Atmospheric Environment 118 (2015) 227–235

Contents lists available at ScienceDirect

Atmospheric Environment

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

Improving spatial nitrogen dioxide prediction using diffusion tubes: A case study in West Central Scotland

Francesca Pannullo [a,*], Duncan Lee [b], Eugene Waclawski [c], Alastair H. Leyland [a]

[a] MRC|CSO Social and Public Health Science Unit, University of Glasgow, 200 Renfield Street, Glasgow, G2 3QB, UK
[b] School of Mathematics and Statistics, University of Glasgow, Glasgow, G12 8QW, UK
[c] Public Health, University of Glasgow, G12 8SJ, UK

HIGHLIGHTS

- Monitors and diffusion tubes that measure NO_2 are spatially sparse.
- Modelled concentrations are also available, but are known to contain biases.
- We utilise all three sources of data to predict fine scale NO_2 concentrations.
- A geostatistical fusion model is utilised implemented within a Bayesian setting.
- Addition of diffusion tubes enhances the predictive performance of fine scale NO_2.

ARTICLE INFO

Article history:
Received 30 March 2015
Received in revised form
9 July 2015
Accepted 3 August 2015
Available online 8 August 2015

Keywords:
Bayesian fusion modelling and prediction
Nitrogen dioxide
Spatial prediction

ABSTRACT

It has been well documented that air pollution adversely affects health, and epidemiological pollution-health studies utilise pollution data from automatic monitors. However, these automatic monitors are small in number and hence spatially sparse, which does not allow an accurate representation of the spatial variation in pollution concentrations required for these epidemiological health studies. Nitrogen dioxide (NO_2) diffusion tubes are also used to measure concentrations, and due to their lower cost compared to automatic monitors are much more prevalent. However, even combining both data sets still does not provide sufficient spatial coverage of NO_2 for epidemiological studies, and modelled concentrations on a regular grid from atmospheric dispersion models are also available. This paper proposes the first modelling approach to using all three sources of NO_2 data to make fine scale spatial predictions for use in epidemiological health studies. We propose a geostatistical fusion model that regresses combined NO_2 concentrations from both automatic monitors and diffusion tubes against modelled NO_2 concentrations from an atmospheric dispersion model in order to predict fine scale NO_2 concentrations across our West Central Scotland study region. Our model exhibits a 47% improvement in fine scale spatial prediction of NO_2 compared to using the automatic monitors alone, and we use it to predict NO_2 concentrations across West Central Scotland in 2006.

1. Introduction

The relationship between air pollution concentrations and ill health has been well documented in the past two decades, with epidemiological studies focussing on the effects of both short-term and long-term exposure. The most common study design is an ecological time series study, such as Omori et al. (2003) and Moolgavkar et al. (2013), which examines the effects of short-term exposure by regressing routinely available air pollution and disease data collected at daily intervals. The health impact of long-term exposure has typically been estimated using cohort studies such as Dockery et al. (1993) and Jerrett et al. (2009), but they are expensive and time consuming to implement due to the length of time required for monitoring the health status of the cohort. Therefore, spatial ecological studies such as Lee et al. (2009) and Haining et al. (2010) have recently been used to estimate the long-term effects of air pollution on human health, which regress variation in disease risk and air pollution across small areal units such

* Corresponding author.
E-mail address: fpannullo@sphsu.mrc.ac.uk (F. Pannullo).

http://dx.doi.org/10.1016/j.atmosenv.2015.08.009

as electoral wards using routinely collected health data.

The health data considered in these studies have included mortality (Kinney and Ozkaynak, 1991) and morbidity, such as hospital admissions (Willocks et al., 2012), for a number of common diseases, such as respiratory (Atkinson et al., 2001) and cardiovascular conditions (Larrieu et al., 2007). Numerous pollutants have been associated with disease in these studies, including carbon monoxide (CO, (Villeneuve et al., 2003)), nitrogen dioxide (NO_2, (Raaschou-Nielsen et al., 2012)), ozone (O_3, (Tao et al., 2012)), particulate matter (such as PM_{10} (Rushworth et al., 2014) and $PM_{2.5}$, (Cesaroni et al., 2013)) and sulphur dioxide (SO_2, (Wong et al., 2008)). The pollution data used in these studies typically come from a small number of automatic monitors, each of which measures concentrations of the above pollutants at a single point in space. However, the number of monitors is few and their geographical positioning is sparse, which does not allow an accurate representation of the spatial variation in pollution concentrations required for the epidemiological studies, particularly cohort and spatial ecological studies. For cohort studies, concentrations are required at the residence of each member in the cohort, while for spatial ecological studies concentrations are required for each spatial unit at which health data are available. These fine scale pollution data are not available, for example in our Glasgow region there are only 16 monitors covering the 368 square kilometre study region. Therefore, inexpensive non-automatic diffusion tubes are also used to measure concentrations of NO_2, and due to their lower cost compared with automatic monitors they are more prevalent. For example, in the Glasgow study region considered here there are 230 diffusion tubes, which thus provides greatly enhanced spatial coverage compared with using the 16 automatic monitors alone.

However, combining these two data sets still does not give complete spatial coverage of a study region, which is illustrated in our case study in Fig. 1. Therefore, modelled concentrations from atmospheric dispersion models are increasingly being used in health studies (Naess et al., 2007), as they provide estimated concentrations on a regular grid and thus have complete spatial coverage of the study region. However, these modelled concentrations are known to contain biases (Berrocal et al., 2010a), and are not as accurate as the measured pollution data. Therefore, there has been recent research interest in statistical fusion models (Berrocal et al., 2010b; Fuentes and Raftery, 2005), which use both the measured and modelled concentrations to predict pollution at fine spatial and temporal scales. There are two main types of statistical fusion models, namely: the regression approach, and the latent process approach. The regression approach, as utilised by Berrocal et al. (2010a, 2010b), Bruno et al. (2013), combines monitored and modelled concentrations by regressing the concentrations from the monitoring stations against the modelled concentrations via a spatially varying linear regression. The latent process approach, utilised by Fuentes and Raftery (2005), Fuentes et al. (2008), McMillan et al. (2009), Sahu et al. (2010), represents the true environmental factor, and drives both the observed and modelled data assuming that both the modelled and observed data provide good information about the same underlying process. These fusion models correct for the biases inherent in atmospheric dispersion models, and also provide associated measures of uncertainty for the predictions, which are typically not available from dispersion models. However, the majority of these models only make use of measured data from automatic monitors, which as previously discussed are spatially sparse.

Therefore in this paper we propose a geostatistical fusion model, that regresses the combined NO_2 concentrations from both automatic monitors and diffusion tubes against modelled NO_2 pollution data from an atmospheric dispersion model. This model is implemented within a Bayesian setting and predicts NO_2 concentrations across the Glasgow region, for use in a future health study. This is thus the first paper to demonstrate the dramatic improvement in fine scale spatial prediction of NO_2 that can be obtained by using abundant diffusion tube data that is relatively inexpensive to collect in addition to the small numbers of automatic monitors.

The remainder of this paper is organised as follows. Section 2 describes the study design of our Glasgow case study, specifically the spatial extent of the region of interest and the NO_2 and covariate data. Section 3 presents the geostatistical fusion model for predicting NO_2 concentrations across the study region proposed here, and discusses its implementation. Section 4 presents the results of our analyses, including a model validation exercise that compares our proposed model against a number of other candidate models, and a fine scale prediction of NO_2 across the Glasgow region. Finally, Section 5 provides a concluding discussion.

2. Study design

2.1. Study region

Our study region is centred around the Greater Glasgow conurbation, which is the largest city in Scotland, UK. The Glasgow conurbation contains just under one quarter of the total Scottish population, equating to around 1.1 million people, with a land area of around 368 km^2. Seven local authorities comprise the study region, namely: East Renfrewshire, Glasgow City, North Lanarkshire, Renfrewshire, South Lanarkshire, and West Dunbartonshire. These local authorities have been selected because they cover the city of Glasgow and the wider area, collectively known as *West Central Scotland*, and include both urban and rural areas leading to a wide variation in pollution concentrations across the study region.

2.2. Air pollution data

The air pollution data comprise annual mean concentrations of nitrogen dioxide (NO_2, measured in μgm^{-3}) in 2006, for which two sources of data are available. The first source is measured data at fixed points in space, which come from both automatic monitoring stations and non-automatic diffusion tubes. NO_2 concentrations from the automatic monitors were downloaded from the Scottish Air Quality website (www.scottishairquality.co.uk), while the non-automatic diffusion tube data were obtained on request from each local authority. The accuracy of the diffusion tubes vary depending on numerous factors, such as the preparation methodology used, and are therefore calibrated using a bias-adjustment factor obtained from co-location studies. These concentrations are measured at 246 sites across the study region, of which 230 are diffusion tubes and 16 are automatic monitors. The locations of these sites within the study region are displayed in Fig. 1, where the diffusion tubes are displayed as crosses and the automatic monitors are presented as triangles. Summary statistics for the measured data are shown in Table 1. These statistics highlight that the distribution of concentrations across the study region are slightly higher for monitors compared to diffusion tubes with median values of 34.55 μgm^{-3} and 29.95 μgm^{-3} respectively. This could be due to the local authorities placing automatic monitors where they have a compliance problem with EU pollution standards.

Fig. 1 shows that the measured data are sparse, and do not provide complete spatial coverage of the study region. In addition, the automatic monitors and diffusion tubes are also likely to be preferentially located in areas where pollution is thought to exceed EU standards, therefore producing an inflated picture of area wide pollution concentrations. Furthermore, the monitoring network often suffers from missing data, arising from monitors and tubes becoming faulty. Nevertheless, as the concentrations are directly