DISPERSION MODELLING FOR DECISION-MAKING: COMBINING SPATIAL STATISTICS WITH MODEL PARAMETERISATION

A. Tiwary*, A. Namdeo, P.S. Goodman, M.C. Bell

School of Civil Engineering and Geosciences, Newcastle University, NE1 7RU, UK

Abstract:
Often air quality is considered as part of a multi-criteria decision making process, requiring dispersion modelling to be conducted alongside a multitude of co-determinants of possible management outcomes. This typically involves diverse information layers, such as socio-demographics, land use, travel patterns, utilising spatial statistics to establish the cause-effect relationship over the study domain. The integrated approach presented in this paper combines spatially-resolved (raster) information on emission sources (road traffic and/or industrial) with the published statistics (socio-demographic, epidemiology, etc.) to gain a priori information on the hotspots. This enables iterative enhancement of dispersion model parameterisation, mainly emissions in specific areas, to refine the air quality predictions.

A case study is presented demonstrating the application of spatial statistics, utilising a combination of weighted-overlay and ordinary least-squares regression analyses for the dissimilar datasets included in the multi-criteria decision problem. All the datasets are generated at the Lower Super Output Area-level for the local administration boundaries of Stockton (UK), spatially mapping the distribution of potential locations of priority areas for model refinement. The study demonstrates the need for robust dataset while associating dispersion modelling outcomes with multi-criteria decision problems.

Key words: dispersion modelling; spatial statistics; weighted-overlay analysis; multi-criteria decision analysis

INTRODUCTION
For holistic environmental impact assessment often air quality is considered as part of a multi-criteria decision making process, requiring dispersion modelling to be conducted alongside a multitude of co-determinants of possible management outcomes (González, A. et al., 2013). This involves application of an integrated approach for assessing the status quo and evaluation of the mitigating effects – combining the underpinning environmental, commercial and social factors - on a spatial platform. The set up features of a dispersion tool can be explored to enhance its predictive capabilities through adequate pre-processing of diverse information layers, utilising spatial statistics to evaluate the strength of the cause-effect associations between these factors.

METHODOLOGY
The application of this integrated approach is demonstrated through a case study for Stockton in the North-East England. The model set up comprises of the following three steps:

i. Emissions modelling:
The calculation of pollutant emissions used in the study were based on vehicle flow and network speed outputs from the strategic transport level Tees Valley Model (TVM) being fed into Newcastle University’s PITHEM (Platform for Integrated Traffic, Health and Emission Modelling) software (Namdeo and Goodman, 2012). This software allowed the development of bespoke mappings between the user classes defined in the TVM with the hierarchical vehicle fleet information within the UK National Atmospheric Emissions Inventory (NAEI) (Venfield and Pang, 2012). The vehicle kilometres travelled and the average speed information for each user class were then combined with the NAEI speed-emissions curves (Boulter et al., 2009) to produce period-based emissions totals for regulated air pollutants. These period totals were then scaled by appropriate diurnal and seasonal flow profiles (DfT, 2012) and normalised by link lengths to produce average annual emissions rates for each link, as well as diurnal hourly emissions profiles. The PITHEM software outputs emissions information in both vector format (ESRI, 2012) for import into GIS, and in a text-based format compatible with the suite of ADMS air quality models (CERC, 2011) used for dispersion calculations.

ii. Dispersion modelling:
The traffic emissions estimated from PITHEM (as described above) were used as input to set-up the air dispersion modelling simulation for the Stockton case study using ADMS-Roads (ADMS-Roads v3.1). The model inputs comprised of traffic-generated emissions of CO, NO₂, PM₁₀, PM₂.₅ and meteorology. Long term (annual average) outputs were obtained for a 250m grid resolution applying an intelligent gridding along the roads. The latter generated additional output points on both sides of the modelled roads where the pollutant concentration gradients are the greatest. This was conducted in two stages – first, in the pre-run stage up to 5000 extra receptor points were added in and around the roads in sets of 4, with at least one set of points added to each road segment that lies within the output grid; second, at the end of the model run, three sets of additional points
were added between the first set of intelligent grid points for generating concentration output by linearly interpolating between the values at the first set of points. The second stage was conducted in ADMS-Roads after the model run by default, to create a regular Cartesian grid within the rectangular region specified (CERC, 2011).

For effective visualisation of the pollutant concentrations the gridded outputs obtained from dispersion model were used to generate interpolated surface maps using an inverse distance weighted technique (IDW) (ArcGIS Spatial Analyst) and overlaid on the existing road network for the study area (Figures 1 a, b).

Figure 1. Hourly average pollutant concentrations: a.) left panel – NO2; b.) right panel – PM2.5 (map data © OpenStreetMap).

iii. Spatial statistics
Two spatial statistics were applied to evaluate the strength of explaining the outcome (dependent variable) through a set of inputs (explanatory variables) – a.) Weighted overlay; b.) Ordinary least squares regression. The weighted-overlay technique allows for an integrated analysis to solve multi-criteria problems (such as site selection and suitability models) by applying a common scale of values to diverse and dissimilar inputs. Different rasters with different value scales can be analysed on a common scale. The cell values for each input raster in the analysis are assigned values from the evaluation scale and reclassified to these values. This makes it possible to perform arithmetic operations on the rasters that originally held dissimilar types of values. Each input raster is weighted, or assigned a percent influence, based on its importance to the model. Any class can also be assigned a restricted value, which means that the corresponding area is unacceptable or cannot be used. Restricted areas are excluded from the analysis. The cumulative influence for all the rasters should sum to 100 percent. The cell values of each input raster are multiplied by the rasters' weights. The input rasters are weighted by importance and added together to produce an output raster scaled on 1 to 10 (with 10 being the most favourable). Values at one end of the scale represent one extreme of suitability (or other criterion); values at the other end represent the other extreme. The Ordinary Least Squares regression (OLS) method allows for developing regression model for quantifying the spatial associations between different attributes with unique identifiers.

RESULTS AND DISCUSSION
In the first step, the spatially-referenced outputs were overlaid to assess the spatial associations using the Statistical and Analytical tools available in ArcGIS v10.1 (ESRI, 2012). This exercise involved combining mapped data for pollution concentration, socio-demographics and hospital admissions (including both respiratory and circulatory) into a single layer at the Lower Super Output Area (LSOA) unitary authority boundary (comprising of minimum population size 1000; mean population size 1500) and then the scatter plot was generated to assess the cross-thematic correlations (Figure 2). It required some pre-processing of the input features in the GIS layers to ensure they had common feature IDs and also that any null/zero values were removed. Conventionally, to assess the association with other attributes the variable of interest (usually the dependent variable) has to be entered as the first or the last in the list of weightings. In this study hospital admission (HospAdm) was chosen as the outcome i.e. dependent variable, assuming it to be explained through spatially-resolved datasets on modelled pollutant concentrations (NO2 and PM2.5) and publically accessible socio-demographic statistics (Yearly Potential Life Loss Indicator, YPLL and the Rank of Index of Multiple Deprivation, RIMD scores provided by the UK Office of National Statistics).
As can be inferred from the plots, the reported hospital admissions (combined for respiratory and circulatory illnesses) in the study area has stronger correlations with the modelled PM$_{2.5}$ than with NO$_2$ concentrations. Also, there seems to be a direct association of hospital admissions with the reported Yearly Potential Life Lost Indicator (YPLL). However, a strong negative correlation is observed for both of these variables with the reported Deprivation index (RIMD). This is expected and solely an artefact of the manner RIMD is reported in an inverse order, lower values reflecting high deprivation areas. For example, in Stockton typical RIMD values in the deprived and affluent regions are in the range of 90-3724 and 24305-31998 respectively. However, this bears interesting implications for the subsequent step in establishing the weightings for the variables (and their scaling on 1 to 10) to be included in the weighted-overlay input list, based on the observed correlations.

Figure 2. Scatter plot of cross-thematic spatially referenced datasets for the study area.

In the next step, weighted-overlay analysis was performed, combining the pollutant concentrations from modelled traffic-related emissions with the reported socio-demographics for the site, acquired from publicly available statistics at LSOA level (http://www.neighbourhood.statistics.gov.uk/). The SOAs in the UK represent the smallest geographic units for disseminating robust census statistics while the confidentiality of individual census returns remains preserved (ONS, 2010). The input layers and their percentage weighting applied to this exercise are shown in Table 1. Essentially, this approach allowed for multi-criteria decision-making, utilising a ranked-score, based on a common scale for traffic emissions, deprivation and reported health implications.

Table 1. Weightings applied to the multi-criteria layers with reference to hospital admissions from the scatter plot

<table>
<thead>
<tr>
<th>Input layer</th>
<th>Data source</th>
<th>Weighting (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM$_{2.5}$</td>
<td>Dispersion model</td>
<td>25</td>
</tr>
<tr>
<td>NO$_2$</td>
<td>Dispersion model</td>
<td>20</td>
</tr>
<tr>
<td>Deprivation (Ranked Index of Multiples Deprivation, RIMD)</td>
<td>ONS (2010)</td>
<td>30</td>
</tr>
<tr>
<td>Potential life lost indicator (Yearly Potential Life Loss Indicator, YPLL)</td>
<td>ONS (2010)</td>
<td>25</td>
</tr>
</tbody>
</table>

The output from weighted-overlay analysis is shown in Figure 3 on a ranking of preference from 1 to 10 (in increasing order). Shown alongside on the same map is the traffic intensity, in terms of the annual average daily traffic (AADT), on the Stockton road network - ranging from blue (2214) to red (57977) vehicles. The weighted-overlay of associated layers provides a LSOA-level spatial map of the spread of potential locations of priority.
areas to improve the traffic-related air quality. The ranked outputs link the explanatory variables (both pollutants and socio-demographics, including RIMD and YPLL) with the outcome (hospital admissions) to locate the regions requiring better understanding of the pollutant exposure.

In the next step, prior to further refinement of the dispersion model parameters, the ordinary least squares regression (OLS) method was used to evaluate the strength of explaining the outcome based on the explanatory variables considered. Similar to the weighted-overlay approach, the OLS assessment treated hospital admissions as the outcome (dependent variable) based on pollutant concentration (PM$_{2.5}$ and NO$_2$) and the reported socio-demographics listed in Table 2 (RIMD and YPLL from the UK Office of National Statistics) as explanatory variables. The residuals plot (Figure 4) shows the regional distribution of under- and over-predictions spatially. The red areas indicate under predictions (i.e. observed values are higher than the predicted); the blue areas show over predictions (i.e. observed values are lower than predicted). Ideally, the over/under predictions for a well-performing outcome should be scattered randomly; any sign of clustering indicates the dependent variable missing key explanatory variables. It was noted from the Adjusted $R^2$ Statistics that only about 18% of the hospital admissions in the region are explained by the explanatory variables studied.

Figure 3. Hotspots from weighted overlay analysis (area) superimposed with traffic information (lines) (© Crown Copyright).

The lower values (lighter tone) reflect sites requiring least interventions while the higher values (darker tone) indicate the sites where management interventions have to be implemented.

Figure 4. Residuals plot from ordinary least-squares regression analysis showing strength of the explanatory variables used (from Table 1) in explaining the reported hospital admissions in the study area (© Crown Copyright).
CONCLUSIONS AND FUTURE WORK
This study has demonstrated application of spatial statistics in refining dispersion model outputs based on a multi-criteria approach using a suite of spatial statistics. The main feature of the methodology is to customise air quality modelling to decision-making through an iterative refinement of the input explanatory variables while evaluating their contributions to a preferred outcome, e.g. hospital admissions in our case study. Our case study suggests that the hospital admissions cannot be fully explained using only a snapshot of the published socio-demographic statistics, as commonly applied in health geography. Both weighted overlay and OLS methods offer robust tools to spatially identify the explanatory variables in the model while associating air quality with decision making, especially human health implications. Future analyses require more rigorous datasets while associating exposure levels with air quality outcomes from dispersion modelling, including weighted information on population age, underlying health conditions, dwelling types, amount of time spent in homes, etc.

REFERENCES