



## The use of GIS to evaluate traffic-related pollution

D J Briggs

*Occup. Environ. Med.* 2007;64:1-2  
doi:10.1136/oem.2006.030080

---

Updated information and services can be found at:  
<http://oem.bmj.com/cgi/content/full/64/1/1>

---

*These include:*

### Rapid responses

You can respond to this article at:  
<http://oem.bmj.com/cgi/eletter-submit/64/1/1>

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

---

### Notes

---

To order reprints of this article go to:  
<http://www.bmjournals.com/cgi/reprintform>

To subscribe to *Occupational and Environmental Medicine* go to:  
<http://www.bmjournals.com/subscriptions/>

GIS for traffic-related pollution

## The use of GIS to evaluate traffic-related pollution

D J Briggs

GIS for exposure modelling

One of the major challenges in epidemiological research is to devise appropriate metrics and methods for exposure assessment. In the context of traffic-related air pollution, this is particularly problematic because of continuing uncertainty about the causal agents, the likelihood of important interactive and cumulative effects from different pollutants, high levels of both spatial and temporal variability in pollutant concentrations and a dearth of monitoring data. Against this background, models that can estimate at unsampled locations are clearly needed. The paper by Morgensten *et al*<sup>1</sup> (see page 8) in this issue presents an example of how geographic information system (GIS) techniques can be used to develop such models for urban-scale analysis, on the basis of readily available data.

The use of GIS methods for exposure modelling in this way has a relatively recent history. Outside epidemiology, the emphasis has mainly been on dispersion modelling, and a range of so-called second-generation models have been developed (eg, AERMOD, ADMS-Urban) to support air pollution management. To date, however, these models have been rather rarely used for epidemiological purposes, partly because of their demanding data requirements, and also, no doubt, because of lack of awareness, lack of understanding or distrust by this research community. By contrast, in epidemiology, the focus has been on developing GIS-based methods. Initially, these mainly involved the extraction of relatively simple distance-based metrics of exposure (eg, based on proximity to source). However, over the past 10 years, attention has turned to GIS-based pollution mapping, using interpolation techniques, such as inverse distance weighting and kriging, and what has become known (perhaps rather misleadingly) as land use regression modelling.<sup>2</sup> Since its original development as part of the Small Area Variations in Air Quality and Health (SAVIAH) study,<sup>3,4</sup> land use regression modelling has attracted particular attention and has been applied in a range of situations and studies, both in Europe and North

America.<sup>5-7</sup> In the 2006 International Society for Environmental Epidemiology/International Society of Exposure Analysis (ISEE/ISEA) conference in Paris, no fewer than 20 papers and posters used the approach as a basis for exposure assessment. In light of this growing popularity, it merits critical consideration, for—like any widely used technique—it has the potential not only to affect the results of epidemiological analyses where it is applied but also to shape the way in which studies are conceived and planned.

In terms of performance, land use regression models have a relatively good record. Where they have been validated against monitored data, coefficients of determination ( $R^2$ ) typically in the range of 0.45–0.7 and standard errors of <20% of the mean<sup>2-8</sup> have been reported—comparable to more sophisticated dispersion modelling<sup>8</sup> and at a much reduced computational cost. On the other hand, the models developed in these studies show considerable variability in both structure and form. In general, variables relating to road sources provide the most important predictors, typically accounting for two thirds or more of the overall  $R^2$ , but the way in which these have been defined and measured often differs. Roads, for example, are classified in different ways and measured at different spatial resolutions, buffer radii vary and, although some studies have used measures of traffic flow, others simply use road length. Such differences mean that the models cannot easily be transferred from one area to another. They also suggest that uniform criteria for model construction are not being applied.

In practice, much greater coherence in these models is almost certainly possible. Briggs *et al*<sup>4</sup> showed that a single land use regression model developed in one city could readily be applied to other locations and other years when recalibrated against data from a few local monitoring sites, with comparable results. Owing to the intercorrelations between many of the variables used in land use regression modelling, substitution of one measure with another, for the

sake of comparability, will rarely alter model performance to any noticeable degree. Differences in data, which to a large extent explain the inconsistencies, can also be resolved in many cases. Often, for example, modellers have resorted to the use of simple, categorical classifications of road type because data on traffic flows (which would provide a far more direct and consistent measure of source activity) seem to be lacking. In reality, although traffic count data are sparse, many (perhaps most) cities now routinely use traffic models for traffic planning, and these can provide well-validated and detailed data, at least for major roads. Equally, different sources of land use data are often used; however, the widespread availability of high-resolution satellite data now offers a far more consistent and ubiquitous source.

On the other hand, the variability found in the models is revealing. They remind us that cities are not all the same, but differ substantially in terms of characteristics such as street and building configuration and fleet composition. Associations with relatively simple land cover and road variables are therefore unlikely to be universal. The fact that such differences arise is also a warning to others, for it implies that data from monitoring sites should not simply be extrapolated to surrounding areas (as has traditionally been done), as the affinity of the sites varies depending on the source distribution, land cover and topography. By the same token, different models should be expected for different pollutants and source types (eg, to reflect different dispersion processes).

At the same time, we need to be aware of the limitations (and potential fallacies) of methods such as land use regression. One of the most important dangers in this context lies in the process of variable selection. As with other regression techniques, unless the selection process is carefully supervised, land use regression can produce highly significant, but nonetheless implausible, models. In the original SAVIAH study,<sup>4</sup> attempts were made to avoid this by allowing variables to enter the model only if their role (as a measure of source intensity or dispersion process) could be predefined, if the sign of the regression coefficient accorded with this expectation, and if the magnitude of the coefficient and the position of the buffer zone was consistent with all others in the model. If land use regression is to become a standard and accepted technique for exposure assessment, the time is perhaps right to develop a clear set of rules for model construction, based on both physical and statistical principles.

Clearly, land use regression methods can also be enhanced in several ways. One

important improvement would be to recognise (and separately model) the different contributions from local and long-range (and primary and secondary) sources. The inability to do this probably accounts for the poorer performance of land use regression models for particulates compared with nitrogen dioxide.<sup>8</sup> Another limitation is the static nature of most land use regression models. The opportunity to use them for shorter-term exposures nevertheless exists, for example, by incorporating time-dependent variables such as wind direction and speed through the use of variable, non-circular kernelling and focalsum techniques. We can even go further if desired, by modelling dispersion in a three-dimensional environment of streets and buildings, or by taking account of population dynamics.

However, herein lies a new dilemma. GIS are powerful tools. Only the surface of their capability for exposure modelling has been scratched; much more sophisticated models could undoubtedly be developed. The question is just how far should we go? Will the quest for greater sophistication (driven

as much by the inevitable curiosity of those involved as any objective need) simply end with the reinvention of a wheel that has already been invented in the form of dispersion modelling? Will we create measures of exposures that are too detailed, and models too complex, to be usable in epidemiology? Or, by helping to show and specify more precisely the spatial and temporal dynamics of exposure processes, will such advances help to strengthen epidemiology by opening up new lines of investigation and enabling new study designs? After many years in which methods for exposure modelling have lagged behind other aspects of epidemiological analysis, GIS are bringing the handmaiden to the fore. These are questions that we therefore now need to deal with.

*Occup Environ Med* 2007;**64**:1–2.  
doi: 10.1136/oem.2006.030080

Correspondence to: Professor D J Briggs,  
Department of Epidemiology and Public Health,  
Imperial College, Norfolk Place, London W2  
1PG, UK; d.briggs@imperial.ac.uk

Accepted 22 September 2006

## REFERENCES

- 1 **Morgenstern V**, Zutavern A, Cyrys J, *et al*. Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children. *Occup Environ Med* 2007;**1**:8–16.
- 2 **Jerrett M**, Arain A, Kanaroglou P, *et al*. A review and evaluation of intraurban air pollution exposure models. *J Expo Anal Environ Epidemiol* 2005;**15**:185–204.
- 3 **Briggs DJ**, Collins S, Elliott P, *et al*. Mapping urban air pollution using GIS: a regression-based approach. *Int J Geogr Inf Sci* 1997;**11**:699–718.
- 4 **Briggs DJ**, de Hoogh C, Gulliver J, *et al*. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. *Sci Total Environ* 2000;**253**:151–67.
- 5 **Brauer M**, Hoek G, van Vliet P, *et al*. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology* 2003;**14**:228–39.
- 6 **Carr D**, von Ehrenstein O, Wagner C, *et al*. Modelling air pollution on the basis of road traffic characteristics. *Environ Res* 2002;**90**:111–18.
- 7 **Gilbert NL**, Goldberg MS, Beckerman B, *et al*. Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model. *J Air Water Manage Assoc* 2005;**55**:1059–63.
- 8 **Briggs DJ**. The role of GIS: coping with space (and time) in air pollution exposure assessment. *J Toxicol Environ Health* 2005;**68**:1243–61.

## ECHO

### Relation between exercise and sick leave



Please visit the *Occupational and Environmental Medicine* website [www.occenvmed.com] for a link to the full text of this article.

**S**trenuous physical activity at least three times a week reduces the amount of time spent on sick leave, a study in the Netherlands has found, but moderate exercise had no noticeable effect. A distinction was made between duration, frequency, and intensity of physical activity. The outcome measure was the numbers of days of sick leave taken.

Physical activity was assessed using two measures from the recommendations of the Centers for Disease Control and Prevention and the American College of Sports Medicine. They included the frequency (in days per week) of having been moderately active for at least 30 minutes a day, and the frequency (in days per week) of having been active at a vigorous intensity for at least 20 minutes a session. Two further variables were created by adding up the duration of activity in tasks such as gardening, chores, and sports to indicate the total amount of moderate and vigorous intensity activity, both expressed in minutes per day.

The data were from three large Dutch databases—two ongoing cross sectional surveys among a representative sample of the Dutch population and one prospective cohort study. There was no difference in the duration of mean sick leave between those who had engaged in moderate physical activity and those who had not.

In two of the three databases, workers meeting the recommendation of vigorous physical activity (active at a vigorous level for at least three times a week) had significantly less sick leave: two thirds of one day over two months, and more than four days over a year.

▲ Proper KI, *et al* *British Journal of Sports Medicine* 2006;**40**:173–178