollutant (proximity to roads (0–200 m), road density, and the use of firewood as a source of fuel) emphasizes the necessity of regular monitoring of indoor and outdoor air pollutant parameters in terms of CRDs prediction in Kandy. Inversely associated socioeconomic predictors confirmed the relationship between poverty and CRD. An economically active and highly educated population is a good indicator of a high standard of living; thus, these results reveal that high standard of living reduced the vulnerability of CRD prevalence in Kandy. Therefore, these findings are important to health care policy makers to initiate appropriate preventive measures for CRD in Kandy district.

The resultant GWR models evidence the presence of spatial heterogeneity within the study area while predicting CRD. The relationships between explanatory variables and CRD incidences are not constant and vary spatially as shown by r^2 values. The direction and strength of the association between explanatory variables and CRD incidence differ in different spatial settings as shown by spatially varying coefficients. This indicates that the prevalence of CRD in a given location is directly related to its physical and social environmental conditions. Kandy is relatively diverse in population, socioeconomic conditions, topography, climate, and environmental conditions. Hence, the examined explanatory variables on CRD cannot be directly generalized for Kandy as a whole, or in the other spatial location even within a national or regional level. In order to assess CRD prevalence in Kandy district, more precise spatial research is essential which can distinguish local

variation. GWR models are, therefore, more important in asthma and COPD surveillance in a district like Kandy. However, spatially and temporally precise data on risk factors that can depict the local variations adequately is needed in local regression models; otherwise, local multi-collinearity, MAUP, might be a problem.

Further, this study discloses the use of GIS capabilities in CRD surveillance, especially to overcome the rudimentary level of available spatial data, incompatibility issues. Geocoding capabilities permit generation of a finer resolution analysis of CRD incidence over space and were able to minimize MAUP in assessing the relationship between CRDs and their risk factors in Kandy district. Moreover, different GIS procedures such as Kriging, IDW, and zonal statistics facilitate making compatible CRD risk data from different sources at different scales to each other. The spatial regression models prove the analytical capabilities that can help to clarify the geographical variability and association of CRD determinants.

However, this study had a few limitations as well. First, this was an ecological study; the models used aggregated data at the GN division level. Thus, individual level exposure to the predictors and impact of personal mobility were underestimated. Therefore, contributing factors responsible for the individual level hospitalization are not predicted by the models applied. Secondly, the study only considered indoor hospitalization at the government hospitals; neither outpatient visits nor private health care units considered. Thus, the actual amount of morbidity is likely underestimated. All of these limitations contribute to modeling errors.

Conversely, the capabilities of GIS and spatial analysis techniques used in this study allow health care policy makers to understand the spatial association of CRD risk factors in Kandy district. Accordingly, they can develop specific mitigation measures that are contingent on the location and nature of risk levels. Moreover, the methodological and theoretical implications raised in this study are important for a country like Sri Lanka, not only asthma and COPD surveillance in Kandy district but more generally for surveillance of any disease that influenced by different socioeconomic and environmental risk factors. Theoretically, the highlighted concepts of spatial non-stationary and models of local regression, etc. are valuable for developing a surveillance strategy for diseases like asthma and COPD, with risk factors that are spatially varied and inextricable. However, continuous experiments are necessary to

investigate the localized impact of risk factors. This study further illustrates the inextricable linkage among the CRD determinants and emphasizes the need for more sophisticated techniques and data that can integrate both spatial and temporal aspects of health-related studies. Further studies are needed that assess the spatiotemporal association between CRD and risk factors to improve the results of this study.

Conclusions and recommendations

Individual level risk factor analysis is not appropriate for CRD morbidity prediction. It is most unlikely to have a single coefficient per explanatory variable. At the same time, the relationship between CRD occurrence and their risk factors is not static across a study area. CRD prediction over space can be established through analysis of a combination of environmental, climatic, and socioeconomic risk factors such as relative humidity, rainfall, proximity to the road (0–200 m), road density, the use of firewood as a source of fuel, a higher level of education, economically active population, and topography (elevation). Both global and local regression models are found to be important in assessing spatially varied relationships of CRD, and they provide necessary insight into the geospatial behavior of environmental and socioeconomic predictors of CRD. Finally, geospatial analytical capabilities used in this study can be used as a surveillance strategy in developing countries, not only for CRD prediction but also for other disease surveillance in different spatial settings.

Acknowledgments The study was conducted with the permission of the Provincial Director of Health Services (PDHS), Central Province and Regional Director of Health Services (RDHS), Kandy, Sri Lanka. The support given by these officials is greatly appreciated. The authors express sincere thanks to the corresponding staff of each health care institution for their support. All research assistants are also gratefully acknowledged for their involvement and support in the data collection process.

Contributions RMKK involved in research design, data collection, analysis, and drafting the manuscript; NKT contributed in conceptualization, research design, modifying the draft, and developing the final version of the manuscript. Both authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Written informed approvals were obtained from the relevant authorities of each health care institution. Further, governmental and institutional regulations concerning the ethical use of patient's records were followed during the study. All study procedures that are relevant to collecting and analyzing individual patient's records were approved by the ethics review committee of the Faculty of Allied Health Sciences, University of Peradeniya, Sri Lanka.

References

- Abeyratne, V., & Ileperuma, O. (2002). Monitoring air pollution levels in Kandy using passive and active gas sampling techniques. *Ceylon Journal of Science, Physical Sciences*, 9(1), 54–61.
- Abramson, M., Koplin, J., Hoy, R., & Dharmage, S. (2015). Population-wide preventive interventions for reducing the burden of chronic respiratory disease. *The International Journal of Tuberculosis and Lung Disease*, 19, 1007–1018.
- Ahmad, S. S., Aziz, N., Butt, A., Shabbir, R., & Erum, S. (2015). Spatio-temporal surveillance of water based infectious disease (malaria) in Rawalpindi, Pakistan using geostatistical modeling techniques. *Environmental Monitoring and Assessment*, 187, 555.
- Ait-Khaled, N., Enarson, D., & Bousquet, J. (2001). Chronic respiratory diseases in developing countries: the burden and strategies for prevention and management. *Bulletin of the World Health Organization*, 79, 971–979.
- Ali-Akbarpour, M., Mohammadbeigi, A., Tabatabaee, S. H. R., & Hatam, G. (2012). Spatial analysis of eco-environmental risk factors of cutaneous leishmaniasis in southern Iran. *Journal of Cutaneous and Aesthetic Surgery*, 5, 30.
- Anselin, L. (1992). *SpaceStat tutorial: a workbook for using SpaceStat in the analysis of spatial data* (p. 263). Urbana-Champaign: University of Illinois.
- Anselin, L. & Rey, S. (2005). GeoDa 1.6. 7 9 March 2015. Exploring spatial data with GeoDa TM: a workbook. Spatial Analysis Laboratory, University of Illinois, Urbana-Champaign.
- Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26, 77–104.
- Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: an introduction to spatial data analysis. *Geographical Analysis*, 38, 5–22.
- Arlian, L. G., Neal, J. S., Morgan, M. S., Vyszynski-Moher, D. L., Rapp, C. M., & Alexander, A. K. (2001). Reducing relative humidity is a practical way to control dust mites and their allergens in homes in temperate climates. *Journal of Allergy and Clinical Immunology*, 107, 99–104.
- Arundel, A. V., Sterling, E. M., Biggin, J. H., & Sterling, T. D. (1986). Indirect health effects of relative humidity in indoor environments. *Environmental Health Perspectives*, 65, 351.
- Ayres-Sampaio, D., Teodoro, A. C., Sillero, N., Santos, C., Fonseca, J., & Freitas, A. (2014). An investigation of the environmental determinants of asthma hospitalizations: an applied spatial approach. *Applied Geography*, 47, 10–19.
- Bacon, S. L., Bouchard, A., Loucks, E. B., & Lavoie, K. L. (2009). Individual-level socioeconomic status is associated with worse asthma morbidity in patients with asthma. *Respiratory Research*, 10, 1.
- Bascuñán, M. M. & Quezada, C. R. (2016). Geographically weighted regression for modelling the accessibility to the public hospital network in Concepción Metropolitan Area, Chile. *Geospatial health*, 11.
- Berkelman, R. L., Sullivan, P. & Buehler, J. W. (2009). Public health surveillance. *Oxford textbook of public health, volume 2: the methods of public health*, 699–715.
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28, 281–298.
- Burnett, R. T., Dales, R. E., Raizenne, M. E., Krewski, D., Summers, P. W., Roberts, G. R., Raadyoung, M., Dann, T., & Brook, J. (1994). Effects of low ambient levels of ozone and sulfates on the frequency of respiratory admissions to Ontario hospitals. *Environmental Research*, 65, 172–194.
- Burney, P. (2017). Chronic respiratory disease—the acceptable epidemic? *Clinical Medicine*, 17, 29–32.
- Burney, P., Kato, B., Janson, C., Mannino, D., Studnicka, M., Tan, W., Bateman, E., Koçabas, A., Vollmer, W. M. & Gislason, T. 2014. Chronic obstructive pulmonary disease mortality and prevalence: the associations with smoking and poverty: a BOLD analysis—authors' reply. *Thorax*, thoraxjnl-2014-205474.
- Charlton, M., Fotheringham, S. & Brunsdon, C. (2009). *Geographically weighted regression*. White paper. National Centre for Geocomputation. National University of Ireland Maynooth.
- Chavarria, J. (2001). Asthma admissions and weather conditions in Costa Rica. *Archives of Disease in Childhood*, 84, 514–515.
- Choi, B. C. (2012). The past, present, and future of public health surveillance. *Scientifica*, 2012.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*, Routledge.
- Comber, A. J., Brunsdon, C., & Radburn, R. (2011). A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. *International Journal of Health Geographics*, 10, 44.
- Corvalán, C., Nurminen, M. & Pastides, H. (1997). Linkage methods for environment and health analysis. Technical guidelines. Geneva: WHO Office of Global and Integrated Environmental Health.
- Cote, T. R., Stroup, D. F., Dwyer, D. M., Horan, J. M., & Peterson, D. E. (1993). Chronic obstructive pulmonary disease mortality: a role for altitude. *Chest*, 103, 1194–1197.
- Crawford, C. A. G. & Young, L. J. (2004). *Combining incompatible spatial data: an introductory overview of statistical issues and methods*. Workshop for the SAMSI Program on Multiscale Model Development and Control Design, Research Triangle Park, NC.
- Cruz, A. A., Bousquet, J. & Khaltayev, N. (2007). *Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach*, World Health Organization.

- Curtis, L. M., Wolf, M. S., Weiss, K. B., & Grammer, L. C. (2012). The impact of health literacy and socioeconomic status on asthma disparities. *Journal of Asthma*, 49, 178–183.
- de Sousa Zanotti Stagliorio Coêlho, M., Luiz Teixeira Gonçalves, F., & do Rosário Dias de Oliveira Latorre, M. (2010). Statistical analysis aiming at predicting respiratory tract disease hospital admissions from environmental variables in the city of São Paulo. *Journal of Environmental and Public Health*, 2010, 209270. <https://doi.org/10.1155/2010/209270>.
- Delamater, P. L., Finley, A. O., & Banerjee, S. (2012). An analysis of asthma hospitalizations, air pollution, and weather conditions in Los Angeles County, California. *Science of the Total Environment*, 425, 110–118.
- Desktop, E. A. (2011). *Release 10*. Redlands, CA: Environmental Systems Research Institute.
- Devries, R., Kriebel, D., & Sama, S. (2016). Low level air pollution and exacerbation of existing copd: a case crossover analysis. *Environmental Health*, 15, 98.
- Ding, P.-H., Wang, G.-S., Guo, Y.-L., Chang, S.-C., & Wan, G.-H. (2017). Urban air pollution and meteorological factors affect emergency department visits of elderly patients with chronic obstructive pulmonary disease in Taiwan. *Environmental Pollution*, 224, 751–758.
- Donaldson, G., Seemungal, T., Jeffries, D., & Wedzicha, J. (1999). Effect of temperature on lung function and symptoms in chronic obstructive pulmonary disease. *European Respiratory Journal*, 13, 844–849.
- Elangasinghe, M. & Shanthini, R. (2009). Determination of atmospheric PM 10 concentration in Kandy in relation to traffic intensity. *Journal of the National Science Foundation of Sri Lanka*, 36.
- Fotheringham, A. S., Brunson, C. & Charlton, M. (2003). *Geographically weighted regression: the analysis of spatially varying relationships*, John Wiley & Sons.
- Gamage, U. (2004). Health information management practice in Sri Lanka. *Health Information Management*, 33, 98–99.
- Gasana, J., Dillikar, D., Mendy, A., Forno, E., & Vieira, E. R. (2012). Motor vehicle air pollution and asthma in children: a meta-analysis. *Environmental Research*, 117, 36–45.
- Gorai, A. K., Tulari, F., & Tchounwou, P. B. (2014). A GIS based approach for assessing the association between air pollution and asthma in New York State, USA. *International Journal of Environmental Research and Public Health*, 11, 4845–4869.
- Guo, Y., Jiang, F., Peng, L., Zhang, J., Geng, F., Xu, J., Zhen, C., Shen, X., & Tong, S. (2012). The association between cold spells and pediatric outpatient visits for asthma in Shanghai, China. *PLoS One*, 7, e42232.
- Hales, S., Lewis, S., Slater, T., Crane, J., & Pearce, N. (1998). Prevalence of adult asthma symptoms in relation to climate in New Zealand. *Environmental Health Perspectives*, 106, 607–610.
- Hansell, A. L., Rose, N., Cowie, C. T., Belousova, E. G., Bakolis, I., Ng, K., Toelle, B. G., Marks, G. B., & Team, C. A. P. S. (2014). Weighted road density and allergic disease in children at high risk of developing asthma. *PLoS One*, 9, e98978.
- Haque, U., Scott, L. M., Hashizume, M., Fisher, E., Haque, R., Yamamoto, T., & Glass, G. E. (2012). Modelling malaria treatment practices in Bangladesh using spatial statistics. *Malaria Journal*, 11, 63.
- Hervás, D., Utrera, J., Hervás-Masip, J., Hervás, J., & García-Marcos, L. (2015). Can meteorological factors forecast asthma exacerbation in a paediatric population? *Allergologia et Immunopathologia*, 43, 32–36.
- Holgate, S. T., Koren, H. S., Samet, J. M. & Maynard, R. L. (1999). *Air pollution and health*, Academic Press.
- Jagger, P., & Shively, G. (2014). Land use change, fuel use and respiratory health in Uganda. *Energy Policy*, 67, 713–726.
- Jarque, C. M. (2011). *Jarque-Bera test. International encyclopedia of statistical science*. Springer.
- Jassal, M. S. (2015). Pediatric asthma and ambient pollutant levels in industrializing nations. *International Health*, 7, 7–15.
- Jerrett, M., Gale, S., & Kontgis, C. (2010). Spatial modeling in environmental and public health research. *International Journal of Environmental Research and Public Health*, 7, 1302–1329.
- Johnston, K., Ver Hoef, J. M., Krivoruchko, K. & Lucas, N. (2001). *Using ArcGIS geostatistical analyst*, Esri Redlands.
- Kala, A. K., Tiwari, C., Mikler, A. R., & Atkinson, S. F. (2017). A comparison of least squares regression and geographically weighted regression modeling of West Nile virus risk based on environmental parameters. *PeerJ*, 5, e3070.
- Kariyawasam, N., Weerasekera, V., Dayaratne, M., Hewapathirana, R., Karunapema, R. & Bandara, I. (2011). eIMMR: the future of health statistics in Sri Lanka. *Sri Lanka Journal of Bio-Medical Informatics*, 1.
- Kauhl, B., Schweikart, J., Krafft, T., Keste, A., & Moskwyn, M. (2016). Do the risk factors for type 2 diabetes mellitus vary by location? A spatial analysis of health insurance claims in Northeastern Germany using kernel density estimation and geographically weighted regression. *International Journal of Health Geographics*, 15, 38.
- Keet, C. A., Matsui, E., McCormack, M. C., & Peng, R. (2017). Urban residence, neighborhood poverty, race/ethnicity and asthma morbidity among children in Medicaid. *Journal of Allergy and Clinical Immunology*, 140, 822–827.
- Kurmi, O. P., Semple, S., Simkhada, P., Smith, W. C. S., & Ayres, J. G. (2010). COPD and chronic bronchitis risk of indoor air pollution from solid fuel: a systematic review and meta-analysis. *Thorax*, 65, 221–228.
- Lee, J. & Wong, D. W. 2001. *Statistical analysis with ArcView GIS*, John Wiley & Sons.
- Li, C., Li, F., Wu, Z., & Cheng, J. (2017). Exploring spatially varying and scale-dependent relationships between soil contamination and landscape patterns using geographically weighted regression. *Applied Geography*, 82, 101–114.
- Li, S., Batterman, S., Wasilevich, E., Elasaad, H., Wahl, R., & Mukherjee, B. (2011). Asthma exacerbation and proximity of residence to major roads: a population-based matched case-control study among the pediatric Medicaid population in Detroit, Michigan. *Environmental Health*, 10, 34.
- Lim, T., Looi, H., & Harun, K. (1991). Asthma and climatic conditions—experience from Kuantan, Malaysia. *The Medical Journal of Malaysia*, 46, 230–234.
- Lindgren, A., Strohm, E., Montnemery, P., Nihlén, U., Jakobsson, K., & Axmon, A. (2009). Traffic-related air pollution associated with prevalence of asthma and COPD/chronic bronchitis. A cross-sectional study in Southern Sweden. *International journal of health geographics*, 8, 2.

- Matthews, S. A., & Yang, T.-C. (2012). Mapping the results of local statistics: using geographically weighted regression. *Demographic Research*, 26, 151–166.
- Mccoy, J., Johnston, K. & Institute, E. S. R. (2001). Using ArcGIS spatial analyst: GIS by ESRI, Environmental Systems Research Institute.
- Mireku, N., Wang, Y., Ager, J., Reddy, R. C., & Baptist, A. P. (2009). Changes in weather and the effects on pediatric asthma exacerbations. *Annals of Allergy, Asthma & Immunology*, 103, 220–224.
- Moore, E., Chatzidiakou, L., Kuku, M.-O., Jones, R. L., Smeeth, L., Beevers, S., Kelly, F. J., Barratt, B., & Quint, J. K. (2016). Global associations between air pollutants and chronic obstructive pulmonary disease hospitalizations. A systematic review. *Annals of the American Thoracic Society*, 13, 1814–1827.
- Mortimer, K., Cuevas, L., Squire, B., Thomson, R. & Tolhurst, R. (2015). Improving access to effective care for people with chronic respiratory symptoms in low and middle income countries. *BMC Proceedings*. BioMed Central, S3.
- Mukaka, M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24, 69–71.
- Muralidharan, V., Salvi, S., Londhe, J., Sambhudas, S., Ingole, V., Brashier, B., Barnes, P., Green, M., & Juvekar, S. (2014). Use of geographical information system (GIS) in prevalence studies. *European Respiratory Journal*, 44, P4156.
- Nandasena, Y. L. S., Wickremasinghe, A. R., & Sathiakumar, N. (2010). Air pollution and health in Sri Lanka: a review of epidemiologic studies. *BMC Public Health*, 10, 1.
- Nandasena, S., Wickremasinghe, A. R., & Sathiakumar, N. (2012). Respiratory health status of children from two different air pollution exposure settings of Sri Lanka: a cross-sectional study. *American Journal of Industrial Medicine*, 55, 1137–1145.
- Nandasena, S., Wickremasinghe, A. R., & Sathiakumar, N. (2013). Indoor air pollution and respiratory health of children in the developing world. *World Journal of Clinical Pediatrics*, 2, 6–15.
- Nkeki, F. N. & Osirike, A. B. (2013). GIS-based local spatial statistical model of cholera occurrence: using geographically weighted regression. *Journal of Geographic Information System*, 2013.
- NWS&DB (2014). Environmental Monitoring Report of Kandy city wastewater management project In: *National water supply and drainage board*, K. (ed.).
- Openshaw, S. (1984). *The modifiable areal unit problem*, Geo Books.
- Ore, T., & Ireland, P. (2015). Chronic obstructive pulmonary disease hospitalisations and mortality in Victoria: analysis of variations by socioeconomic status. *Australian and New Zealand Journal of Public Health*, 39, 243–249.
- Pan, D., Qian, Y., Wang, C., Guo, Y., Bian, Z., Xie, K., Chen, L., Zhang, Y., & Li, Q. (2016). Interaction between smoking and indoor air pollution on chronic obstructive pulmonary diseases. *Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi*, 37, 1444.
- Park, Y. M., & Kim, Y. (2014). A spatially filtered multilevel model to account for spatial dependency: application to self-rated health status in South Korea. *International Journal of Health Geographics*, 13, 1.
- Peled, R., Reuveni, H., Pliskin, J. S., Benenson, I., Hatna, E., & Tal, A. (2006). Defining localities of inadequate treatment for childhood asthma: a GIS approach. *International Journal of Health Geographics*, 5, 3.
- Phillips, M. J., Smith, E. A., Mosquin, P. L., Chartier, R., Nandasena, S., Bronstein, K., Elledge, M. F., Thornburg, V., Thornburg, J., & Brown, L. M. (2016). Sri Lanka pilot study to examine respiratory health effects and personal PM2.5 exposures from cooking indoors. *International journal of environmental research and public health*, 13, –791.
- Premasiri, H. D. S., Samarasinghe, I. H. K. & Lakmali, K. M. N. (2012). Population exposure risk assessment to air pollution in Kandy city area. NBRO symposium on Sustainable Disaster Resilient Technologies.
- Prescott, E., & Vestbo, J. (1999). Socioeconomic status and chronic obstructive pulmonary disease. *Thorax*, 54, 737–741.
- Qiu, H., Yu, I. T.-S., Tse, L. A., Chan, E. Y., Wong, T. W., & Tian, L. (2015). Greater temperature variation within a day associated with increased emergency hospital admissions for asthma. *Science of the Total Environment*, 505, 508–513.
- Rao, M. (2016). Investigating the potential of land use modifications to mitigate the respiratory health impacts of NO₂: a case study in the Portland-Vancouver metropolitan area.
- Released, S. I. (2007). SPSS for Windows, Version 16.0.
- Robertson, C., & Nelson, T. A. (2010). Review of software for space-time disease surveillance. *International Journal of Health Geographics*, 9, 16.
- Rosenshein, L., Scott, L. & Pratt, M. (2011). *Exploratory regression: a tool for modeling complex phenomena*. ESRI. Redlands, 1.
- Rushton, G., Armstrong, M. P., Gittler, J., Greene, B. R., Pavlik, C. E., West, M. M. & Zimmerman, D. L. (2007). *Geocoding health data: the use of geographic codes in cancer prevention and control, research and practice*, CRC Press.
- Scherber, K., Langner, M., & Endlicher, W. (2014). Spatial analysis of hospital admissions for respiratory diseases during summer months in Berlin taking bioclimatic and socioeconomic aspects into account. *DIE ERDE-Journal of the Geographical Society of Berlin*, 144, 217–237.
- Schwartz, J., Slater, D., Larson, T. V., Pierson, W. E., & Koenig, J. Q. (1993). Particulate air pollution and hospital emergency room visits for asthma in Seattle. *American Review of Respiratory Disease*, 147, 826–831.
- Scott, L. M. & Janikas, M. V. (2010). Spatial statistics in ArcGIS. *Handbook of applied spatial analysis*, 27–41.
- Seys, S. F., Daenen, M., Dilissen, E., Van Thienen, R., Bullens, D. M., Hespel, P., & Dupont, L. J. (2013). Effects of high altitude and cold air exposure on airway inflammation in patients with asthma. *Thorax*, 68, 906–913.
- Silva, P., Rodrigues, P., Ignotti, E., Rosa, A., & Hacon, S. (2011). Hospitalization of children for asthma in the Brazilian Amazon: trend and spatial distribution. *Epidemiology*, 22, S169.
- So, F. F. & Lai, P. (2007). *Spatial epidemiology of asthma in Hong Kong*. GIS for Health and the Environment. Springer.
- Soysal, Ö. M., Schneider, H., Shrestha, A., Guempel, C. D., Li, P., Donepudi, H., Kondoju, N. K. & Sekeroglu, K. (2012). Zonal statistics to identify hot-regions of traffic accidents. Proceedings of the international conference on modeling, simulation and visualization methods (MSV). The Steering Committee of The World Congress in Computer Science,

- Computer Engineering and Applied Computing (WorldComp), 1.
- Stevanović, I., Jovašević-Stojanović, M., & Stošić, J. J. (2016). Association between ambient air pollution, meteorological conditions and exacerbations of asthma and chronic obstructive pulmonary disease in adult citizens of the town of Smederevo. *Vojnosanitetski Pregled*, 73, 152–158.
- Thishan Dharshana, K., & Coowanitwong, N. (2008). Ambient PM10 and respiratory illnesses in Colombo city, Sri Lanka. *Journal of Environmental Science and Health Part A*, 43, 1064–1070.
- Tonne, C., Basagaña, X., Basile, C., Maud, H., Perry, H., Nawrot Tim, S., Remy, S., Roel, V., Jennifer, W., & Mark, N. (2017). New frontiers for environmental epidemiology in a changing world. *Environment International*, 104, 155–162.
- Tseng, C.-M., Chen, Y.-T., Ou, S.-M., Hsiao, Y.-H., Li, S.-Y., Wang, S.-J., Yang, A. C., Chen, T.-J., & Perng, D.-W. (2013). The effect of cold temperature on increased exacerbation of chronic obstructive pulmonary disease: a nationwide study. *PLoS One*, 8, e57066.
- Uphoff, E., Cabieses, B., Pinart, M., Valdés, M., Antó, J. M., & Wright, J. (2015). A systematic review of socioeconomic position in relation to asthma and allergic diseases. *European Respiratory Journal*, 46, 364–374.
- Uragoda, C. (1980). Respiratory disease in tea workers in Sri Lanka. *Thorax*, 35, 114–117.
- Van Miert, E., Sardella, A., Nickmilder, M., & Bernard, A. (2012). Respiratory effects associated with wood fuel use: a cross-sectional biomarker study among adolescents. *Pediatric Pulmonology*, 47, 358–366.
- Van Vliet, P., Knape, M., De Hartog, J., Janssen, N., Harssema, H., & Brunekreef, B. (1997). Motor vehicle exhaust and chronic respiratory symptoms in children living near freeways. *Environmental Research*, 74, 122–132.
- Villeneuve, P. J., Leech, J., & Bourque, D. (2005). Frequency of emergency room visits for childhood asthma in Ottawa, Canada: the role of weather. *International Journal of Biometeorology*, 50, 48–56.
- Wang, W., Ying, Y., Wu, Q., Zhang, H., Ma, D., & Xiao, W. (2015). A GIS-based spatial correlation analysis for ambient air pollution and AECOPD hospitalizations in Jinan, China. *Respiratory Medicine*, 109, 372–378.
- Ward, M. D. & Gleditsch, K. S. (2007). An introduction to spatial regression models in the social sciences. Manuscript at <http://www.faculty.washington.edu/mdw>. Last visited August, 8, 2007.
- Wheeler, B. W., & Ben-Shlomo, Y. (2005). Environmental equity, air quality, socioeconomic status, and respiratory health: a linkage analysis of routine data from the Health Survey for England. *Journal of Epidemiology and Community Health*, 59, 948–954.
- Wickramasinghe, A., Karunaratne, D., & Sivakanesan, R. (2011). PM 10-bound polycyclic aromatic hydrocarbons: concentrations, source characterization and estimating their risk in urban, suburban and rural areas in Kandy, Sri Lanka. *Atmospheric Environment*, 45, 2642–2650.
- Wong, T. W., Lau, T. S., Yu, T. S., Neller, A., Wong, S. L., Tam, W., & Pang, S. W. (1999). Air pollution and hospital admissions for respiratory and cardiovascular diseases in Hong Kong. *Occupational and Environmental Medicine*, 56, 679–683.
- Wubuli, A., Xue, F., Jiang, D., Yao, X., Upur, H., & Wushouer, Q. (2015). Socio-demographic predictors and distribution of pulmonary tuberculosis (TB) in Xinjiang, China: a spatial analysis. *PLoS One*, 10, e0144010.
- Xiong, Q., Zhao, W., Gong, Z., Zhao, W., & Tang, T. (2015). Fine particulate matter pollution and hospital admissions for respiratory diseases in Beijing, China. *International Journal of Environmental Research and Public Health*, 12, 11880–11892.
- Xirasagar, S., Lin, H.-C., & Liu, T.-C. (2006). Seasonality in pediatric asthma admissions: the role of climate and environmental factors. *European Journal of Pediatrics*, 165, 747–752.
- Yin, P., Feng, X., Astell-Burt, T., Qi, F., Liu, Y., Liu, J., Page, A., Wang, L., Liu, S., & Wang, L. (2016). Spatiotemporal variations in chronic obstructive pulmonary disease mortality in China: multilevel evidence from 2006 to 2012. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 13, 339–344.
- Zhang, Z., Xiao, R., Shortridge, A., & Wu, J. (2014). Spatial point pattern analysis of human settlements and geographical associations in eastern coastal China—a case study. *International Journal of Environmental Research and Public Health*, 11, 2818–2833.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.