Spatial and Spatio-temporal Epidemiology 000 (2016) 1–14

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

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

Non-stationary spatio-temporal modeling of traffic-related pollutants in near-road environments

Owais Gilani [a,b], Veronica J. Berrocal [a,*], Stuart A. Batterman [b]

[a] Department of Biostatistics, University of Michigan, School of Public Health, Ann Arbor, MI 48109, United States
[b] Department of Environmental Health Sciences, University of Michigan, School of Public Health, Ann Arbor, MI 48109, United States

A R T I C L E I N F O

Article history:
Received 8 November 2015
Revised 5 March 2016
Accepted 24 March 2016
Available online xxx

Keywords:
Spatial dependence
Non-stationary covariance function
Gaussian processes
Mixture model
Covariates-in-covariance function
MCMC algorithm

A B S T R A C T

A problem often encountered in environmental epidemiological studies assessing the health effects associated with ambient exposure to air pollution is the spatial misalignment between monitors' locations and subjects' actual residential locations. Several strategies have been adopted to circumvent this problem and estimate pollutants concentrations at unsampled sites, including spatial statistical or geostatistical models that rely on the assumption of stationarity to model the spatial dependence in pollution levels. Although computationally convenient, the assumption of stationarity is often untenable for pollutants concentration, particularly in the near-road environment. Building upon the work of Fuentes (2001) and Schmidt et al. (2011), in this paper we present a non-stationary spatio-temporal model for three traffic-related pollutants in a localized near-road environment. Modeling each pollutant separately and independently, we express each pollutant's concentration as a mixture of two independent spatial processes, each equipped with a non-stationary covariance function with covariates driving the non-stationarity and the mixture weights.

1. Introduction

A major source of air pollution in urban environments is combustion of fossil fuels by automobiles. Several studies have observed high concentrations of pollutants on major roads, particularly on urban highways, that rapidly decline within 30–200 m downwind of the highway (Baldwin et al., 2015; Shi et al., 1999; Zhu et al., 2002a; 2002b). Living within close proximity to highways has been associated with numerous adverse health effects in children and adults, including heart rate variability (Adar et al., 2007), cardio-pulmonary mortality (Hoek et al., 2002), acute myocardial infarction (Tonne et al., 2007), development of asthma (Nicolai et al., 2003), reduced lung function (Gauderman et al., 2007), acute respiratory symptoms (Boezen et al., 1999), and DNA damage (Brugge et al., 2007). The associations have generally been attributed to ambient exposure to high levels of traffic-related pollutants concentrations near major roadways and have been corroborated by other epidemiological studies that have found weak or no association between ambient exposure to background concentrations of pollutants and adverse health outcomes (Hoek et al., 2002; Lewis et al., 2005; Venn et al., 2000).

A crucial step in epidemiological studies analyzing the effects of traffic-related air pollutants on health outcomes is assigning pollutants exposure to study participants. Given the high costs of monitoring pollutants concentrations at a fine spatial and temporal resolution, monitoring stations are generally spatially sparse, and limited in the number of pollutants they measure. Thus, to assign an exposure to study participants, it is necessary to develop approaches that estimate pollutant concentrations at unsampled locations, typically subject residences or

* Corresponding author.
E-mail address: berrocal@umich.edu (V.J. Berrocal).

http://dx.doi.org/10.1016/j.sste.2016.03.003
1877-5845/

workplaces, *given* the available monitoring data. These approaches include methods such as assigning pollutants concentrations to participants using either measurements from the nearest monitoring site or averages of measurements from the closest sites (Escamilla-Nuñez et al., 2008; Naeher et al., 1999), or applying interpolation techniques such as inverse distance weighting (Hoek et al., 2002; Jerrett et al., 2009). While these approaches have the advantage of not being extremely complicated, they do not provide a measure of the uncertainty associated with the estimates.

To address these limitations, other approaches have been developed. Land use regression (LUR) (Briggs et al., 2000; Brook et al., 2008; Gehring et al., 2002; Jerrett et al., 2009; Morgenstern et al., 2007) is a method often used in epidemiological studies that consists of expressing a pollutant concentration at a given site as a function of local Geographic Information System (GIS) covariates, such as surrounding land use and traffic characteristics, through a multiple regression framework. While LUR has the ability to capture the mean pollution trend in an intra-urban setting relatively well, being based on a multiple linear regression framework, it implicitly assumes that pollution concentration levels at various sites are independent of each other. In an air pollution context, this assumption might not be realistic, particularly in the near-road environment where typical land use covariates do not exhibit much variability. An approach that addresses this potential limitation of LUR is the geostatistical method of universal kriging. In universal kriging, a pollutant concentration is modeled as a spatial process whose mean trend is expressed as a function of covariates, for example the same GIS covariates used in LUR, and whose covariance function accounts for the spatial dependence in pollution levels at different sites (Finkelstein et al., 2003; Künzli et al., 2005; Pikhart et al., 2001; Son et al., 2010). Typically, the covariance function used to model the spatial dependence in a pollutant concentration is assumed to be stationary, implying that the correlation between pollution levels at two locations is only a function of their separation. Although computationally convenient and widely used, the assumption of stationarity might be untenable for concentrations of traffic-related pollutants, as there might be characteristics of the locations themselves that affect the covariance between pollution levels at any two sites.

Non-stationary spatial modeling, and specifically non-stationary covariance function, has been an active area of research in spatial statistics for the past 20 years. Several methods have been proposed in the literature, with the most used and cited methods including: the "deformation" approach of Sampson and Guttorp (1992), initially presented in a frequentist non-parametric setting and later extended to a Bayesian framework by Damian et al. (2001) and Schmidt and O'Hagan (2003); process convolution (Calder et al., 2002; Higdon, 1998; Higdon et al., 1999; Paciorek and Schervish, 2006); basis function expansions (Holland et al., 1999; Katzfuss, 2013; Matsuo et al., 2011; Nychka et al., 2002; Pintore and Holmes, 2005); kernel mixing (Banerjee et al., 2004; Fuentes, 2001; Fuentes et al., 2005; Fuentes and Smith, 2001); Markov random field models using stochastic partial differential equations (Lindgren et al., 2011); and dimension expansion (Bornn et al., 2012). More recently, in the past 10 years, there has been an increasing effort to develop models for non-stationary covariance functions that incorporate spatial covariates in the covariance function to help identify the factors that drive the non-stationarity. Working within the convolution approach framework of Higdon (1998), Calder (2008) used covariate information to determine and fix the kernel parameters, Neto et al. (2014) proposed a method to include directional variables in the spatially varying kernels, and Risser and Calder (2015) included covariates in the kernels through covariance regression-based methods. On the other hand, taking as starting point the kernel mixing approach of Fuentes (2001), Reich et al. (2011) showed how covariates could be included in the weights of the kernels, while, building on the Bayesian deformation approach of Schmidt and O'Hagan (2003), Schmidt et al. (2011) included covariate information in the covariance function of the spatial process. Finally, Ingebrigtsen et al. (2014) proposed to add covariates in the method proposed by Lindgren et al. (2011).

Applications of the above-mentioned methods for non-stationary covariance functions have generally been devoted to environmental processes for which data are available over a large geographic region (e.g. Higdon, 1998; Fuentes and Smith, 2001; Higdon, 2002; Paciorek and Schervish, 2006; Calder, 2007) that might include natural boundaries such as land and oceans (e.g. Fuentes et al., 2005). However, the need for non-stationary covariance functions can also arise when considering processes within a small, localized spatial domain, such as pollution concentrations across major highways in an urban setting. Due to the effects of wind, traffic-related pollutants concentration gradients are more prominent with distance from highways at downwind sites, so that the covariance at these sites is expected to have a larger effective range as compared to the covariance between sites that are upwind.

In this paper, our focus is on modeling concentrations of three traffic-related pollutants, nitrogen oxides (NO and NO_x) and black carbon (BC), in a localized near-road environment. Previous efforts to model and estimate these pollutants concentrations have generally relied on LUR or spatial statistical models with stationary covariance functions. For example, Aguilera et al. (2007), Beelen et al. (2007), Henderson et al. (2007), Madsen et al. (2007), and Shi and Harrison (1997), Stedman et al. (1997), Madsen et al. (2007), Wilton et al. (2010), and Dons et al. (2013), Montagne et al. (2015) use a LUR to model NO, NO_x and BC, respectively, over large geographical areas. On the other hand, still covering a large spatial domain, Schmidt and Gelfand (2003) jointly model daily average concentrations of NO, carbon monoxide (CO) and nitrogen dioxide (NO_2) over California using a spatial coregionalization model with underlying stationary covariance functions, while, considering a relatively smaller geographical region, Gryparis et al. (2007) and Bliznyuk et al. (2014) propose to use latent processes to combine multiple data sources on BC concentrations in Boston, Massachusetts.

Here, we propose a spatio-temporal model for the concentrations of NO, NO_x, and BC in a near-road environment using data from sites located both upwind and downwind

Please cite this article as: O. Gilani et al., Non-stationary spatio-temporal modeling of traffic-related pollutants in near-road environments, Spatial and Spatio-temporal Epidemiology (2016), http://dx.doi.org/10.1016/j.sste.2016.03.003