ARTICLE IN PRESS

Environmental Modelling & Software xxx (2018) 1–9

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Modelling background air pollution exposure in urban environments: Implications for epidemiological research

Álvaro Gómez-Losada [a, *], José Carlos M. Pires [b], Rafael Pino-Mejías [c]

[a] European Commission, Joint Research Centre (JRC), Edificio Expo. C/ Inca Garcilaso 3, 41092, Seville, Spain
[b] LEPABE, Departamento de Engenharia Química, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias s/n, 4200-465, Porto, Portugal
[c] Department of Statistics and Operational Research, University of Seville, Avda. Reina Mercedes s/n, Seville, Spain

A R T I C L E I N F O

Article history:
Received 23 January 2017
Received in revised form
16 January 2018
Accepted 13 February 2018
Available online xxx

Keywords:
Clustering techniques
Background pollution
Air quality
Time-series analysis
Exposure
Health risk

A B S T R A C T

Background pollution represents the lowest levels of ambient air pollution to which the population is chronically exposed, but few studies have focused on thoroughly characterizing this regime. This study uses clustering statistical techniques as a modelling approach to characterize this pollution regime while deriving reliable information to be used as estimates of exposure in epidemiological studies. The background levels of four key pollutants in five urban areas of Andalusia (Spain) were characterized over an 11-year period (2005–2015) using four widely-known clustering methods. For each pollutant data set, the first (lowest) cluster representative of the background regime was studied using finite mixture models, agglomerative hierarchical clustering, hidden Markov models (*hmm*) and k-means. Clustering method *hmm* outperforms the rest of the techniques used, providing important estimates of exposures related to background pollution as its mean, acuteness and time incidence values in the ambient air for all the air pollutants and sites studied.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Determining the population's health risks due to ambient air pollution is critical to the development of effective risk management policies and strategies (Samet and Krewski, 2007). To better understand the adverse health effects associated with air pollution, accurate exposure assessment is essential. Epidemiological studies have provided a substantial body of evidence linking daily concentrations of outdoor air pollution to adverse effects on a range of health outcomes. Studies have tended to focus on the mass concentrations of particles and selected gaseous pollutants, but more insight is required regarding the most harmful sources and components of the air pollution mixture to inform focused public health protection policies (Atkinson et al., 2016).

Background concentration is the ambient level of pollution that is not affected by local sources of pollution (WHO, 1980; Menichini et al., 2007). There are two motivations for focusing on this regime: (i) to better understand the contribution of local sources to total pollutant concentrations; and (ii) to allow the assessment of new pollutant sources that are introduced into the area of study and their impact on local air quality. However, up until now research has not significantly addressed this lowest fraction of pollution as representative of a permanent concentration of ambient air pollution to which the population is chronically exposed. This work focuses on this specific fraction of pollution.

Han et al. (2015) classify the methods to determine the background pollution using four categories: (i) physical methods to identify the regional and local pollution processes via atmospheric variables; (ii) chemical methods to identify the chemical composition of air pollutants; (iii) numerical simulations methods using trajectory models; and (iv) statistical methods. Regarding the latter, Langford et al. (2009) used principal component analysis to describe the local background O_3 concentrations recorded during 76 days in 30 monitoring sites in Texas. Tchepel et al. (2010) study the contributions to background pollution of PM_{10} from different sources in four monitoring sites in Lisbon (Portugal) during two days, through air quality time series via spectral analysis. Other authors have used clustering techniques to characterize regimes in air pollution. Austin et al. (2012) classify air pollution daily data during six years performing k-means (*km*) and hierarchical clustering for identifying profiles in them. Beaver and Palazoglu (2006) used an aggregated solution of *km* to characterize classes of ozone

* Corresponding author.
E-mail addresses: alvaro.gomez-losada@ec.europa.eu, alvaro.gomez.losada@gmail.com (Á. Gómez-Losada).

https://doi.org/10.1016/j.envsoft.2018.02.011
1364-8152/© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article in press as: Gómez-Losada, Á., et al., Modelling background air pollution exposure in urban environments: Implications for epidemiological research, Environmental Modelling & Software (2018), https://doi.org/10.1016/j.envsoft.2018.02.011

episodes occurring in the San Francisco Bay. Considerable effort has been made to characterize profiles of key air pollutants (Carslaw and Ropkins, 2012; Carslaw and Beevers, 2013) since the threshold values that can be considered safe for human health is still under debate. Pioneering research work explored this relationship for O_3 and PM_{10} (Koop and Tole, 2006), and for $PM_{2.5}$ (Kiesewetter et al., 2015). Background profiles of CO and NO_x were studied by Venegas and Mazzeo (2006) in the city of Buenos Aires, and for NO_x, NO_2 and O_3 in the California South Coast Air Basin by Pournazery et al. (2014).

This study proposes the use of statistical clustering techniques as a methodology for the estimation of background pollution in urban environments. To that end, four well-known clustering methods were compared using data obtained from monitoring sites, namely: finite mixture models (*fmm*), agglomerative hierarchical clustering (*hc*), hidden Markov models (*hmm*) and *km*.

This study aims to: (i) evaluate the best clustering statistical method to estimate the background pollution; and (ii) provide model-derived exposure estimates from the best method as inputs for epidemiological research. The best clustering method was assessed according to its ability to cluster the lowest concentrations of ambient air pollution in a consistent manner. To that end, data sets from key pollutants CO, NO_2, O_3 and PM_{10} from five monitoring sites in Andalusia (south of Spain) were studied over 11 years.

2. Data and methods

2.1. Air pollution data

Air quality data (hourly average concentrations of CO, NO_2, O_3 and PM_{10}) were collected from 2005 to 2015 as independent yearly series for each pollutant. These data were obtained at five monitoring sites exhibiting different typology (suburban, urban) and predominant emission sources (background, traffic). Since monitored data were available on an average hourly basis, daily mean concentrations were calculated when at least 80% of the data were available. A total of 200 yearly data sets, each one consisting of daily average values for a single pollutant and complete years were studied, resulting from 40, 55, 50 and 55 data sets corresponding to the air pollutants CO, NO_2, O_3 and PM_{10}, respectively (Table 1). In order to favour the heterogeneity both of data and range of pollutant concentrations to study, monitoring sites were selected in three different cities of Andalusia (Córdoba, Jaén and Seville) with different meteorological conditions governing the local air pollutant behaviour. The standard monitoring methods established in European Directive 2008/50/EC (Directive, 2008) were used for air pollutants CO, NO_2 and O_3, and beta attenuation monitoring was applied for PM_{10}. Air quality monitoring networks are subject to an intense maintenance program to ensure accurate values. Prior to undergoing analysis, the data obtained were validated by the Regional Ministry of Environment and Land Planning of Andalusia.

2.2. Background pollution estimation

For each independent yearly data set with measurements of a single pollutant a clustering technique was applied. For a clustering result, each cluster represents ranges of concentration values (profiles or regimes of pollution) for a given pollutant that can be associated to an emission source of pollution. This view is based on the Lenschow approach (Lenschow et al., 2001) that assumes that the air pollutant concentrations at a monitoring site correspond to the sum of regional, urban background and local nature contributions. This approach has been used as a prior analysis in source apportionment studies (Belis et al., 2013), and may be applied to urban areas with negligible impact from industrial emissions, as in case of Córdoba, Seville and Jaén.

The concentration measured at a traffic site corresponds to the sum of local traffic, urban and regional background contributions. With regard to an urban or suburban background site, the contributions that explain the ambient pollution correspond to those from the background levels of the city or metropolitan area, respectively, and those of the regional background.

Being a univariate clustering process, the resulting clusters represent certain categorization of the original variables into a set

Table 1
Analysed pollutants, classification of monitoring sites and period of study where data were obtained: S-Suburban, U-Urban, B-Background, T-Traffic. Locations are given in X,Y ETRS89-UTM coordinates, zone 30.

City	Site	Location		Type	Main pollution source	Pollutant	Annual periods	Number of data sets
		X	Y					
						CO	2005–07	9
							2010–15	
Córdoba	Asomadilla	343546	4196517	U	B	NO_2		11
						O_3	2005–15	11
						PM_{10}		11
						CO	2007–15	9
Jaén	Bailén	431261	4216416	U	B	NO_2	2005–15	11
						O_3	2010–15	6
						PM_{10}	2005–15	11
						NO_2		11
	Aljarafe	230473	4137017	S	B	O_3		11
						PM_{10}		11
						CO		11
Seville	Bermejales	236063	4137554	S	B	NO_2		11
						O_3	2005–15	11
						PM_{10}		11
						CO		11
	Torneo	234151	4142873	U	T	NO_2		11
						O_3		11
						PM_{10}		+11
								200

Download English Version:

https://daneshyari.com/en/article/6961976

Download Persian Version:

https://daneshyari.com/article/6961976

Daneshyari.com