Atmospheric Environment 42 (2008) 5933–5937

Contents lists available at ScienceDirect

Atmospheric Environment

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

Technical note

Determining the spatial scale for analysing mobile measurements of air pollution

Christy Lightowlers *, Trisalyn Nelson, Eleanor Setton, C. Peter Keller

Spatial Sciences Research Laboratory, Geography Department, University of Victoria, PO BOX 3050 STN CSC, Victoria, BC, Canada V8W 3P5

ARTICLE INFO

Article history:
Received 15 October 2007
Received in revised form
13 March 2008
Accepted 17 March 2008

Keywords:
Spatial autocorrelation
Semivariogram
Air pollution
Spatial scale
Mobile monitoring

ABSTRACT

When dealing with spatial data or modelling in a geographical context, identifying an appropriate scale for analysis is a critical precursor; however, it is difficult to determine due to limited availability of data at an adequate spatial resolution. This paper describes a mobile monitoring method to collect spatially representative measurements of woodsmoke particulates in support of spatial modelling. A geostatistical technique is described to characterize the spatial scale of woodsmoke particulates collected for 19 evenings over two heating seasons in Victoria, British Columbia, Canada. Semivariograms were applied to 20 data sets (19 evenings and a combined data set) to characterize the appropriate spatial-analysis scale as defined by the semivariogram range, the maximum distance of spatial dependence. Typically, the semivariogram range occurred at 2673 m. This method can be used to identify an optimal sampling interval for woodsmoke data collection, to define the neighbourhood size for performing spatial analyses, and to produce robust model variables and parameters by characterizing the degree of spatial autocorrelation in the data set.

1. Introduction

Woodsmoke contains substances known to harm human health and is a major contributor to air pollution in many parts of the world, yet knowledge on the degree of risk, the carcinogenic effects, the impacts of long-term exposure and the biological mechanisms linking woodsmoke to health outcomes is limited (Boman et al., 2003, 2006; Naeher et al., 2007; Zelikoff et al., 2002). Spatial modelling of exposure to woodsmoke is a practical approach to addressing this lack of understanding by enabling epidemiological analysis, risk assessment, and providing a tool to inform air quality management and planning.

Air pollution varies geographically due to the location of emission sources, topography, and weather conditions; yet most studies of air pollution and health are aspatial (Jerrett and Finkelstein, 2005). The inability to incorporate spatial variability is cited as a major deficiency in studies of air pollution and health (Brauer et al., 2003; Briggs et al., 2000; Hoek et al., 2002). Studies examining the relationship between air pollution and health typically characterize exposure for a population using measurements from a few sparsely located air quality monitoring stations, and often only one. Mobile monitoring is a viable response to this issue by providing the potential for measuring air pollution at many locations (Robinson et al., 2007; Larson et al., 2007) resulting in an extensive spatial data set.

As spatial data on air pollution become more available, new challenges for analysis of these data must be addressed. For instance, since the results of spatial studies are dependent on the scale of spatial analysis, it is essential to identify an appropriate scale when characterizing air pollution exposure. Failure to do so can obscure the underlying spatial processes linked to

* Corresponding author. Tel.: +1 250 472 4624; fax: +1 250 721 6216.
E-mail address: c.lightowlers@shaw.ca (C. Lightowlers).

1352-2310/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2008.03.033

health outcomes (Jerrett and Finkelstein, 2005). Impacts of scale variations on results in epidemiological analysis has been demonstrated by Jerrett et al. (2005) in Los Angeles where health effects analysed at the local census scale were three times greater when compared with the air pollution and health relationship characterized at the city-wide scale. These results suggest that air pollution and health studies conducted at an inappropriate scale provide misleading conclusions. At present, little is known about the spatial scale of air pollutants; however, the spatial scale of particulate matter is known to vary depending on the source and ranges from 50 m to 4 km for primary combustion sources to hundreds of kilometers for sulphate particulates (Gilliland et al., 2005).

Spatial analysis such as hot-spot detection, kriging or spatial modelling, employ a set search radius, or neighbourhood size. Neighbourhoods in spatial analyses are often defined by distance and therefore should be informed by the scale of the process under investigation to avoid pitfalls associated with working with spatial data (O'Sullivan and Unwin, 2003). For example, spatial dependence (or spatial autocorrelation), where observations in close proximity are more related than those further away, is inherent in spatial data. When spatial dependence is ignored, it has been shown to affect the prediction accuracy of spatial models and in determining significant model variables (Anselin, 2005; O'Neill et al., 2003).

This paper describes a mobile monitoring method to support spatial modelling at a fine resolution and describes a method for quantifying the spatial scale of woodsmoke for spatial analysis using semivariogram analysis. Methods for quantifying spatial relationships using Geographic Information Systems (GIS) and semivariogram analysis are applied to woodsmoke particulate data collected in Victoria, British Columbia (BC). The influence of weather on spatial dependence is examined by analysing different meteorological conditions.

2. Data description and methods

2.1. Data

Victoria is located on Vancouver Island, BC, Canada (Fig. 1) where wood burning for residential heating has been identified as a health concern. Measurements of light scatter-off particulate matter $<2.5\,\mu$m in diameter ($PM_{2.5}$) were collected with a Radiance Research M903 nephelometer installed in a passenger vehicle for 19 evenings during the winter heating seasons (November to the end of March) of 2004/05 and 2005/06. At the same time, positional data were collected with a Delorme Bluelogger Global Positioning System (GPS) receiver.

Measurements were taken at 15 s intervals to capture the spatial and temporal variability of woodsmoke. The nephelometer logged the average values measured in the preceding 15 s. With travelling speeds of 40–60 km h^{-1}, the logged value represents an average measured over 165–255 m, a sufficient resolution for this regional study.

Mobile nephelometer and GPS data were collected during evenings (between 9 and 11 pm) in the burn seasons (November through March), using a route that covers the area expected to be impacted by smoke (Fig. 1 displays the sampling routes). At the end of each sample evening, nephelometer and GPS data were downloaded and combined with GPS data based on time. The sample evening data sets were converted to point data in a GIS. Regional temperature, precipitation, and wind speed data used to determine the stability of weather conditions were obtained from the Environment Canada airport station.[1] The number of measurements per evening ranged from $n = 280$ to 509 and was typically around 350. Light-scatter measurements ranged from 0.000005 to 0.00207.

A nephelometer was stationed alongside a $PM_{2.5}$ fixed monitor (TEOM) for 1 week to measure the correlation between nephelometer and $PM_{2.5}$ measurements ($R = 0.87$, $\alpha = 0.01$). Converting the light-scatter value (b_{sp}) to a $PM_{2.5}$ concentration was calculated using the following formula ($R^2 = 0.76$, $p < 0.00$):

$$b_{sp} = 2.96\text{E}-06 + 4.17\text{E}-06(\text{TEOM PM}_{2.5}) \quad (1)$$

Stationary Partisol Model 2000 Air Samplers with R&P $PM_{2.5}$ Cyclone inlets were placed throughout the study region to measure levoglucosan, a chemical unique to the presence of woodsmoke (Larson et al., 2007). Sixteen sampling periods of 12 h durations occurred at four different sites (eight from 7 am to 7 pm, and eight from 7 pm to 7 am) to measure differences between day and nighttime levels of levoglucosan (Fig. 2). Twelve-hour periods were chosen to ensure sufficient levoglucosan levels were deposited on the filter for analysis. Levoglucosan results demonstrated elevated levoglucosan levels in $PM_{2.5}$ in the evenings which is reflected in the diurnal pattern of $PM_{2.5}$ in the winter heating season compared to the summer non-heating season (Fig. 3).

2.2. Semivariogram analysis

Semivariograms were applied to identify the distance where data are no longer spatially autocorrelated. The semivariogram is a function of distance and is based on the average sum of squared differences in attribute values for all pairs of points that are a defined distance apart. It is estimated by

$$\hat{\gamma}(d) = \frac{1}{2n(d)} \sum_{\mathbf{s}_i - \mathbf{s}_j = d} (z_i - z_j)^2 \quad (2)$$

where γ is the symbol for a semivariogram, and z_i and z_j are the attribute values of points s_i and s_j. The summation is over all pairs of points that are separated by a distance d and $n(d)$ is the number of these pairs (O'Sullivan and Unwin, 2003). To reduce the number of points on the semivariogram, pairs of locations are binned based on their distance from each other. The number of pairs in each lag were never <40 in this study to ensure statistical reliability (Burrough and McDonnell, 1998).

[1] http://www.wunderground.com/history/airport/CYYJ/.

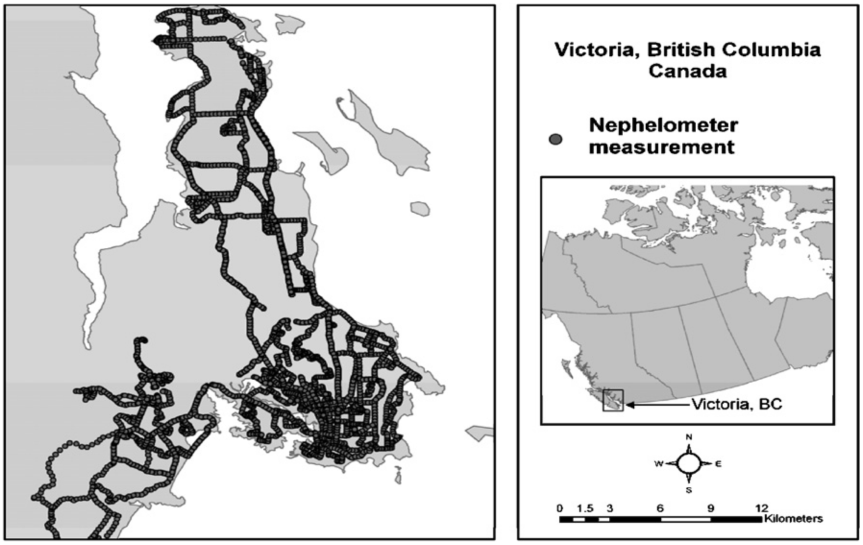

Fig. 1. The study area and nephelometer measurements.

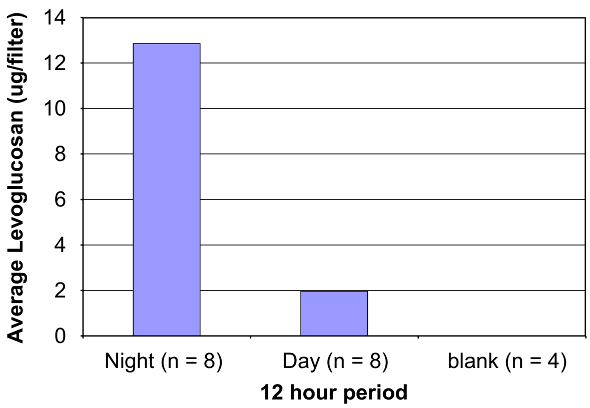

Fig. 2. Average levoglucosan levels for 12 h sampling periods from Partisols located throughout Victoria, BC during March 2007.

Spatial dependence is summarized by fitting a model to the empirical semivariogram. Semivariance increases with distance reaching a level called the sill. The range is the distance where the sill is reached and describes the distance within which the variable is dependent (Diem, 2003). Beyond the range, there is no spatial structure in the data, and for woodsmoke particulates, the range defines the analysis scale.

Semivariograms were applied to point data from the nephelometer. Semivariograms are sensitive to the input parameters. Changing the bin distance can have an impact on the range value and defining the maximum semivariogram distance impacts the goodness of fit for the model. In order to compensate for this sensitivity, semivariograms were calculated at various lags and distances to determine the impact on semivariogram-model parameters. The range was estimated for each sample evening and for the combined data set (all 19 evenings were appended into one GIS layer). To investigate the effect of weather conditions on the range of spatial dependence fitted-model parameters were compared between calm evenings (wind speed $<8\,\mathrm{km\,h^{-1}}$), calm evenings with clouds/fog (wind speed $<8\,\mathrm{km\,h^{-1}}$ plus clouds/fog) and non-calm conditions (wind speed $>8\,\mathrm{km\,h^{-1}}$).

3. Results

With the exception of two evenings, the models fit to semivariograms computed with several different input combinations produced similar model parameters. According to Englund and Sparks (1991), there is confidence in model parameters that fit semivariograms computed at different lag intervals. Varying the input parameters (lag size and maximum distance) made little difference to the fitted-model parameters, which allowed the selection of common input parameters to calculate semivariogram-model parameters for each evening and the evenings combined into one data set. The results included in this section are for models calculated using a bin distance of 500 m up to a maximum distance of 5 km. A spherical model was chosen because it provided the best fit as indicated by the objective value goodness of fit test.

The semivariogram model for the combined data set is shown in Fig. 4. The range is 2750 m, indicating the spatial scale of analysis for the heating season. Fig. 5 shows a typical semivariogram and fitted spherical model for a sample evening (8 February 2005). Summary statistics for the 19 sample evenings are provided in Table 1. Results show an average range of 2673 m. This distance is similar to the range of the combined data set suggesting that spatial processes on individual evenings reflect those for the heating season. Because air pollutant concentrations are affected by atmospheric stability (Larson et al., 2007), the range of different weather conditions were also examined. There is a slight decrease in the range as stability decreases; however, there are not enough evenings to determine if the decrease is significant

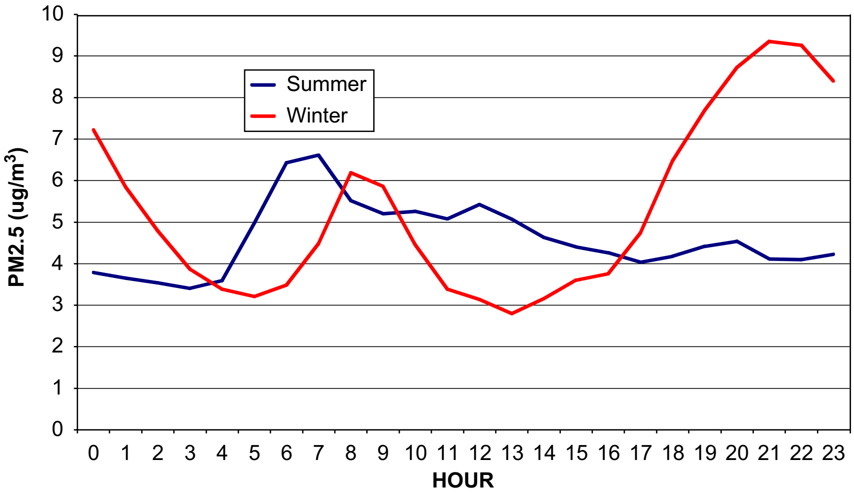

Fig. 3. Diurnal $PM_{2.5}$ concentration during the winter heating season and summer non-heating season in Victoria, BC (data are from the BC Ministry of Environment fixed monitoring network and represent the hourly averages for 2003/05, 2004/05, and 2005/06).

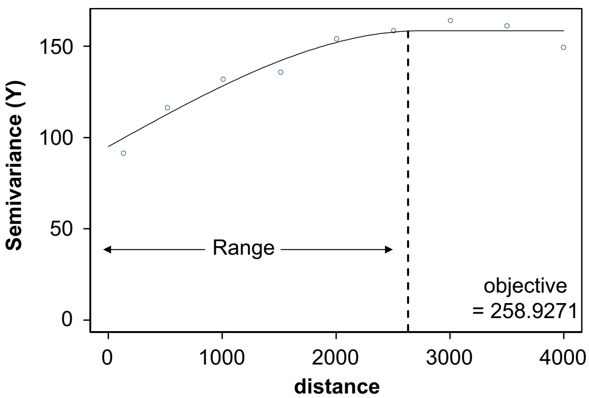

Fig. 4. Semivariogram for combined data set fitted with a spherical model.

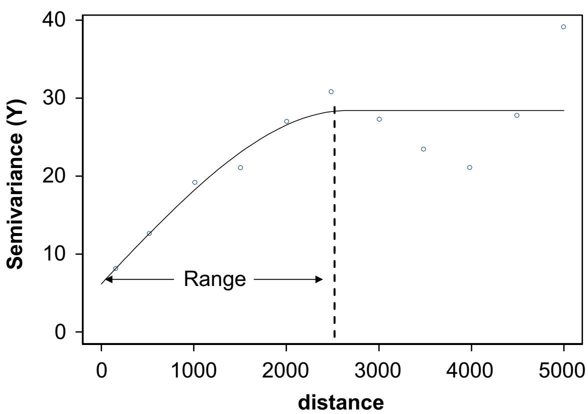

Fig. 5. Semivariogram for a typical evening fitted with a spherical model.

(Table 2). As atmospheric stability decreases, the maximum distance at which spatial dependence occurs appears to become smaller. In other words, windy

Table 1
Summary statistics for fitted model parameters calculated using 500 m lags to a maximum distance of 5000 m ($n = 19$)

Summary statistics	Range (m)	Nugget (γ)	Sill (γ)
Mean	2673	10.31	40.08
Coefficient of variation	0.36	1.07	0.79
Median	2593	7.83	28.62
Min	932	1.27	6.78
Max	4222	43.8	130.93

Table 2
Mean-model parameter values for different weather conditions

Weather conditions	Range (m)	Coefficient of variation
Not calm ($n = 5$)	2486	0.34
Calm plus fog/clouds ($n = 9$)	2696	0.41
Calm ($n = 5$)	2811	0.34

conditions cause the spatial pattern of woodsmoke to be more localized.

4. Discussion and conclusions

Identifying an appropriate scale for analysis of air pollution data is a critical, but much neglected component of spatial modelling. Knowledge of the spatial scale can be applied to identify an optimal sampling interval for woodsmoke data collection, define the neighbourhood size for performing spatial analyses, and facilitates the production of robust models by characterizing the degree of spatial autocorrelation in the data set.

The semivariogram analysis revealed that spatial dependence exists up to approximately 2673 m, when averaged across individual evenings and heating seasons, and did not vary substantially between different weather conditions. This analysis scale provides a defensible neighbourhood size for performing spatial analysis of the data, such as kriging. For example, if the distance separating an unmeasured site and a measured observation is greater than the range, the measured observation should not contribute to the estimation of that unmeasured location (Burrough and McDonnell, 1998). The analysis scale also helps identify social and environmental variables with similar scales which are most likely to affect woodsmoke levels, and therefore be important to include in predictive or explanatory models. In addition, modelling at an appropriate scale lends credibility to model performance and model variables which are affected by spatial autocorrelation.

Acknowledgements

The research was supported in part by Health Canada via an agreement with the British Columbia Centre for Disease Control to the Border Air Quality Study as well as the Vancouver Island Health Authority, the Capital Regional District and the BC Ministry of Environment.

References

Anselin, L., 2005. Exploring Spatial Data with GeoDa: A Workbook. Centre for Spatially Integrated Social Science, Urbana, IL.

Boman, B., Forsberg, A., Jarvholm, B., 2003. Adverse health effects from ambient air pollution in relation to residential wood combustion in modern society. Scandinavian Journal of Work, Environment and Health 29 (4), 251–260.

Boman, C., Forsberg, B., Sandstrom, T., 2006. Shedding new light on wood smoke: a risk factor for respiratory health. European Respiratory Journal 27 (3), 446–447.

Brauer, M., Hoek, G., VanVliet, P., Meliefste, K., Fischer, P., Gehring, U., Heinrich, J., Cyrys, J., Bellander, T., Lewne, M., Brunekreef, B., 2003. Estimating long-term average particulate matter air pollution concentrations: application of traffic indicators and geographic information systems. Epidemiology 14 (2), 228–239.

Briggs, D.J., de Hoogh, C., Guiliver, J., Wills, J., Elliott, P., Kingham, S., Smallbone, K., 2000. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. Science of the Total Environment 253 (1–3), 151–167.

Burrough, P., McDonnell, R., 1998. Principles of Geographical Information Systems. Oxford University Press, New York.

Diem, J.E., 2003. A critical examination of ozone mapping from a spatial-scale perspective. Environmental Pollution 125 (3), 369–383.

Englund, E., Sparks, A., 1991. Geostatistical Environmental Assessment Software User's Guide. United States Environmental Protection Agency, Las Vegas.

Hoek, G., Meliefste, K., Cyrys, J., Lewne, M., Bellander, T., Brauer, M., Fischer, P., Gehring, U., Heinrich, J., van Vliet, P., Brunekreef, B., 2002. Spatial variability of fine particle concentrations in three European areas. Atmospheric Environment 36 (25), 4077–4088.

Gilliland, F., Avol, E., Kinney, P., Jerrett, M., Dvonch, T., Lurmann, F., Buckley, T., Breysse, P., Keeler, G., de Villiers, T., McConnell, R., 2005. Environmental Health Perspectives 113, 1447–1454.

Jerrett, M., Finkelstein, M., 2005. Geographies of risk in studies linking chronic air pollution exposure to health outcomes. Journal of Toxicology and Environmental Health—Part A—Current Issues 68 (13–14), 1207–1242.

Jerrett, M., Burnett, R., Ma, R., Pope, C., Krewski, D., Newbold, K., Thurston, G., Shi, Y., Finkelstein, N., Calle, E., Thun, M., 2005. Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology 16 (6), 727–736.

Larson, T., Su, J., Baribeau, A., Buzzelli, M., Setton, E., Brauer, M., 2007. A spatial model of urban winter woodsmoke concentrations. Environmental Science and Technology 41 (7), 2429–2436.

Naeher, L., Brauer, M., Lipsett, M., Zelikoff, J., Simpson, C., Koenig, J., Smith, K., 2007. Woodsmoke health effects: a review. Inhalation Toxicology 19 (1), 67–106.

O'Neill, M., Jerrett, M., Kawachi, L., Levy, J., Cohen, A., Gouveia, N., Wilkinson, P., Fletcher, T., Cifuentes, L., Schwartz, J., 2003. Health, wealth, and air pollution: advancing theory and methods. Environmental Health Perspectives 111 (16), 1861–1870.

O'Sullivan, D., Unwin, D., 2003. Geographic Information Analysis. Wiley Inc., New Jersey.

Robinson, D.L., Monro, J.M., Campbell, E.A., 2007. Spatial variability and population exposure to $PM_{2.5}$ pollution from woodsmoke in a New South Wales country town. Atmospheric Environment 41 (26), 5464–5478.

Zelikoff, J., Chen, L., Cohen, M., Schlesinger, R., 2002. The toxicology of inhaled woodsmoke. Journal of Toxicology and Environmental Health—Part B—Critical Reviews 5 (3), 269–282.