Science of the Total Environment 562 (2016) 740–750

Contents lists available at ScienceDirect

Science of the Total Environment

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

Geostatistical uncertainty of assessing air quality using high-spatial-resolution lichen data: A health study in the urban area of Sines, Portugal

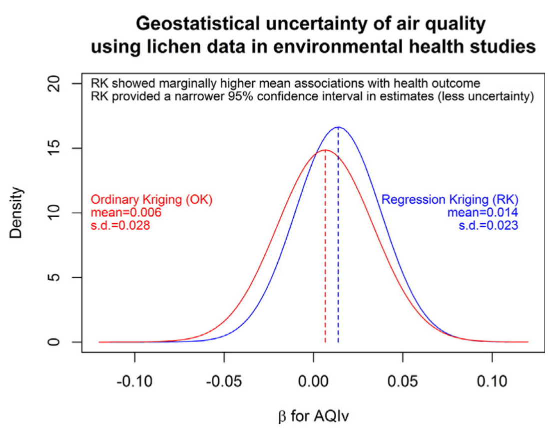
Manuel C. Ribeiro [a,*], P. Pinho [a,b], C. Branquinho [b], Esteve Llop [b,c], Maria J. Pereira [a]

[a] Centro de Recursos Naturais e Ambiente, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001, Lisbon, Portugal
[b] CE3C, Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Campo Grande Bloco C2 5° Piso, 1749-016 Lisbon, Portugal
[c] Departament de Biologia Vegetal, Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain

HIGHLIGHTS

- The health outcome and air quality in a small city were correlated using a novel spatial approach.
- Lichens ($n = 83$) served as air-quality proxies, due to the lack of air-monitoring stations ($n = 1$).
- The ordinary Kriging (OK) and regression Kriging (RK) methods were used to predict exposures.
- OK and RK were used in sequential simulation to address the spatial uncertainty of exposures.
- RK was a more effective method for correlating health with air quality in an urban environment.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 7 December 2015
Received in revised form 7 April 2016
Accepted 11 April 2016
Available online xxxx

Editor: J. P. Bennett

Keywords:
Lichens
Geostatistics
Simulation
Health exposure
Birth weight
Air quality

ABSTRACT

In most studies correlating health outcomes with air pollution, personal exposure assignments are based on measurements collected at air-quality monitoring stations not coinciding with health data locations. In such cases, interpolators are needed to predict air quality in unsampled locations and to assign personal exposures. Moreover, a measure of the spatial uncertainty of exposures should be incorporated, especially in urban areas where concentrations vary at short distances due to changes in land use and pollution intensity. These studies are limited by the lack of literature comparing exposure uncertainty derived from distinct spatial interpolators.
Here, we addressed these issues with two interpolation methods: regression Kriging (RK) and ordinary Kriging (OK). These methods were used to generate air-quality simulations with a geostatistical algorithm. For each method, the geostatistical uncertainty was drawn from generalized linear model (GLM) analysis. We analyzed the association between air quality and birth weight. Personal health data ($n = 227$) and exposure data were collected in Sines (Portugal) during 2007–2010. Because air-quality monitoring stations in the city do not offer high-spatial-resolution measurements ($n = 1$), we used lichen data as an ecological indicator of air quality ($n = 83$). We found no significant difference in the fit of GLMs with any of the geostatistical methods. With RK, however, the models tended to fit better more often and worse less often. Moreover, the geostatistical uncertainty results showed a marginally higher mean and precision with RK. Combined with lichen data and land-use data of high spatial resolution, RK is a more effective geostatistical method for relating health outcomes with air quality in

* Corresponding author.
E-mail address: manuel.ribeiro@tecnico.ulisboa.pt (M.C. Ribeiro).

http://dx.doi.org/10.1016/j.scitotenv.2016.04.081

urban areas. This is particularly important in small cities, which generally do not have expensive air-quality monitoring stations with high spatial resolution. Further, alternative ways of linking human activities with their environment are needed to improve human well-being.

1. Introduction

Poor air quality is a pertinent public health issue, as the incidence of adverse health outcomes (such as respiratory diseases and lung cancer) increases with exposure at different life stages, ranging from prenatal period to adult life (European Environment Agency, 2013). These major public health concerns are commonly related to anthropogenic sources of air pollution such as motor vehicles and industrial facilities, or to natural sources such as forest fires. These sources are responsible for emissions of air pollutants such as particulate matter, benzo(a)pyrene, carbon monoxide, ozone, nitrogen dioxide, and sulfur dioxide. Developing models to assess air pollution exposure in environmental epidemiology studies at the urban scale has been prioritized (Brunekreef and Holgate, 2002). This is because the majority of the global population is expected to live in urban areas in the coming decades, as is currently the case three-fourths of the European population (European Environment Agency, 2012). With the recent advances in geographic information systems (GIS) and spatial modeling methods, several air pollution models for assessing individual exposure in environmental epidemiology studies at the urban scale have been considered (Liao et al., 2006; Nuckols et al., 2004).

Proximity models use distance buffers to measure the associations between health outcomes and sources of air pollution emission, but this approach is limited by the assumption that individuals within a given buffer distance of an emission source are equally exposed (Ryan et al., 2006). Dispersion models generally rely on Gaussian plume equations and incorporate data on meteorology and topography; however, they need a relatively expensive data input (Jerrett et al., 2005). Land-use regression employs least square regressions to combine geographic data collected on air pollution and on predictor variables (e.g., land use, length of roads, or traffic intensity) spatially distributed in the study area to construct an air pollution map. This represents a relatively cheap and practical approach for predicting air pollution exposure in urban settings. Thus, these maps have been increasingly used in the past few years for environmental epidemiology studies at the urban scale (Briggs et al., 2000; de Hoogh et al., 2014). However, land-use regression models fail to provide an uncertainty measure of predictions throughout the unsampled spatial domain.

The Kriging geostatistical methods (Lee et al., 2012) are interpolation methods that were originally developed for fields such as mining engineering and other geosciences (Cressie, 1993). These methods can account for spatial trends and spatial autocorrelation of air pollution concentrations measured at the available network of sampling sites in the study area. The most frequent type of Kriging model used in practice is ordinary Kriging (OK) (Webster and Oliver, 2007), which assumes a constant trend and a spatially homogeneous variation of air pollution concentrations. At the national and regional scales, this method has been used to assess air pollution exposure in environmental epidemiology studies (Augusto et al., 2012; Paustenbach, 2010). However, some authors have reported the limitations of using this interpolator at an urban, local scale (Aguilera and Sunyer, 2008; Gulliver et al., 2011), as OK tends to over-smooth air pollution predictions in an environment expected to abrupt variations in concentrations at short distances. In urban areas, air pollution concentrations are highly susceptible to changes in land use over very short distances, whereas OK modeling only takes into consideration spatial autocorrelation of air pollution concentrations generally collected from a sparse network of monitoring stations (O'Leary and Lemke, 2014).

To overcome the limitation imposed by the lack of air-quality monitoring stations, lichen diversity biomonitoring programs can be considered. Lichens are living symbiotic organisms consisting of fungi and algae or cyanobacteria, which tend to decline in frequency and diversity in polluted areas as a physiological response to the long-term cumulative adverse effects of both detected and undetected air pollutants. In urban areas, lichens have been mostly used as biomonitors of air quality (Llop et al., 2012; Paoli et al., 2015), and their associations with human health have been explored by Cislaghi and Nimis (Cislaghi and Nimis, 1997) and Ribeiro et al. (Ribeiro et al., 2013). However, both studies were conducted at a regional scale, so the correlations in urban settings remain unexplored. With an increased sample density provided by a lichen diversity biomonitoring network, a more flexible geostatistical approach can be achieved in urban settings using regression Kriging (RK) (Hengl et al., 2007). The RK method can incorporate predictor variables (e.g., land use) to model the relationship with air pollution data using multiple linear regression, while modeling the residual as a stationary random spatial function (Fortin et al., 2012) to predict air pollution concentrations at unsampled locations, using known values of predictors at those locations. Moreover, although Kriging methods do not provide a measure of spatial uncertainty (Isaaks and Srivastava, 1989), they can be used for simulations that provide several maps from which a measure of spatial uncertainty can be drawn (Waller and Gotway, 2004).

The aim of this study was to evaluate alternative cost-effective and useful ways of assessing spatial exposure uncertainty with high-resolution mapping of air quality using a relevant ecological indicator, to measure associations with the health outcome birth weight. Birth weight is usually used as a proxy for infant mortality or morbidity (Kramer, 2003); it has been comprehensively used to measure associations with air pollution. In Europe, the impact of air pollution on birth weight tends to be weak, as the exposure levels are generally below the standards set by national environmental agencies to prevent adverse health effects in the population (Dockery, 1993). Nevertheless, some recent studies (Pedersen et al., 2013; Winckelmans et al., 2015) analyzing the exposure of pregnant women to air pollutants have revealed such impacts on birth weight. Environmental and individual health data used in this study were collected in Sines (Portugal) under the Gestão Integrada da Saúde e Ambiente project (GISA project) developed during 2007–2010, to measure the associations between air quality and birth weight in a much broader area. To overcome the limitation imposed by the low spatial resolution of air-quality data in the city (only one air-quality monitoring station is located within the city limits), we used a sampling network of lichens ($n = 83$) as ecological indicators of air quality. To predict air quality in unsampled locations, we used OK and RK models. To incorporate variations in concentrations at very short distances, we included land-use data, to be used with the RK model. We also incorporated a measure of spatial uncertainty of predicted exposures with a geostatistical simulation algorithm (Sequential Gaussian Simulation (Deutsch and Journel, 1998)). To compare the interpolation results of the two geostatistical methods, we measured the mean error (ME) and root mean square error (RMSE). For both methods, the uncertainty of exposure was drawn from generalized linear models (GLM) fitted to data controlling the effect of other health covariates known to be associated with variations in birth weight.