A geographically weighted regression approach to investigate air pollution effect on lung cancer: A case study in Portugal

Diogo Cardoso,[1] Marco Painho,[1] Rita Roquette[1,2]

[1]NOVA IMS Information Management School; [2]"Doutor Ricardo Jorge" National Institute of Public Health (INSA), Lisbon, Portugal

Abstract

The risk of developing lung cancer might to a certain extent be attributed to tobacco. Nevertheless, the role of air pollution, both form urban and industrial sources, needs to be addressed. Numerous studies have concluded that long-term exposure to air pollution is an important environmental risk factor for lung cancer mortality. Still, there are only a few studies on air pollution and lung cancer in Portugal and none addressing its spatial dimension. The goal was to determine the influence of air pollution and urbanization rate on lung cancer mortality. A geographically weighted regression (GWR) model was performed to evaluate the relation between particle matter$_{10}$ (PM_{10}) emissions and lung cancer mortality relative risk (RR) for males and females in Portugal between 2007 and 2011. RR was computed with the BYM model. For a more in-depth analysis, the urbanization rate and the percentage of industrial area in each municipality were added. GWR efforts led to identifying three variables that were statistically significant in explaining lung cancer relative risk mortality, PM_{10} emissions, urbanization rate and the percentage of industrial area with an adjusted R^2 of 0,63 for men and 0,59 for women. A small set of 8 municipalities with high correlation values was also identified (local R^2 above 0,70). Stronger relationships were found in the north-western part of mainland Portugal. The local R^2 tends to be higher when the emissions of PM_{10} are joined by urbanization and industrial areas. However, when assessing the industrial areas alone, it was noted that its impact was lower overall. As one of the first communications on this subject in Portugal, we have identified municipalities where possible impacts of air pollution on lung cancer mortality RR are higher thereby highlighting the role of geography and spatial analysis in explaining the associations between a disease and its determinants.

Correspondence: Diogo Cardoso, NOVA IMS Information Management School, Campus of Campolide, 1070-312 Lisbon, Portugal.
Tel.: +35.1916886142.
E-mail: diogo.luzcardoso@gmail.com

Key words: Air pollution; Geographically weighted regression; Lung cancer mortality; PM_{10}; Portugal.

See online Appendix for additional Figures.

Contributions: DC developed the study, undertook the statistical analysis and wrote the manuscript. MP and RR critically reviewed the manuscript and provided helpful feedback. All authors read and approved the final manuscript.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: none.

Availability of data and materials: all data generated or analyzed during this study are included in this published article and its references. The only exception is Lung Cancer Mortality RR data, which belong to the Ministry of Health and Rita Roquette, and it is not available to the public.

Received for publication: 21 April 2018.
Revision received: 24 October 2018.
Accepted for publication: 25 October 2018.

Geospatial Health 2019; 14:701
doi:10.4081/gh.2019.701

Introduction

The beginning of the Industrial Revolution in the United Kingdom in the 18th century brought about air pollution as a serious problem which has impacted human society ever since. Until then, air pollution was limited to volcanic activities, mining and some domestic tasks involving the use of fuels such as coal (Stern, 2014). The rapid urbanization and industrialization around the world, resulting in substantial increases in emissions of pollutants from burning fossil fuels, is not a problem that exclusively affects the urban populations. In rural areas, local populations are exposed to the use of solid fuels such as vegetal, wood and coal, among others (Arbex *et al.*, 2012). Several studies have shown the link between particulate and gaseous pollutants emitted by different sources, and the symptoms of respiratory diseases and the consequent demand for health services and hospitalizations (Arbex *et al.*, 2012). In 2016, the World Health Organization (WHO) estimated that global air pollution (indoor and outdoor) killed 7 million people worldwide (one in nine deaths), a number that has doubled past estimates, making air pollution a major health risk in the world (WHO, 2016).

One of the main diseases linked to this type of pollution is lung cancer, the type of cancer that kills the most and the fifth leading cause of death in the world: 1.69 million deaths in 2015 (WHO, 2017). At the end of the 19th century and beginning of the

20th century, it was an almost unknown disease. At the time it represented only 1% of all cancers identified in autopsies at the Institute of Pathology at the University of Dresden in Germany in 1878, rising to about 10% in 1918 and more than 14% in 1927 (Witschi, 2001). Over the course of the century, this number increased further and several causes were identified, including increased air pollution due to the higher number of motor vehicles on the roads and exposure to different types of noxious gases during World Wars I and II, while smoking was only shown to be associated decades later (Witschi, 2001). However, studies associating these two factors would only appear in the middle of the 20th century, after three momentous events of air pollution, or smog, as they later became known. Such events in 1938 in the Meuse River Valley, Belgium; in 1948 in Donora, Pennsylvania; and the gravest of all, in 1952 in London, which killed about 12,000 people, arousing the authorities to the true consequences of prolonged exposure to harmful material in the air (Witschi, 2001). Even though some authors claim that given the high percentage of tobacco-related cases it is difficult to carry out studies proving other factors that cause the disease (Zamboni, 2002), numerous studies conducted in various parts of the world have concluded that long-term exposure to air pollution is an important environmental risk factor for lung cancer mortality and other respiratory diseases (Pope III *et al.*, 2002; Katanoda *et al.*, 2011; Rückerl *et al.*, 2011; Hamra *et al.*, 2014). Diverse methods were used in these studies to analyze the effects of air pollution on lung mortality. The choice of methods depended directly on the type of epidemiological study carried out; some of the adopted methods were: Cox proportional hazards model (Pope III *et al.*, 2002; Katanoda *et al.*, 2011); Hierarchical logistic regression models (Hystad *et al.*, 2013) and log-linear regression model (Samet *et al.*, 2000).

The International Agency for Research on Cancer (IARC) estimates for 2010 indicate that of the total deaths in the world from lung cancer, around 223,000 (near 13% of the total), were directly related to air pollution (IARC, 2013). Even though the relative risk of developing cancer as a result of exposure to air pollution is low, the attributed risk (the relative risk multiplied by the number of exposed persons) is high, making air pollution the most significant environmental risk for lung cancer (Fajersztajn *et al.*, 2013). With the available study results, in 2013, IARC classified air pollution as a carcinogen to humans, being included in group 1, a category used *when there is strong evidence of carcinogenicity in humans. Exceptionally, an agent may be placed in this category when the evidence of carcinogenicity in humans is less, but there is sufficient evidence of carcinogenicity in experimental animals and strong evidence in exposed humans where the agent acts through a relevant carcinogenicity mechanism* (IARC, 2013).

Given the current economic and social situation, especially the increase in urban areas and a context where economic growth is encouraged, it is to be expected that more and more people will be exposed to air pollution. Thus, it is essential to determine the contribution of risk factors in the development of the disease, so that preventive measures may be implemented. With access to information, people living in high-risk areas make pressure on public authorities and stakeholders to act in the name of public health and the environment. It is important to note that by the year 2050, exposure to external air pollution is expected to become the leading cause of premature environmental deaths in the world (OECD, 2012), overcoming malaria and deaths associated with poor water quality. While everything points to a decrease in deaths because of these two last causes, pollution should continue to make more victims each year. Since air pollution is intrinsically associated with different economic activities, it is crucial to define strategies that can bring together the various stakeholders in order to solve the health effects of air pollution.

The Air Quality Report for 2016 estimated 6,640 deaths from air pollution in Portugal (Cristina *et al.*, 2016). The same report mentions that Portugal should reduce emissions of nitrogen oxide and ozone concentrations, especially in urban areas. It also addresses the excessive use of individual transport as the main aggravating factor for air quality problems and responsible for the high levels of air pollution in the cities of Lisbon and Oporto. The Portuguese Environment Agency (APA) also identifies emissions from road vehicles as one of the primary sources of PM, highlighting *the combustion of biomass by the domestic sector (burning of fuels such as wood and coal)* (Ferreira *et al.*, 2015). However, there are satisfactory indicators in some points, such as particle matter $(PM)_{2.5}$ values lower than those recommended by the WHO, one of the sharpest decreases in PM_{10} (Cristina *et al.*, 2016) as well as a downward trend in the remaining pollutants under analysis (Ferreira *et al.*, 2015).

The objective of this study is to investigate the relation between the relative risk (RR) of lung cancer mortality in mainland Portugal from 2007-2011 and air pollution. Through a spatial analysis, we intend to identify municipalities and regions where air pollution has more impact on the RR of lung cancer mortality. Using the urbanization rate and the percentage of industrial area we intend to assess if the source of the pollutants affects the results or not. This article is organized in five sections. Section 1 (Introduction) introduces the background of this study and includes a review of the theme and the situation in Portugal. Section 2 (Materials and Methods) consists of three parts: i) delimitation of the study area; ii) data acquisition and procedures; iii) methods of data analysis; section 3 includes results; section 4 a discussion about the study and its findings and section 5 presents the conclusions.

Materials and Methods

Study area

The model was built for mainland Portugal with data from the 2016 Official Administrative Charter of Portugal (CAOP), which includes all of its 278 municipalities, the second-level administrative subdivision of Portugal (Figure 1).

For data analysis purposes, boundaries of the Nomenclature of Territorial Units for Statistics level (NUTS) II was used. Lisbon Metropolitan Area (MA) is simultaneously NUTS II and MA. Oporto MA, in turn, integrates the NUTS II in the North (the Norte region).

Pollution

We focus on the PM_{10} approach given the various studies that have linked it to lung cancer and because it is defined as a carcinogenic agent with sufficient evidence in humans (IARC, 2013). Other agents like nitrogen dioxide, diesel and arsenic are also identified with carcinogenic capabilities (Cogliano *et al.*, 2011), but they were not included in this study due to the lack of data for Portugal. At first, the pollution data used were provided by the European Environment Agency (EEA) on its website (EEA, 2018),

based on the values of the stations around mainland Portugal managed by the APA and some in Spain along the Portuguese border. These values, referring to the concentration of PM_{10} expressed as $\mu g/m^3$, although reliable, were not homogeneous at the temporal level. Many stations only had values for just a few years, and many of the interior municipalities did not have a nearby station that could serve as a reference, leading to the concentration of data on coastal and urban areas.

We worked with the APA report where the values of PM_{10} emissions per municipality were available for 2009 (APA, 2011). These values are represented in ton per square km (t/km^2) and include natural sources (the difference between the inclusion of natural sources and its exclusion is practically nil). The majority of the literature gives values in $\mu g/m^3$ (Guo *et al.*, 2016; Jerrett *et al.*, 2013), but since the values on the report are from 2009, that is, the middle of our temporal analysis (between 2007 and 2011), and encompass all the municipalities of mainland Portugal, we chose to work with these values in order to guarantee a better consistency of the results and consequently a more accurate analysis. It is essential to make a distinction between the emissions and atmospheric concentrations of PM: emissions refer solely to the emission of particulate matter by sources like industry, traffic or agriculture. The concentrations, most often expressed as $\mu g/m^3$ in open air, are determined by those emissions and meteorological conditions. There is, however, no linear relation between the emissions and the concentrations of particulate matter (Fierens *et al.*, 2015).

Relative risk

Lung cancer mortality data were made available as RR by municipality, for males and females. In this study, the RR was calculated using the Bayesian model of BYM (Besag *et al.*, 1991), a model often used to estimate spatial risk patterns in the hierarchical mapping of diseases (Gerber, 2013). This method appears as the best option when the disease is specific enough. Some fluctuations that may arise in smaller counts imply that maps based only on raw data may be difficult to interpret and misleading. There are advantages in applying some form of smoothing, which may or may not involve a spatial component and provide point and interval estimates for hazards (Besag *et al.*, 1991).

The BYM model is based on the Poisson regression, where the observed cases are the dependent variable, the expected cases the offset and two types of terms of random effects that take into account both spatial continuity and spatial heterogeneity (López-Abente *et al.*, 2014). It was computed with INLA (Rue *et al.*, 2009), using mortality and population data, obtained from Statistics Portugal (INE). We adopted codes C00 to C97 in data cancer collection, according to the 10th revision of International Statistical Classification of Diseases and Related Health Problems (10^{th} ICD). In terms of period, we considered data aggregated in a five-year period, 2009-2013, given that it is recommended to use large populations, and data grouping in several years (Jensen, 1991). To calculate the reference population, we used 2011 data, the central year of the period and the Census year, multiplied by five (the number of years). Data were disaggregated by sex and eighteen age groups were considered: 0-4; 5-9; 10-14; 15-19; 20-24; 25-29; 30-34; 35-39; 40-44; 45-49; 50-54; 55-59; 60-64; 65-69; 70-74; 75-79; ≥80 years old. In the model construction, we adopted the *Besag* model, the *Laplace* option and neighbourhood based on spatial contiguity.

Urbanization rate

For the urbanization rate, we used the Land Use and Land Cover Map (COS) of 2010, made available by the Directorate-General of Territory (DGT) (DGT, 2016). The level one class was used – Artificial Territories – and all its sub-classes, except the sub-class 1.4.1 corresponding to urban green spaces. Then, the percentage of artificial land use was calculated for all the municipalities of the country, using the following formula (Eq. 1):

$$UR = \frac{x}{y} \times 100 \quad \text{Eq. 1}$$

where *UR* is the urbanization rate, the artificialized area of each municipality andthe total area of each municipality. Afterwards, the urban area was divided into two: urban area *per se* and industrial area. The urban area includes the following subclasses: continuous urban fabric (1.1.1), discontinuous urban fabric (1.1.2), road and rail networks and associated spaces (1.2.2), port areas (1.2.3) and airports (1.2.4). In the industrial areas the following classes were encompassed: industry (1.2.1.01), energy production infrastructures (1.2.1.05), opencast mines (1.3.1.01), quarries (1.3.1.02), landfill sites (1.3.2.01) and dumps and scrap yards (1.3.2.02). The aim was to separate the continuous and discontinuous urban fabric and the transport networks from the industrial areas and the production and extraction of aggregates sites, potentially more polluting.

Methods of exploratory data analysis

Exploratory data analysis methods were applied to consider spatial autocorrelation within spatial data. The first approach involved the computation of Ordinary Least Squares (OLS) and Global Moran's *I*, which is widely used in Geographic Information Systems (GIS), having a rather large usefulness in the geographical analysis of variables in health and epidemiology (Getis *et al.*, 1992; Bui *et al.*,

Figure 1. Mainland Portugal Nomenclature of Territorial Units for Statistics level (NUTS) level II and Oporto Metropolitan Area (MA) which belongs to the Norte Region. Data source: Official Administrative Charter of Portugal (CAOP), 2016.

2017). It serves as a complement to a cluster analysis, since the existence of a geographical pattern may indicate that another geographical phenomenon may explain the events under study. The OLS, a linear regression model, shows the deviation of the actual results from the expected results. However, it presents some limitations, especially with regard to the spatial question, since it uses a single equation for all geographic areas (Gutierrez *et al.*, 2012). To explore the local spatial heterogeneity of the potential relationships between PM_{10} and lung cancer, the Geographically Weighted Regression (GWR) model appears as the best option. Unlike the OLS, the GWR defines a different equation for each of the geographical areas as it takes the local geographic variation into account (Gutierrez *et al.* 2012), since a relation (or pattern) that is applied to one area does not necessarily apply to the rest (Comber *et al.*, 2011). However, OLS can be a good tool to indicate a potential problem with global or local multicollinearity; if the Variance Inflation Factor (VIF) value for each explanatory variable is large (above 7,5), global multicollinearity is preventing GWR from a good performance (Mitchel, 2005). Using GWR, each data point is a regression point that is weighted by the distance of the point itself. A spatial kernel map fits the data, and a kernel bandwidth indicates the distance beyond which neighbouring regions no longer have an influence on local estimates (Sassi, 2010). The GWR is then an improvement of the classical regression models (Foody, 2003). GWR extends the global regression technique by allowing local parameters to be estimated, instead of global, therefore making it possible to model regional variations within the data (Fotheringham *et al.*, 2003) (Eq. 2):

$$y_i = \beta_0(u_i, v_i) + \sum_{v=1}^{m} \beta_i(u_i, v_i)\, x_i + \varepsilon_i \qquad \text{Eq. 2}$$

where (u_i, v_i) are the coordinates for every i^{th} point in space, allowing a continuous surface of parameter values. An important aspect of GWR is that spatial autocorrelation is present within the sampled data. As a result, it is assumed that data near point *i* will have more influence regarding the estimation of the continuous function at point *i* than data further away from *i*. This method has a high importance because it addresses one of the fundamental principles of geographical analysis: to evaluate the possibility of spatial variability in the statistical models (Comber *et al.*, 2011). The choice of bandwidth tends to be very demanding, since *n* regressions can be used at each step (Charlton *et al.*, 2009). In the development of this model, an adaptive kernel type was used instead of a fixed type. With the adaptive type, the bandwidth distance will change according to the spatial density of each feature in its input. The bandwidth thus becomes a function of the number of the closest neighbours, each local estimate being based on the same number of features. Instead of a specific distance, the number of neighbours used for the analysis is taken into account (Charlton *et al.*, 2009). For the analysis of the GWR models, we used the adjusted R^2 results. R^2 assumes that each variable explains the variation in the dependent variable, also indicating the percentage of variation explained only by the independent variables that actually affect the dependent variable. The adjusted R^2 value is always lower than R^2, as it reflects the complexity of the model (the number of parameters) relative to the data. As so, the adjusted value of R^2 is a more accurate and reliable measure of model performance (Bui *et al.*, 2017). As for the individual analysis of each municipality, the chosen method of analysis fell on the local R^2. The local R^2 in the GWR model indicates how well the local regression model fits the observed values of y. Very low values indicate that the local model performs poorly and may need more variables to better explain the causes. On the other hand, higher values indicate a causal relationship. Mapping the local R^2 values to see where the GWR predicts well and where it predicts poorly can provide clues about important variables that may be lacking in the developed model. All results were determined by the Geographically Weighted Regression modelling tool within the ESRI's ArcGIS Software.

Results

PM_{10}, urbanization rate and relative risk of lung cancer mortality maps

Figure 2A shows the map with the emission values of PM_{10} t/km^2 in mainland Portugal. There is a clear distinction between coastal areas and the interior, particularly in the two main urban centres and more densely populated areas of the country. This map almost coincides with the map with the urbanization rate (Figure 2B), where the regions with the highest percentage of urbanized area correspond to the two metropolitan areas. Regarding lung cancer mortality RR (Figure 3), there are several differences between males and females. Although both exhibit high values in the two metropolitan areas, they differ in some parts of the country: the values for women are smaller in the interior, except in some municipalities in the Centro region, while in men these values are higher in particular in the southern regions (Alentejo and Algarve). A comparison with the data on the percentage of smokers in Portugal is only available at the NUTS II level (Figure 4). The regions with the highest percentage of male smokers match with those where the mortality RR values are higher in the southern regions; in women, this percentage is higher in Lisbon MA, where mortality RR values are also higher and in the South. It should be noted that given the considerable size of the administrative regions at which variables are available, a more precise spatial analysis is difficult, so they should only be seen as a complement to the analysis. For instance, in the Norte region, there is a considerable difference in mortality RR between the coastal area and the interior of the region, but it is not possible to make a detailed analysis regarding the percentage of smokers.

Ordinary least squares model results

In our study, we modelled the relation of the RR of lung cancer mortality with PM_{10} emissions, the urbanization rate and the percentage of industrial area. VIF results for each variable were: 1,323011 for the PM_{10} emissions, 1,655088 for the urbanization rate and 1,072336 for the industrial area, which indicates that the model would not be affected by multicollinearity. Akaike Information Criterion Corrected (AICc) values, presented in Table 1, show that

Table 1. Ordinary least squares Akaike Information Criterion Corrected values.

	RR & PM_{10}	RR, PM_{10}, & TU	RR, PM_{10}, & PAI	RR, PM_{10}, TU & PAI
Women	-352	-17	-246	-56
Men	-253	-59	-329	-80

RR, relative risk; PM, particle matter. TU, urbanization rate; PAI, percentage of industrial area.

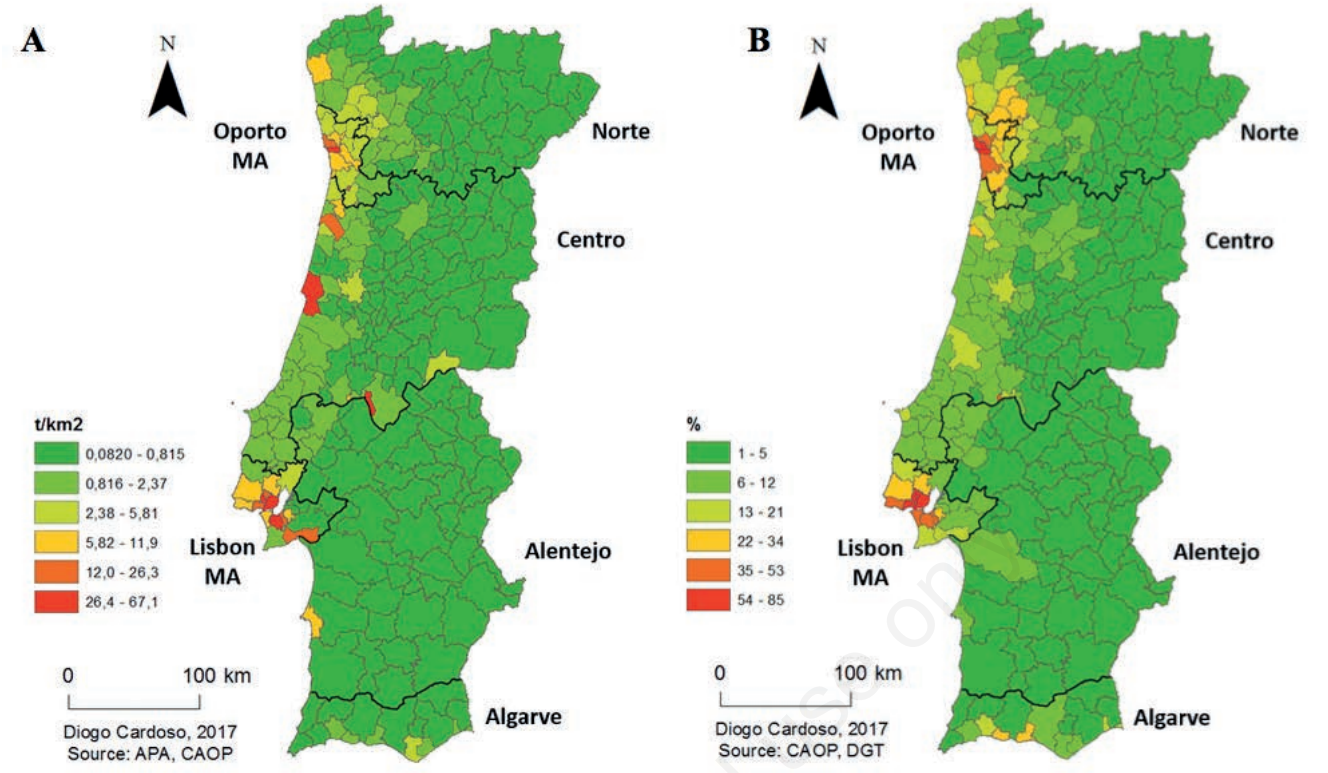

Figure 2. (A) Particle matter$_{10}$ emissions (t/km^2); (B) Urbanization rate. MA, metropolitan area.

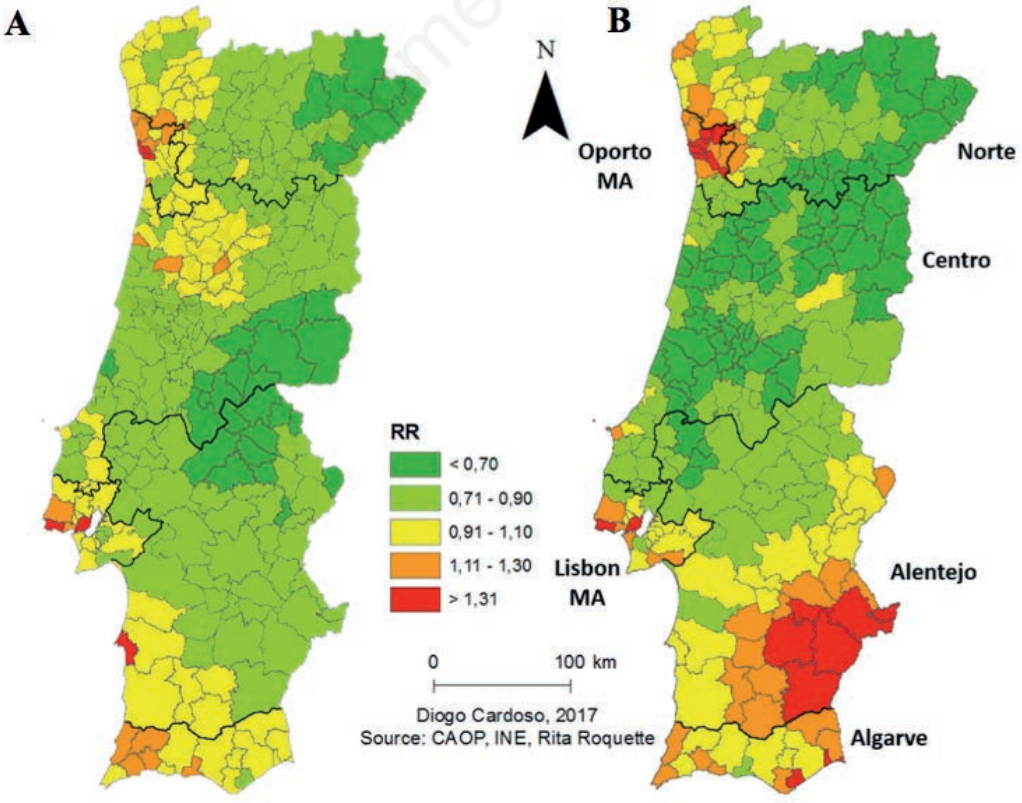

Figure 3. Lung cancer mortality relative risk for women (A) and men (B) between 2007-2011. MA, metropolitan area.

the best model for women considers only PM_{10}, while for men it considers PM_{10} and percentage of industrial area.

Relative risk with PM_{10} emissions

The local coefficients of PM_{10} and RR were positive in all municipalities (Appendix Figure A1). For men, higher values were found to be located in South and northwest. For women, the cluster in South was smaller than that for men, while the cluster in Northwest was higher.

This model had an overall adjusted R^2 result of 0,62 for men and 0,58 for women, that is, tends to be successfully predicting about 64% of the mortality RR variability in men and 59% in women. However, when the map of local R^2 (Figure 5) was overlaid with the RR map (Figure 2B), we found some differences; looking at the values for men there was an area to the Southeast with high RR values that did not have the correspondence with PM_{10} values. On the other hand, there seemed to exist a relation in the Northwest area, were both maps had higher values. For women, the highest values were scattered, but as for men, they were more concentrated to the Northwest, so as the RR values.

Relative risk with PM_{10} emissions and urbanization rate

In this model, the local coefficients of PM_{10} and RR were positive in all municipalities (Appendix Figure A2). As in the previous model, clusters of higher values were located in the South and Northwest and were more extensive for men than women. It should be noted that the municipalities of the Lisbon MA was not classified in the higher class of values, contrary to what happened with

Figure 6 shows that the model tends to successfully predict about 65% of the relative risk variability for men and 58% for women, *i.e.* a similar result as the situation before: an adjusted R^2 of 0,65 and 0,59, respectively. Concerning men, the map revealed higher values in both MAs and their surroundings, but the southeastern hotspot remained without a relationship with the variables under analysis. The values for women were again higher in the northern coastal region, with a cluster in the northern central region. Values around Lisbon MA were lower than with the previ-

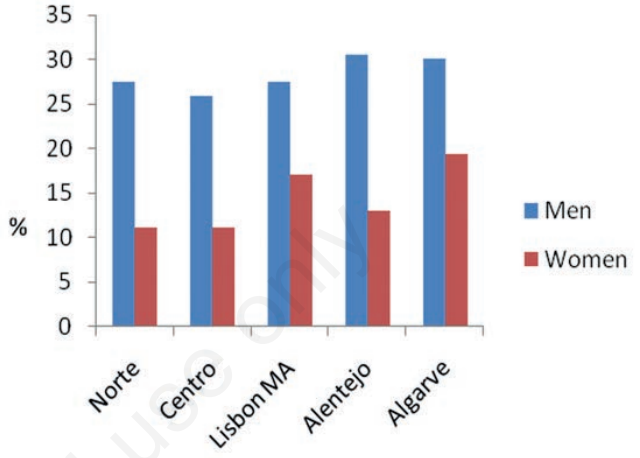

Figure 4. Percentage of smokers in mainland Portugal by Nomenclature of Territorial Units for Statistics II. Data source: Statistics Portugal (INE), 2014. MA, metropolitan area.

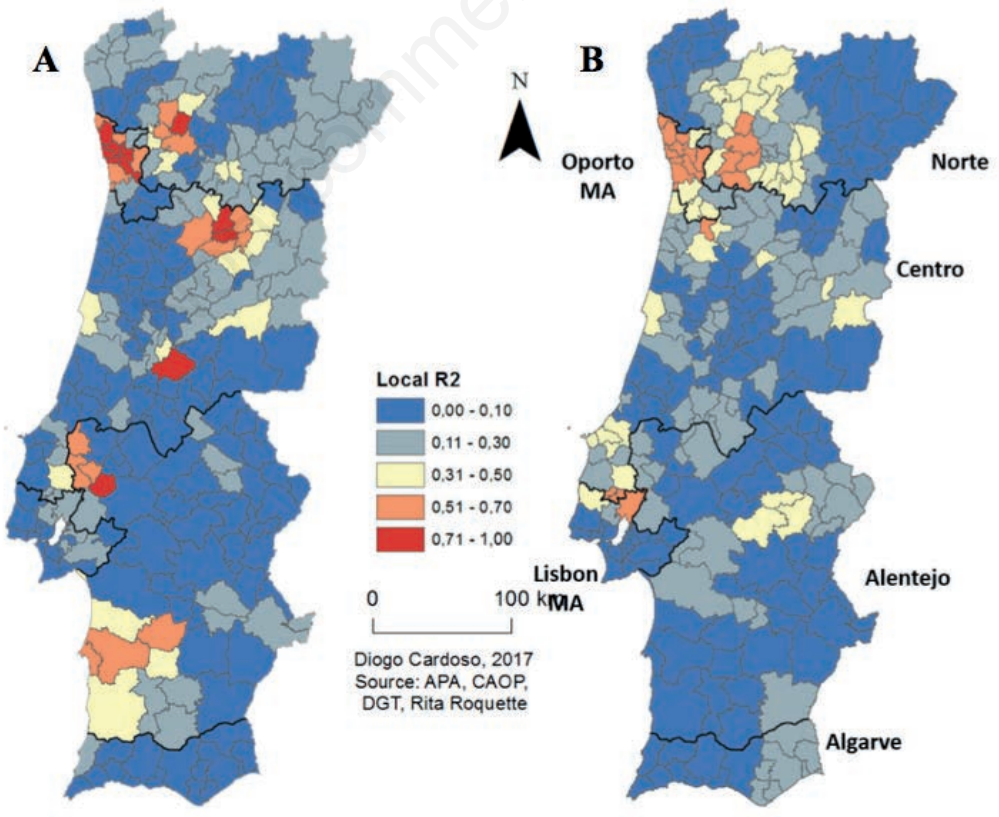

Figure 5. Geographically weighted regression model with particle matter$_{10}$ emissions values for women (A) and men (B). MA, metropolitan area.

ous model. With these two variables, 18% of the municipalities had values above 0,50 for men and 15% for women.

Relative risk with PM_{10} emissions, urbanization rate and percentage of industrial area

The industrial area was separated from urban area (Figure 7 and Appendix Figure A3), which now only includes continuous and discontinuous urban area, and the transport infrastructures. Adjusted R^2values were 0,59 for women and 0,63 for men. Regarding men, higher values were found in both MAs, but for women they were between 0,30 and 0,50 in Lisbon MA. Using these variables, 20% of municipalities had values above 0,50 both for men and women. Also, 3% of the municipalities had values above 0,70 for men.

Relative risk with PM_{10} emissions and percentage of industrial area

With the objective of measuring the possible impact of industry, quarries and aggregate extraction sites on mortality RR values (in particular the hotspot in the south-eastern region in men's RR) the urban area was separated from the industrial area (Figure 8 and Appendix Figure A4). With this model, we were faced with a low percentage of the municipalities with a local R^2 above 0,50; only 8% of municipalities for men and 13% for women, and an adjusted R^2 of 0,48 and 0,46 respectively, the lowest result taking into account all the variables.

Model comparison

We compared the models in terms of standard errors and AICc criterion. Standard errors (Appendix Figures A5-A8) measure the reliability of each coefficient estimate. Confidence is higher when standard errors are small in relation to the actual coefficient values. Moran's *I* values ranged between 0,2107 (the lowest) and 0,6591 (the highest), and while values over 0,5 point to a trend towards clusters, it is possible to say that in general, the model does not tend to clustering (Perrino, 2010; Zhang *et al.*, 2017).

Table 2 presents AICc values for each model. The lowest val-

Table 2. Geographically weighted regression Akaike Information Criterion Corrected values.

	RR & PM_{10}	RR, PM_{10}, & TU	RR, PM_{10}, & PAI	RR, PM_{10}, TU & PAI
Women	-596	-653	-510	-623
Men	-465	-468	-457	-467

RR, relative risk; PM, particle matter; TU, urbanization rate; PAI, percentage of industrial area.

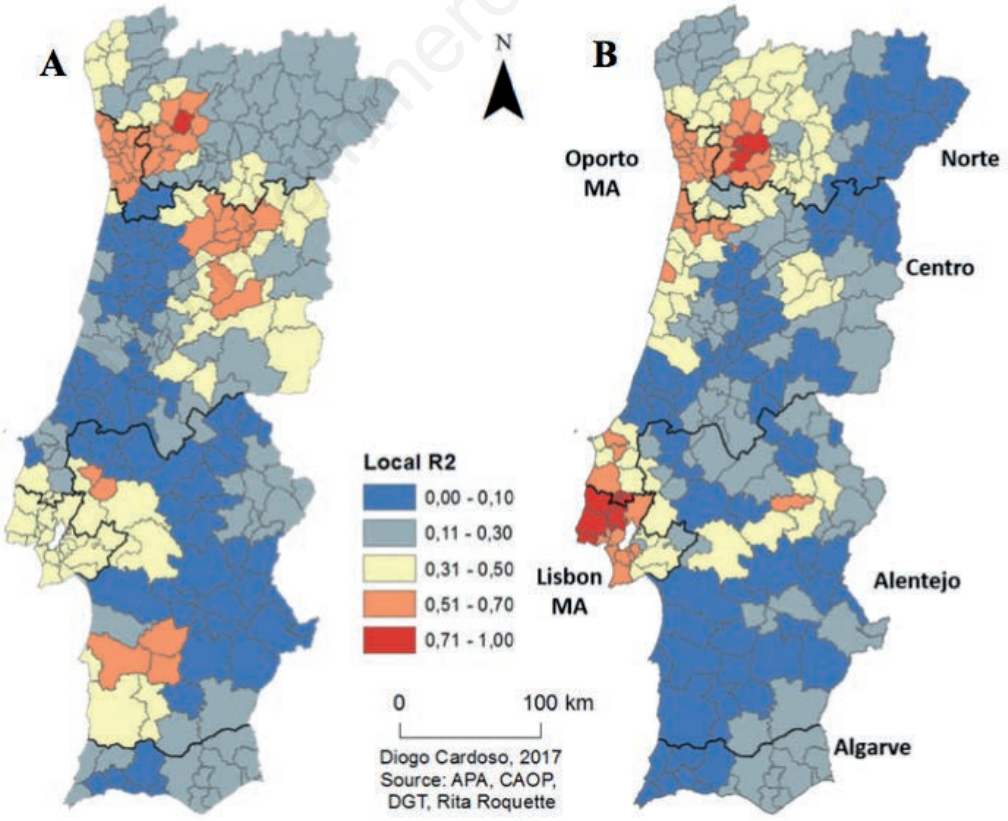

Figure 6. Geographically weighted regression model with particle matter$_{10}$ emissions and urbanization rate values for women (A) and men (B). MA, metropolitan area.

ues correspond to the GWR model, which considers PM_{10} and the urbanization rate. However, the model which considers all variables (PM_{10} Emissions, Urbanization Rate and Percentage of Industrial Area) presented very similar values.

Discussion

The results of this study are in line with the hypotheses initially set. The highest values of correlation are mainly concentrated around the two metropolitan areas of the country, which are the regions with higher PM_{10} emissions, urbanization rate and percentage of industrial area. Nonetheless, those values are higher in the North-western part of mainland Portugal. It was expected that some of the values of lung cancer mortality RR (Figure 3) were not related to those obtained in the GWR models mainly in the southern regions, since smoking is a significant risk factor of lung cancer (about 90% of cases in men and 55% to 80% of cases in women are attributed to cigarette smoking; Levi, 1999). Moreover, even though air pollution is also identified as an important cause of the disease, its impact is expected to be lower. The GWR results between the percentage of industrial area and lung cancer mortality RR represent a lower correlation than the other two variables. A more detailed analysis is necessary in this matter, but this may mean that the impact of the industries is smaller in the emission of PM_{10} with the origin from motor vehicles being higher. As mentioned in the introduction, the use of individual transport is one of the main aggravating factors for the high levels of air pollution in both metropolitan areas (Cristina *et al.*, 2016). The use of urbanization rate is not new in the study of lung cancer mortality, but they tend to have a focus on rural-urban differences and socioeconomic aspects (Riaz *et al.*, 2011; Singh *et al.*, 2012), a different approach from that carried out in this study.

A geographic approach with remote sensing can help to fill in data gaps that hamper current efforts to study air pollution. A study by Hu and Baker (2017) shows that there is a significant positive association between mortality from this type of cancer and $PM_{2.5}$. This result was achieved using data from the MODIS satellite sensor and MISR Annual Global Grid $PM_{2.5}$ data (Hu and Baker, 2017). Nonetheless, the statistically significant association between lung cancer mortality and presence of $PM_{2.5}$ may be indicative of a potential effect of air pollution; the authors suggest that the same association would require a toxicological approach in order to observe the adverse biological mechanism of $PM_{2.5}$ pollution (Hu and Baker, 2017). The model developed in this study yielded satisfactory results and is in line with other similar studies using GWR models (Fu *et al.*, 2015; Ren *et al.*, 2016), or other spatial analyst tools (Bilancia *et al.*, 2009; López-Cima *et al.*, 2011). Even though the study of cancer's spatial epidemiology has had a greater emphasis in the last decade (Roquette *et al.*, 2017), there are only a few studies on the relationship between lung cancer and air pollution in Portugal (Slezakova *et al.*, 2011). This study repre-

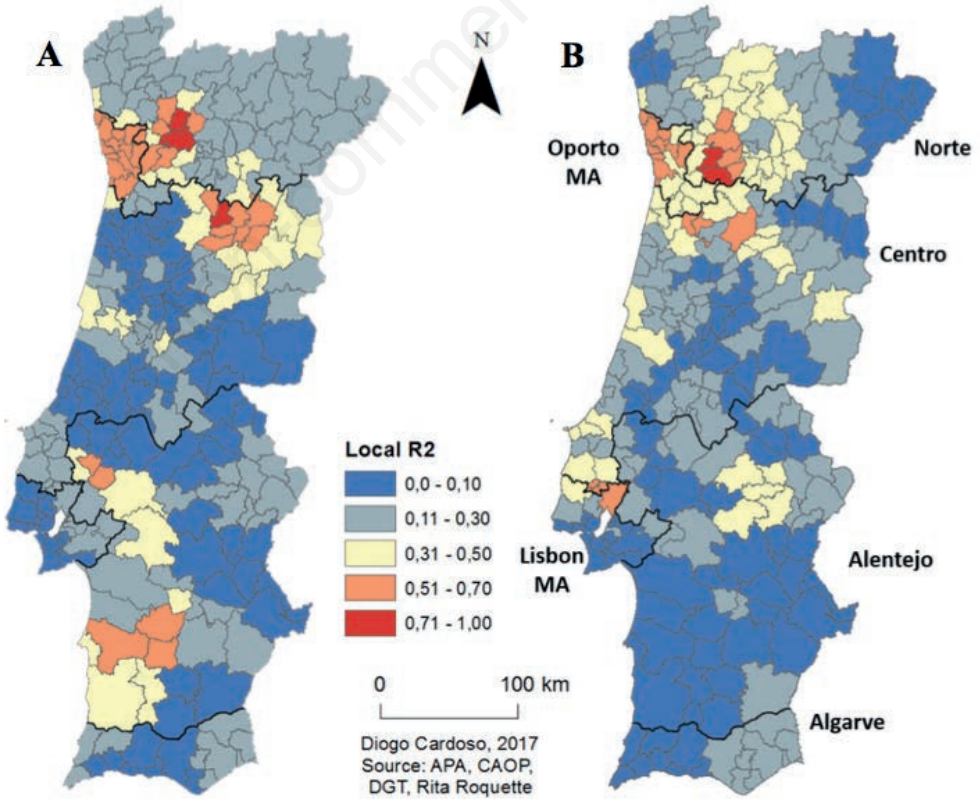

Figure 7. Geographically weighted regression model with particle matter$_{10}$, urbanization rate and percentage of industrial area values for women (A) and men (B).

sents one of the first to use RR mortality data along with spatial regression analysis tools to explore a possible relationship between both factors in mainland Portugal.

One of the advantages of using GWR is that it accounts for spatial autocorrelation in the residuals that are usually found in global modelling. Further, it is possible that a variable that is insignificant at the global level might be important locally (Fotheringham *et al.*, 2008). When relationships are consistent across a study area, an OLS model fits neatly into these relationships; it creates equations that best describe general relationships of data in each area. However, it is not always like that, so often these relationships have different behaviours throughout space. When the exploratory variables exhibit non-stationary relationships (regional variation), the model tends to fail, unless robust models adapted to this problem are applied. The GWR model addresses this issue precisely (Mitchel, 2005).

Identifying causes and effectively addressing them can lead to significant savings in health spending. With the implementation of rigorous legislation on gaseous emissions, health expenditure directly linked to air pollution in Europe has been decreasing. It is estimated that from € 803 billion spent in 2000, it will decrease to € 537 billion in 2020 (Brandt *et al.*, 2013). The convergence of the many studies carried out so far has led to a reconsideration and updating of health standards and guidelines, leading to new long-term research programs in order to analyze the effects of particulate pollution on health (Brandt *et al.*, 2013). However, it is known that these changes always have economic and social impacts, sometimes facing great opposition in certain sectors of society.

Our results sustain the hypothesis that air pollution might be a risk factor for lung cancer. Indeed, it indicates a higher lung cancer mortality RR among municipalities where both urban and industrial areas are also superior. It demonstrated the benefits of GWR, both in respect to model performance and by allowing spatial analysis of the data. Lung Cancer mortality RR was found to be heterogeneously related to human factors at the municipality level in mainland Portugal. Our findings may assist local authorities when assessing risks, and by helping public health entities allocate resources and address the issue according to the specific conditions of each region.

Conclusions

With this research, our objective was not to find the municipalities where people are most likely to die from lung cancer, but rather to assess the impact of the PM_{10} emissions in each municipality and to understand the influence of the urbanization rate and the percentage of industrial area in these values. Furthermore, including two variables that address the land use may be a new method of approaching this subject and generate a more realistic model. As a result, this study contributes to the knowledge of the effects of air pollution on lung cancer and on the use of local spa-

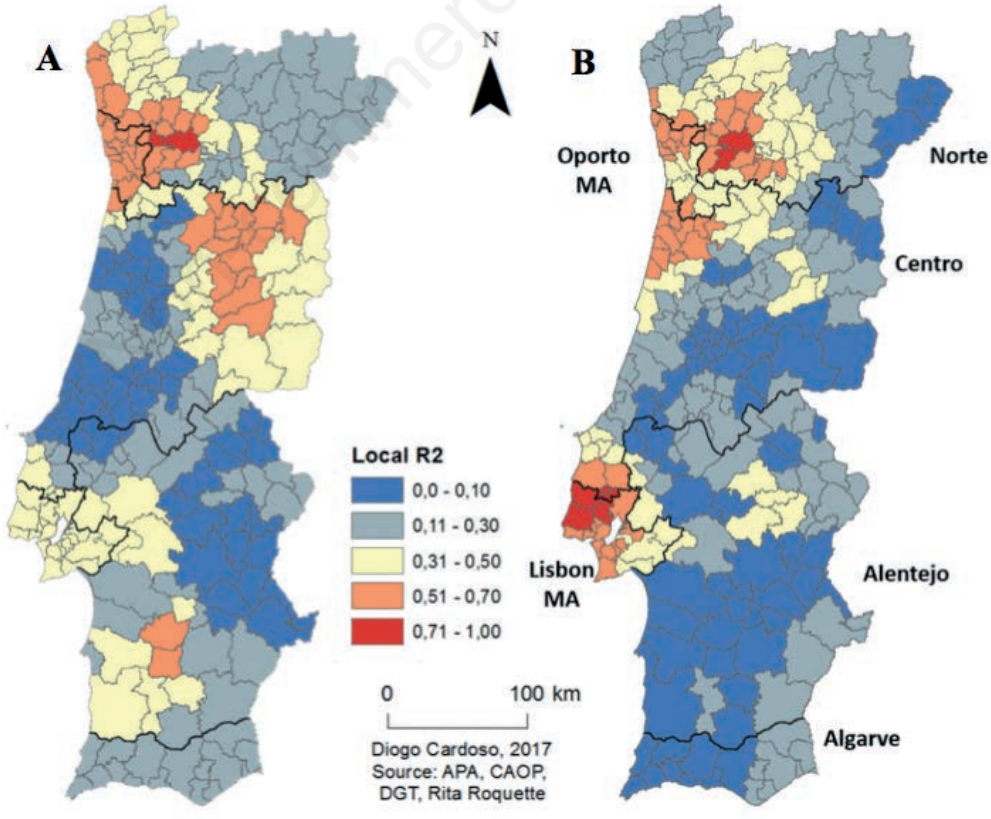

Figure 8. Geographically weighted regression model with particle matter$_{10}$ and percentage of industrial area values for women (A) and men (B). MA, metropolitan area.

tial analyses in epidemiological studies. Such information can be used in urban planning to reduce air pollution.

No evident homogeneous pattern distribution association was found between PM_{10} emissions and lung cancer mortality RR in municipalities across mainland Portugal. There is. The relation between PM_{10}, urban area and industrial areas and lung cancer mortality rates varies spatially, and there are other agents that may influence the lung cancer mortality rate in different areas of mainland Portugal, but we can say that it has a focus on the two metropolitan areas. Several municipalities tend to show values of R^2 always above 0,50 in all models which represents a positive relation. It is pertinent to state that the emission of PM_{10} as the urbanization rate and percentage of industrial area affect the lung cancer mortality RR values in those municipalities. The relation of lung cancer turned out to be higher when the emissions of PM_{10} were joined by the urbanization rate and the percentage of industrial area (R^2 value of 0,63 for men and 0,59 for women). However, when assessing the industrial areas alone, it was noted that their impact is lower in the overall results (R^2 equal to 0,48 for men and 0,46 for women).

Spatial variation in the relations between lung cancer RR and air pollution means that in some places PM_{10} and urbanization rate have a greater effect on mortality than in other places. In the municipalities where the values are high, local authorities should step in to minimize the effects of air pollution and carry out better planning in order to benefit the public health of the local populations. We note that the problem is complex, and that further investigation is needed for a full understanding of this issue.

References

APA (Portuguese Environment Agency), 2011. Emissões de Poluentes Atmosféricos por Concelho 2009: Gases acidificantes e eutrofizantes, precursores de ozono, partículas, metais pesados e gases com efeito de estufa. Available from: https://www.apambiente.pt/_zdata/DPAAC/INERPA/Emissoes%20Concelho%2020111109.pdf

Arbex MA, Paula Santos U, Martins LC, Saldiva PHN, Pereira LAA, Braga ALF, 2012. A poluição do ar e o sistema respiratório. J Bras Pneumol 38:643-55.

Besag J, York J, Mollié A, 1991. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Statist Math 43:1-20.

Bilancia M, Fedespina A, 2009. Geographical clustering of lung cancer in the province of Lecce, Italy: 1992-2001. Int J Health Geogr 8:40.

Brandt J, Silver JD, Christensen JH, Andersen MS, Bønløkke JH, Sigsgaard T, Geels C, Gross A, Hansen AB, Hansen KM, Hedegaard GB, Kaas E, Frohn LM, 2013. Assessment of past, present and future health-cost externalities of air pollution in Europe and the contribution from international ship traffic using the EVA model system. Atmosph Chem Phys 13:7747-64.

Bui DT, Do AN, Bui HB, Hoang ND, 2017. Advances and applications in geospatial technology and earth resources. Proceedings of the International Conference on Geo-Spatial Technologies and Earth Resources 2017. Springer, New York, NY.

Charlton M, Fotheringham S, Brunsdon C, 2009. Geographically weighted regression. White paper. National Centre for Geocomputation, National University of Ireland Maynooth, Ireland.

Cogliano VJ, Baan R, Straif K, Grosse Y, Lauby-Secretan B, El Ghissassi F, Bouvard V, Benbrahim-Tallaa L, Guha N, Freeman, C, Galichet L, Wild CP, 2011. Preventable exposures associated with human cancers. J Nation Canc Institute 103:1827-39.

Comber AJ, Brunsdon C, Radburn R, 2011. A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. Int J Health Geogr 10:44.

Cristina G, Ortiz AG, Leeuw F, Viana M, Horálek J, 2016. Air quality in Europe - 2016 report. Publication Office of the European Union, Luxembourg.

Directorate-General of Territory (DGT), 2016. Especificações técnicas da Carta de uso e ocupação do solo de Portugal Continental para 1995, 2007 e 2010: Relatório Técnico. Available from: http://www.dgterritorio.pt/cartografia_e_geodesia/cartografia/cartografia_tematica/cartografia_de_uso_e_ocupacao_do_solo__cos_clc_e_copernicus_/

European Environment Agency (EEA), 2018. Pollution data. Available from: https://www.eea.europa.eu/data-and-maps/data/aqereporting-8

Fajersztajn L, Veras M, Barrozo LV, Saldiva P, 2013. Air pollution: a potentially modifiable risk factor for lung cancer. Nat Rev Cancer 13:674-8.

Ferreira F, Seixas, J, Barroso, JE, Fortes P, Tente H, Monjardino J, Dias L, Gomes P, Miranda AI, Monteiro A, Ferreira J, Martins H, Ribeiro I, Fernandes AP, Boavida F, Jardim D, Martins C, Marques F, Anacleto T, 2015. Estratégia Nacional Para O Ar 2020 – Emissões Atmosféricas e Qualidade do Ar Ambiente: Enquadramento e Diagnóstico. Available from: apambiente.pt/_zdata/DAR/Ar/ENAR_IV_LinhasEstrategicas_CP.pdf

Fierens F, Vanpoucke C, Trimpeneers E, Peeters O, Quidé S, de Vos T, Maetz P, Hutsemékers V, 2015. Annual Report: Air Quality in Belgium 2015. Available from: http://www.irceline.be/en/documentation/publications/annual-reports/annual-report-2011/view

Foody G, 2003. Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI–rainfall relationship. Remote Sens Environ 88:283-93.

Fotheringham AS, Brunsdon C, Charlton M, 2003. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons, New York, NY.

Fotheringham AS, Rogerson PA, 2008. The SAGE handbook of spatial analysis. Sage, London, UK.

Fu J, Jiang D, Lin G, Liu K, Wang Q, 2015. An ecological analysis of PM 2.5 concentrations and lung cancer mortality rates in China. BMJ Open 5:e009452.

Gerber F, 2013. Disease mapping with the Besag-York-Mollié model applied to a cancer and a worm infections dataset. Master's thesis. University of Zurich, Switzerland.

Getis A, Ord JK, 1992. The analysis of spatial association by use of distance statistics. Geogr Anal 24:189-206.

Guo Y, Zeng H, Zheng R, Li S, Barnett AG, Zhang S, Zou S, Huxley R, Chen W, Williams G, 2016. The association between lung cancer incidence and ambient air pollution in China: A spatiotemporal analysis. Environ Res 144:60-65.

Gutierrez L, Sassi M, 2012. Spatial and non spatial approaches to agricultural convergence in Europe. Econ Diritto Agroalim 17:9.

Hamra GB, Guha N, Cohen A, Laden F, Raaschou-Nielsen O, Samet JM, Vineis P, Forastiere F, Saldiva P, Yorifuji T, Loomis D, 2014. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. Environ Health Perspect 122:906.

Hu Z, Baker E, 2017. Geographical analysis of lung cancer mortality rate and PM2. 5 Using Global Annual Average PM2. 5 Grids from MODIS and MISR Aerosol Optical Depth. J Geosci Environ Protect 5:183.

Hystad P, Demers PA, Johnson KC, Carpiano RM, Brauer M, 2013. Long-term residential exposure to air pollution and lung cancer risk. Epidemiol 24:762-72.

International Agency for Research on Cancer (IARC), 2013. IARC: Outdoor air pollution a leading environmental cause of cancer deaths. Available from: www.iarc.fr/en/media-centre/iarcnews/pdf/pr221_E.pdf

Jensen OM, 1991. Cancer registration: principles and methods. Vol. 95. IARC Scientific Publications, Lyon, France.

Jerrett M, Burnett RT, Beckerman BS, Turner MC, Krewski D, Thurston, G, Martin RV, van Donkelaar A, Hughes E, Shi Y, Gapstur SM, Thun MJ, Pope CA 3rd, 2013. Spatial analysis of air pollution and mortality in California. Am J Respir Crit Care Med 188:593-9.

Katanoda K, Sobue T, Satoh H, Tajima K, Suzuki T, Nakatsuka H, Takezaki T, Nakayama T, Nitta H, Tanabe K, Tominaga S, 2011. An association between long-term exposure to ambient air pollution and mortality from lung cancer and respiratory diseases in Japan. J Epidemiol 21:132-43.

Levi F, 1999. Cancer prevention: epidemiology and perspectives. Eur J Canc 35:1912-24.

López-Abente G, Aragonés N, García-Pérez J, Fernández-Navarro P, 2014. Disease mapping and spatio-temporal analysis: importance of expected-case computation criteria. Geospat Health 9:27-35.

López-Cima MF, García-Pérez J, Pérez-Gómez B, Aragonés N, López-Abente G, Tardón A, Pollán M, 2011. Lung cancer risk and pollution in an industrial region of Northern Spain: a hospital-based case-control study. Int J Health Geogr 10:10.

Mitchel A, 2005. The ESRI Guide to GIS analysis, volume 2: spartial measurements and statistics. ESRI Guide to GIS analysis, Redlands, CA.

OECD, 2012. OECD environmental outlook to 2050:[the consequences of inaction]: OECD. Available from http://www.oecd.org/env/indicators-modelling-outlooks/oecd-environmental-outlook-1999155x.htm

Perrino C, 2010. Atmospheric particulate matter. Biophys Bioeng Letter 3.

Pope III CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD, 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Jama 287:1132-41.

Ren H, Cao W, Chen G, Yang J, Liu L, Wan X, Yang G, 2016. Lung cancer mortality and topography: A Xuanwei Case Study. Int J Environ Res Public Health 13:473.

Riaz SP, Horton M, Kang J, Mak V, Lüchtenborg M, Møller H, 2011. Lung cancer incidence and survival in England: an analysis by socioeconomic deprivation and urbanization. J Thoracic Oncol 6:2005-10.

Roquette R, Painho M, Nunes B, 2017. Spatial epidemiology of cancer: a review of data sources, methods and risk factors. Geospat Health 12:23-35.

Rue H, Martino S, Chopin N, 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J Royal Statist Society: Series B (Statist Methodol) 71:319-92.

Rückerl R, Schneider A, Breitner S, Cyrys J, Peters A, 2011. Health effects of particulate air pollution: a review of epidemiological evidence. Inhalation Toxicol 23:555-92.

Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL, 2000. Fine particulate air pollution and mortality in 20 US cities, 1987-1994. N Engl J Med 343:1742-9.

Sassi M, 2010. OLS and GWR approaches to agricultural convergence in the EU-15. Int Adv Econ Res 16:96-108.

Singh GK, Siahpush M, Williams SD, 2012. Changing urbanization patterns in US lung cancer mortality, 1950-2007. J Commun Health 37:412-20.

Slezakova K, Castro D, Begonha A, Delerue-Matos C, Alvim-Ferraz MC, Morais S, Pereira MC, 2011. Air pollution from traffic emissions in Oporto, Portugal: health and environmental implications. Microchem J 99:51-9.

Stern AC, 2014. Fundamentals of air pollution. New York, NY: Elsevier.

World Health Organization (WHO), 2016. Ambient air pollution: A global assessment of exposure and burden of disease. Available from: http://apps.who.int/iris/bitstream/handle/10665/250141/ 9789241511353-eng.pdf; jsessionid= E8BFEA2AF75742453FD37C8B156F0C03?sequence=1

World Health Organization (WHO), 2017. Global health observatory (gho) data: Top 10 causes of death. Available from http://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/

Witschi H, 2001. A short history of lung cancer. Toxicol Sci 64:4-6.

Zamboni M, 2002. Epidemiologia do câncer do pulmão. J Pneumol 28:41-7.

Zhang W, Li F, Gao W, 2017. Traffic-related air pollution and lung cancer: A meta-analysis. Thorac Cancer 8:546.