International Journal of Hygiene and Environmental Health 218 (2015) 107–116

Contents lists available at ScienceDirect

International Journal of Hygiene and Environmental Health

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

Modelling personal exposure to particulate air pollution: An assessment of time-integrated activity modelling, Monte Carlo simulation & artificial neural network approaches

A. McCreddin, M.S. Alam, A. McNabola*

Dept of Civil, Structural & Environmental Engineering, Trinity College, Dublin, Ireland

ARTICLE INFO

Article history:
Received 18 March 2014
Received in revised form 19 August 2014
Accepted 28 August 2014

Keywords:
Air pollution modelling
PM_{10}
Personal exposure
Monte Carlo simulation
Neural networks
Time–activity models

ABSTRACT

An experimental assessment of personal exposure to PM_{10} in 59 office workers was carried out in Dublin, Ireland. 255 samples of 24-h personal exposure were collected in real time over a 28 month period. A series of modelling techniques were subsequently assessed for their ability to predict 24-h personal exposure to PM_{10}. Artificial neural network modelling, Monte Carlo simulation and time–activity based models were developed and compared. The results of the investigation showed that using the Monte Carlo technique to randomly select concentrations from statistical distributions of exposure concentrations in typical microenvironments encountered by office workers produced the most accurate results, based on 3 statistical measures of model performance. The Monte Carlo simulation technique was also shown to have the greatest potential utility over the other techniques, in terms of predicting personal exposure without the need for further monitoring data. Over the 28 month period only a very weak correlation was found between background air quality and personal exposure measurements, highlighting the need for accurate models of personal exposure in epidemiological studies.

Introduction

The personal exposure of individuals or population sub-groups to air pollution is a multifaceted function of air pollution sources, physical and chemical atmospheric processes, as well as duration of personal activity patterns. As such the direct prediction of personal exposure to air pollution is a complex and challenging problem. In practice it may be impractical to attempt to model such a myriad of complex processes directly, and more usefully statistical or stochastic models of personal exposure can be developed.

The personal exposure of an individual to air pollution is a key indicator of the resulting health impacts. However, as noted in literature the monitoring of personal exposure is often not feasible due to the costs involved, in terms of expense, time and resources (Hoek et al., 2008; Molter et al., 2012; Yann et al., 2014). As a cheaper and more readily available alternative to personal exposure measurements, background air quality data is often used to represent the exposure of individuals or groups. This assumption has been the basis for a significant body of epidemiological evidence on the links between air pollution exposure and health impacts (Dockery et al., 1993), and has produced significant results in the field of air pollution science.

However, it is often highlighted in literature that problems exist in the use of ambient air quality data to represent the personal exposure of individuals (Boudet et al., 2001; Steinle et al., 2013). Potential differences between the assumed background air quality measurements and actual personal exposure at the individual level may present a weakness in current epidemiological models. Yann et al. (2014), recently illustrated the differences between the use of various air quality models to represent personal exposure data, and the use of background monitoring data, on the associations between air pollution exposure among pregnant women and infant birth weight. The investigators concluded that differing spatio-temporal approaches to air pollution exposure assessment provided well correlated predictions but these resulted in different rankings of the study subjects regarding their exposure, and that this may lead to different conclusions regarding associations with health outcomes. To improve this situation, without the need for additional expensive and resource intensive personal exposure monitoring, more reliable and accurate models of personal exposure are required.

Numerous approaches to the prediction of personal exposure to air pollution using models have been published in literature. Common approaches have included the use of time-integrated activity

* Corresponding author. Tel.: +353 1 896 3837; fax: +353 1 6773072.
E-mail address: amcnabol@tcd.ie (A. McNabola).

http://dx.doi.org/10.1016/j.ijheh.2014.08.004
1438-4639/

modelling, where the total 24-h personal exposure is modelled as a sum of the product of the time spent in each microenvironment and its corresponding pollutant concentration. Chang-Fu et al. (2005) developed models of the personal exposure of children to $PM_{2.5}$ based on data collected using personal nephelometers. A 'time–space model' based on the concentration of PM at a fixed site monitor and factors representing time activity patterns, season and distance from home produced the highest R^2 achieved of 0.41. Gerharz et al. (2009) proposed a novel personal exposure modelling system for $PM_{2.5}$ based on activity diaries, GPS tracks and existing outdoor and indoor air pollution models. Data collected from 6 subjects showed that it was possible to model personal exposure using a combination of existing models and measurement networks.

Similar approaches were taken by Molter et al. (2012) in developing and assessing a micro-environmental model of personal exposure to NO_2. Here the use of a combination of land use regression modelling and indoor air quality modelling, together with temporal variation factors were used to predict the exposure of children. Based on 60 participants the exposure model produced good agreement with measured data.

Stochastic techniques such as the use of artificial neural networks (ANN) have also been used to predict personal exposure based on the analysis of historic data records. Ibarra-Berastegi et al. (2008) used ANNs to predict hourly concentrations of five urban pollutants in Bilbao up to 8 h ahead of background measurements. The performance of these models varied depending on pollutant type and the background monitor in question (R^2 = 0.15 to 0.88). Karakitsios et al. (2007) used ANNs to predict the exposure patterns of filling station workers based on personal exposure data, meteorology, traffic and other activity factors. Very high levels of agreement between the modelled and measured data (R^2 > 0.81).

This paper presents the results of an investigation into alternative methods of personal exposure modelling. Based on personal exposure data collected from a number of office workers using real-time nephelometers the performance of three techniques of personal exposure modelling were compared: (i) time-integrated activity modelling, (ii) Monte Carlo simulation, and (iii) neural network modelling.

Methodology

Volunteer recruitment

A 24-h personal exposure monitoring campaign was undertaken for a period of 28 months from February 2009 to June 2011. A total of 59 subjects were recruited on a voluntary basis completing 255 24-h sampling periods. The recruitment of subjects was restricted to office workers living and working in the Greater Dublin Area, in order to limit the extent of variation in personal exposure among the sample population. The study population was 57% male and 43% female. The majority (48%) of subjects were aged 26 to 35 years, with 27% in the 18 to 25 years category and the remainder between 36 and 55. Approximately 12% of subjects declared themselves as smokers of some degree, or were in residence with a smoker. Samples were only recorded on weekdays.

Sampling method

Sampling of personal exposure, activity and location of subjects was carried out using a real time particulate matter (PM_{10}) sampling device (Met One Aerocet-531 particle profiler), GPS tracking equipment (Garmin GPSMAP® 60CSx), and a personal activity diary. Particulate Matter (PM) was chosen as the main pollutant to be monitored due to its health significance, its multisource nature (indoor and outdoor environments), and the ability to record its concentration using real-time monitors that are small and mobile whilst maintaining sufficient resolution and accuracy.

Table 1
The input parameters for the FFNN and GRNN models.

Activity-Time Budget Inputs	Categorical inputs
Bus	Gender
Café/restaurant	Smoker
Car	
Cycling	Meteorological inputs
Home cooking	Barometer level pressure (hPa)
Other indoor	Global radiation (j/cm^2)
Other outdoor	Rainfall (mm)
Pub	Relative humidity (%)
Recreation/sport	Sea level pressure (hPa)
Shopping	Temperature (°C)
Sleeping	Wind direction (Cosine)
Train	Wind direction (Sine)
Tram	Wind speed (m/s)
Walking	
Working	

The Aerocet-531 is a real-time automatic particulate matter monitor capable of recording concentrations of PM_{10} at 2 min intervals. The instrument used a laser diode with a right angle scatter method at 0.78 μm. The light travelled at a right angle to the collection sensor and detector, and the instrument used the information collected from the scattered particles to calculate a mass per unit volume. A mean particle diameter was recorded and was used to calculate a volume in cubic meters, which was then multiplied by the number of particles and a generic density that was representative of typical aerosols. The calculated mass was then divided by the volume of air sampled to obtain mass per unit volume (μg/m^3).

The GPS device used as part of this study was chosen because of its high sensitivity receiver which meant it could easily and quickly obtain GPS satellite signal in an urban landscape. It was also a small handheld device which made it convenient for volunteers to carry on their person along with the Aerocet-531 instrument, bringing the total weight of the sampling equipment to approximately 1.1 kg.

The activities of the volunteers were also monitored through use of an activity log. Each volunteer was instructed to record, in as much detail as possible, the time of day they partook in a certain activity or were in a specific location. This information was then used to divide up the particulate concentrations recorded by the Aerocet-531 and assign them to defined activity groups. A list of the 15 main microenvironments/activity groups encountered in the study is given in Table 1. The sampling campaign and resulting data analysis are described in detail in McCreddin et al. (2013). In addition, a description of the quality control procedures adopted during this sampling campaign is also included in the Supplementary information section of this paper.

Model development

There were three approaches taken to personal exposure modelling which were based on (i) time–activity weighted, (ii) Monte Carlo simulation, and (iii) neural network modelling techniques. The performance of each of the models was validated by splitting the personal exposure dataset recorded into a model development dataset and a validation dataset. The model development dataset consisted of 230 24-h samples, which represented 90% of the overall dataset, while the validation dataset consisted of the remaining 10%. The validation data was chosen from the main dataset using by a specially developed algorithm in *MATLAB*. The *MATLAB* code randomly chose and removed 25 sampling days from the main dataset and stored them in separate file for use later during testing.