



## **University of Huddersfield Repository**

Collins, Susan

A GIS approach to modelling traffic related air pollution

### **Original Citation**

Collins, Susan (1998) A GIS approach to modelling traffic related air pollution. Doctoral thesis, University of Huddersfield.

This version is available at <http://eprints.hud.ac.uk/id/eprint/4843/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: [E.mailbox@hud.ac.uk](mailto:E.mailbox@hud.ac.uk).

<http://eprints.hud.ac.uk/>

**A GIS APPROACH TO MODELLING TRAFFIC  
RELATED AIR POLLUTION**

**SUSAN COLLINS**

**A thesis submitted to the University of Huddersfield  
in partial fulfilment of the requirements for  
the degree of Doctor of Philosophy**

**The University of Huddersfield in collaboration with the National  
Institute of Public Health and Environmental Protection,  
Bilthoven, The Netherlands, and the London School  
of Hygiene and Tropical Medicine, London, UK**

**June 1998**

## **ABSTRACT**

There is increased concern regarding the effect of traffic related pollution on public health. As the number of vehicles on the roads continues to rise, it is becoming increasingly more important to identify areas where the population may be at a greater risk to raised levels of pollution and areas where the implementation of policy to control and monitor levels of pollution would be beneficial.

Traditionally, levels of air pollution have been established through dispersion modelling or monitoring. However, for modelling traffic related pollution for large populations, these methods have proved inappropriate.

Three new approaches have been developed to model traffic related air pollution and are reported in this thesis. The approaches have been developed in a Geographical Information System (GIS) and involve generating detailed maps of the pollution surface from monitored data and information about the pollution sources. The new methods are compared against the geostatistical technique kriging.

The first approach combines spatial interpolation from monitoring sites and dispersion modelling, linking the dispersion model to the GIS, the second combines GIS techniques for filtering data and spatial interpolation, and the third uses a combination of GIS techniques for filtering and statistical techniques.

The three approaches are tested and validated by predicting levels of pollution at monitoring sites not used to develop the models. It was found that the new approaches provided more reliable estimates of pollution at unsampled locations than kriging, with the last of these proving to be the most effective. The adjusted  $r^2$  values for kriging, interpolation and dispersion, interpolation and filtering, and filtering and statistics were found to be 0.44, 0.63, 0.67 and 0.82 respectively.

The approaches therefore have clear potential in the areas of air pollution management and epidemiology, where the maps can be used to help identify locations where levels of pollution exceed air quality standards, assess the relationship between air pollution and health outcome and examine the risk of exposure to raised levels of pollution.

## **ACKNOWLEDGEMENTS**

I would like to take this opportunity to thank my supervisor, Professor David Briggs, now at Nene Centre for Research, Nene College, Northampton, for his continued support and advice. I would also like to thank Professor David Butcher, now at the University of Central Lancashire, for guidance and taking on the role of second supervisor during the latter stages of the thesis.

The EU-funded project, Small Area Variations in Air Quality and Health (SAVIAH), provided the starting point for the research presented in this thesis. Thanks are due to all participants of the SAVIAH project and especially the project co-ordinator, Professor Paul Elliott, Imperial College, London (formerly at the London School of Hygiene and Tropical Medicine).

## **CONTENTS**

<b>Abstract</b>	
<b>Acknowledgments</b>	i
<b>Contents</b>	iv
<b>List of Figures</b>	vii
<b>List of Tables</b>	viii
<b>List of Appendices</b>	

<b>Chapter 1 Introduction</b>	1
1.1 Urban Air Pollution	1
1.1.1 Sources of Air Pollution	3
1.1.2 Processes	4
1.1.3 Trends in Air Pollution	6
1.2 Air Pollution and Health	9
1.2.1 Acute Verses Chronic Health Effects	12
1.3 Trends in Respiratory Illness	13
1.3.1 Time-Series Studies	17
1.3.2 Spatial Studies	19
1.3.2.1 Vehicle Emissions and Respiratory Illness	20
1.4 Environmental and Health Policy	25
1.5 Aims and Objectives	30
<b>Chapter 2 Mapping Air Quality</b>	32
2.1 Spatial Variations in Urban Air Pollution	32
2.2 Dispersion Modelling	33
2.3 Spatial Interpolation	38
2.4 Adapted Statistical Methods	44
2.5 Geographical Information Systems	48
2.6 Producing Detailed Maps of Air Pollution	49
<b>Chapter 3 Data Collection</b>	52
3.1 The SAVIAH Project	52

3.2	Measuring Air Pollution	54
3.3	Survey Design	56
3.4	The Huddersfield Study	60
3.4.1	Monitored Data	60
3.4.2	Road Network	64
3.4.3	Land Cover	66
3.4.4	Altitude	66
3.4.5	Field Data	69
3.4.6	Health Data	69
<b>Chapter 4</b>	<b>Kriging</b>	74
4.1	Kriging Routines	74
4.2	Kriging the Huddersfield NO <sub>2</sub> Data	76
4.3	Kriging the Poznan SO <sub>2</sub> Data	77
4.4	Applying Kriging to Traffic Related Pollution	90
<b>Chapter 5</b>	<b>Automatic Modelling of Traffic Related Air Pollution (AMTRAP)</b>	92
5.1	Modelling Near-Source Air Pollution	93
5.1.1	Generating Pollution Concentrations for a Set of Pre-Defined Variables	96
5.1.2	Description of the TRAFFPOL Routine	98
5.1.3	Applying the TRAFFPOL Routine to the SAVIAH Data	107
5.2	Background Pollution	111
5.3	Validation of the Method	116
5.4	The Application of the Meteorological Data	121
5.5	Applying the AMTRAP Model to Traffic Related Pollution	124
<b>Chapter 6</b>	<b>The Moving Window Approach</b>	125
6.1	Moving Windows	126
6.2	Modelling Near-Source Pollution	129
6.2.1	Defining the Air Pollution Template	129

6.2.2	Applying the Air Pollution Template	131
6.3	Background Pollution	134
6.4	Validation of the Method	134
6.5	Applying Moving Windows to Traffic Related Pollution	136
<b>Chapter 7</b>	<b>The Regression Approach</b>	138
7.1	Identifying the Independent Variables	138
7.2	Calculating the Independent Variables	140
7.2.1	Traffic Volume	140
7.2.2	Land Cover	144
7.2.3	Topography	146
7.2.4	Height of the Sampler and Topex	147
7.4	The Regression Equation	147
7.4.1	Generating the Air Pollution Map	148
7.5	Results and Validation	148
7.6	Applying Regression Mapping to Traffic Related Pollution	153
<b>Chapter 8</b>	<b>Evaluation and Comparison of the Different Approaches</b>	155
8.1	Comparison of the Pollution Maps	155
8.2	Estimating the Population at Risk	162
8.3	Advantages and Disadvantages	168
8.3.1	Kriging	169
8.3.2	AMTRAP	171
8.3.3	Moving Window	173
8.3.4	Regression	174
8.3.5	Summary	175
8.4	Discussion	176
<b>Chapter 9</b>	<b>Conclusion</b>	179
9.1	Future Research Issues	183
<b>References</b>		185

## **FIGURES**

### **Chapter 1**

Figure 1.1.	Private cars currently licensed: 1955 - 1995.	8
Figure 1.2	Percent prevalence of wheeze in the last year, British surveys 1965-1992.	14
Figure 1.3	Hospital admission rates.	14

### **Chapter 3**

Figure 3.1	The Huddersfield study area.	61
Figure 3.2	Annual mean NO <sub>2</sub> concentrations at the 80 permanent monitoring sites.	63
Figure 3.3	The road network with average daily traffic volume.	65
Figure 3.4	Land cover.	67
Figure 3.5	Altitude.	70
Figure 3.6	The place of residence of the Children.	73

### **Chapter 4**

Figure 4.1	The semivariogram.	75
Figure 4.2	Map of annual mean NO <sub>2</sub> by kriging.	78
Figure 4.3	Variance for the kriging estimates of NO <sub>2</sub> .	79
Figure 4.4	Monitored annual mean NO <sub>2</sub> against kriging predictions for the 8 consecutive monitoring sites.	80
Figure 4.5	The Poznan study area and main emission sources of SO <sub>2</sub> (chimneys).	81
Figure 4.6	Average SO <sub>2</sub> at the 68 permanent monitoring sites.	82
Figure 4.7	Map of Average SO <sub>2</sub> by Kriging: GSLIB/Splus.	84
Figure 4.8	Map of average SO <sub>2</sub> by kriging: ARC/INFO.	85
Figure 4.9	Monitored average SO <sub>2</sub> against kriging predictions by the ARC/INFO and GSLIB/Splus methods for the 8 consecutive monitoring sites.	86
Figure 4.10	Kriging predictions for GSLIB/Splus against kriging predictions for ARC/INFO for the 8 consecutive monitoring sites.	87
Figure 4.11	Location of the randomly generated points.	88
Figure 4.12	Estimated SO <sub>2</sub> values for Kriging by the GSLIB/Splus method against Kriging by the ARC/INFO method for the 18 random locations.	89

### **Chapter 5**

Figure 5.1	Location of receptors.	98
Figure 5.2	Distance to the nearest road.	101
Figure 5.3	Value of the nearest road by type of road.	102
Figure 5.4	Orientation of the nearest road with respect to North.	103
Figure 5.5	Perpendicular direction to the nearest road.	104
Figure 5.6	Day time traffic volume on the nearest road.	105
Figure 5.7	The main emission sources.	108

Figure 5.8	The surface roughness class.	110
Figure 5.9	Near-source NO <sub>2</sub> concentrations.	112
Figure 5.10	Annual mean NO <sub>2</sub> at the background monitoring sites.	113
Figure 5.11	Map of background NO <sub>2</sub> by kriging.	114
Figure 5.12	Variance for the kriging estimates of NO <sub>2</sub> .	115
Figure 5.13	Map of annual mean NO <sub>2</sub> by the AMTRAP method.	117
Figure 5.14	The structure of the AMTRAP model.	118
Figure 5.15	Monitored annual mean NO <sub>2</sub> against AMTRAP predictions for the 8 consecutive monitoring sites.	119
Figure 5.16	Monitored annual mean NO <sub>2</sub> against AMTRAP predictions at the 54 near-source monitoring sites.	121
Figure 5.17	Distribution of the residuals at the 54 near-source monitoring sites.	120
Figure 5.18	Monitored annual mean NO <sub>2</sub> against AMTRAP predictions (with average meteorology conditions) at the 54 near-source monitoring sites.	123
Figure 5.19	Predicted values with full meteorology against predicted values with average meteorology conditions.	123
<b>Chapter 6</b>		
Figure 6.1	An example of a grid showing the location of two roads, A and B.	126
Figure 6.2	A 3x3 window passes over a grid a) and b) calculating the mean value and placing the result in a new grid c) and d).	127
Figure 6.3	Moving window in the outermost ring of the grid.	128
Figure 6.4	An example of a weighted template.	128
Figure 6.5	The decay of concentration over distance.	130
Figure 6.5	Distance between the central point of the cell and the centre of the template.	131
Figure 6.6	The weighted template.	132
Figure 6.7	A sample grid and template.	133
Figure 6.8	Map of annual mean NO <sub>2</sub> by the moving window method.	135
Figure 6.9	Monitored annual mean NO <sub>2</sub> against moving window predictions for the 8 consecutive monitoring sites.	136
Figure 6.10	Monitored annual mean NO <sub>2</sub> against moving window predictions at the 54 near-source monitoring sites.	137
<b>Chapter 7</b>		
Figure 7.1	Emission sources and dispersion patterns for two different monitoring sites.	139
Figure 7.2	A sample of the traffic volume grid and a 50m filter.	141
Figure 7.3	Creating a 350m mask for traffic volume around the monitoring sites.	143
Figure 7.4	Total industrial land in the 0-300m zone.	145
Figure 7.5	Map of annual mean NO <sub>2</sub> by the regression method.	149
Figure 7.6	Monitored values against predicted values at the 80 permanent monitoring sites.	150
Figure 7.7	Normal scores for the residuals.	151
Figure 7.8	First, second and third order nearest neighbour residuals.	152

<b>Figure 7.9</b>	Monitored annual mean NO <sub>2</sub> against regression predictions for the 8 consecutive monitoring sites.	153
<b>Chapter 8</b>		
<b>Figure 8.1</b>	Annual mean NO <sub>2</sub> in the residential area of Huddersfield by kriging.	157
<b>Figure 8.2</b>	Annual mean NO <sub>2</sub> in the residential area of Huddersfield by the AMTRAP approach.	158
<b>Figure 8.3</b>	Annual mean NO <sub>2</sub> in the residential area of Huddersfield by the moving window approach.	159
<b>Figure 8.4</b>	Annual mean NO <sub>2</sub> in the residential area of Huddersfield by the regression approach.	160
<b>Figure 8.5</b>	Monitored annual mean NO <sub>2</sub> against kriging, AMTRAP, moving windows and regression predictions at the 8 consecutive monitoring sites.	161
<b>Figure 8.6</b>	Histograms of pollution scores for the kriging, AMTRAP, moving window and regression approaches.	164

## Tables

### Chapter 1

Table 1.1	Major urban air pollutants and their sources.	2
Table 1.2.	UK Emissions by type of fuel (1994).	3
Table 1.3.	Sources of the Principal Pollutants in the UK (1994).	3
Table 1.4	Increases in UK Estimated Emissions from Road Transport (1980-1990).	6
Table 1.5	The health effects of the major pollutants.	11
Table 1.6	The United Kingdom National Air Quality Strategy standards for the major pollutants.	26

### Chapter 3

Table 3.1	Number of returned questionnaires and response rates.	53
Table 3.2	Geographical data and associated attributes.	54
Table 3.3	Coefficient of variation for the Willems badge and Palmes tube.	57
Table 3.4	Mean diffusional resistance ( $R_d$ ) and coefficient of variation of $R_d$ for Willems badges and Palmes tubes.	58
Table 3.5	Land cover classification scheme.	68
Table 3.6	Health outcomes.	72

### Chapter 4

Table 4.1	Annual mean $\text{NO}_2$ values and kriging estimates from ARC/INFO for the 8 consecutive monitoring sites ( $\mu\text{g}/\text{m}^3$ ).	76
Table 4.2	Average $\text{SO}_2$ values and kriging estimates from ARC/INFO and GSLIB and Splus for the 8 consecutive monitoring sites.	86
Table 4.3	Correlation coefficients.	86
Table 4.4	Kriging estimates of $\text{SO}_2$ from ARC/INFO and GSLIB/Splus combination for a random set of points.	89

### Chapter 5

Table 5.1	Codes and reference values for wind direction, wind speed, atmospheric stability, surface roughness and mixing zone.	97
Table 5.2	Sample data from the concs.dat data file.	98
Table 5.3	Wind direction classification scheme for the weather data.	106
Table 5.4	Wind speed classification scheme for the weather data.	106
Table 5.5	Road type classification scheme.	109
Table 5.6	Surface roughness classification scheme.	109
Table 5.7	Annual mean $\text{NO}_2$ values and AMTRAP predictions for the 8 consecutive monitoring sites.	116

### Chapter 6

Table 6.1	Annual mean $\text{NO}_2$ values and filtering predictions for the 8 consecutive monitoring sites.	134
-----------	--	-----

<b>Chapter 7</b>		
Table 7.1	Regression coefficients, confidence intervals, significance values, standard errors and t values for the variables in the regression equation.	148
Table 7.2	Annual mean NO <sub>2</sub> values and regression predictions for the 8 consecutive monitoring sites.	152
<b>Chapter 8</b>		
Table 8.1.	Adjusted r <sup>2</sup> , constant, slope coefficients, standard error and range for the kriging, AMTRAP, moving window and regression approaches.	161
Table 8.1.	Adjusted r <sup>2</sup> , constant, slope coefficients, standard error and range for the kriging, AMTRAP, moving window and regression approaches.	164
Table 8.3.	Percentage of children with a pollution score greater than the National Air Quality Strategy standard and the study average monitored value for the kriging, AMTRAP, moving window and regression approaches.	166
Table 8.4	Odds ratios and 95% confidence intervals for the four pollution approaches against health outcome.	167
Table 8.5	Processing times and data requirements.	169
Table 8.6	Advantages and disadvantages.	176

## **APPENDICES**

- Appendix 1    The *hydro* program**
- Appendix 2    The Health Questionnaire**
- Appendix 3    The *consect* Program**
- Appendix 4    The *TRAFFPOL* Program**
- Appendix 5    The *adirect* Program**
- Appendix 6    The *weather* Program**
- Appendix 7    The *rel\_relief* Program**

## **CHAPTER 1 INTRODUCTION**

The last decade has seen a renewed and growing concern about the effects of air pollution on human health and the environment. Accurate and detailed estimates of the spatial variations in air quality across major cities are essential to help monitor and control levels of air pollution. This thesis presents research that has been undertaken to evaluate existing techniques and develop new methods of modelling air quality in complex urban environments.

The term air pollution could potentially embrace a number of aspects - from indoor air pollution to greenhouse gases; the research presented here, however, is primarily concerned with outdoor air pollution that could have an adverse effect on public health.

The vast majority of people live in cities. Atmospheric pollution is a direct result of human activity and consequently higher in areas where the majority of the population is concentrated. The Commission for the European Communities (1992) estimate that at least 70% of the population of Europe live in cities with greater than 20,000 people and therefore may be subject to potentially high levels of atmospheric pollution.

### **1.1 Urban Air Pollution**

Urban air pollution derives from a wide range of emission sources, including motor vehicles, industry and domestic activity. The main pollutants tend to be airborne particles (for example, black smoke), sulphur dioxide, nitrogen dioxide, carbon monoxide, ozone

and volatile organic compounds (VOCs), for example benzene. Emission sources of the major urban air pollutants, as identified by Godlee (1991), are summarized in Table 1.1.

*Table 1.1 Major urban air pollutants and their sources.*

Pollutant	Sources
Airborne particulates	Emitted from diesel exhausts and burning of fossil fuels.
Sulphur dioxide	Burning of fossil fuels, emissions from power stations and diesel exhausts.
Nitrogen dioxide	Motor vehicles and power stations
Carbon monoxide	Incomplete combustion of fossil fuels and tobacco smoke
Ozone	Photochemical reaction between nitrogen oxides and hydrocarbons
Benzene	Emissions and evaporation from petrol engine

(Source: Godlee, 1991)

Pollutants that are emitted directly into the atmosphere (for example, carbon monoxide and sulphur dioxide) are collectively termed primary pollutants; pollutants formed by chemical reactions in the atmosphere (for example, ozone) are referred to as secondary pollutants. Some pollutants, such as nitrogen dioxide, can be both a primary and a secondary pollutant. In the case of nitrogen dioxide, the pollutant can be emitted (from vehicle exhausts or power stations) or it can be produced from the oxidation of nitric oxide in the air.

As the Quality of Urban Atmospheric Review Group (1993) illustrates, this distinction has an important bearing on the development of control strategies. In general, levels of primary pollutants are proportional to rates of emissions, and therefore direct control of primary pollutants is possible through emission control. Reductions in secondary pollutants, however, are more difficult to achieve because of their more complex histories.

### 1.1.1 Sources of Air Pollution

Urban air pollution arises from a variety of processes. For most pollutants (with the exception of VOCs) the combustion of fossil fuels is currently the main process leading to urban air pollution. Emissions vary depending upon the type of fossil fuel used. Emissions by type of fuel in the UK are presented in Table 1.2. At a national level, as Table 1.3 shows, road transport represents the single most important source of many of the more common air pollutants.

*Table 1.2. UK Emissions by type of fuel (1994)*

Pollutant	Type of Fuel				
	Coal	Petrol	DERV	Fuel Oil	Other
Sulphur Dioxide	72%			21%	
Black Smoke	24%		55%		
Nitrogen Oxides	24%	29%	20%		
Carbon Monoxide		84%			
Volatile Organic Compounds		26%			67%

(Figures from Department of the Environment 1996 - where other indicates a type of fuel other than coal or a form of petroleum).

*Table 1.3. Sources of the Principal Pollutants in the UK (1994)*

Source	% Total Emissions				
	Sulphur Dioxide	Black Smoke	Nitrogen Oxides	Carbon Monoxide	Volatile Organic Compounds
Road Transport	2	58	49	88	29
Electricity Supply Industry	65	4	24	0	-
Other Industry	16	3	6	1	19
Domestic	3	22	3	7	2
Other	14	13	18	4	50

(Figures from Department of the Environment, 1996)

This thesis is concerned with emissions from road transport and, as will be discussed in Chapter 3, concentrates on nitrogen dioxide as a proxy for the complex of traffic related pollutants.

The major sources of man-made emissions of nitrogen oxides into the atmosphere are the combustion of fossil fuels in stationary sources (for example, heating, power generation) and in mobile sources (for example, the internal combustion engines of motor vehicles). There are, however, other industrial and domestic contributors, including specific non-combustion industrial processes, such as the manufacture of nitric acid and the use of explosives and welding processes, and indoor sources - for example smoking, gas-fired appliances and oil-stoves. Nevertheless, the majority of NO<sub>x</sub> entering the atmosphere from pollution sources today does so as NO generated from internal combustion engines. Variations in NO<sub>2</sub> between different countries are mainly attributed to differences in the fossil-fuel consumption (World Health Organisation, 1987).

NO<sub>2</sub> is also produced naturally by bacterial and volcanic action, as well as by lightning. Globally, far more NO<sub>2</sub> is produced by natural processes than by human activity; however, since this is spread across the surface of the Earth, it is reduced to low background concentrations.

### 1.1.2 Processes

Many of the processes by which the air pollutants are produced are dependent upon reactions with other gases and elements in the atmosphere. Some form of energy, such as sunlight or heat, is often necessary to activate the reaction. Nitrogen dioxide, for example, is the product of a reaction with nitric oxide and ozone in the presence of sunlight.

Nitric oxide and nitrogen dioxide are collectively known as oxides of nitrogen ( $\text{NO}_x$ ). Nitric oxide is colourless and odourless and is the primary form of  $\text{NO}_x$  emitted into the atmosphere. Nitrogen dioxide on the other hand is a red-brown pungent gas and is the secondary pollutant of  $\text{NO}_x$ .

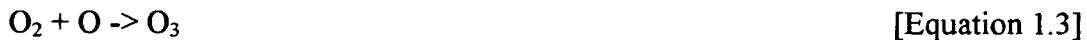
In the case of road transport, at very high temperatures the combustion of petroleum causes nitrogen to combine with oxygen to give nitric oxide (NO). Nitric oxide is emitted from vehicle exhausts and transformed into  $\text{NO}_2$  in the atmosphere through oxidation with gases such as ozone (equation 1.1).



In the presence of hydrocarbons,  $\text{NO}_2$  is then transformed photochemically to NO and atomic oxygen (O) (equation 1.2).



Oxygen then combines with the atomic oxygen to form ozone (equation 1.3).



It can therefore be seen that nitrogen dioxide is a product of a reaction between ozone and the nitrogen oxide emitted from car exhausts. Ozone itself is formed by the action of sunlight on nitrogen dioxide. In urban areas, large sources of nitric oxide (NO) convert a significant fraction of the ambient ozone concentration to  $\text{NO}_2$ . Since  $\text{NO}_2$  is readily photolysed to NO, the urban excess of  $\text{NO}_2$  is often regarded as a reservoir of photochemical oxidants (Photochemical Oxidants Review Group, 1993). Consequently, ozone concentrations in city centres are often lower than would otherwise be expected due to its reaction with nitrogen oxide emitted from car exhausts.

### 1.1.3 Trends in Air Pollution

Over recent decades, the profile of urban air pollution has changed considerably, with marked variations in levels of air pollution, due to the combined effect of environmental policy, technical innovation and changes in levels of human activity.

A major change has been the reduction in the levels of sulphur dioxide and black smoke. An important influence in this context was the 1956 Clean Air Act. This introduced controls on emissions from power stations and made it an offence to emit smoke from households in most urban areas. Largely as a result, levels of sulphur dioxide and black smoke in the UK fell by over 50% between 1970 and 1993 (Department of the Environment, 1996).

At the same time, however, emissions from road vehicles have increased (Table 1.4 shows increased emissions in the UK between 1970 and 1994) leading to a marked rise in levels of traffic-related pollutants such as NO<sub>x</sub>, CO, O<sub>3</sub>, airborne particulates and VOCs.

*Table 1.4 Increases in UK Estimated Emissions from Road Transport (1970-1994)*

Pollutant	% Increase
Carbon monoxide	31
Nitrogen oxides	72
Volatile organic compounds	29
Black smoke	147
Sulphur dioxide	43

(Figures from Department of the Environment, 1996)

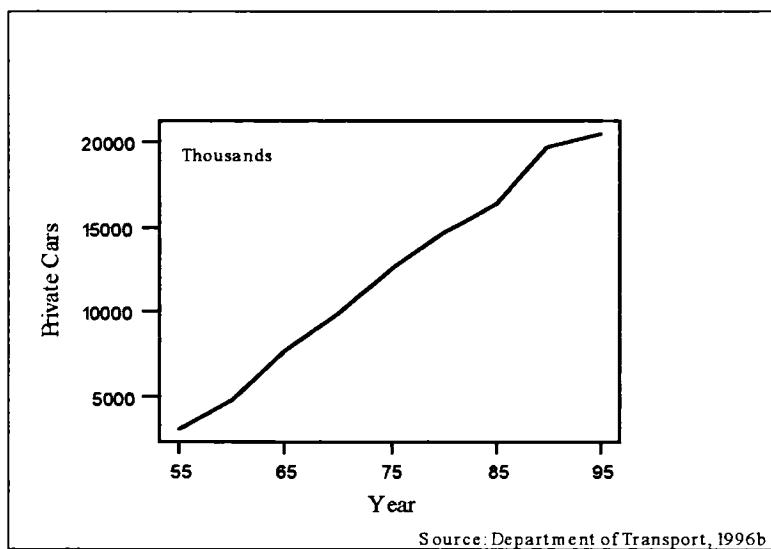
In the case of NO<sub>x</sub>, over recent years, decreases in industrial emissions in urban areas have been offset by increases in emissions from motor vehicles (Eggleston *et al*, 1992). Road transport emissions in the UK increased by 72% during the period 1981-1991, contributing 51% of the NO<sub>x</sub> emissions in 1991, while power station emissions declined by

about 14% and represented only 26% of the UK total emissions in 1991 (Photochemical Oxidants Review Group, 1993). However, emissions of NO<sub>x</sub> have declined in recent years due to the increasing number of diesel cars, which emit less NO<sub>x</sub> than non-catalyst petrol-fueled cars, and also due to increases in road fuel duty (Department of the Environment, 1996). Projected trends in emissions of NO<sub>x</sub> show that while emissions have decreased in recent years, predicted increases in the number of cars on the roads will result in a stabilisation of emissions by the year 2010 (Quality of Urban Atmospheric Review Group, 1993) and are likely to increase thereafter (Committee on the Medical Effects of Air Pollutants, 1995).

Internal combustion engines in vehicles produce large amounts of reactive hydrocarbons and nitrogen oxides, as well as fine particulates and carbon monoxide. Since 1971, the European Union has introduced a wide range of legislation to control these. This has helped to stimulate the development of both vehicle and fuel technologies, aimed at reducing emission levels - for example, unleaded gasoline, catalytic converters, low sulphur diesel fuel and advanced electronic engine management systems have all been introduced. Catalytic converters are designed to perform a dual role: to reduce NO in the exhaust gas and to oxidise hydrocarbons and CO. They nevertheless tend to be relatively inefficient - and indeed may actually increase emissions - at low engine temperatures and low speeds. Since the majority of car journeys in Britain are less than 5km (Department of Transport, 1996a), this means that their impact on emissions has been relatively limited.

The Department of Transport (1996b) estimated that the number of private cars licensed in the UK rose from 3.1 million vehicles in 1955 to over 20 million in 1995 (Figure 1.1); the total number of all vehicles on UK roads increased by 32% between 1980 and 1995. In 1989, the UK Government predicted that traffic volume would have increased by between 83% and 142% by the year 2025 (Department of the Environment, 1989). These increases are not restricted to the UK. In the European Union, the number of vehicles on the roads has more than doubled since 1970, and is set to increase by a further 40-50% in the next 20 years (Commission of the European Communities, 1992).

Future increases in the number of vehicles will be at least partly offset by continuing improvements in vehicle and fuel technology. This is likely to limit the rise in pollution levels to some extent.



*Figure 1.1. Private cars currently licensed: 1955 - 1995.*

Changes in the number of vehicles on the road do not tell the full story, for emissions also depend on the level of usage of the vehicles - itself a reflection of lifestyle. Since the early 1980s there has been a distinct change in lifestyle, partly related to increased earnings, but also related to a higher proportion of women in work and a greater emphasis on recreational activities. These changes have had a direct effect on numbers of vehicles and vehicle usage. The number of cars per household has increased. According to the Census, for example, the percentage of total households with 2 or more cars in England increased from 15.9% to 24.0% between 1981 and 1991. The use of public transport, such as coaches and buses, has decreased. Many people prefer to use their own car rather than public transport due primarily to convenience, comfort and reliability, but also increasingly to cost.

Since more women are working full-time they are, therefore, more likely to take their children to school by car and go shopping after work in larger, often out-of-town, stores. According to the Department of Transport (1996a) in the last decade there has been a 32% increase in miles travelled for shopping and a 108% increase in miles travelled to escort children to school. With the development of late-night shopping, people are more willing to travel further for shopping and leisure facilities, especially with many of the larger complexes combining the two. Health and fitness awareness campaigns have resulted in many people travelling to sports centres or to the countryside. Furthermore, the increased availability of company cars has contributed to greater use and longer travel to work distances. All this activity has had an impact on congestion, which is now not only a problem in the city centres but is also moving into the surrounding areas (Royal Commission on Environmental Pollution, 1994).

As a result of these changes, level of vehicle use in the UK rose from 77 billion vehicle kilometres in 1955 to 430 billion vehicle kilometres in 1995 in the UK (Department of Transport, 1996b). According to the Department of Transport (1996a) over the last 20 years the average daily distance travelled by car in Britain has risen by 57% to nearly 22 km. At the same time, the number of journeys made by car has risen by 45%. Although the number of journeys and the daily distance travelled has increased, it is interesting to note that 94% of all car journeys are less than 40 km (Department of Transport, 1996a). This has a direct impact on levels of emissions in urban environments.

## **1.2 Air Pollution and Health**

Against this background, there is growing concern about the effect of traffic related pollution on public health. This concern has been reflected in the media. The article 'Gasping for Breath' in the 'Independent on Sunday' in October 1993, for example, reported anxieties about the effect of exhaust fumes on children's health. The 'Independent on Sunday' published a review entitled 'You Hardly Dare Breathe: A Special

'Report on Air Pollution in Britain' (3rd March 1995), highlighting the issues of pollution for ozone, nitrogen dioxide and sulphur dioxide and the rising prevalence in asthma. In addition, these issues were also raised in a programme in Channel 4's Cutting Edge 'Fighting for Breath', on the 9th October, 1995. More recently, there have been a number of national and regional health surveys focusing on asthma and respiratory illness, for example, the Health Survey for England 1996, commissioned by the Department of Health and carried out by the Joint Surveys Unit of Social and Community Planning Research and the Department of Epidemiology and Public Health at University College.

Concern is far more wide-spread than the UK. Traffic related air pollution is now of international concern, not least because of the health effects of exposure, but also because of the longer-term, and potentially hazardous, implications on the environment and society. Road transport contributes to two major global atmospheric pollution issues: global warming and stratospheric ozone depletion. These issues may well, in the future, have substantial implications for public health.

In Europe alone, concern is reflected through a heightened research agenda for air pollution and health. The European Commission, for example, initiated the STEP, ENVIRONMENT, and Environment and Climate research programmes (European Commission, 1995) which all have air quality components. The World Health Organisation produced the Air Quality Guidelines for Europe (World Health Organisation, 1987). In addition, there have been a number of workshops and symposia: the Workshop on Air Pollution Epidemiology [Basel 1991], PHARE Symposium on Environment and Health [Bilthoven 1993], Workshop on Health Risk Assessment and Air Pollution Epidemiology [Brussels 1994] are just a few examples. These have coincided with the introduction of new annual conferences on the theme of air pollution and health, including 'Air Pollution' which is now in its fifth year, 'International Society for Exposure Analysis' which is in its sixth year and eight annual meetings of the 'International Society for Environmental Epidemiology'.

Much research has been undertaken in recent years to examine the association between exposure to outdoor air pollution and health. The major pollutants and their health effects are summarized in Table 1.5.

*Table 1.5 The health effects of the major pollutants.*

Pollutant	Groups at risk (levels above NAAQS)*	Clinical consequences
Airborne particulates (e.g. black smoke)	Children	Increased respiratory symptoms, increased respiratory illness, decreased lung function
	Chronic lung/heart disease	Excess mortality
	Asthmatics	Increased asthma exacerbations
Nitrogen dioxide	Healthy adults	Increased airway reactivity
	Asthmatics	Decreased lung function
	Children	Increased respiratory symptoms
Carbon monoxide	Healthy adults	Decreased exercise capacity
	Patients with ischemic heart disease	Decreased exercise capacity, angina pectoris
Ozone	Healthy adults and children	Decreased lung function, increased airway reactivity, lung inflammation, increased respiratory symptoms
	Athletes, outdoor workers	Decreased exercise capacity
Acid aerosols	Healthy adults	Altered mucociliary clearance
	Children	Increased respiratory illness
	Asthmatics and others	Decreased lung function
Lead	Children	Altered neurobehavioral function
	Adults	Increased blood pressure
Sulphur dioxide	Healthy adults and patients with chronic obstructive pulmonary disease	Increased respiratory symptoms, increased respiratory mortality and increased hospital visits for respiratory disease
	Asthmatics	Decreased lung function

(\* 1991 air quality data. Does not reflect exposure indoors or to brief peak levels of pollutants outdoors. NAAQS: National Air Quality and Emissions Trends Report, US Environment Protection Agency, 1991. Source: Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society, 1996.)

### **1.2.1 Acute Verses Chronic Health Effects**

The public health effects (respiratory and non/respiratory) of air pollution fall into two categories:

- *acute* - health effects caused by exposure to short-term (often extreme) exposure
- *chronic* - health effects caused by long-term, possibly cumulative, exposure to pollution

Studies looking at acute health effects often consider the association between short-term, specific pollution events and the incidence of illness. Most of these studies analyse daily time-series data and observe temporal associations between pollution events and measures of illness, such as hospital admissions or deaths. Acute health effects often occur when exposure triggers an existing illness, which brings forward symptoms of the illness (for example, asthma attack). Consequently, acute health effects are of greater risk, but relatively small numbers of people are affected.

Chronic health effects, on the other hand, are typically examined through population studies looking at the effects of long-term exposure on prevalence of health outcome, and may be associated with initial sensitisation to pollutants. Over long periods of time (i.e. one year or more), the cumulative effect of exposure may result in the onset of illness. In comparison to acute health effects, there is a low individual risk to chronic health effects, but large numbers of people can be affected, and therefore there is a greater public health concern.

Both types of health effects are difficult to quantify due to other factors that may be associated with the incidence of illness. These factors are known as confounding factors, and in order accurately to measure the association between pollution and ill health, they need to be taken into consideration and controlled for. In the case of acute health effects

confounding factors might include seasonal and daily variations in weather conditions, pollen levels and temperature. Chronic health effects may be associated with variations in population-based lifestyle factors such as socio-economic status, smoking habits, education (or parent's education), personal (or parental) history of asthma and housing conditions. It is important that information about confounders is collected as well as information on health effects and that the results of any analysis are adjusted for the confounders. Sophisticated statistical techniques are used to measure the association between air pollution and ill health, especially in the presence of confounding factors.

### **1.3 Trends in Respiratory Illness**

One of the major health concerns is the link between air pollution and respiratory illness, particularly asthma. At the same time as increases in many of the air pollutants have been recorded, increases in respiratory disorders and asthma have also been observed. Much of the knowledge about prevalence of asthma and its association with air pollution, however, comes from epidemiological studies.

Over the past 20 years, the proportion of children in the UK showing wheezing symptoms in the last year has increased by about 50%. Furthermore, currently 4-6% of adults and 10% of children in the UK have been diagnosed as having asthma and there may also be many others that remain undiagnosed (Committee on the Medical Effects of Air Pollutants, 1995). Figure 1.2 shows the percent prevalence of wheeze in the last year in schoolchildren in British surveys between 1965 and 1992. In addition, there has been a three-fold increase in hospital admissions for asthma since 1965. Figure 1.3 shows trends in hospital admissions for asthma in England and Wales between 1965 and 1985 and for England between 1985 and 1992.

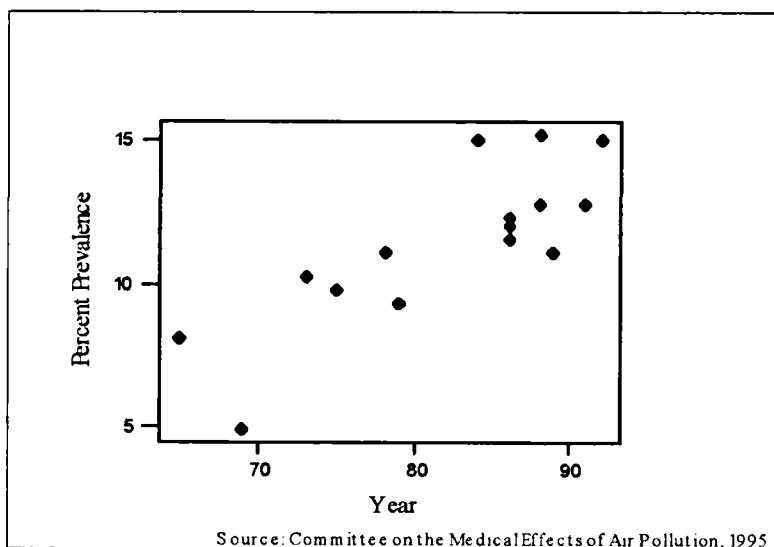


Figure 1.2 Percent prevalence of wheeze in the last year, British surveys 1965-1992.

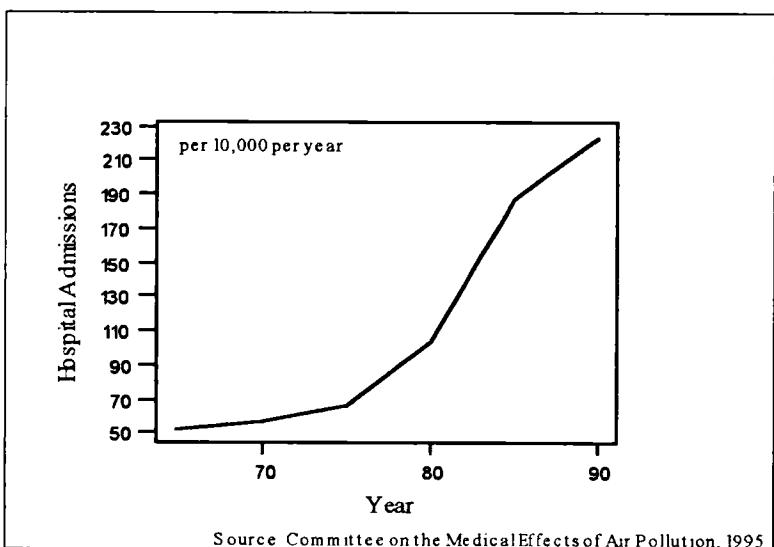


Figure 1.3 Hospital admission rates.

In recent years, several studies in the UK have shown increased admissions to hospital for respiratory diseases. Burr (1987), for example, reported increases in deaths from asthma in England and Wales between the 1970s and 1980s. Deaths among young persons aged 5 to 24 years have not only been rising since the mid 1970s, but have nearly doubled. In Wales, hospital admissions showed that the number of children between 5 and 14 years

admitted to hospital with a diagnosis of asthma increased fourfold from 1978 to 1984. Trends in hospital admissions for asthma in the South West Thames Region have also been shown to be on the increase. Between 1970 and 1985, the number of admissions rose by 186% for children aged 0 to 4 years and 56% for children aged 5 to 14 years (Anderson, 1989).

Other studies have looked at the prevalence of asthma over time by comparing results from two different studies a number of years apart, but in the same geographical location. Anderson *et al* (1994), for example, compared two studies, 13 years apart, in the London Borough of Croydon. The two studies were based on children aged between 7.5 and 8.5 in 1978 and 1991. Comparison of the two studies showed a 16% relative increase in the prevalence of wheezing in the last 12 months.

Another study (Nystad *et al*, 1997), compared two studies of schoolchildren aged between 6 and 16 years in Oslo. The two studies were undertaken in 1981 and 1994, with samples of 1772 and 2577 children respectively from school classes selected at random. The children's parents were interviewed by questionnaire, with questions focusing on asthma and respiratory symptoms. The lifetime prevalence of physician-diagnosed asthma was found to have increased nearly three-fold, from 3.4% to 9.3%, in the 13 year period between the two studies. At the same time, attacks of wheezing in the last three years had also increased from 3.7% to 6.8%.

Changes in the prevalence of asthma over time have also been observed in Busselton, Western Australia (Peat *et al*, 1992). A study of men and women aged 18-55 in 1981 was compared with a similar study in the same area in 1990. In the 1981 study 553 people were selected at random and in the 1990 study 1028 people were selected at random. The authors found that over the time period recent wheeze in the subjects had increased from 17.5% to 28.8% and diagnosed asthma from 9.0% to 16.3%. These increases were found to be greatest among people under 30 years old.

Many epidemiological studies have been applied to examine the link between increases in asthma and increases in air pollution. Epidemiological studies examine the association between exposure and public health effects, usually for sample populations. However, in some cases, for example in the study of rare diseases, large numbers of people are required to achieve statistical significance. The health effects of air pollution have important implications for the public as a whole, as the Committee on the Medical Effects of Air Pollutants (1995, p7) proclaim, ‘outdoor air pollution is a public concern, since all citizens are potentially exposed and outdoor air is a community resource’. The Committee on the Medical Effects of Air Pollutants (1995) splits epidemiological studies into two categories: temporal and spatial.

Temporal studies can be panel or time-series and provide information about acute effects of pollution on respiratory health. Panel studies are prospective studies where individuals are selected in advance depending upon the purpose of the study. Individuals can be healthy or suffer from respiratory disorders and are closely monitored, usually making a record of their daily activities and movements during the study period in diaries. The associations between health outcomes, such as lung function, respiratory symptoms or medication use, and air pollution are examined. Time-series studies observe the relationship between fluctuations in air pollution and fluctuations in hospital admissions and mortality due to respiratory illness over time.

In spatial studies, the association between health outcomes and levels of pollution for sample populations in different geographical locations is examined. The geographical locations are often chosen to reflect differences in levels of pollution and information pertaining to health outcomes is usually obtained through questionnaires. The measures of pollution used in such studies are normally the long-term averages - typically one year or more. Spatial studies are therefore often implemented to examine the chronic, cumulative, effects of exposure on health and are often termed ecological or cross-sectional studies.

### **1.3.1 Time-Series Studies**

Time-series studies generally examine acute health effects - predominantly the short-term temporal association between lung function or respiratory symptoms, measured in terms of hospital admissions or asthma attacks, and raised concentrations or pollution episodes. Most studies of these studies use daily time-series data. The relationship between air pollution and ill health is usually analysed through regression modelling. However, one of the problems associated with using daily time-series data is that many days have very small numbers of admissions - and in some cases no admissions at all. This results in a strongly skewed distribution with a lot of zero values. Thus one additional case, which may be random, can represent an apparently large increase in risk. This problem is often overcome by using a regression model with Poisson errors which assigns less weight to days with zero counts. Another problem is the presence of autocorrelation. This occurs when the respiratory symptoms and air pollution measurements exhibit cyclical fluctuations throughout the year which may result in unusually high daily correlations. This long-term trend can either be removed from the data before analysis or entered as an additional term in the regression model. Alternatively, an autoregressive model could be applied to the data to account for the autocorrelation (Committee on the Medical Effects of Air Pollutants, 1995).

There has been an abundance of time-series studies looking at the relationship between air pollution and respiratory disorders, and the methods and techniques are now well established and documented. A summary of articles pertaining to time-series studies can be found in Pope *et al* (1995b). There are also many more recent articles. For example, Ponka *et al* (1996) report on a study analysing daily hospital admissions for asthma in relation to ambient levels of SO<sub>2</sub>, NO<sub>2</sub>, total suspended particles (TSP) and O<sub>3</sub> in Finland, Helsinki, between 1987 and 1989, and found significant positive associations for O<sub>3</sub> after adjusting for confounders.

Wordley *et al* (1997) assessed the relationship between particulate matter with a diameter less than 10 µm ( $PM_{10}$ ) and hospital admissions and mortality for all respiratory diseases in Birmingham, UK, between April 1992 and March 1994. The results were adjusted for confounders and significant associations were found.

A further study looking at hospital admissions is reported by Ponce de Leon *et al* (1996). The authors examined the relationship between hospital admissions for respiratory disease and black smoke,  $SO_2$ ,  $O_3$  and  $NO_2$  in London, UK, during two periods: 1987 -1988, and 1991 - 1992. After taking into account the confounders, significant associations were found with  $O_3$ .

Other studies have focused on pre-selected sample populations showing symptoms of respiratory illness. Peters *et al* (1997) report on one such panel study to examine the short-term effects of  $SO_2$ , TSP,  $PM_{10}$ , particle strong acidity (PSA) and fine particle sulphate concentration ( $SO_4$ ) on children with asthma or related chronic respiratory disease in Sokolov, Czech Republic, during the winter of 1991 to 1992. After adjusting for confounders, decreased peak respiratory flow rates, increased respiratory symptoms and increased medication use were found to be associated with elevated levels of air pollution.

In another study, the incidence of acute childhood wheezy episodes, measured in terms of hospital admissions for 1025 children, over a period of one year between March 1992 and February 1993 in the London Borough of Hillingdon, UK, were examined in relation to  $O_3$ ,  $SO_2$  and  $NO_2$  (Buchdahl *et al*, 1996). Strong associations were found with  $O_3$  and  $SO_2$ , and weaker, but significant, associations with  $NO_2$  after controlling for confounding factors.

One of the main limitations in the majority of these studies is that the pollution data used in the studies is often collected for just one site. It is therefore assumed that variations at that site reflect variations in the wider geographical area, and in levels of exposure of the

study population as a whole. However, levels of pollution vary spatially, as well as temporarily, due to changes in pollution sources and features on the surface of the Earth over distance. Pollution measured at a site is therefore unlikely to be representative of pollution at locations further away from the site. Thus levels of exposure will also vary spatially.

### **1.3.2 Spatial Studies**

In comparison to studies looking at acute health effects, very few studies have been undertaken to look at chronic health effects and consequently relatively little is known about the long-term health effects of air pollution.

In the USA, Dockery *et al* (1993) examined the relationship between mortality rates and SO<sub>2</sub>, O<sub>3</sub> and suspended sulphates in six US cities (controlling for confounding factors). Results suggested that particulate air pollution was positively associated with death from lung cancer and cardiopulmonary disease, and that mortality was most strongly associated with fine particulates, including sulphates.

Schwartz (1993) reported on a study undertaken in 53 urban areas in the U.S. The author examined the relationship between rates of chronic respiratory illness and TSP. After controlling for confounders, TSP was found to be associated with increased risk of chronic bronchitis and a respiratory diagnosis by an examining physician.

In another study, the relative risk of mortality for half a million adults in 151 U.S. metropolitan areas was assessed with respect to sulphates and fine particulates through a prospective study between 1982 and 1989 (Pope *et al*, 1995a). Particulate air pollution was found to be associated with cardiopulmonary and lung cancer mortality after adjusting for confounding factors.

### **1.3.2.1 Vehicle Emissions and Respiratory Illness**

Although these studies have concentrated on particulate air pollution in general, attention is increasingly being directed towards the specific relationship between traffic related pollution (i.e. vehicle emissions) and respiratory disorders. There is evidence to suggest that emissions from road traffic are having an effect on health. In an attempt to clarify this association a number of spatial studies have been undertaken that have analysed the relationship between respiratory disorders and indicators of exposure to vehicle emissions (for example, distance to major roads or traffic density) for sample populations. Several of the studies reported positive and significant associations between the indicators and the prevalence of some respiratory disorders.

Edwards *et al* (1994), for example, report on a study of pre-school children admitted to hospital for asthma and an association with traffic flow in Birmingham, UK. Pre-school children (i.e. under the age of 5) were chosen mainly because they spend much of their time at home, but also because asthma is most prevalent in this age group. Three groups of children were chosen: cases, hospital controls and community controls. The cases correspond to hospital admissions for asthma, the hospital controls to hospital admissions for non-respiratory cases (excluding road accidents) and the community control consisted of a random sample of pre-school children. The authors state that in the county of West Midlands, which includes the city of Birmingham, hospital admissions for all cases has been found to be closely related to deprivation and therefore the hospital control was included partially to account for socio-economic confounding. The postcode was used to locate the place of residence for all groups and each child classified according to distance from the nearest major road (0-200m, 200-500m and >500m). Traffic flow data were attributed to all major roads and each child also classified according to traffic flow on the nearest major road. The authors found that children admitted to hospital for asthma were significantly more likely to have high traffic flow along the nearest segment of main road compared to the two controls. They also found a significantly greater risk of hospital

admissions for asthma amongst children living less than 500m from main roads, but only compared to the hospital control.

The impact of road transport on residents living along main busy roads has been studied in Japan (Murakami *et al*, 1990). The 4 day averages of indoor and outdoor concentrations of nitrogen dioxide and suspended particulate matter were measured five times during four seasons. The measurements were taken at the same time in 200 houses where the households consisted of families. Members of the families were interviewed by questionnaire. Distance from the nearest main road was used to classify the houses into three groups, based upon the following distance bands: 0-20m, 20-50m and 50-150m. Prevalence rates of respiratory symptoms were found to be higher in families (children and parents) located within 20m of a main road than those in the 20-50m and 50-150m bands. Prevalence rates were found to be as much as 50% higher in these locations for asthma-like symptoms such as wheezing.

Another study, reported by Ishizaki *et al* (1987), looked at the prevalence of allergic rhinitis among people with different exposures to vehicle emissions and cedar pollen in the Nikko-Imaichi district of Japan. In a questionnaire survey, 3133 children and parents from 631 families provided information on the prevalence of cedar pollinosis (hay fever) during the months of March and April. A statistically significant difference between the prevalence of cedar pollinosis in different areas of the district was observed. In areas close to tree-lined inter-city highways, the prevalence of cedar pollinosis was 13.2%, approximately 1.5 times greater than the prevalence in city farming areas close to, and far away from, cedar forests. In forest areas with little traffic there was a prevalence of 5.1% and prevalence was as low as 1.7% in mountainous areas above the treeline (area descriptions and figures from Committee on the Medical Effects of Air Pollutants, 1995).

Results from these studies present a strong argument to suggest an association between health and respiratory disorders; however, the results could be misleading due to the lack of information about socio-economic confounding. Factors such as socio-economic

status, housing conditions (dampness, overcrowding), pets, parental asthma and parental cigarette smoking have not always been taken fully into account. Furthermore, the study reported by Ishizaki and colleagues does not provide any quantitative estimate of exposure to vehicle emissions, and in several of the studies the exposure indicator is relatively weak. Nevertheless, other studies which have adjusted for potential confounders have still displayed positive associations.

In Bochum, Germany, for example, schoolchildren were interviewed by questionnaire in June and July 1991, focusing on respiratory disorders and related indicators (Weiland *et al*, 1994). The schoolchildren were 7th and 8th graders from 13 schools, chosen at random from 38 schools in the area. The questionnaire concentrated on wheezing during the last 12 months (a video was used to aid correct diagnosis) and included questions on socio-economic characteristics and exposure to traffic density. On the questionnaire, exposure was defined in terms of two indicators of traffic density: type of street (main road or side street) and frequency of trunk traffic (never, seldom, frequent or constant). Results were adjusted for identified confounders and an increase in the prevalence of wheezing and allergic rhinitis for children living on main streets, as opposed to side streets, was observed. A positive and significant association was also observed between the frequency of trunk traffic and both the prevalence of wheezing and allergic rhinitis. Results from this study, however, are not wholly conclusive due the nature of data collection, in that measures for social characteristics and traffic exposure were obtained through children's self-reporting, which cannot always be considered reliable.

In another study in Germany, Wjst *et al* (1993) examined the prevalence of allergic and asthmatic diseases in children aged 9-11 in Munich in relation to traffic volumes. The study took place during 1989 and 1990. A questionnaire was sent to the parents of 7445 children (with a response rate of 88%) and included questions on respiratory symptoms, for example, physician diagnosis of asthma and wheezing, along with demographic data. All the children, selected in random order, participated in lung function tests throughout the year. Children with a nationality other than German, who had been resident in Munich

for less than 5 years, or with acute respiratory infections, were excluded from the study. In Munich there are 117 primary school districts and the results from the questionnaire and lung function tests were aggregated for these districts. Using data from the traffic census survey, a measure of traffic volume was also attributed to each district based upon the street with the highest volume of traffic in that district. Results were controlled for the following confounders: parental history of asthma, parental school education, number of people in the house, use of gas or coal for cooking or heating, month of the survey, number of cigarettes smoked in the home and who compiled the questionnaire. Associations were observed between traffic load and reduced pulmonary function and increased respiratory symptoms. Lifetime prevalence of recurrent wheezing was found to be significantly and positively related to traffic volume.

Nitta *et al* (1993) report on a study where a total of 4822 women were interviewed in three separate surveys, between 1979 and 1983, and the responses were compared with distance of the home to the nearest major road. All the women were aged between 40 and 59 and lived in Tokyo, but had not recently moved there. After adjusting for confounders the prevalence of long-term wheeze was found to be significantly greater in women living within 20m of a busy road in two of the surveys.

A further study in Germany collected data for 3 areas in the Baden-Württemberg region: Stuttgart, Tübingen/Reutlingen and Freudenstadt (Wichmann *et al*, 1989). Parents of 8420 children aged 6 years were interviewed by questionnaire, with a 93% response rate. Crude prevalence for croup syndrome and obstructive bronchitis was found to be 9%. Pollution data was only available for Stuttgart and after adjusting for confounders, prevalence of croup syndrome was found to be slightly greater in streets with high traffic load. For childhood asthma, the authors also observed a correlation with traffic-dependent pollutants (NO<sub>2</sub>, NO, CO) and traffic load.

Oosterlee *et al* (1996) report on a study researching the chronic respiratory symptoms of children and adults living on streets with high traffic volume in Horalem, The Netherlands.

Parents of 291 children and 1485 adults were interviewed by questionnaire with questions relating to respiratory illness, including chronic cough, wheeze, dyspnoea, doctor's diagnosis of asthma and medication. The children and adults either lived on streets with heavy traffic or on quiet control streets. Selected streets were identified from environmental traffic maps. Results from the study showed higher prevalences for most respiratory conditions of children who live along busy traffic roads compared to the control after adjusting for confounders. In adults, only dyspnoea was found to be significant.

Lung function in children living near motorways in the Netherlands was examined by Brunekerp *et al* (1997). Children living in one of six areas near motorways and in one of 13 schools were selected for the study. The lung function of the children was measured in the schools and exposure to traffic-related air pollutants assessed through traffic counts of automobiles and trucks on the motorways. In addition, concentrations of PM<sub>10</sub> and NO<sub>2</sub> were measured in the schools. Information about confounders was collected through a questionnaire. All children living within 1000m of the motorway, with a valid lung function test and a usable questionnaire were included in the analysis; 778 children in total. After adjusting for confounders, lung function was found to be associated with truck traffic volume and the association was stronger in children living nearer to the motorway. Lung function was also found to be associated with black smoke in schools.

These studies have identified a range of possible associations between air pollution and health. In epidemiology, however, one of the major problems associated with examining the health effects of exposure to air pollution is estimating an individual's level of exposure. Many different factors will affect an individual's exposure, for example, time spent indoors and outdoors, daily activities such as travel to work or travel to school, age and health of the individual.

Against this background, doubts must be raised about the accuracy of many of the exposure indicators used in studies examining the relationship between traffic-related

pollution and health. Although estimates of exposure such as distance to the nearest main road or traffic volume on the nearest main road may give an impression of community exposure, the relationship between air pollution and health is difficult to interpret, as they are unlikely to provide reliable estimates of individual exposure. As a result, the true effect of exposure to traffic-related pollution on health is not known with any degree of certainty. As the Advisory Group on the Medical Aspects of Air Pollution Episodes (1993, p119) stress, ‘there are no systematic measurements of the levels to which people are actually exposed’.

Despite these problems, there continues to be widespread concern about the relationship between air pollution and health. One solution to help overcome these problems would be to provide better estimates of individual exposure. Measuring exposure at a personal level is likely to provide the most reliable estimates of individual exposure. However, in spatial studies, where the association between health outcome and pollution for populations in different geographical locations is examined, personal monitoring would be extremely difficult to achieve due to the size of the study population. An alternative approach to personal monitoring - air pollution mapping - may provide more reliable estimates of exposure for sample populations than the exposure indicators described in this section.

#### **1.4 Environmental and Health Policy**

In response to the increasing levels of air pollution and the possible link with respiratory illness, many agencies and government bodies have established committees, expert panels and advisory groups to report on the health issues of air pollution and present standards, legislation and recommendations for levels of air pollution and emissions.

In 1985, The Commission of the European Communities established air quality standards and published limit values for pollution in their 85/203/EEC directive. The limit values are levels of pollution which must not be exceeded without subsequent action designed to

prevent its recurrence. In 1987 the World Health Organisation also published guidelines on levels of pollution based upon the protection of human health. The guidelines relate to concentrations above which health effects may be discernible, and have formed the basis for many national air quality standards.

More recently, in the UK, the Department of the Environment has set standards and objectives for the main air pollutants representing the Government's present judgment of air quality targets. The standards are reproduced in Table 1.6 - the different pollutants have different averaging times depending upon whether chronic or acute exposure is considered important. The United Kingdom National Air Quality Strategy (1997) now recommends that local authorities identify air quality management areas if air pollution exceeds these standards.

*Table 1.6 The United Kingdom National Air Quality Strategy standards for the major pollutants.*

Pollutant	Standard	
	Concentration	Measured as
Benzene	5 ppb	running annual mean
1,3-Butadine	1 ppb	running annual mean
Carbon monoxide	10 ppm	running 8-hour mean
Lead	0.5 µg/m <sup>3</sup>	annual mean
Nitrogen dioxide	104.6 ppb 20 ppb	1 hour mean annual mean
Ozone	50 ppb	running 8-hour mean
PM <sub>10</sub>	50 µg/m <sup>3</sup>	running 24-hour mean
Sulphur dioxide	100 ppb	15 minute mean

(Source: The United Kingdom National Air Quality Strategy, 1997)

In the case of traffic-related emissions, the vast majority of policy is directed towards maintaining these standards and ensuring that pollution levels do not exceed these standards, primarily through strategies to reduce and control the volume of traffic on the roads. As the Royal Commission on Environmental Pollution (1994, p234) states, one of the main objectives for a sustainable transport policy is 'to achieve standards of air quality that will prevent damage to human health and the environment'. The aim is to achieve this by reducing the dominance of cars and lorries and providing alternative means of transport, and therefore improving the quality of life in towns and cities.

As previously discussed, one solution to help meet these standards will be the implementation of technological changes in vehicle design to try and reduce emissions. However, as the Royal Commission on Environmental Pollution (1994, p233) suggest in their conclusions and recommendations,

‘even allowing for technical improvements in vehicle design, the consequences of growth on such a scale [the Government’s 1989 figures for increases in traffic volume - between 70% and 140% by the year 2025] would be unacceptable in terms of emissions, noise, resource depletion, declining physical fitness and disruption of community life’.

Improved technology will, nevertheless, be necessary, but clearly other actions are also essential. The most important of these will be the introduction of new traffic schemes designed to ease congestion and thereby reduce emission-generating stops and starts; this might include new roads, such as orbital routes or by-passes. However, these will need to be introduced in conjunction with policies to reduce the number of vehicles on the roads, for example, policies to encourage alternative means of transport, such as park and ride schemes, improvements to the rail network, and pedestrian, cycle and public transport priorities.

In order to introduce these actions it is important to understand where such schemes will have the greatest impact and, as highlighted in the Design Manual for Roads and Bridges (1994, 6/1), it is necessary to ‘identify areas where it is likely that air quality will be improved as a result of reduced traffic flows, changes in traffic speed or reduced congestion or queuing times’. To date, there have been very few studies looking at the effect of traffic management strategies on vehicle emissions. It is, however, imperative that detailed studies are undertaken before and after schemes are implemented to monitor the impact on levels of pollution. This is difficult in cities, where measurements of air quality are based upon a limited number of monitoring stations and in some cases, just one monitoring station which is assumed to be representative of air pollution across the city.

In order to implement effectively new traffic schemes it is important to understand the spatial distribution and variations of city-wide air pollution. One of the most efficient methods of providing this information is to produce a map of the pollution surface. Variations in urban traffic-related air pollution occur over very small distances, reflecting changes in the local and global environments, such as meteorology, topography, housing density, industry and traffic volume. To help establish accurate and reliable city-wide predictions of air quality it is essential that detailed maps are produced.

The research presented in this thesis examines the problem of predicting variations in air pollution and producing detailed maps of the pollution surface in urban environments.

Pollution surfaces are very complex - varying in both spatial and temporal dimensions. A model, which is a simplified representation of the pollution surface, is therefore applied to portray the variations in air pollution. The pollution surface is then visualised through mapping. Maps can be used to help investigate spatial patterns of air pollution, identify hot spots and areas where pollution levels exceed guide-lines and standards. Analysis of the spatial distribution of air quality can aid the design of monitoring networks and the selection of new locations for monitoring stations.

Maps have an important role from a health perspective. Epidemiologists looking at the geographical distribution of health are looking at sample populations in different geographical locations. The epidemiologists require accurate estimates of human exposure to air pollution in order to identify at-risk populations and quantify the levels of exposure involved. Studies of the links between air pollution and health have been hampered by the lack of accurate and precise estimates of exposure. Maps offer an effective method of estimating exposure at the individual and small area scale and thus helping to establish associations between pollution and health.

Against this background it is vital that a true picture of the pattern of air pollution is drawn. It is clear, however, that the accuracy of the maps will have a direct impact on the success of control strategies. It is therefore essential that reliable predictions of air quality can be derived. To achieve this the spatial variations in air quality that exist at street level, in urban environments, need to be identified and modelled.

Understanding the pattern and pathways of pollution at ground level is of great value to the policy makers and planners. As the Quality of Urban Atmospheric Review Group (1993, p12) illustrates,

'predicting air quality in urban areas serves a number of important roles. At a local level it allows the impact of a new schemes to be assessed, e.g. a new road or industrial plant. At a broader level, modelling can be used to help policy formulation by testing the impact of various policy options on air quality'.

Proposals to amend the legislation on air pollution are likely to benefit from information on the distribution and character of major emission sources and levels of pollution. Identifying the main contributors to air pollution and areas that do not comply with EC directives will aid the development of targeted control strategies. Models can be used to forecast future (and past) levels in air quality, thus informing the design and implementation of control strategies. The models can then be used to monitor the effect

of control strategies on levels of pollution after implementation. Furthermore, modelling air quality will help identify and define effective indicators of emissions thereby helping to establish more reliable predictions in areas where the sampling density is low or non-existent.

## **1.5 Aims and Objectives**

Against this background, the overall aim of this thesis is to examine the problem of producing detailed maps of pollution surfaces in urban environments. In order to achieve this, the research presented in the thesis is based upon the following specific objectives:

- to develop a range of different methods for mapping air pollution within a Geographical Information System (GIS) environment;
- to apply these methods to examine air pollution in urban environments, in particular in Huddersfield, UK;
- to compare and evaluate the performance of the different methods as a basis for informing policy management and epidemiology investigations.

To encompass the overall aim and objectives, the thesis is structured as follows:

- |           |  |
|-----------|--|
| Chapter 1 | reviews the issues associated with traffic-related air pollution and respiratory illness, and introduces the problem of pollution mapping; |
| Chapter 2 | reviews the traditional approaches to air pollution mapping and introduces the application of GIS and three new methods;                   |
| Chapter 3 | describes the data used to develop, test and validate the different traffic-related air pollution mapping methods;                         |
| Chapter 4 | demonstrates the application of the geostatistical technique, kriging;   |

- Chapter 5** describes the development of an automatic approach based upon a combination of dispersion modelling and kriging;
- Chapter 6** describes the development of a moving window approach based upon GIS neighbourhood operations and kriging;
- Chapter 7** describes the development of a regression based model using GIS neighbourhood operations and statistical techniques;
- Chapter 8** compares and evaluates the different methods;
- Chapter 9** presents conclusions and further research.

## **CHAPTER 2 MAPPING AIR QUALITY**

It was established in Chapter 1 that there is a need for accurate and reliable area-wide estimates of air quality and one way of achieving this is through mapping. The production of maps of the pollution surface involves estimating pollution concentrations at unsampled locations.

### **2.1 Spatial Variations in Urban Air Pollution**

Deriving estimates at unsampled locations is a difficult task due to the complex nature of urban air pollution - which, at a local level, reflects patterns in emission sources, the dispersion environment, turbulence and chemistry of the pollutants. Levels of pollution in urban areas change in response to local variations in emission sources. In the case of vehicle emissions, density of the road network, type of road, composition of traffic, volume and speed of traffic, will all influence emission patterns. The environment through which the pollutants then move affects the rates and pathways of dispersion, for example the density of buildings adjacent to the sources and the climatic conditions. As a result, variations in the levels of air pollution in urban environments occur over very small distances.

In the case of NO<sub>2</sub>, for example, research undertaken by Laxen and Noordally (1987) to investigate the distribution of NO<sub>2</sub> in street canyons suggested that NO<sub>2</sub> concentrations decline very rapidly from the center of roads, with concentrations close to local background levels at a distance of 30m. The research also showed that at some locations nearly two-fold variations in NO<sub>2</sub> were present over distances less than 100m. High

degrees of local variation in urban areas have also been found by Hewitt (1991), who examined spatial variations of NO<sub>2</sub> in the city of Lancaster, UK (based on diffusion tube measurements at 49 sites).

Motor vehicle emissions dominate NO<sub>2</sub> concentrations in cities and near to roads, with the heights of emissions from the vehicles strongly influencing the spatial patterns in NO<sub>2</sub> concentrations at ground level. As Wardlaw (1993) points out, 'outdoor concentrations are very localised being highest by the kerbside and falling rapidly away from the pavement and with height above the pavement. Concentrations correlate with density of cars, maximum concentrations occurring during rush hours'.

Furthermore, as Larssen *et al* (1993) suggest, 'the relationship between air pollution exposure and health effects associated with it are often hampered by the inaccuracy of determination of the air pollution concentrations that the group or individuals being studied are actually exposed to'. It is evident, therefore, that air pollution needs to be modelled and mapped at a resolution that reflects these spatial variations.

There are two main approaches to mapping air quality:

- ***dispersion modelling*** - estimating pollution concentration from emissions data based upon a model of dispersion processes (generally Gaussian)
- ***spatial interpolation*** - estimating patterns of pollution from point data derived from field monitoring

## 2.2 Dispersion Modelling

Air quality dispersion models can be characterised as either point, line or area dispersion models. The earliest dispersion models were developed to model the path of pollution from industrial point sources. As highlighted in Chapter 1, in recent years concern has

begun to change from industrial pollution to vehicle pollution. As a result, a number of line dispersion models have been developed for road transport. Both line and point models are source-receptor models and can be used either to calculate pollution concentration at one, or a number, of specific locations (e.g. the place of residence of an individual subject), or to produce an array of points as a means of producing a pollution map.

The models use information on two main sets of factors: characteristics of the emission source and characteristics of the dispersion environment. Point dispersion models use information about the source of pollution - for example stack height, stack diameter and emissions data - and meteorological data. The model calculates the plume of dispersion around the stack using this data. The width and depth of the plume depends primarily on the wind speed and turbulence of the atmosphere. The pattern of dispersion over distance from the source is described by fitting a model, such as a Gaussian model.

Line dispersion models provide estimates based on emissions from one or more linear sources, such as roads. The linear sources are represented by links, where a link is usually characterised as a straight segment of uniform conditions. The models use information about emission source (e.g. composition of traffic, traffic volume, emission rates, traffic idling time, traffic speed) and the dispersion environment (e.g. meteorology, mixing height, information about the surface adjacent to the road).

An early example of a line dispersion model is described by Dabberdt *et al* (1973) and Johnson *et al* (1973), who report on the development of a sub-model to the APRAC model. APRAC was a multipurpose urban diffusion model for predicting concentrations of CO. The sub-model, STREET, describes the microscale street effects (i.e. the effects of street canyons) and was developed because 'cross-street CO concentrations often differ by a factor of 2 to 3' (Dabberdt *et al*, 1973). Other dispersion models based on the street canyon theory have since been developed to try and improve on the STREET model.

One such example is the Canyon Plume-Box Model (Yamartino and Wiegand, 1986) developed in Germany for modelling the dispersion of NO<sub>x</sub>, NO<sub>2</sub> and CO.

In the late 1970s several models to predict pollution near roadways were developed in the USA (Rodden *et al*, 1982). Over time a number of these models have grown into commercially available products, of which the most notable and widely used is the CALINE suite of models (Quality of Urban Atmospheric Review Group, 1993). CALINE4 is the latest version of the California Line Source Dispersion Model, developments on which were reported as early as 1972 (Benson, 1992), and predicts CO, NO<sub>2</sub> and aerosols. CALINE4, and its predecessors, have been widely used across the USA and now also in the UK.

More recently a model has been developed in The Netherlands to determine city street air quality, the CAR model (Eerens *et al*, 1993). This is widely used throughout the Netherlands by municipalities and central government. The model was developed to take account of circumstances when streets do not behave like street canyons - for example where there are open spaces in cities, such as parks, squares and gaps between houses - that the authors felt other models had neglected. As Eerens *et al* (1993) note, 'at most locations near city traffic, including intersections of streets, the buildings and trees do not form a street canyon'. The CAR model calculates the annual percentile values and average concentrations at the kerbside of roads for non-reactive air pollutants and NO<sub>2</sub>. The model is predominately used for implementation of national air quality decrees, traffic and environmental planning, and the evaluation of current policy and the consequences of new policy proposals (Eerens *et al*, 1993).

Over the last few years, changes in computer technology have resulted in the adoption of Windows based software products, encouraging the development of dispersion modelling packages such as ADMS. ADMS is an expansion of the Highways Agency Design Manual for Road and Bridges (DMRB) model developed by Cambridge Environmental Research Consultants Ltd. Using windows functionality, ADMS models the dispersion of

pollution from point, line, area or volume sources for a variety of pollutants and the results can be graphically displayed within the system.

Line dispersion models are relatively effective at estimating pollution levels for small localised situations, for example, at a road junction. The models, however, tend to be extremely data intensive and it is often the case that all the data are not available. Using point source models to create maps is relatively straightforward, but this is not the case for line source pollution. Variations in the characteristics of a road change quite rapidly in urban areas - for example due to changes in the housing density at the side of a road, variations in traffic volume and changes in the direction of the road. Consequently, in area-wide studies, a great many road segments need to be modelled. Where a whole city has to be modelled, the demands on data are therefore very high. This tends to render the method inappropriate for area-wide studies.

Furthermore, line dispersion models can only accurately model pollution levels close to the source, for example up to a distance of 35m in the case of the CAR model and about 200m in the CALINE and DMRB models. In the case of the CAR model, a city-wide background concentration is calculated based on the regional background concentration and the radius of the city. Most line dispersion models, however, provide no estimate of background concentrations and the use of a constant for the background concentration fails to reflect the variation in pollution levels which may occur.

Another disadvantage of dispersion models is that they only model dispersion from those sources which have been identified. As Table 1.3 demonstrated, a high proportion of the pollutants associated with road transport also derive from other sources, such as industrial or domestic. In order to establish area-wide estimates of pollution, it is necessary to model pollution from these different sources. This is often achieved by undertaking a full emissions inventory for a study area. All sources of pollution are identified and the total emissions recorded for grid cells superimposed over the study area. An alternative is to use models that predict dispersion from a combination of pollution sources. In practice

this is often achieved by aggregating emissions to a grid and then using an area-based dispersion model, such as a box-model. The grid dimensions for this type of model are usually of the order of 1 to 10 km (Russell, 1988).

Simpson *et al* (1990), for example, calculated NO<sub>x</sub> concentrations in the UK at a 10 km resolution using a source-receptor dispersion model, the results from which were combined with the European contribution of NO<sub>x</sub>, calculated using the EMEP NO<sub>x</sub> model (with UK emissions set to zero) at 150 x 150 km resolution. The source-receptor dispersion model included two sub-models: one to describe the long range transport,  $\geq$  5km from the source, and the other to describe the shorter range transport of NO<sub>x</sub>. The model used data on emissions from motor vehicles (obtained from the national database at Warren Springs Laboratory on a 10 x 10 km grid), data related to point sources (i.e. power stations) and meteorological data.

Her Majesty's Inspectorate of Pollution (1993) undertook a study to look at the effect on levels of NO<sub>x</sub> in the East Thames corridor in the UK, if new industrial sources of NO<sub>x</sub> were introduced. Concentrations of NO<sub>2</sub> were estimated on a 5 x 5 km resolution grid, based upon modelled concentrations, using the PLUMES dispersion model and emissions data from present industrial sources in the East Thames corridor, and the spatial distribution of NO<sub>2</sub> on a 5 x 5 km resolution from the national database at Warren Springs Laboratory.

Alexopoulos *et al* (1993) developed a model to estimate emissions from traffic in Athens at a 1 km resolution. A grid was superimposed over Athens and the contribution of emissions from line sources within each cell was established, using information on traffic composition, type of road and road length.

Although these area-wide methods are useful on a regional or national scale - to give a generalised impression of the spatial distribution of air quality (for regional or national comparisons) - from a policy or health perspective they do not reflect the true variations

that occur at ground level, and consequently are not readily applicable to city-wide studies. This is especially true for vehicle-related pollution where the major variation occurs at street level, close to the sources. As Russell (1988) points out, there are problems associated with grid-based air quality modelling at such coarse resolutions, due primarily to 'a mismatch between the high concentrations that in fact do exist near the sources versus the lower concentrations computed by a model that immediately mixes those emissions throughout a grid cell of several kilometres'. Exposure estimates are thus rendered homogenous across the grid cell.

### **2.3 Spatial Interpolation**

An alternative to dispersion modelling in area-wide studies is spatial interpolation from monitoring sites. Pollution at any one location is the product of combinations of pollution from all sources within the surrounding area. The advantage of using monitored data is that the different sources of pollution at a location do not need to be identified.

Air quality is a spatially continuous variable. The locations of the monitoring sites and the air pollution values measured at those sites are used to describe the spatial distribution of air pollution across the whole of the study area. Predictions of values at points where the attribute has not been sampled can then be established (Bailey and Gatrell, 1995). Spatial interpolation is the term applied to techniques employed to derive predictions at unsampled locations.

A variety of spatial interpolation techniques exist, differing in terms of the underlying algorithm used to describe spatial patterns and each with its own advantages and disadvantages. The most commonly used techniques include distance weighted averages, Delaunay triangulation, splines, trend surface analysis and kriging. Visual representation of the results is usually achieved by contouring. Reviews and comparisons of the various techniques can be found in El Abbass *et al* (1990), Lam (1983), Dubrule (1984),

Burrough (1986), Laslett *et al* (1987) and Knotters *et al* (1995). Many of these authors conclude that the choice of technique is largely dependent upon the application, the required accuracy of the end result and the spatial distribution and number of sampling points.

For distance weighted averages, estimates at unsampled locations are found by averaging a set of neighbouring sample points which are weighted by distance, so that nearer points have greater influence on the estimate than distant points. In the case of Delaunay triangulation, triangles, as close to equilateral as possible, are created between sample points. If the values at the vertices of triangles are assumed to be heights, then the value at an unsampled location can be calculated by simple geometry. An alternative to Delaunay triangulation, albeit conceptually less sophisticated, is the Theissen (or Voronoi) polygon, whereby every location is assigned the value of its nearest sample point. Splines are computationally more complex: a series of polynomials, usually cubic or quadratic, are fitted to small neighbouring sub-sets of the total sample. The polynomials are constrained to give a continuous surface.

The local estimators, described above, could be used to predict values at unsampled locations. However, as Bailey and Gatrell (1995) highlight, 'the justification for doing so is weak. The techniques do not involve an explicit statistical model for the data under consideration and make no attempt to incorporate explicitly the possibility of spatial dependence. As a result none of the methods provide any estimate of the errors that can be expected in the results.'

One method that does involve an explicit statistical model is trend surface analysis. This is by far the most commonly applied technique for describing global trends in the data. Polynomials are fitted by least squares regression on the spatial co-ordinates. This has a number of disadvantages, as Oliver and Webster (1990) have identified. In the first instance, the powerful smoothing inherent in the technique results in a loss of detail; any

outliers or errors in the data cause instability in the model and variation in one part of the region will affect the fit of the surface everywhere.

Oliver and Webster (1990) explain that 'the nature of most spatial properties is such that it defies simple mathematical description. Rather, properties of the natural environment, at least, seem to behave as random variables'. A technique that will take this into account must therefore involve an explicit statistical model and provide estimates of the errors. A group of geostatistical techniques, known collectively as kriging, has been developed for this purpose.

Kriging is an optimal interpolator the estimates from which are unbiased and have known minimum variance (Oliver and Webster, 1990). The technique is based upon the theory of regionalized variables and utilises the spatial structure of the data. The theory of regionalized variables is discussed in detail in Journel and Huijbregts (1978). Kriging involves the construction of a semi-variogram and the fitting of an appropriate model.

The most common and widely-used techniques are ordinary and universal kriging - where an estimate is found for a point of a grid. Other more complex techniques, such as block or co-kriging, are more difficult to apply and hence are often only used in specific applications. In block kriging the estimates are established for areas, as opposed to points. In co-kriging, other variables which have been sampled at the same location, but often with a higher density of sample locations, are used to help calculate the estimate by incorporating the correlation between the variables into the calculations. As a means of describing the kriging technique in further detail, however, the remainder of the discussion will focus on the more commonly used techniques.

In kriging, the first order component of the data is removed: in the case of ordinary kriging this is the mean and in the case of universal kriging this is the trend in the data. The remaining variance, the second order component, is the local (random spatially dependent) variation in the data. The semi-variogram is then applied to the second order

component to explore the spatial dependence of the variance. The semi-variogram is found using equation 2.1.

$$\hat{\gamma}(h) = \frac{1}{2M(h)} \sum_{i=1}^{M(h)} \{z(x_i) - z(x_i + h)\}^2 \quad [\text{Equation 2.1}]$$

in which  $x_i$  and  $x_i + h$  are two points with values  $z$ ,  $h$  distance apart, and  $M(h)$  is the number of pairs  $h$  distance apart, referred to as the lag. In most cases, there are only a small number of observations available and therefore distances are often grouped into classes of equal intervals. The semi-variogram typically increases with distance (lag) until a point where it levels out, known as the sill; the distance to the sill is referred to as the range. At zero distance, the semi-variogram can be greater than zero, due to noise in the data (for example when two different observations have been recorded at the same location) and is known as the nugget effect. Where the semi-variogram is increasing with distance (i.e. where the distance is less than the range), a correlation or covariance exists between pairs of points, reflecting spatial dependency in the data. Beyond the range (i.e. when the semi-variogram levels out) covariance is zero and therefore there is no spatial dependence (Schaug *et al*, 1993).

A model, in the form of a smooth curve, is fitted to the semi-variogram by generalised least squares to obtain an overall description of the covariance structure. Generally, spherical, exponential or gaussian models are fitted. More information about fitting models to the semi-variogram can be found in McBratney and Webster (1986).

The model provides values for the intercept, the sill and the range. This information is then used in the kriging calculations. The kriging interpolation estimates are found by local weighted averaging, using equation 2.2, where the weights ( $\lambda_i$ ) have been determined by the variogram and the configuration of the data (Oliver and Webster, 1990). The estimates are calculated for a series of points, usually a grid across the study

area. In the kriging calculation the Lagrange multiplier is used to ensure that the variance errors are minimised by setting the sum of the weights to 1.

$$\hat{z}(B) = \sum_{i=1}^n \lambda_i z(x_i) \quad [\text{Equation 2.2}]$$

Spatial dependence, the variogram and kriging are covered in greater detail in Webster and Oliver (1990) and Bailey and Gatrell (1995). Further mathematical description of spatial prediction and kriging can be found in Cressie (1993).

Kriging has been widely used in the environmental sciences. For example, Gambolati and Volpi (1979) used kriging to create contour maps of groundwater in an attempt to help understand and control land subsidence in Vience. Three major aquifers underlying the Venetian lagoon were sampled on three separate occasions, two in 1973 and one in 1977; kriging was then applied to the sample observations. The sampling density for the three aquifers ranged from 27 to 40 observations.

Wood *et al* (1990) mapped the salinity of the soil in the Bet Shean Valley of Israel by applying disjunctive kriging to 201 measurements of electrical conductivity. Measurements were undertaken in two separate surveys in two consecutive years. The maps were subsequently used to observe spatial variations over time, and to identify areas where critical thresholds were exceeded or where deficits occurred.

Another application of kriging is demonstrated by Vauclin *et al* (1983) who used kriging and co-kriging to study the variability in sand, silt, clay, available water content (AWC) and water stored at the 1/3 bar (pF2.5) and the relationships between them. The study was undertaken in Tunisia and 40 samples were taken on a 10m grid. The kriging techniques were used to calculate estimates of the sampled data at a finer grid. It is interesting to note that co-kriging was found to be the better interpolator.

*Knotters et al* (1995) examined the application of auxillary variables for estimating soil horizon depths. The authors compared kriging with co-kriging and kriging combined with regression. Soft layer depth and soil electrical conductivity were sampled at 117 random locations and 539 locations in a regular grid in Lickebaert-polder in the Netherlands. The 539 samples were used for interpolation (kriging and co-kriging) and the 117 samples for analysing the accuracy of the interpolation. The 117 samples were also used for the regression analysis; the regression model was used to predict the depth of soft layers from soil electrical conductivity at the 539 regular grid locations. Kriging was then applied to the prediction error. The use of auxillary variables was found to improve the kriging estimates and kriging combined with regression provided better estimates than co-kriging.

*Cassiani and Medina* (1997) used co-kriging to estimate aquifer transmissivity of a land fill site contaminated by leachate from medical and chemical wastes in Duke Forest, Durham, North Carolina, USA. Transmissivity measurements were taken at 27 wells and auxillary measurements of vertical electrical surrounding were taken at 10 of the 27 wells and also at 111 other sites. The auxiliary data were found to improve the known estimations of transmissivity at the wells.

Recently, kriging has been applied in studies of air quality, where area-wide levels of air pollution have been estimated from values recorded at monitoring sites. For example, *Atkins and Lee* (1995) used kriging to look at spatial and temporal variations in rural levels of NO<sub>2</sub> for the UK. NO<sub>2</sub> was recorded at 57 rural sites during two surveys, one in April 1987 and the other in June 1988.

*Liu et al* (1995) examined the use of kriging and co-kriging to estimate levels of ozone concentration in Toronto, USA. Ozone was measured at a total of 40 outdoor residential locations between June and August 1992, though measurements were only taken at 4 of the locations at any one time. Within the study area of Metropolitan Toronto there are 19 ambient monitoring sites measuring ozone. For selected days, between June and August 1992, kriging was applied to the 19 monitoring sites. On those selected days, the

measurements at the 4 residential locations where ozone was also measured were compared with the interpolated values. Kriging was found to provide better estimates of measured ozone at the residential locations than selecting the nearest monitoring site. Co-kriging, taking into account traffic conditions, provided slightly better estimates.

The differences between modelled values of sulphate and nitrate in precipitation and sulphate in aerosols and measurements at EMEP sites have been mapped using kriging, as reported in Schaug *et al* (1993). The modelled values were calculated using the Lagrangian model - which calculates values for sites in the EMEP network and points on a regular grid. The maps were used to identify areas of systematic under- or over-estimation in modelled values across Europe. Points on the regular grid could then be adjusted to compensate for these systematic differences.

## 2.4 Adapted Statistical Methods

An alternative to the traditional methods of mapping air pollution is regression mapping. If two variables are related, i.e. they are significantly correlated, then known values of one of the variables can be used to predict unknown values of the other variable. Where more than one known variable is used to predict the unknown variable, then the statistical technique is termed multiple regression and the regression equation thus takes the form of Equation 7.1. The known variables are termed the independent variables and the variable to be predicted is termed the dependent variable. The coefficient of determination ( $r^2$ ) can be used as an estimate of the amount of variation in the dependent variable that is statistically explained by the independent variable.

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad [\text{Equation 7.1}]$$

The regression equation describes the relationship between the dependent variable and the independent variables. In the case of regression mapping, the dependent and independent

variables have a spatial element related to the location of sample sites. In the environmental sciences, the sample sites are usually locations on the surface of the earth. If the independent variables have been sampled to a higher density than the dependent variable then the regression equation can be applied to predict the values of the dependent variable at the locations where it has not been sampled. This is very useful in situations where the independent variables are a lot easier to measure than the dependent variable.

It is, however, essential that an adequate number of sample sites are established to provide statistical rigour and that meaningful variables are chosen to optimise the predictive power of the relationship. One of the main restrictions of regression mapping is that the spatial locations of the sample sites are not taken into consideration. This has important implications for the residuals of the regression analysis. Spatial autocorrelation - where residuals at sample locations closer together are more likely to resemble each other than residuals at sample locations further apart - may be present in the data. Spatial autocorrelation is an indication of spatial dependence in the data and its presence, therefore, implies that the residuals are not random and that explanatory variation in the dependent variable remains. The problem is discussed further in Chapter 7 and detailed in Haining (1990) and Bailey and Gatrell (1995).

The application of regression as a predictive tool has been used extensively in the environmental sciences. Tyler *et al* (1996), for example, applied multiple regression to predict General Yield Class (GYC) of Douglas fir, Japanese larch and Scots pine in Scotland. Data on GYC, soil, climate and topographic factors were collected at sample sites and the regression model used to estimate mean productivity at regional and national levels. Regression has also been applied to predict the presence of the red squirrel in fragmented habitats in the Netherlands based upon the area per wood-lot covered with conifers (van Apeldoorn *et al*, 1994).

Regression has also been widely used in the field of remote sensing, not just for prediction, but also to help classify and interpolate images. In remote sensing, samples on the ground

are compared with training sites on the image (i.e. areas of pixels corresponding to areas on the ground) and regression is applied to establish the relationship between the two. The linear relationships are then used to classify other pixels in the image.

Examples of the application of regression in remote sensing are numerous. Pratt *et al* (1997) used remotely sensed data to estimate the areas of different crops in North-east Nigeria. Regression was applied to calibrate the classification based on sample field data. George (1997) estimated the concentration of chlorophyll in lakes in the Lake District, UK, from the relationship between maximum reflectance wavelength band of imagery and concentration of chlorophyll at sample locations across the lake. Rasmasses (1997) established a linear relationship between operational millet yield in the field and remotely sensed data to help forecast millet yield in Senegal from AVHRR data. Guo and O'Leary (1997) derived estimates of suspended solids in Waitemata Harbour in Auckland from the relationship between samples of suspended solids and reflectance on scanned aerial photographs. Trotter *et al* (1997) applied regression to help estimate timber volumes from remotely sensed imagery and forest inventory data to support the design of harvesting strategies in New Zealand.

There are, however, very few examples of regression applied to air pollution modelling. Campbell *et al* (1994), applied a statistical model to map air quality in the UK. Levels of NO<sub>2</sub> for the UK were estimated for 5 x 5 km grid squares. NO<sub>2</sub> was measured at 363 urban locations during a six month period in 1991. Regression analysis was applied to the six month mean concentrations of NO<sub>2</sub> and population density for all 5 x 5 km grid squares containing a monitoring site. Population density was used as an indication of vehicle density - it was assumed that population density was highly correlated with vehicle density. The rural background was removed from the NO<sub>2</sub> values before the regression equation was applied. The rural background was found by applying kriging to 39 rural sites. When the corrected NO<sub>2</sub> values were regressed against population, population was found to explain 56% of the variation in NO<sub>2</sub>. The equation was then used to produce a

concentration map for the UK from the spatial distribution of population and the interpolated rural background.

This technique has since been used by AEA Technology to map other pollutants including SO<sub>2</sub>, CO and non-methane VOC's (Stedman, 1996).

Given the size and complexity of traffic-related pollution surfaces in urban environments, however, it is unlikely that the techniques described above - dispersion modelling, spatial interpolation and regression - will provide reliable estimates of pollution in unsampled locations at this scale. Recent developments in information systems and spatial analysis provide new opportunities to develop approaches to modelling air pollution. Geographical Information Systems (GIS), for example, have the capacity to analyse large volumes of spatially referenced data, in a timely fashion, and produce high quality maps of the resulting pollution surfaces.

All the techniques discussed above could benefit from the use of GIS. Indeed, point source dispersion models, spatial interpolation techniques and regression for remote sensing applications have already been implemented into the more sophisticated GIS, for example in the GIS ARC/INFO. For general regression applications, GIS have also been used in combination with statistical packages.

Mattson and Godfrey (1994), for example, used a combination of regression and GIS techniques to examine road salt contamination in streams in Massachusetts, United States. Samples of sodium were collected from 162 streams - the watersheds of which fell within the boundaries of Massachusetts. The GIS was used to calculate the total length of four classes of road in each watershed. The sodium values were regressed against the four classes of road and values for sea spray precipitation. All the variables were found to be significant. The regression equation was then used to predict samples in other streams that were not used in the model.

Iverson *et al* (1997) also used a combination of GIS and regression to predict forest productivity and species composition for forests in Ohio, USA. GIS was used with forest-plot data to develop an integrated moisture index (IMI) based upon several landscape features. Regression was then used to estimate the relationship between IMI and oak site index and the percentage of composition of oak and yellow poplar. The statistical relationships were then applied in the GIS to create maps of site index and composition.

## 2.5 Geographical Information Systems

GIS have the capacity to store large volumes of spatially referenced data, at a fine resolution, and also have the capability to analyse the data in a spatial manner. This might simply involve overlaying two different data sets (e.g. a point-in-polygon operation) or more complex analyses of the data on the basis of data in neighbouring locations (e.g. neighbourhood operations). GIS are also very flexible systems - many have their own languages to help customise existing functions and operations, and to help develop new ones - for example, ARC/INFO's Advanced Macro Language (AML). The results of any analysis or operation can then be presented visually as a map, either on screen or as a hard copy.

An important role of GIS is the ability to bring together data from a variety of sources, such as health, social-economic data and environmental data, within a common framework. Once a map of air pollution has been created in the GIS it can be integrated with other information, such as the locations of sample populations and their associated health outcomes. GIS can usefully be applied to help produce more realistic maps of the spatial distribution of air quality and hence, more reliable estimates of exposure.

The data can be stored and analysed in either vector or raster (grid) format. The vector format stores information as x and y co-ordinates, whereas grid format stores information as an array of cells - where each cell contains some information. In the case of a road

network, for example, vector format describes the roads as a series of co-ordinates, whereas in grid, one numerical value indicates that a road exists at a location and another numerical value indicates that it does not. When the resolution of the grid gets smaller (i.e. to provide greater detail) then the grid structure begins to take up far more space than the vector structure. In the past vector format was often chosen in preference to grid due to its ability to store data at a finer resolution and with less space. Today, however, advances in computer technology (i.e. larger discs and faster processors) have, to some extent, reduced this problem and although grid format at a fine resolution involves large amounts of data, it does have its advantages. In particular, it is extremely useful for modelling environmental phenomena, such as air quality, which do not vary according to administrative and political boundaries, or some other aggregation.

The grid functionality is a powerful tool for modelling spatial variations. Once the data are stored in grid format, then techniques such as moving windows or kernel estimators can easily be applied to analyse the data. Here, at any location, the data are analysed on the basis of the neighbouring cells, perhaps calculating the local average or the local sum. Neighbourhood techniques are often used to smooth data and remove random variation. Although neighbourhood techniques are not restricted to data stored as a regular grid and can also be applied to point data or irregular area data, the tools are not readily available, due to the complex nature of regions (i.e. variations in shape and size) and the irregular distribution of points. The techniques and their applications are discussed in greater detail in Haining (1990) and Bailey and Gatrell (1995).

## **2.6 Producing Detailed Maps of Air Pollution**

There are many different ways in which GIS could be applied to produce detailed maps of air pollution. In order to examine the capacity of GIS to model air pollution and produce reliable predictions at unsampled locations three new approaches have been developed by the author. The research presented in this thesis demonstrates the development, testing

and comparison of the methods. All three methods have been developed in a GIS framework and have been applied and tested in Huddersfield, UK, using data that is described in Chapter 3.

The first method presented is an automatic approach to modelling traffic related air pollution. The approach applies a combination of the two traditional pollution mapping techniques: dispersion modelling and spatial interpolation. A line dispersion model was adapted to work automatically within a GIS to model near-source pollution from road traffic. The near-source pollution was combined with background concentrations, which were generated by kriging pollution data measured at monitoring sites in background locations. While many of the spatial interpolation techniques, including kriging, have now been incorporated within the more sophisticated GIS, there has been no real attempt to link line dispersion models with GIS.

The second method is also a combination of two different techniques: spatial analysis and spatial interpolation. Within GIS, a moving window is used to estimate near-source pollution concentrations from road traffic. In a similar manner to the first method, the background pollution is generated by kriging and the near-source concentrations are added to the background concentrations. As previously discussed, the moving window technique has primarily been employed to smooth data; it has not, however, been applied to air pollution mapping.

The third method is a regression approach that employs spatial analysis and statistical techniques. Variables that provide measures of the emission sources and dispersion patterns of pollution are identified and defined within the GIS. The regression analysis is applied outside the GIS. The relationship between these variables and the monitored data is then used to estimate concentrations within the GIS at unsampled locations. Although regression has been used for mapping purposes, the technique has rarely been applied to air pollution mapping, and in particular, when used in combination with GIS techniques for analysing spatial data.

Since kriging is the most sophisticated and widely-used of the traditional mapping techniques, kriging was chosen as a control against which to compare the performance of the new methods.

The four approaches - kriging, automatic modelling, moving window and regression mapping - are discussed in detail in Chapters 4 to 7 and are reviewed and compared in Chapter 8.

## **CHAPTER 3 DATA COLLECTION**

The data used to undertake this research were collected as part of the Small Area Variations in Air Quality and Health (SAVIAH) project.

### **3.1 The SAVIAH Project**

The SAVIAH project was an EU-funded methodological study aimed to apply, test and evaluate new and emerging methodologies in the fields of epidemiology, geography, air pollution modelling and small area health statistics, as a basis for informing environmental health policy (Elliott *et al*, 1995). The SAVIAH project was undertaken in four study areas: Huddersfield (UK), Amsterdam (The Netherlands), Prague (Czech Republic) and Poznan (Poland). The project was designed to look at the long-term, chronic health effects of pollution. Further information pertaining to the project can be found in Briggs *et al* (1997), Lebret *et al* (1997) and Pikhart *et al* (1997).

Air quality monitoring was undertaken in each area, using nitrogen dioxide as the measured pollutant in Huddersfield, Amsterdam and Prague, and sulphur dioxide in Poznan. Although the profile of air pollution in Western and Central Europe has changed from industrial to traffic-related pollution over the last few decades, as discussed in Chapter 1, in certain areas of Eastern Europe, such as Poznan in Poland, industrial pollution is still considered to have greater long-term chronic health effects than traffic-related pollution. Consequently, in Poznan sulphur dioxide was selected as a proxy for industrial pollution. In Huddersfield, Amsterdam and Prague, nitrogen dioxide was selected as a proxy for the complex of traffic related pollutants. Both pollutants were

chosen for their ease of measurement and the fact that they are the predominant pollutants measured in similar studies, which allows comparisons to be made.

In each study area all children in a selected age band between 7 and 11 were surveyed by questionnaire. The questionnaires were sent to the parents / guardians of the children via the schools in Huddersfield, Prague and Poznan, and by post in Amsterdam. The number of returned questionnaires and response rates for each study area are shown in Table 3.1. The questionnaire included questions relating to health, with particular reference to respiratory disorders, and socio-economic conditions. The questionnaire is discussed in further detail later in this Chapter.

*Table 3.1 Number of returned questionnaires and response rates.*

	Prague	Amsterdam	Poznan	Huddersfield
Age band	7-10	7-11	7-8	7-9
Number of eligible returns	3680	3811	4633	5027
Response rate	88.1%	62.6%	95.6%	90.7%

One component of the project was the collection of geographical data for each study area. In Huddersfield, the data were collected and stored in the GIS ARC/INFO version 6.0.2, running on a UNIX platform. All the data sets were spatially referenced to the National Grid co-ordinate system within the GIS. Geographical data included road network, land cover, altitude, monitoring sites and the children's place of residence. All the data were stored as either point, line, polygon or grid coverages with associated attributes (Table 3.2) and are described in more detail later in the Chapter.

The SAVIAH study was a large project and many people were involved in the air quality monitoring, the health survey and the collection and capture of geographical data in the GIS. In Huddersfield, this included members of staff at the Institute of Environmental and

Policy Analysis, and other researchers at the University of Huddersfield. Part of the purpose of the SAVIAH project was to develop, test and apply new methodologies for air pollution modelling. The author was primarily responsible for this aspect of the project, which included the development of the GIS (i.e. the bringing together of the different data sets within a consistent geographical framework).

*Table 3.2 Geographical data and associated attributes.*

Geographical data	Coverage	Attributes
Road network	Line	Road type, traffic volume
Land cover	Polygon	Land cover class
Altitude	Grid	Height above sea level
Monitoring sites	Point	Measured NO <sub>2</sub> values
Place of residence	Point	Responses to the health questionnaire

### **3.2 Measuring Air Pollution**

Air quality monitoring in the UK is the responsibility of a number of different agencies and is undertaken through a range of different networks, using different devices. In the UK there is a network of 47 automatic stations which measure O<sub>3</sub>, NO<sub>2</sub>, SO<sub>2</sub>, CO, PM<sub>10</sub> and speciated HC; NO<sub>2</sub> is measured at 26 of these sites (Bower and Vallance-Plews, 1995). Automatic monitoring stations consist of pump samplers and chemiluminescent analysers. These enable short-term measurements of exposure to be undertaken, and allow specific pollution events to be identified and analysed. Unfortunately, they are very expensive to purchase and operate, and consequently there are very few sites - although more sites are planned for the future. The stations therefore provide a great deal of temporal information but fail to reflect spatial variations.

Agencies, such as local authorities, also undertake fixed site sampling, however, the data are not held centrally, nor in the public domain, and therefore it can be very difficult to obtain the data. Even if the data were readily available, issues of comparability between

data readings from different automatic monitors are raised. The samples and analysers at the different monitoring stations and sites are not calibrated against a standard or against each other. This type of measurement error is very difficult to quantify.

An inexpensive alternative to the automatic station and fixed site sampler is the diffusion tube. Measurements using diffusion tubes allow for a higher density of sample locations, thus providing a more detailed spatial resolution of the concentration surface. Through government funded national monitoring programmes there are now 1200 sample sites using diffusion tubes for NO<sub>2</sub> as part the UK monitoring networks (Bower and Vallance-Plews, 1995). In addition, many local authorities are using diffusion tubes for local monitoring surveys.

The main disadvantage to using diffusion tubes is that short-term measurements - for example, hourly measurements - cannot be recorded and hence the acute effect of short term pollution events cannot be determined. Diffusion tubes do, nevertheless, allow the long term, chronic, effects of pollution to be investigated.

A diffusion tube is a passive sampler which measures the time weighted average concentration of airborne oxides in ambient air (such as NO<sub>2</sub> and SO<sub>2</sub>). The most commonly used sampler is the Palmes Tube, designed to measure NO<sub>2</sub> (Palmes *et al*, 1976), which is a passive diffusion device consisting of an acrylic tube open at one end with coated stainless steel screens at the closed end. The airborne oxides diffuse through the air in the tube and are trapped as nitrite ions in the triethanolamine which coats the screens (Miller, 1988). The tubes are exposed for specific periods, usually 1 to 4 weeks, and measure the mean concentration over that period (Her Majesty's Inspectorate of Pollution, 1993). The diffusion tubes are not really sensitive for periods shorter than one week. Long term monitoring can be undertaken by successively replacing the tube at the end of the period of measurement. The uptake of airborne oxides may vary with changes in pollutant concentration, exposure time, atmospheric temperature, humidity and turbulence.

Palmes tubes have been tested by the U.S. National Bureau of Standards and Warren Spring Laboratory in the UK where accuracy was determined to be better than +or- 10% and precision was better than 4 ug/m<sup>3</sup> for a 1-week sampling period (Boleij *et al*, 1986). Recent studies, however, have suggested that Palmes diffusion tubes overestimate NO<sub>2</sub> concentrations by as much as 28% (Heal and Cape, 1997) and a 1.26 mean ratio of concentration measured by diffusion tube over that by chemiluminescent monitor has been reported (Campbell *et al*, 1994). One possible explanation for the overestimation is turbulence generated in the mouth of the tube, which Atkins and Lee (1995) suggest could be avoided by careful selection of the sampling sites. The accuracy and precision of two different types of passive samplers - a tube-type sampler and a badge-type sampler - were examined as part of the SAVIAH study.

Despite uncertainty about the accuracy of the concentrations, diffusion tubes have been widely used to measure ambient air pollution both indoors (Boleij *et al*, 1986 and Fisher *et al*, 1989) and outdoors (Atkins *et al*, 1986; Atkins and Lee, 1995; Bower *et al*, 1991; Campbell *et al*, 1994; Heal and Cape, 1997; Hewitt 1991 and Laxen and Noordally, 1987). The continued use of diffusion tubes is primarily due to their low-cost and mobility, which makes it possible to undertake relatively detailed long-term surveys and enables surveys to be designed to meet a particular purpose. Even with the inherent uncertainty in measurements provided by diffusion tube sampling, the spatial resolution achieved is very valuable in identifying areas exposed to large concentrations (Photochemical Oxidants Review Group, 1993)

### **3.3 Survey Design**

In the SAVIAH project NO<sub>2</sub> and SO<sub>2</sub> were measured as two-weekly averages during four surveys throughout a twelve month period: NO<sub>2</sub> in Amsterdam, Prague and Huddersfield, and SO<sub>2</sub> in Poznan. In the different study areas the surveys were undertaken at slightly different times - dependent upon local logistics - but within the following time framework:

Survey 1 - June-July 1993

Survey 2 - October-November 1993

Survey 3 - February-March 1994

Survey 4 - May-June 1994

The first survey acted as a pilot for the project and was used to compare the performance of two different types of passive samplers and select the most suitable device to be used in the remainder of the surveys. The two passive samplers compared were the Willems badge (Willems and Hofschreuder, 1991), a badge-type sampler, and the Palmes tube.

In Amsterdam, Prague and Huddersfield the Willems-badge was placed at 60 sites and the Palmes-tube at 20 sites. Duplicate badges and tubes were placed at each site. Precision of the duplicates was established by calculating the combined coefficient of variation (Table 3.3) - this was found to be 11.0% for the Willems-badge and 7.7% for the Palmes-tube (Reeuwijk *et al*, 1997). Precision of the tubes is within the 5 to 8% range reported by Atkins *et al* (1986).

*Table 3.3 Coefficient of variation for the Willems badge and Palmes tube*

	Amsterdam	Prague	Huddersfield
Willems badge	7.1%	13.3%	12.4%
Palmes tube	7.3%	8.2%	7.6%

Accuracy of the badges and tubes was determined using reference methods. The total diffusional resistance ( $R_t$ ) of the two types of samplers is calculated using NO<sub>2</sub> values recorded at chemiluminescent monitors. The accuracy of the tubes and badges is then expressed as the variability of the calculated  $R_t$  of the two designs at the reference sites. In Amsterdam, Prague and Huddersfield duplicates of both types of samplers were located at 3, 4 and 2 real-time chemiluminescent monitors respectively. The mean  $R_t$  values and

the coefficient of variation for the badges and tubes at the reference sites are presented in Table 3.4.

*Table 3.4 Mean diffusional resistance ( $R_t$ ) and coefficient of variation of  $R_t$  for Willems badges and Palmes tubes.*

		Amsterdam	Prague	Huddersfield
Willems badge	Mean $R_t$	913	1564	957
	CV $R_t$	6%	40%	20%
Palmes tube	Mean $R_t$	4595	6878	5165
	CV $R_t$	12%	25%	20%

As can be seen from the table, the mean  $R_t$  values are quite different for badges and tubes - due to variations in the design of the samplers. Further details can be found in Reeuwijk *et al* (1997). Briefly, the badges have an entrance filter which introduces diffusional resistance, resulting in a non-fixed sampling rate. For each survey period, therefore, the sampling rate needs to be determined empirically by the reference method. The  $R_t$  value calculated at the monitoring stations is then used to calculate badge concentrations at other sites. For the comparison of concentrations measured by samplers with that by chemiluminescent monitors, the mean  $R_t$  value for Amsterdam was chosen as the reference value for the calculation of badge concentrations due to the relatively small CV of  $R_t$ . The mean ratio of the concentration measured by the badges compared to the reference methods was found to be 0.95 and 0.58 for Huddersfield and Prague respectively. For the tubes, the mean ratio of concentration measured by the tubes compared to chemiluminescent monitors was 1.16, 1.03 and 0.77 for Amsterdam, Huddersfield and Prague respectively (Reeuwijk *et al*, 1997).

The Palmes tubes were ultimately chosen in preference to the Willems-badge due to their greater precision, durability and robustness (i.e. fewer values were lost due to damaged entrance filters).

In Poznan two-week averages of SO<sub>2</sub> concentrations were measured at 16 sites with the Willems-badge and at 69 sites with the Palmes-tube. Again duplicates were placed at each site and precision for the Willems-badge was found to be 7% and 26% for the Palmes-tube. The R<sub>t</sub> values for the two samplers was established using reference measurements at six fixed-site manual monitoring stations. Variation in the R<sub>t</sub> values at the six sites was found to be 60% for the Palmes-tube and 15% for the Willems-badge (Reeuwijk *et al*, 1995). The Willems-badge was therefore adopted for monitoring SO<sub>2</sub>.

In Amsterdam, Prague and Huddersfield, for each individual survey, NO<sub>2</sub> was measured at 80 permanent monitoring sites, i.e. sites for which the location remained the same for each survey.

Throughout the study period, approximately 10% of the readings were lost through stealing, vandalism, damage or local factors (e.g. temporary road works) rendering the reading invalid. To compensate for the loss of data, mean concentrations were established for the permanent sites using multi-level modelling techniques at RIVM in the Netherlands (Lebret *et al*, 1995). Multi-level modelling was applied using the mixed model - PROC MIXED - in the statistical package SAS, to model the variation between sites and between surveys. The model takes into account repeated measurements at the same locations and dependency between measurements. Within the model, new, adjusted concentrations are calculated for each site using maximum likelihood estimation. Thus, concentrations can be established for sites with missing data. The model is based upon equation 3.4.

$$X_{ijk} = u + a_i + b_j + e_{ijk} \quad [\text{Equation 3.4}]$$

where X<sub>ijk</sub> = adjusted mean concentration at site i, survey j, duplicate k

u = overall mean

a<sub>i</sub> = site-effect, i = 1 to 80

b<sub>j</sub> = survey-effect, j = 1 to 4

e<sub>ijk</sub> = residual error, k = 1,2 (pairs)

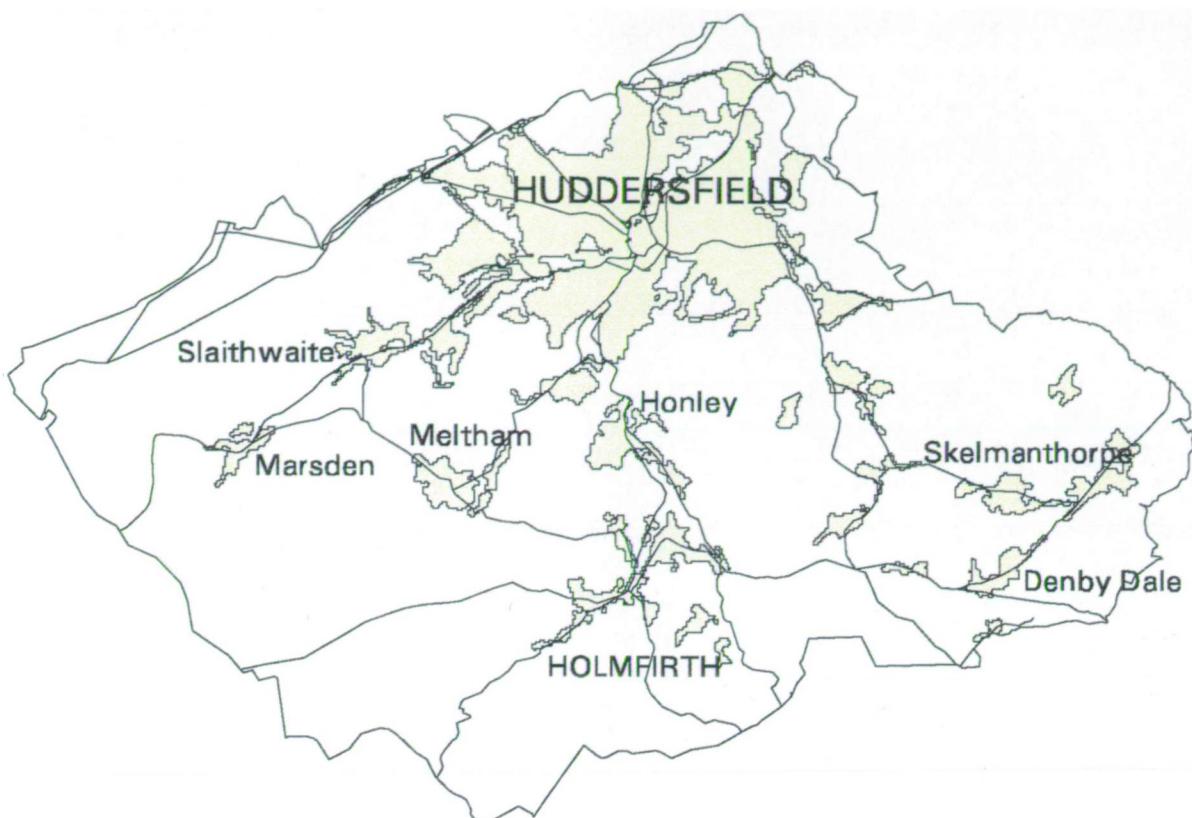
In Poznan, SO<sub>2</sub> was measured at 68 permanent sites for each individual survey. However, there is a substantial difference between SO<sub>2</sub> concentrations in summer and winter - with winter concentrations been far greater than summer concentrations due to the burning of fossil fuels for domestic heating. In Poznan therefore only the 2nd and 3rd surveys, which were undertaken in the winter months (November 1993 and February 1994 respectively), were taken into consideration and the mean of these two surveys was calculated for all 68 sites.

### **3.4 The Huddersfield Study**

The Huddersfield study area (Figure 3.1) represents the Huddersfield District Health Authority as it was at the time of the SAVIAH study. The boundaries of the Health Authority were chosen to facilitate the collection of health data. Huddersfield Health Authority covers an area of 304 Km<sup>2</sup> and has a population of approximately 160,000. The area is characterised by a mixture of dense urban land and rural farmland and moorland. The altitude ranges from 33m in the east to 580m above sea level in the west.

#### **3.4.1 Monitored Data**

In Huddersfield, the SAVIAH surveys were undertaken in June 1993, October 1993, March 1994 and June 1994. For each survey the tubes were placed at 80 permanent sites and 40 variable sites. Variable sites were relocated for each survey in order to examine specific patterns and sources of variation. In addition, there were 8 consecutive monitoring sites, which were exposed for the full duration of the study period, i.e. from June 1993 to June 1994. Tubes at consecutive sites were exposed on a monthly basis, apart from when the individual surveys were undertaken, when they were exposed for the same period as the survey tubes (i.e. two weeks). The consecutive sites were used purely for validation of the air pollution modelling. The permanent sites were used to develop



## LEGEND

**Urban areas**

**Main roads**



5 km

*Figure 3.1 The Huddersfield study area.*

the models and the variable sites were used for further validation and to provide additional information. Two tubes were exposed at all sites to provide a measure of the at-site variation and also to provide insurance against loss or damage of tubes.

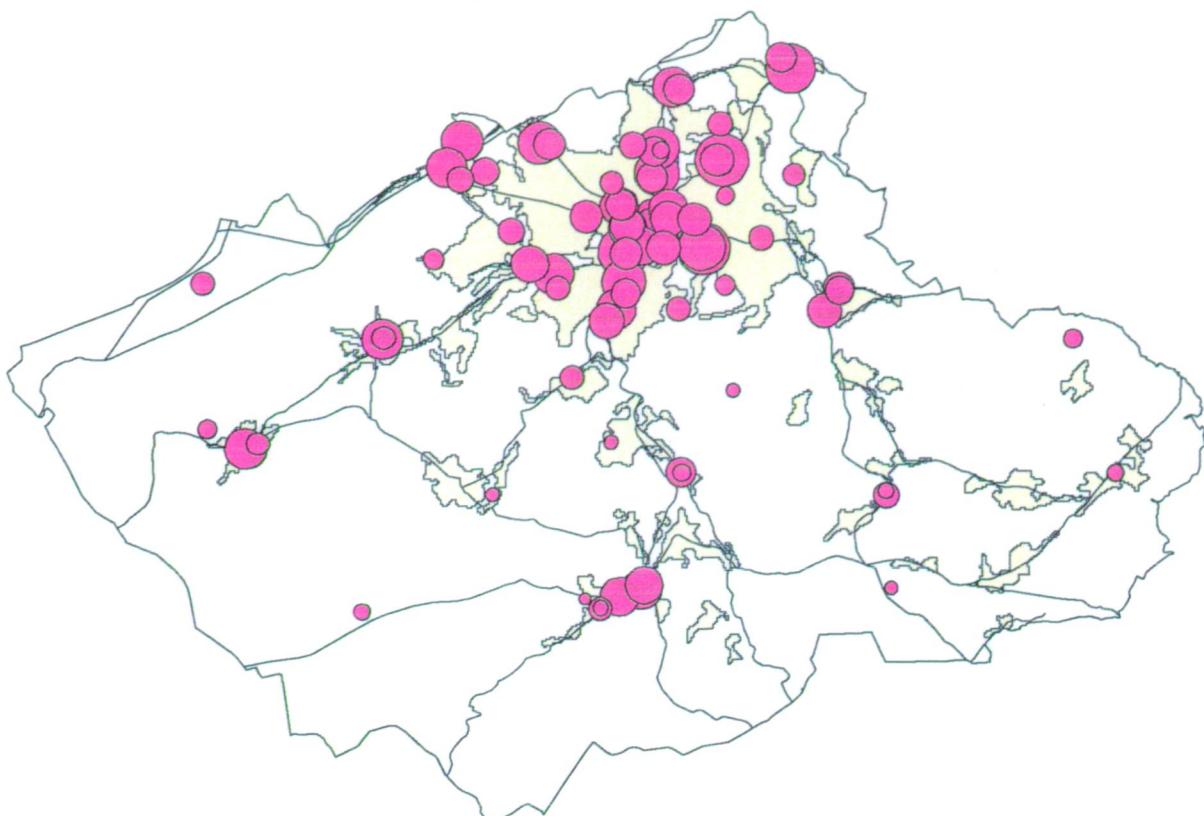
The geographical locations of the monitoring sites were chosen to reflect differences in rural and urban areas, and differences in NO<sub>x</sub> emissions from road vehicles and industrial sources. Additional sites were chosen in background locations, away from the sources of pollution, to a) act as a control and b) help establish the background level of NO<sub>2</sub>. Locations therefore fell into 6 categories:

- urban areas - near traffic sources
- urban areas - near industrial sources
- urban background
- rural areas - near traffic source
- rural areas - near industrial source
- rural background

Further consideration was also given to the spatial distribution of the sites, in so far as:

- where possible the sites should be evenly spread across the study area and avoid excessive distance between neighbouring points
- the sites should be denser in the more densely populated areas
- at most roadside locations one site should be placed at the kerbside and another approximately 50-100m back from the kerbside

The adjusted mean concentrations (section 3.3) for the permanent sites in Huddersfield can be seen in Figure 3.2.



## LEGEND

□ Urban areas

■ Main roads

ug/m<sup>3</sup>

• 17.8

● 57.6



5 km

**Figure 3.2** Annual mean NO<sub>2</sub> concentrations at the 80 permanent monitoring sites.

### **3.4.2 Road Network**

The road network was digitised from Ordnance Survey 1:10,000 topographic map sheets. The road network was classified according to road type:

- Motorway
- Trunk roads
- Main roads
- Secondary roads
- Road > 4m width
- Road < 4m width
- Small residential road, drive or track

Traffic volumes were attributed to individual road segments. For the motorways the traffic volume data were obtained from the Department of Transport, Leeds, based upon automatic traffic counts. Where possible, traffic flow for the remaining roads was obtained from automatic traffic counts from West Yorkshire Highways and Technical Joint Committee, Leeds, and manual traffic counts from Kirklees Highways Services, Huddersfield. For the major roads, where information was not available, the traffic volumes were interpolated from the known points. The minor roads were assigned traffic volumes by the author based upon local knowledge of the study area as follows: road > 4m width 200 vehicles per day; road < 4m width 100 vehicles per day and small residential road, drive or track 50 vehicles per day. Traffic flow for all road types were expressed as 24 hour counts in the first instance. The road network and average daily traffic volume by type of road can be seen in Figure 3.3.

Data presented in the Kirklees Traffic Monitoring Report (1992) were used to establish the day-time hourly traffic volumes on all roads. The report indicated that day-time traffic flows lasted for 16 hours, starting at 6 a.m. in the morning and finishing at 10 p.m. in the evening. The report presented a profile of daily (24 hour) traffic volume, on an hourly



## LEGEND

Road type	Vehicles/hour
Motorways	4594
Trunk road	1576
Main road	694
Secondary road	267
Road > 4m wide	11
Road < 4m wide	5
Other road	2



5 km

Figure 3.3 The road network with average daily traffic volume.

basis. The total day-time and night-time traffic volumes were calculated from this profile. The proportion of day-time to night-time traffic volume was found to be 93.5% to 6.5% respectively. For each road segment the average day-time hourly traffic volume was therefore calculated using equation 3.5:

$$dhtv_i = (tv_i * p) / h \quad [\text{Equation 3.5}]$$

where  $dhtv_i$  is the average day-time hourly traffic volume for road segment i

$tv_i$  is the 24h traffic volume for road segment i

p is the proportion of day-time traffic (0.935)

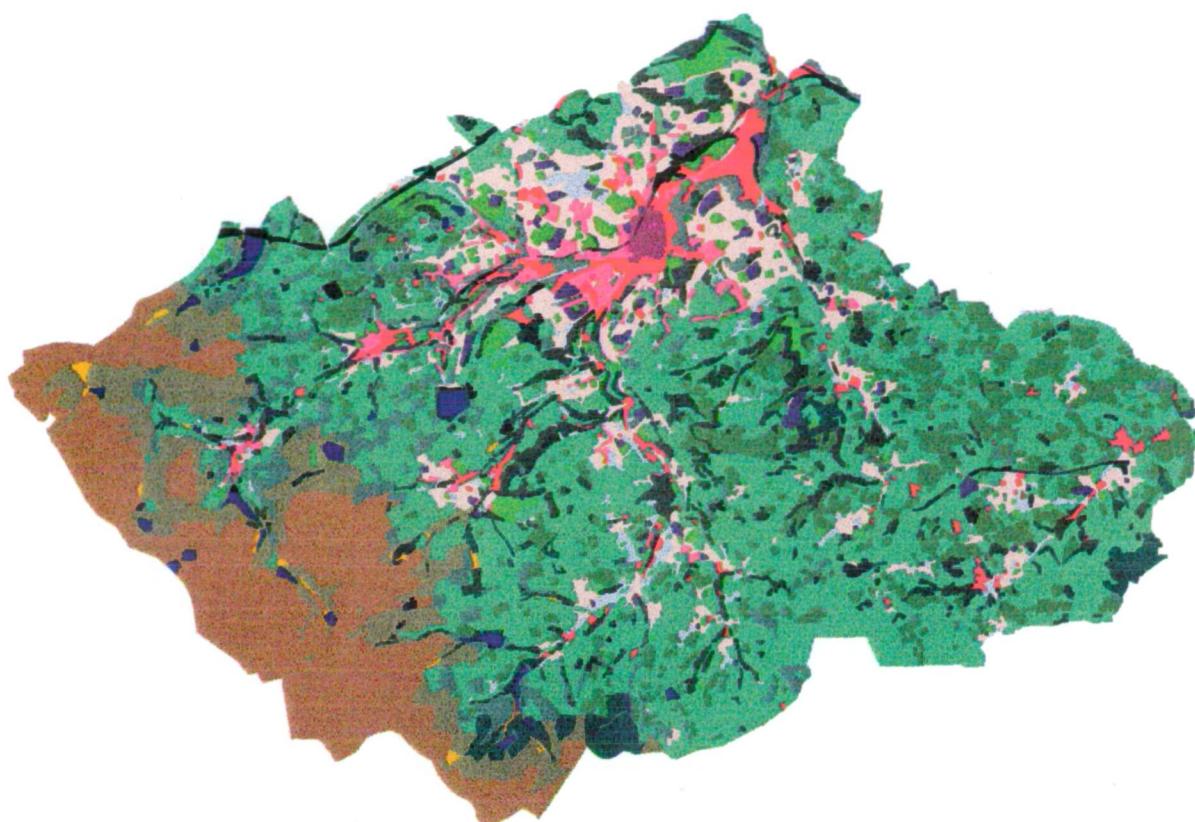
h is the number of day-time hours (16)

### 3.4.3 Land Cover

Information about the land cover in the study area was obtained from aerial photographs (at a scale of 1:10,000). The different land cover types identified on the aerial photographs were classified into categories according to the land cover classification scheme shown in Table 3.5 (Smallbone, 1998). Sheets of transparent film were overlaid on top of Ordnance Survey 1:10,000 maps - which were used as a reference. Parcels of land were identified on the aerial photographs, classified and the boundaries drawn on the transparent film. The parcels of land were then digitised (Figure 3.4).

### 3.4.4 Altitude

The altitude data was held in the GIS as a Digital Terrain Model (DTM). The DTM was created from height data supplied by the Institute of Hydrology. The data were supplied in the Hydrological DTM Data Transfer Format in an ASCII text file. The heights were



## LEGEND

[Light Blue Box]	<b>Very low density housing</b>	[Dark Blue Box]	<b>Reservoirs</b>
[Light Orange Box]	<b>Low density housing</b>	[Dark Green Box]	<b>Broad leaf &amp; mixed trees</b>
[Pink Box]	<b>High &amp; very high density housing</b>	[Medium Green Box]	<b>Coniferous trees</b>
[Dark Red Box]	<b>High density commercial</b>	[Teal Box]	<b>Rough grass</b>
[Dark Purple Box]	<b>Public Institutions</b>	[Dark Teal Box]	<b>Pasture</b>
[Red Box]	<b>Industry</b>	[Dark Green Box]	<b>Arable land</b>
[Dark Green Box]	<b>Disused and sequestered land</b>	[Brown Box]	<b>Peat sedge</b>
[Dark Green Box]	<b>Recreational and urban green space</b>	[Grey Box]	<b>Moor grass &amp; braken</b>
[Dark Green Box]	<b>Quarries</b>		



5 km

*Figure 3.4 Land cover.*

*Table 3.5 Land cover classification scheme.*

Category	Description
industry	any industry
public institutions	hospitals, schools, colleges
high density commercial	office blocks, shopping centres
very high density housing	tower blocks
high density housing	> 60% area housing
low density housing	25-60% area housing
very low density housing	<25% area housing
urban green space	allotments, cemeteries
recreation	sports grounds, parks
disused and sequestered	wasteland, derelict, sequestered
quarries	used and disused quarries
broad leaf trees	approx 80% broadleaf
coniferous trees	approx 80% coniferous
mixed trees	30-70% broadleaf and coniferous
agriculture	crop land at any time of survey
pasture	permanent pasture
rough grass	unimproved pasture
bracken	75% bracken
moor grasses	moor grasses, molinia, juncus
peat sedge	peat hags with thin vegetation/bare peat

for 1km square blocks with a header for each block that included the Easting and Northing of the south-west corner of the block. The heights were recorded at 50m intervals and therefore each block contained 20 rows x 20 columns of data items. One row of data is stored on two lines of the text file. A FORTRAN program was written by the author (see Appendix 1) to transform the data from the Hydrological DTM Data Transfer Format into IDRISI (IDRISI is a PC based GIS). The data were then transferred to ARC/INFO. The program was written in the early stages of the research when FORTRAN for DOS was the only version of FORTRAN available. Under these circumstances it was more efficient to transform the data into IDRISI and then convert the data in IDRISI to a format compatible with ARC/INFO. The result was a 50m resolution DTM representing heights

above sea level (Figure 3.5). All other FORTRAN programs presented in this research have been written in FORTRAN for UNIX.

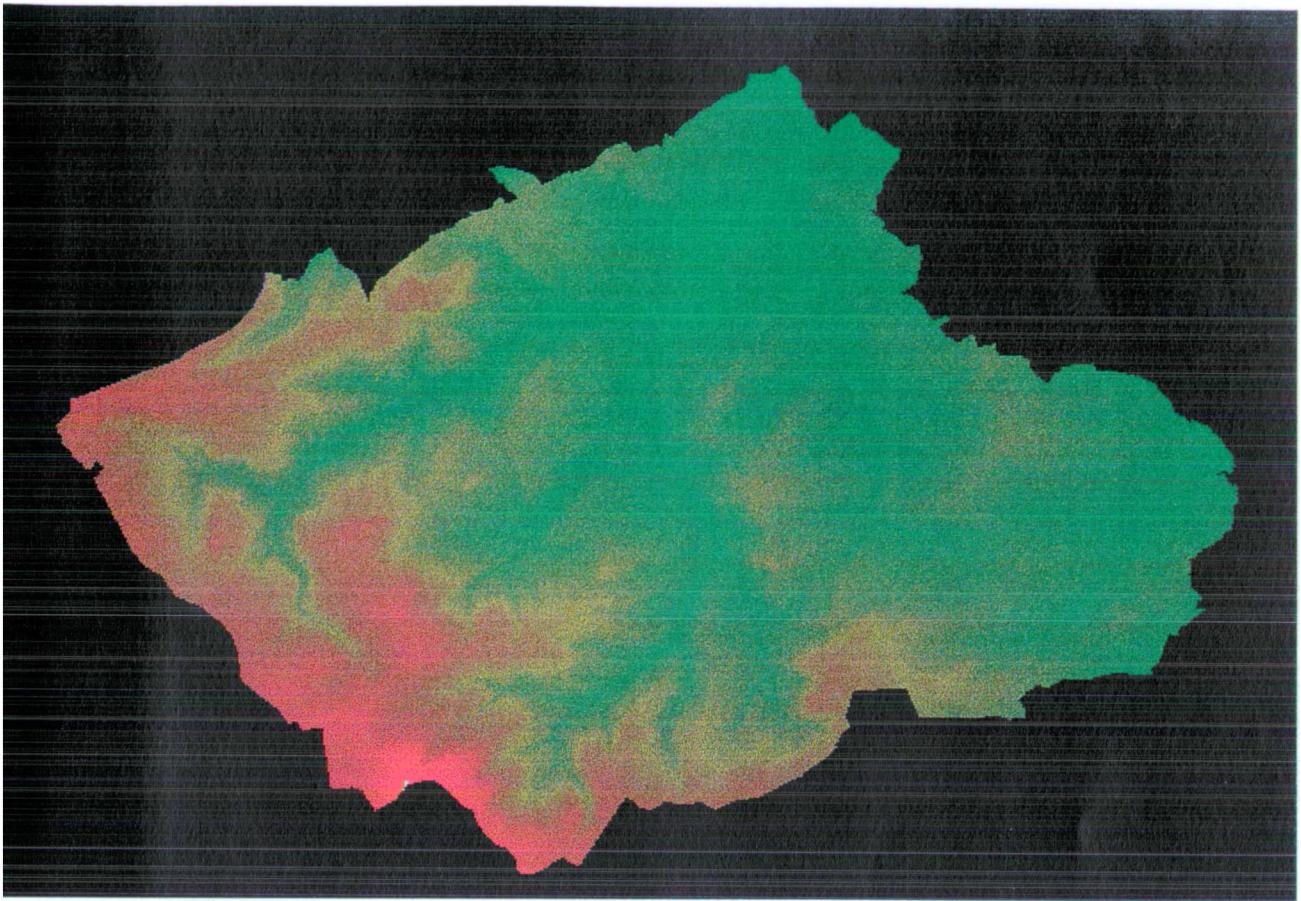
### **3.4.5 Field Data**

During each survey, data were also collected in the field for each site. The data included the height of the sampler above the ground and the mean angle to the horizon (topex). The protocol for monitoring in the SAVIAH study gave a guideline for the height of the samplers at 2 to 3 m above the ground. This height was attained at most sites. At some urban sites the sampler had to be placed slightly higher to counteract the threat of vandalism which had proved a potential threat in two pilot surveys. Topex provides a measure of the topographic exposure of the site and was calculated as the mean angle to the visible horizon from the site, in eight directions - north, north-east, east, south-east, south, south-west, west and north-west - measured with a clinometer.

### **3.4.6 Health Data**

Data on health were collected through a questionnaire survey by researchers at the Institute of Environmental and Policy Analysis under the direction of the SAVIAH co-ordinator, Prof. Elliott, then from the London School of Hygiene and Tropical Medicine. The questionnaire was sent to parents of all children aged between 7 and 9 living within and attending a school in the Huddersfield study area.

The questionnaire was designed by the SAVIAH project team, but incorporated questions previously used by the World Health Organistation and the International Study of Asthma and Allergies in Childhood (Asher *et al*, 1995).



#### LEGEND



44 m

580 m



5 km

*Figure 3.5 Altitude.*

The questionnaire (Appendix 2) included a range of questions relating to the child, broadly falling into the following categories:

- place of residence - address and postcode, length of time at that address, previous address if valid
- school of the child - address and postcode
- home exposures - heating, damp, pets, perceived exposure to traffic fumes and noise
- parents - education, smoking habits, respiratory disorders
- health of the child - with particular reference to respiratory disorders.

The questionnaire included 16 questions related to the health of the child, with some specific questions on respiratory wheeze and cough, such as:

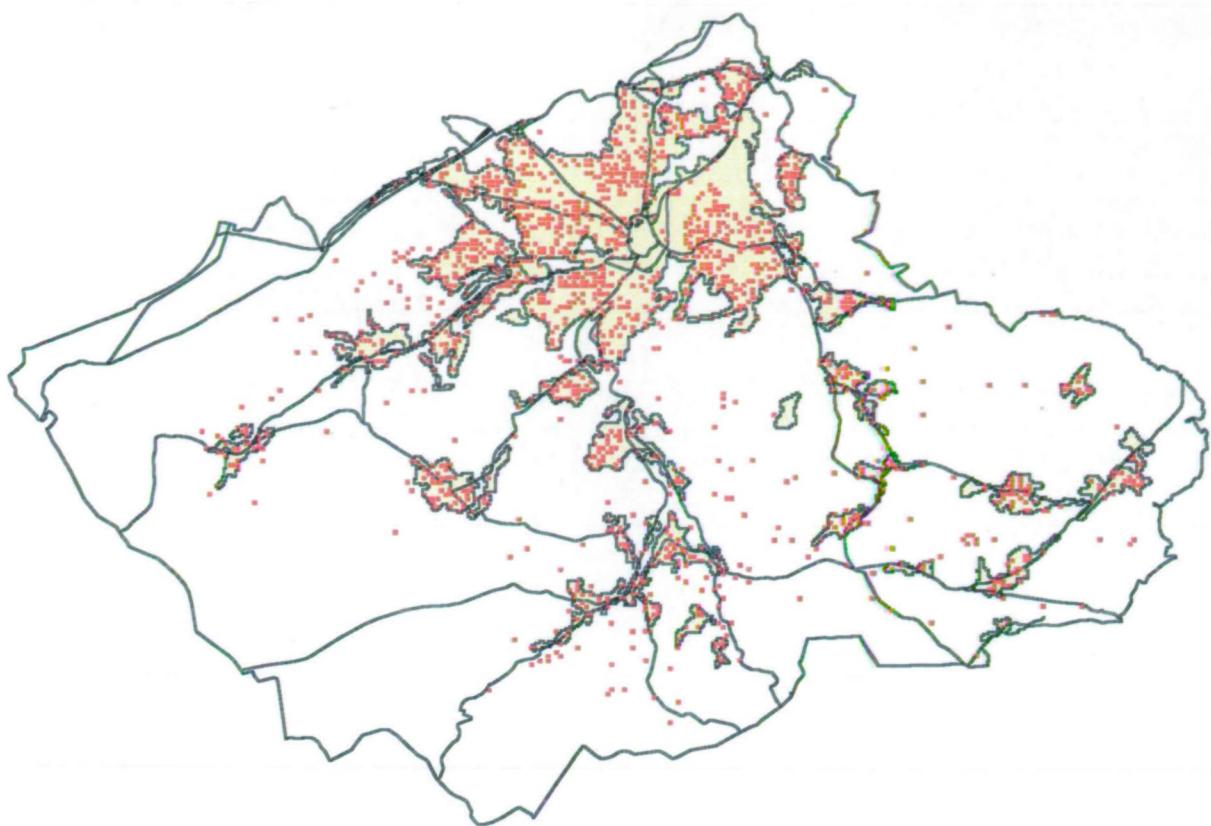
- ‘Has the child *ever* had wheezing or whistling in the chest *at any time* in the past?’
- ‘Has this child *in the last 12 months* had wheezing or whistling in the chest?’
- ‘Has the child had a dry cough at night *in the last 12 months* apart from a cough associated with a cold or chest infection?’
- ‘Does the child *usually* cough in the morning in the autumn-winter season?’

Questions were designed to represent measures of increasing severity of symptoms, over the two time-scales of concern - the previous 12 months and the full life of the child. Table 3.6 shows the health outcomes for two of the questions on the questionnaire; lifetime prevalence of wheezing or whistling and prevalence of wheezing or whistling in the last twelve months (Kriz *et al*, 1995).

*Table 3.6 Health outcomes.*

	Prague	Amsterdam	Poznan	Huddersfield
Life-time prevalence of wheezing or whistling	26.5%	22.8%	29.1%	30.0%
Prevalence of wheezing or whistling in the last twelve months	11.5%	10.2%	12.7%	17.8%

For inclusion in the GIS, the addresses of the children were geo-coded by colleagues at London School of Hygiene and Tropical Medicine using the postcode. The process of geo-coding enables national grid co-ordinates to be linked to the individual addresses of the children. The co-ordinates can then be used to map the geographical location of the place of residence. The postcodes were matched to postcodes in the Office for National Statistics (ONS) Central Postcode Directory (CPD). The CPD contains the 100m centroids of the postcodes, these were extracted for the matched postcodes and then imported into ARC/INFO (Figure 3.6).



## LEGEND

Urban areas

Main roads

- Place of residence



5 km

*Figure 3.6 The place of residence of the Children.*

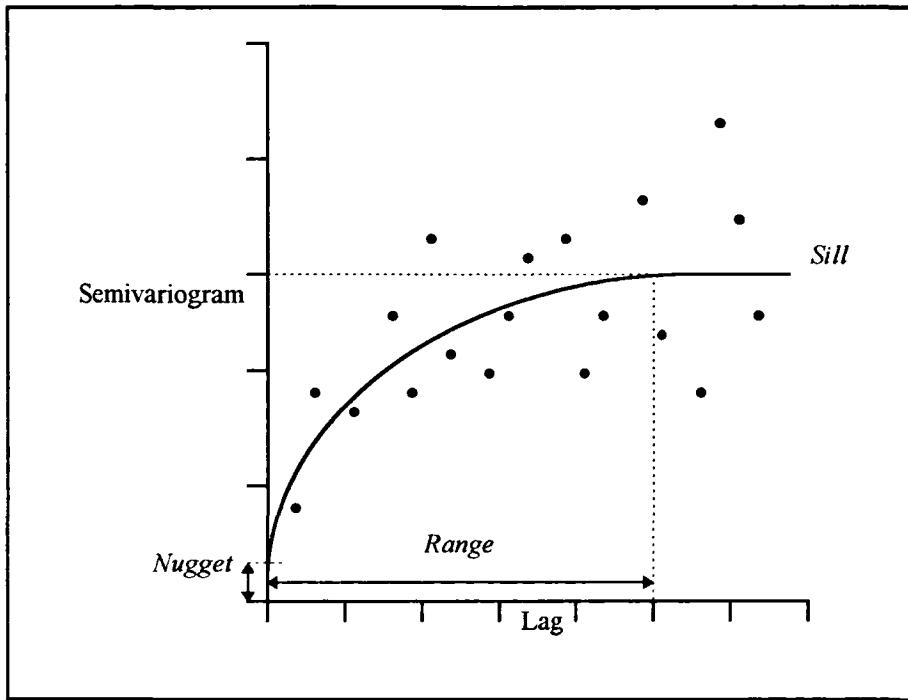
## **CHAPTER 4 KRIGING**

As discussed in Chapter 2, kriging was chosen as a control against which to compare the new methods. Kriging was applied to the Huddersfield adjusted annual mean NO<sub>2</sub> data in ARC/INFO. However, as will be seen, the ARC/INFO kriging routine does have limitations and is not very flexible.

In the SAVIAH study, in addition to kriging the NO<sub>2</sub> data, the author also applied kriging in ARC/INFO to the Poznan SO<sub>2</sub> data. Applying kriging to this data set had an important asset - kriging was also applied to the Poznan SO<sub>2</sub> data at RIVM (Bilthoven) using the FORTRAN library gslib and the spatial statistical package Splus (Dekkers, 1995). The SO<sub>2</sub> data, kriged by the two different methods, therefore afforded the opportunity to validate the ARC/INFO routine. In the case of the Huddersfield NO<sub>2</sub> data, access was not available to the gplib and Splus packages so only the ARC/INFO version was used.

### **4.1 Kriging Routines**

Many different kriging programs exist. As a general rule, the programs allow the user to calculate the semi-variogram for a data set, based upon a user-defined lag. The semi-variogram will then be displayed on the screen as a graph. A number of models will often be available (for example, spherical, exponential, gaussian): the user can select a model and the program will fit the model to the data (Figure 4.1). The user can interactively change the parameters of the model (e.g. the intercept of the curve, the range and the sill) until the model provides the best fit for the data - where the goodness of fit is often selected visually. The chosen parameters are then saved or recorded and used in the kriging calculations to provide estimates for a regular array of points.



*Figure 4.1 The semivariogram.*

Some of the programs come in the form of subroutines of FORTRAN or C code that are called from libraries. An example of one such library of subroutines is GSLIB (Deutsch and Journel, 1992). There are also packages especially designed to produce the semi-variogram, such as VARIOWIN (Pannatier, 1994), which has a Windows interface, or packages designed to undertake a range of spatial statistical functions, including kriging, for example, InfoMap (Bailey and Gatrell, 1995). Other packages, such as the statistical packages Splus, contain graphic and statistical functions. Splus also has its own language which allows new functions to be written. Kriging functions for Splus are available as user-developed code (e.g. Venables and Ripley, 1994). Finally, kriging functions have also been integrated with other types of packages, for example ARC/INFO.

Not all programs, however, have the same degree of flexibility. The kriging routine in ARC/INFO, for example (which includes a simple kriging function and variogram plot), does not allow the user interactively to change the parameters of the model.

Changing the parameters enables the user to modify the model until a model that best reflects the overall structure of the variogram is attained. If the parameters cannot be altered, then it is more difficult to achieve a good fit model for the variogram, in which case, the result may not be as statistically robust as the best fit model.

#### **4.2 Kriging the Huddersfield NO<sub>2</sub> Data**

Kriging was applied to the annual mean adjusted concentrations (section 3.3) for the 80 permanent monitoring sites. The spatial distribution and NO<sub>2</sub> value at each site can be seen in Figure 3.2.

In ARC/INFO, the NO<sub>2</sub> data were stored as a point coverage which holds information about the locations of the monitoring sites and the attribute values. The kriging routine was applied to the point coverage and values are entered for the lag, search radius and also the model. In the case of ARC/INFO the model could be spherical, exponential, gaussian or linear. The variogram was calculated and the distance and semi-variance values written to an INFO data file. The parameters of the fitted model are written to the display screen and stored internally in ARC/INFO. The data file and parameters are then used by ARC/INFO to draw the variogram and fitted model in a graphics window. The kriging routine is run for different models and lags until the best fit is visually attained. A spherical model for the Huddersfield annual mean NO<sub>2</sub> data was chosen with the following parameters:

Nugget	39.9 ug/m <sup>3</sup>
Range	2830 m
Sill	87.9 ug/m <sup>3</sup>
Lag	550 m

The kriging routine can then be used to calculate estimates for an array of grid points. A prediction of the variance (a measure of confidence) at each grid point can also be

calculated. The pollution map and the variance can be seen in Figures 4.2 and 4.3 respectively.

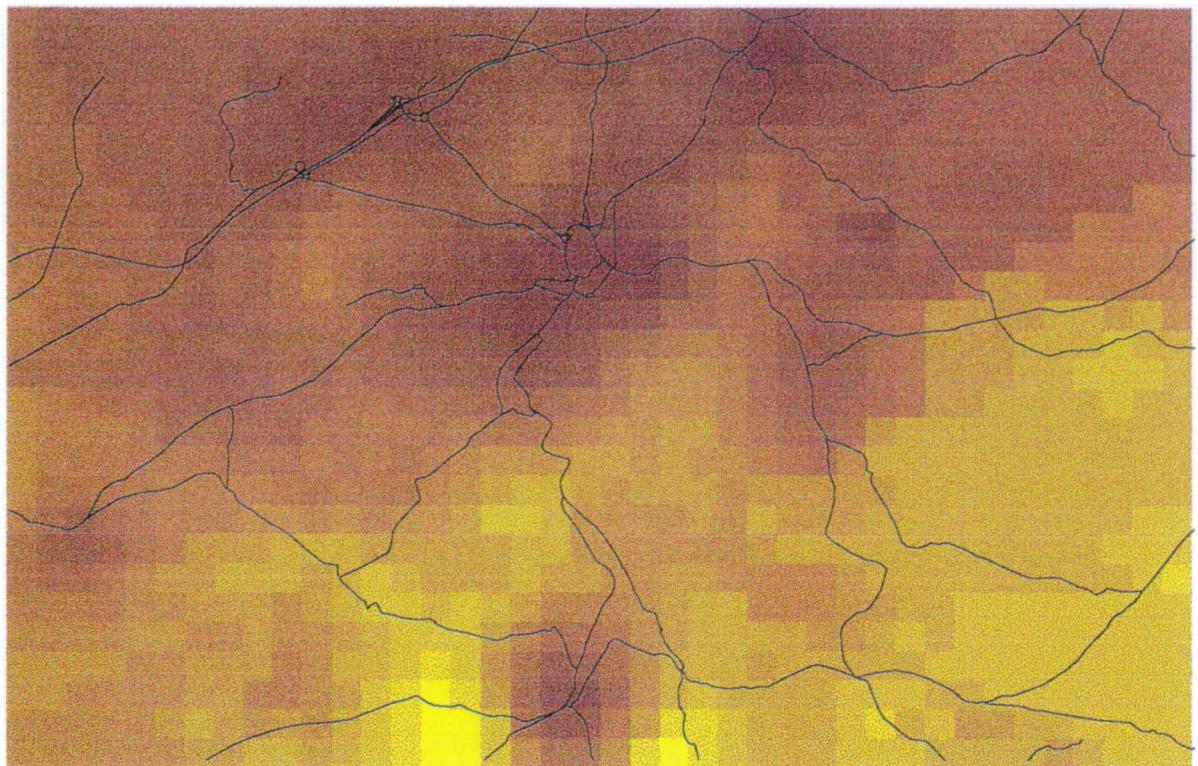
In Huddersfield, an additional 8 sites were monitored continuously throughout the study period. These 8 sites were not used in the kriging calculations. Estimates of NO<sub>2</sub> were derived for these 8 sites from the kriged pollution map. The annual mean measured values and estimates for the 8 sites can be seen in Table 4.1 and the scatter plot can be seen in Figure 4.4. The graph indicates that kriging may be overestimating for the lower NO<sub>2</sub> values and underestimating for the higher NO<sub>2</sub> values. The adjusted  $r^2$  value between the annual mean and the kriged estimates was found to be 0.439 (se = 6.45 ug/m<sup>3</sup>).

*Table 4.1 Annual mean NO<sub>2</sub> values and kriging estimates from ARC/INFO for the 8 consecutive monitoring sites (ug m<sup>3</sup>).*

Site-id	Annual Mean NO <sub>2</sub>	Kriging	Difference
103	31.3	33.4	-2.1
104	28.6	32.3	-3.7
105	34.3	31.7	+2.6
106	19.9	23.5	-3.6
107	21.7	29.1	-7.4
108	44.6	40.9	3.7
109	31.0	24.5	6.5
110	41.5	31.0	10.5

### 4.3 Kriging the Poznan SO<sub>2</sub> Data

In Poznan the predominant source of pollution is SO<sub>2</sub> from industrial point sources (i.e. chimneys). The geography and distribution of major chimneys in Poznan can be seen in Figure 4.5.



## LEGEND

ug/m<sup>3</sup>



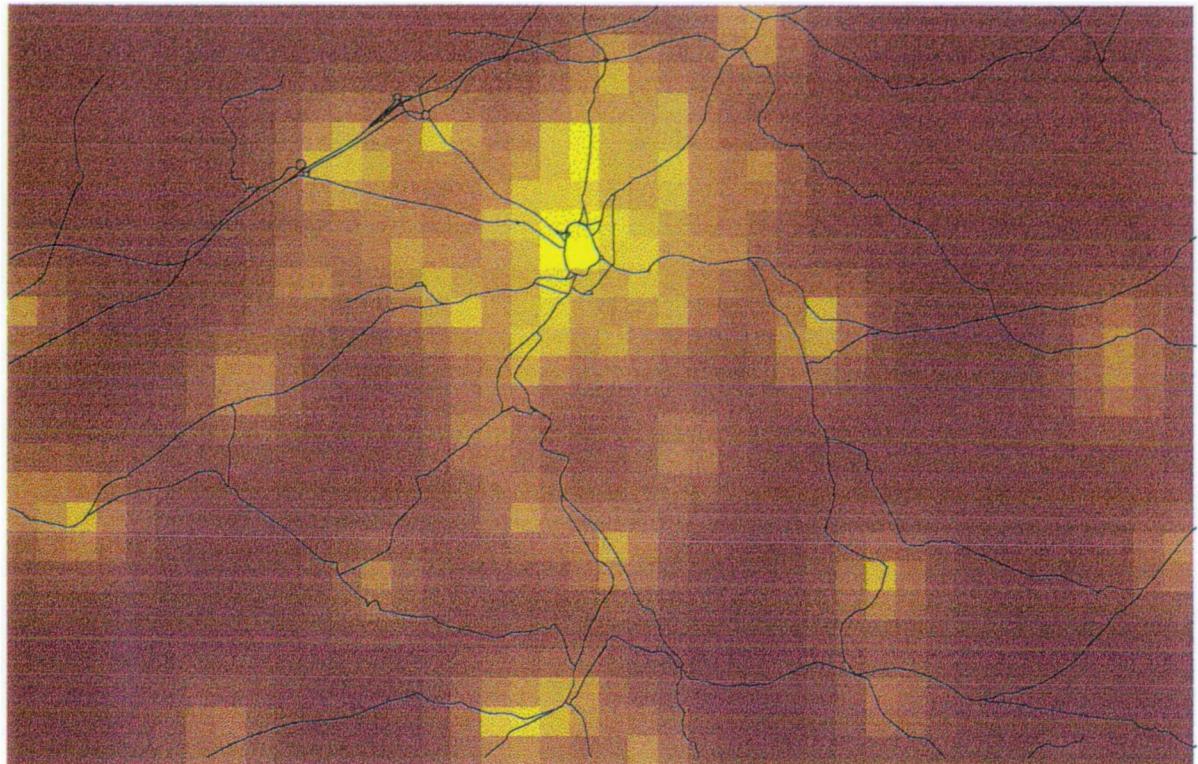
21

42



5 km

*Figure 4.2 Map of annual mean NO<sub>2</sub> by kriging.*



## LEGEND

### Variance



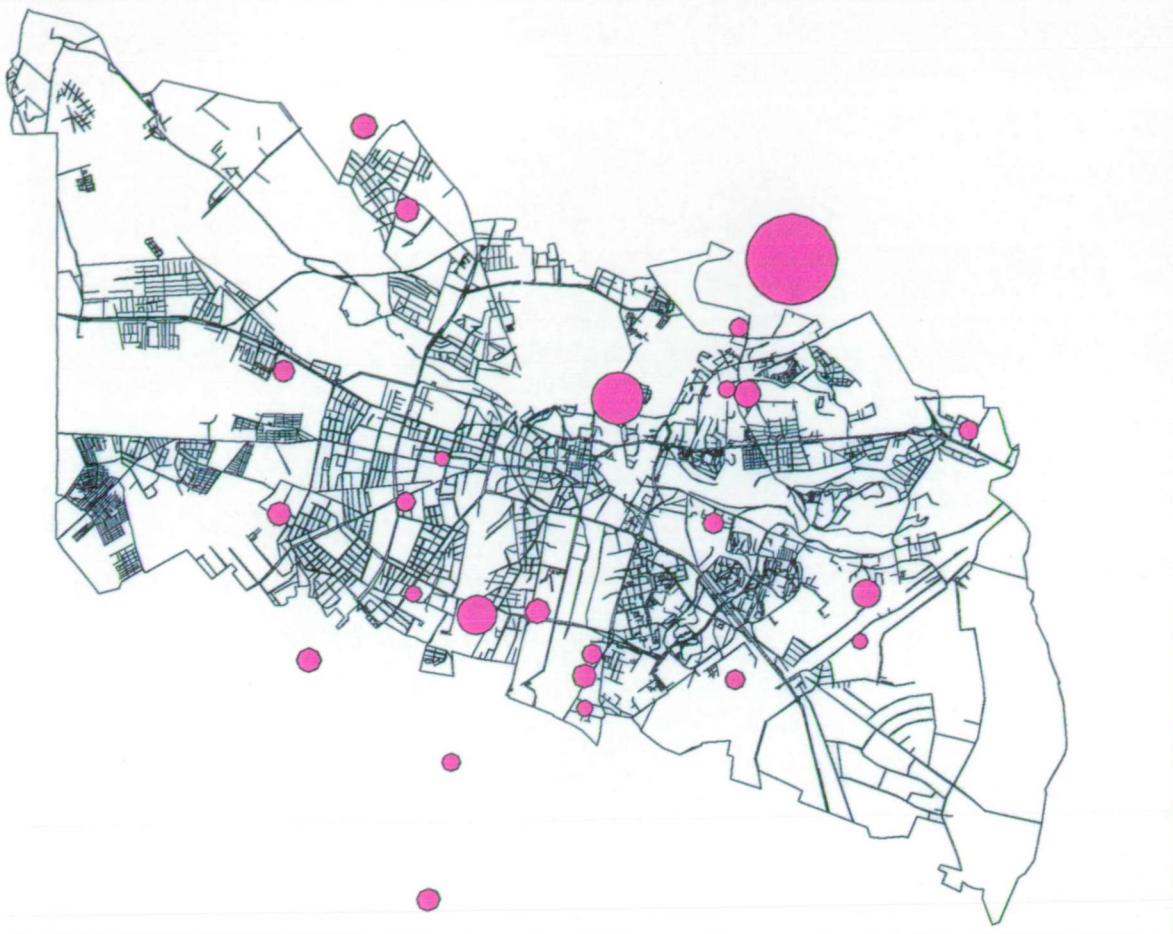
48

103



5 km

*Figure 4.3 Variance for the kriged estimates of NO<sub>2</sub>.*



## LEGEND

□ Roads

Tons/year

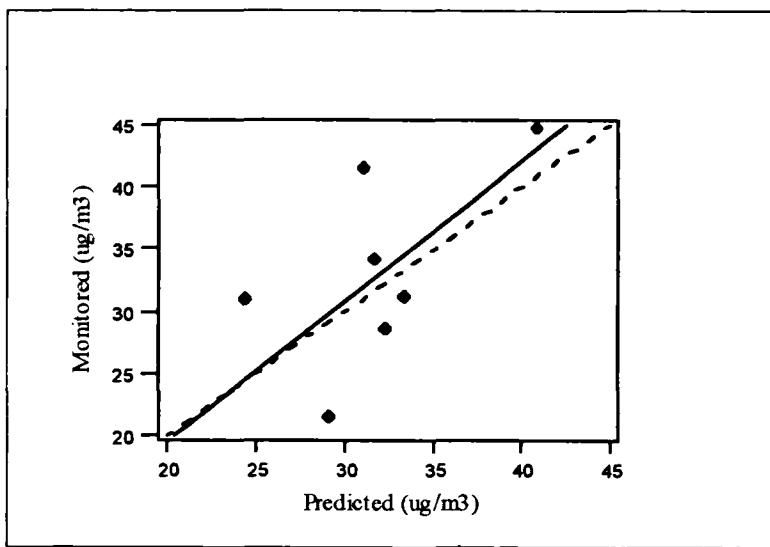
• 14

7072



5 km

Figure 4.5 The Poznan study area and main emission sources of  $\text{SO}_2$  (chimneys).

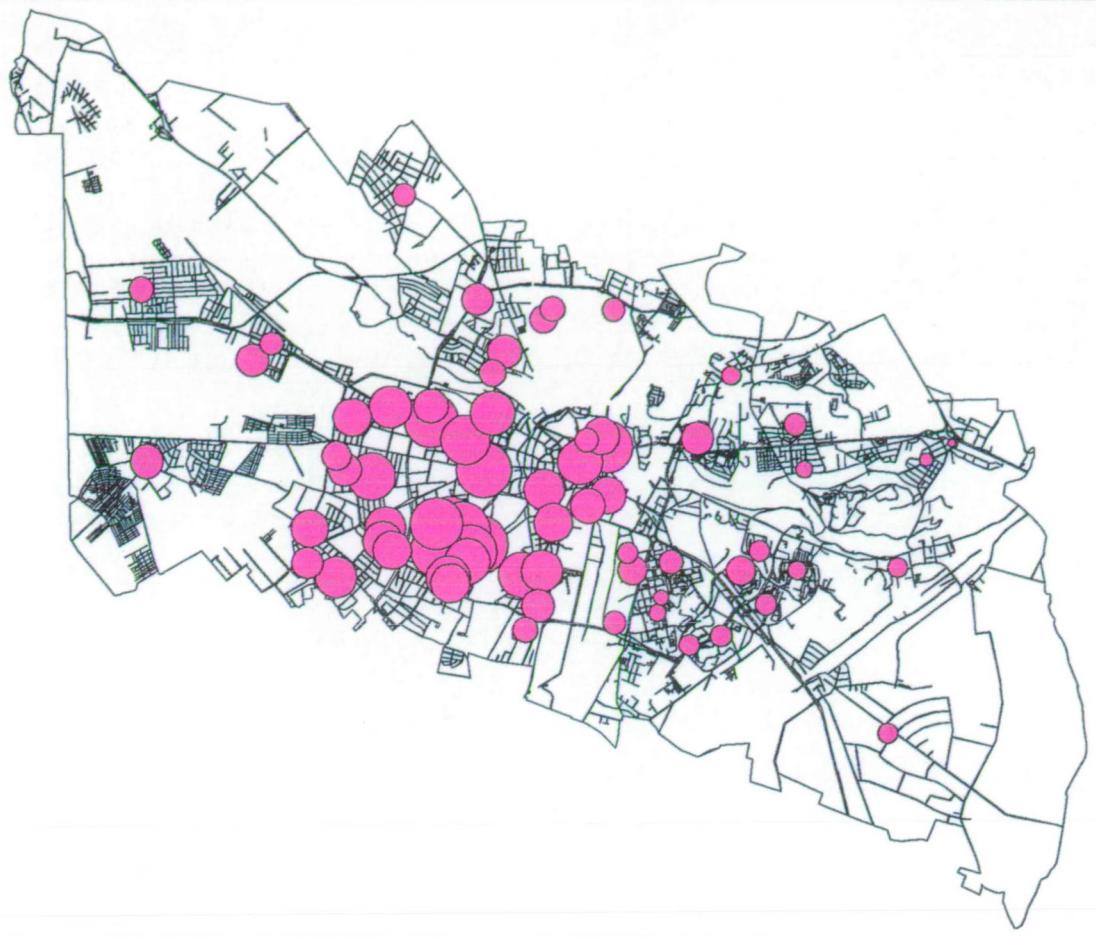


*Figure 4.4 Monitored annual mean NO<sub>2</sub> against kriging predictions for the 8 consecutive monitoring sites.*

As discussed in Chapter 3, SO<sub>2</sub> was measured at 68 locations (Figure 4.6). For each site, the average concentrations from the second and third surveys (November 1993 and February 1994) were calculated. Kriging was applied to the Poznan data at RIVM using GSLIB and Splus.

An analysis of the data showed that there was no anisotropy in the data (i.e. the data behaved the same in all directions) so the data could be described by one variogram, rather than modelling separate components in different directions. A spherical model was fitted to the semi-variogram with the following parameters:

Nugget	149 ug/m <sup>3</sup>
Range	3140 m
Sill	704 ug/m <sup>3</sup>
Lag	200 m



## LEGEND

■ Roads

$\mu\text{g}/\text{m}^3$

• 41.9

● 162.6



5 km

*Figure 4.6 Average  $\text{SO}_2$  at the 68 permanent monitoring sites.*

Ordinary kriging was then used to calculate estimates of SO<sub>2</sub> on a 100x100m grid across the Poznan study area using the spherical model.

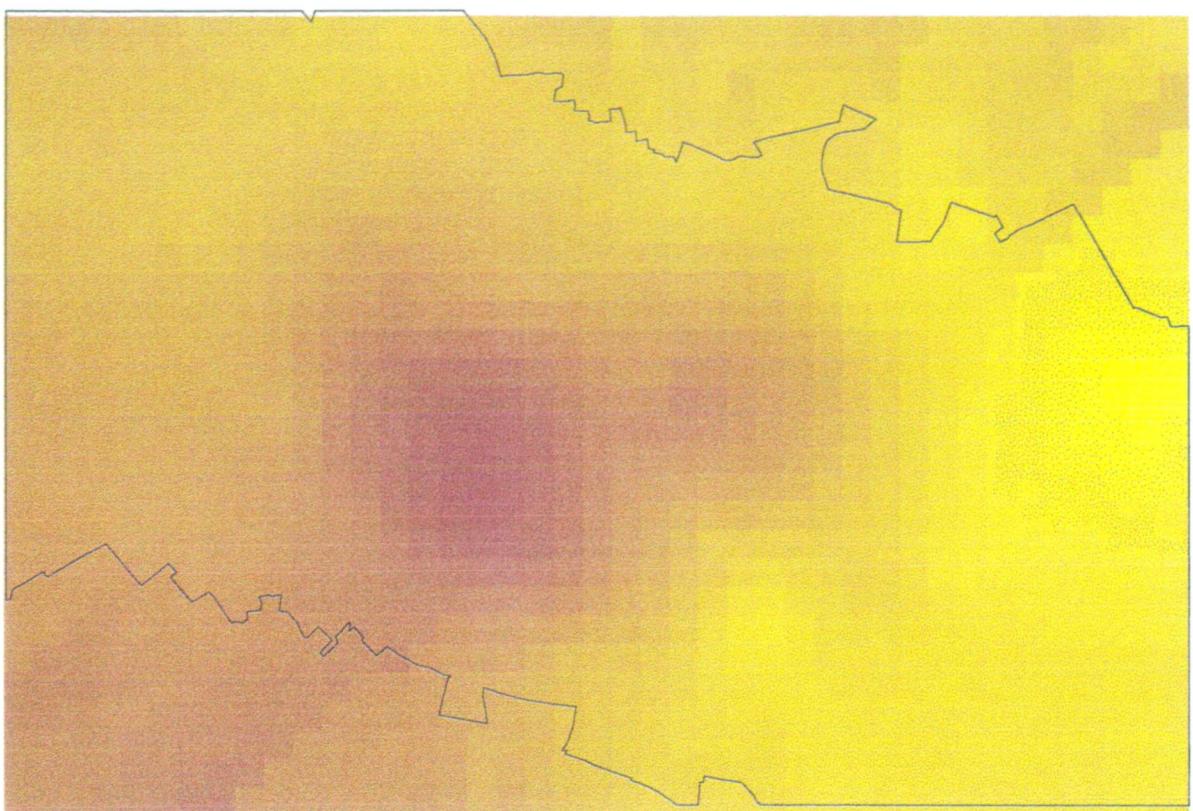
In Huddersfield, the author applied kriging to the same SO<sub>2</sub> data set in ARC/INFO. The kriging routine was run for different models and lags. A spherical model with the following parameters was found to visually provide the best fit:

Nugget	201 ug/m <sup>3</sup>
Range	4320 m
Sill	981 ug/m <sup>3</sup>
Lag	350 m

The maps representing the kriging estimates from ARC/INFO and the GSLIB/Splus combination can be seen in Figures 4.7 and 4.8 respectively.

In Poznan, as in Huddersfield, an additional 8 sites, not used in the kriging calculations, were monitored continuously throughout the study period. Both methods were used to calculate estimates for the 8 sites as can be seen in Table 4.2 and Figure 4.9. At low concentrations, both kriging methods provide reasonable predictions of SO<sub>2</sub>; at high concentrations, however, kriging appears considerably to underestimate levels of SO<sub>2</sub>. The two methods were compared by calculating the correlation coefficients between the measured values (the average of surveys 2 and 3) and the estimated values. The correlation coefficients can be seen in Table 4.3. As can be seen in the table and Figures 4.9 and 4.10, the two methods provide very similar estimates at the 8 sites.

As Figure 4.10 and Tables 4.2 and 4.3 show, there is very little difference in the ability of the two kriging routines to estimate the measured values - both provide very similar estimates at the 8 points with similar degrees of variation between the measured values and the estimated values. The correlation coefficients between the ARC/INFO kriging and the GSLIB/Splus combination and the measured SO<sub>2</sub> values are almost identical - 0.733 and 0.730 respectively.



#### LEGEND

$\mu\text{g}/\text{m}^3$



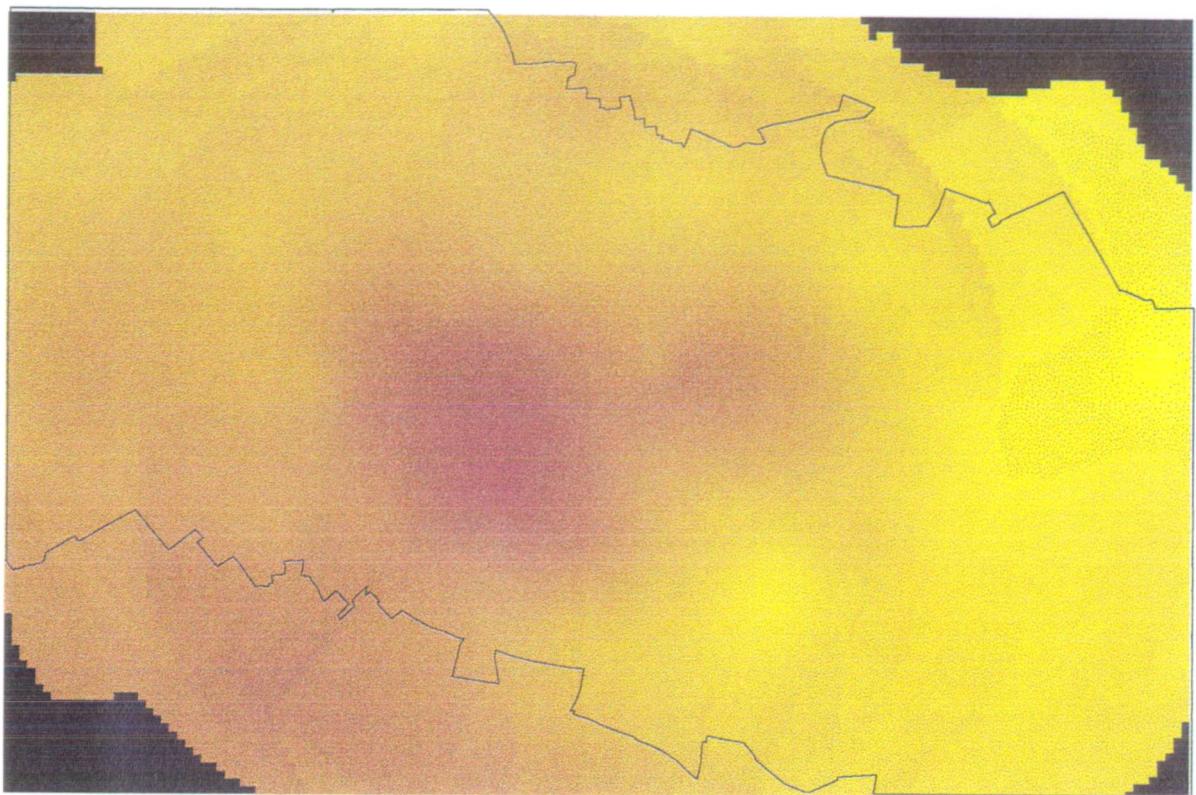
44

143



5 km

*Figure 4.7 Map of Average  $\text{SO}_2$  by Kriging: ARC/INFO.*



#### LEGEND

ug/m<sup>3</sup>



43

144

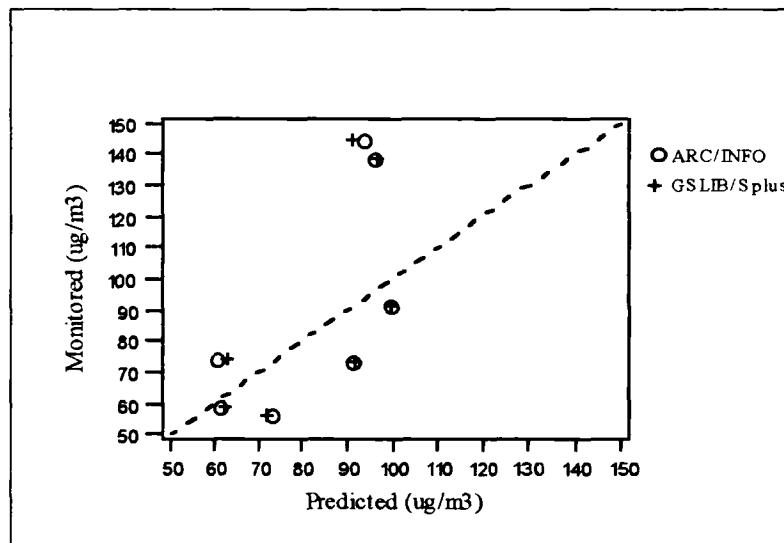


5 km

*Figure 4.8 Map of average SO<sub>2</sub> by kriging: GSLIB Splus.*

*Table 4.2 Average SO<sub>2</sub> values and kriging estimates from ARC/INFO and GSLIB and Splus for the 8 consecutive monitoring sites.*

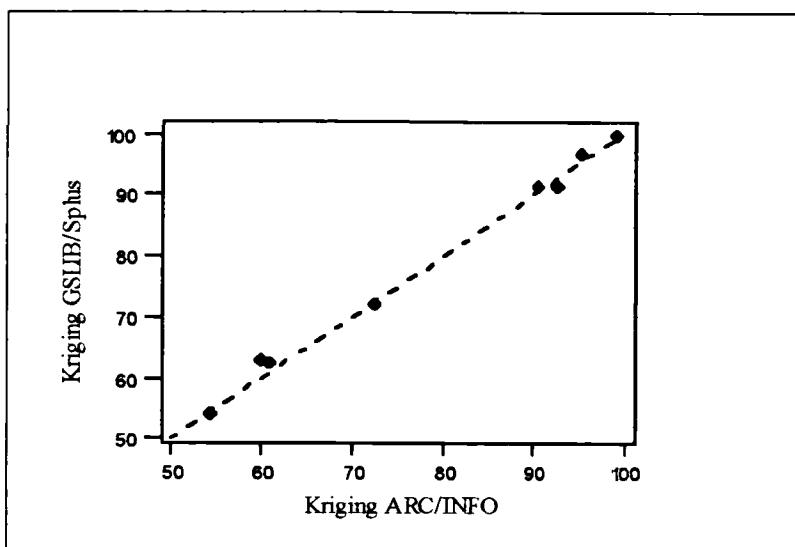
Site-id	Average SO <sub>2</sub> (ug/m <sup>3</sup> )	Kriging ARC/INFO	Difference	Kriging GSLIB/ Splus	Difference
171	91.7	99.2	-7.5	99.9	-8.2
172	144.7	92.8	51.9	91.5	53.2
173	73.3	90.7	-17.4	91.6	-18.3
174	58.7	60.7	-2.0	62.4	-3.7
185	74.5	59.8	14.7	63.0	11.5
186	48.6	54.3	-5.7	54.2	5.6
187	56.3	72.3	-16.0	72.2	-15.9
188	138.8	95.4	43.4	96.8	42.0



*Figure 4.9 Monitored average SO<sub>2</sub> against kriging predictions by the ARC/INFO and GSLIB/Splus methods for the 8 consecutive monitoring sites.*

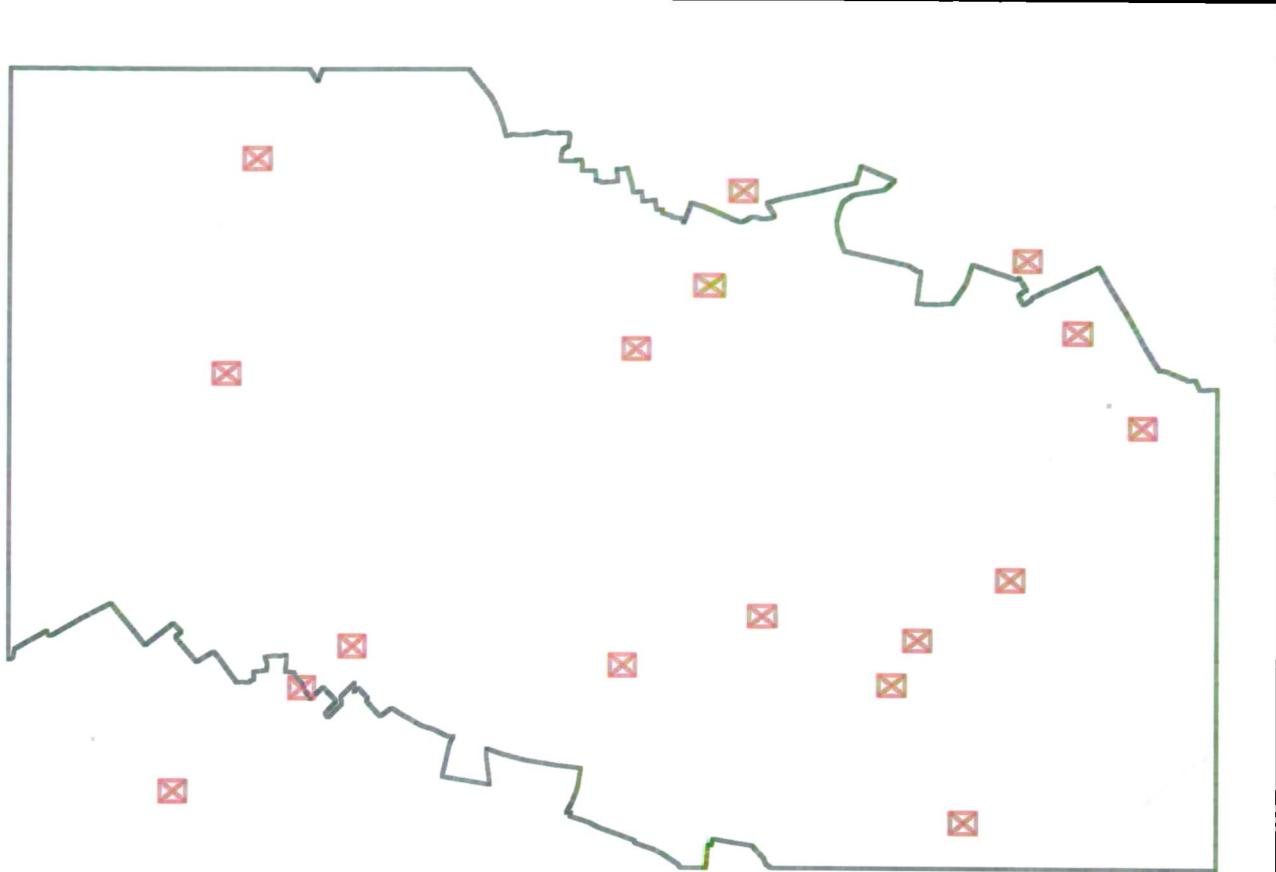
*Table 4.3 Correlation coefficients.*

Correlation Coefficient	Average SO <sub>2</sub> (ug/m <sup>3</sup> )	Kriging ARC/INFO	Kriging GSLIB/Splus
Average SO <sub>2</sub>	1.000	0.733 ( $\rho = 0.039$ )	0.730 ( $\rho = 0.040$ )
Kriging ARC/INFO		1.000	0.997 ( $\rho = 0.000$ )
Kriging gslib/Splus			1.000



*Figure 4.10 Kriging predictions for GSLIB/Splus against kriging predictions for ARC/INFO for the 8 consecutive monitoring sites.*

The degree to which the two routines differ, with respect to providing estimates at unsampled location, can be tested further by taking a random sample of points across the study area and comparing the two estimates for those points. To this end, 20 sets of co-ordinates within the Poznan study area were randomly generated (Figure 4.11). Estimates of the pollution surface from the ARC/INFO kriging and the GSLIB/Splus combination were established for all 20 points - 2 of the points were subsequently found to lie outside the extents of the grids created by the two kriging routines. The estimated SO<sub>2</sub> values for the remaining 18 sites are presented in Table 4.4. The correlation coefficient between the estimates from the ARC/INFO kriging routine and the GSLIB/Splus combination for the 18 points was found to be 0.993 ( $\rho = 0.000$ ) and the scatter plot can be seen in Figure 4.12. As the graph shows, there is a strong linear relationship between the two estimates, with the slope of the regression line very close to 1. In this case, therefore, the lack of flexibility in the ARC/INFO kriging routine does not appear significantly to affect the ability of the routine to provide estimates at unsampled locations compared to the more interactive routines of GSLIB and Splus.



## LEGEND

☒ Random point

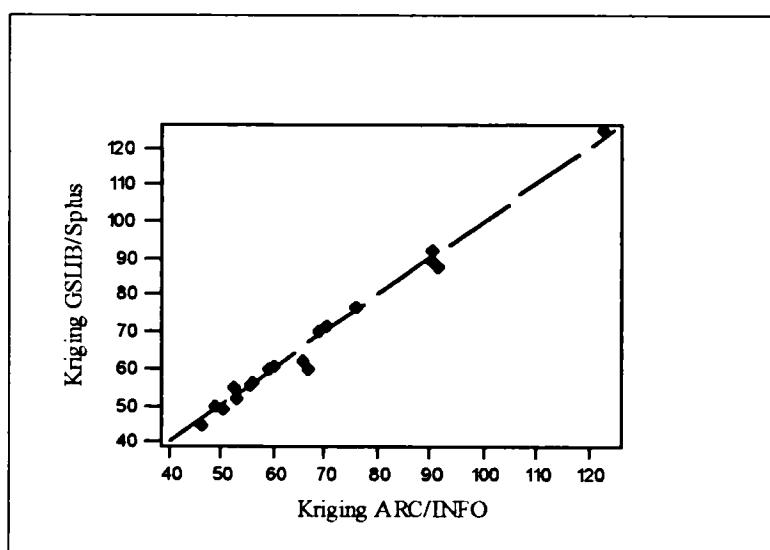


5 km

*Figure 4.11 Location of the randomly generated points.*

*Table 4.4 Kriging estimates of SO<sub>2</sub> from ARC/INFO and GSLIB/Splus combination for a random set of points.*

Random point ID	Kriging ARC/INFO	Kriging GSLIB/Splus	Difference
2	65.7	62.1	3.6
3	90.6	88.8	1.8
4	53.0	52.1	0.9
5	52.5	54.6	-0.1
6	70.3	71.9	-1.6
8	68.9	70.3	-1.4
9	59.3	59.9	-0.6
10	56.2	56.1	0.1
11	50.5	49.4	1.1
12	91.4	87.6	3.8
13	46.5	44.9	1.6
14	49.1	49.7	-0.6
15	67.0	60.1	6.9
16	76.0	76.4	-0.4
17	122.8	124.5	-1.7
18	90.5	91.9	-1.4
19	55.7	55.5	0.2
20	60.3	60.4	-0.1



*Figure 4.12. Estimated SO<sub>2</sub> values for Kriging by the GSLIB Splus method against Kriging by the ARC/INFO method for the 18 random locations.*

#### **4.4 Applying Kriging to Traffic Related Pollution**

Kriging has been applied to model two different pollutants: NO<sub>2</sub> and SO<sub>2</sub>. The pollutants are products of two very different sources of pollution. In Huddersfield, the major source of pollution is NO<sub>2</sub> from road transport, whereas in Poznan the major source of pollution is SO<sub>2</sub> from industrial chimneys and domestic heating. The two pollution surfaces are therefore likely to be very different.

Emissions from road transport occur at ground level, with peaks of NO<sub>2</sub> reflecting the linear patterns of the road network. In urban environments, NO<sub>2</sub> has a short range of transport due to the effect of height and density of buildings on either side of the road. In comparison, emissions from chimneys occur above ground level and the pattern of dispersion is less affected by buildings, and therefore the pollutant has a longer range of transport than NO<sub>2</sub>.

Although kriging, which uses local variations in the data to describe the surface, is a sophisticated interpolation technique, emissions from road traffic are much more difficult to represent by interpolation methods. The nature of traffic related pollution is that it will peak strongly close to main roads and decline rapidly with distance away from the roads. Unless there is a dense network of sites, which adequately represents the relatively small near-source areas as well as more distant, background areas, interpolation will tend to smooth the variation that occurs between the points. To model the complexity of spatial variation in urban areas without smoothing would require an unrealistically large number of sites.

As discussed in Chapter 2, kriging has been widely used to model and map variations in air pollution. Where only monitored data are available, kriging is the most appropriate tool with which to generate maps of the pollution surface. However, where other information is available, then other techniques, such as co-kriging (section 2.2), which uses additional information sampled at the same location and also at other locations (providing a denser network of sample points), may help overcome the problems associated with a sparse network of monitoring sites. Nevertheless, as a

sophisticated interpolation technique, that also provides estimates of the errors, kriging remains the most widely applied technique for producing maps of pollution surfaces and is therefore a good standard against which to compare other methods.

## CHAPTER 5 AUTOMATIC MODELLING OF TRAFFIC RELATED AIR POLLUTION (AMTRAP)

As was noted in Chapter 2, two main approaches to air pollution mapping can be defined: spatial interpolation and dispersion modelling. Individually, neither might be considered ideal in order to represent the complex pollution fields seen in urban environments: as has been seen with kriging, without an overly dense network of monitoring sites spatial interpolation techniques tend to smooth the surface, whilst the available line dispersion models may only be considered accurate close to the emission source. For this reason, a hybrid approach was developed as part of this research - AMTRAP (Automated Model for Traffic Related Air Pollution). AMTRAP is the only method developed that uses daily meteorological data. The method is based upon the principle that patterns of air pollution in urban environments can be described by two different, but related, components of variation (Collins, *et al.*, 1995):

- ***near-source variation*** - related to the dispersion processes associated with distinct point or line sources
- ***background variation*** - reflecting differences in diffuse sources, broader (e.g. local topographic) controls on dispersion and long-distance transport of pollutants across the study area

The two components were predicted separately within a GIS environment using different techniques. The near-source pollution was modelled by adapting the line dispersion model CALINE3 to operate automatically within the GIS. This provides a close link between the GIS and the dispersion model. The advantages of linking the dispersion model to the GIS in this manner are (Collins, 1998):

- high resolution modelling.
- many road segments can be modelled automatically - this is especially useful

for large study areas.

- the user does not need to transfer data from one system to another - running the dispersion model outside the GIS and then transferring the results back into the GIS for mapping.
- the user does not need to learn two different software packages (probably running on two different operating systems).

The background variation is modelled using the kriging routine available in the GIS. The two separate components were then additively combined to produce the final air pollution map.

### **5.1 Modelling Near-Source Air Pollution**

The near-source pollution changes in response to local variations in emission sources, influenced primarily by:

- road type
- traffic volume
- traffic composition
- traffic speed
- emission rates

and the dispersion environment, defined by:

- the rate of dispersion over distance from the source
- the surface roughness of the land adjacent to the roads
- wind direction
- wind speed
- atmospheric stability

A line source dispersion model was chosen to model the near-source component of pollution which could be adapted to work within the GIS. As discussed in Chapter 2, there are many different line-source dispersion models available; in this case, the dispersion model CALINE3 was chosen for the AMTRAP approach. Although CALINE3 is not the most sophisticated model, it was chosen in preference to more complex models due to its ease of use, its relatively simple data requirements and the fact that the input data were available. CALINE3 predicts concentrations of the traffic related pollutant, CO or other inert gases, at road-side locations based on meteorology and traffic flow. The transport and dispersion element of the model is a revised version of the Gaussian point source plume dispersion model and applies vertical and horizontal dispersion curves modified for the effects of surface roughness, averaging time and vehicle-induced turbulence (Benson, 1992).

CALINE4 is the most recent version of the CALINE group of dispersion models and does offer a number of improvements over the CALINE3 model, including new provisions for lateral plume spread and vehicle-induced thermal turbulence, an intersect option and options for modelling NO<sub>2</sub> and aerosols, as well as CO. However, the new provisions for lateral plume spread and vehicle-induced thermal turbulence were found to be ineffective in complex terrain, the intersect option is impracticable in large study areas due to the additional data requirements for deceleration time, acceleration time, cruising speed and idling time, and the NO<sub>2</sub> option is not recommended where parallel wind conditions occur and it also requires some measure of the background ozone level (Benson, 1992). The Huddersfield study area is very hilly and requires many road segments to be modelled under different wind conditions. Against this background, running CALINE4 was felt to offer no significant advantage over running the simpler CALINE3 model. However, in the CALINE3 model, NO<sub>2</sub> has to be treated as an inert gas, which may introduce some uncertainty into the measurements.

The input variables for CALINE3 include: traffic volume, emission rate, surface roughness, mixing zone (width of traffic lanes plus 3m on either side of the road), wind speed, wind direction, stability class, source height, receptor height, averaging time, the start and end co-ordinates of a straight length of road and the co-ordinates of the

receptors. CALINE3 predicts pollution concentrations for road links (road segments of uniform conditions) and can model up to 20 links and 20 receptors at any one time. It should be noted, however, that these constraints can be adjusted in the source code, but to do so adds considerably to the processing time required to run the model. Each link is a straight segment of constant width, traffic volume and emission rate. Real world situations can be approximated by analysing multiple links, but the background pollution value must be specified by the user. Surface roughness, atmospheric stability, wind speed and direction are assumed to be geographically constant over the study area. The links and the receptors are oriented within a co-ordinate reference system. As mentioned in Chapter 2, once concentrations have been established at the receptors, the values could be entered into a mapping package and a pollution surface generated.

In area-wide detailed studies, road networks are extremely complex, with constantly changing conditions - reflecting changes in emission sources, the surface adjacent to the road and temporal variations in meteorology. In these situations, CALINE3 would have to be run a great many times in order to model the many different conditions - every change in conditions along the road network would have to be represented by a new link in the model. Furthermore, changes in the surface roughness, atmospheric stability, wind speed and direction would need to be modelled in separate runs of CALINE3. For a large study area, running these many links through CALINE3 and then creating a pollution map would take excessive amounts of time and effort. In the light of this complexity, an automated approach to pollution modelling was developed by adapting the CALINE3 model to operate within the GIS.

The automated method in GIS was used to predict pollution concentrations in near-source areas - a 200m band adjacent to the road links. As opposed to predicting pollution concentrations for specified receptor locations, the automated method generates a prediction every 10m, within this 200m band. To achieve this, a new command was created by the author in ARC/INFO using a combination of AML (Advanced Macro Language), which is ARC/INFO's own language for developing customised routines, and compiled FORTRAN programs. The command is described in more detail later in this Chapter.

Within AMTRAP, the input variables are characterised at every location on a 10m grid and a concentration, pre-defined in CALINE3, is extracted for those characteristics. The pre-defined values were generated by running CALINE3 for a simplified set of conditions, chosen to represent the full range of possible weather conditions, surface roughness, road width and distance from the road (as explained in the next section). For example, in CALINE3, a value for wind direction could be entered between 0 and 360 degrees: this was simplified to 12 classes, each class representing 30 degree intervals. The pre-defined values were generated using a single (reference) traffic flow. The generated concentrations could then be proportionally increased or decreased to reflect the actual traffic volume.

As previously mentioned, the near-source concentrations are then added to the background concentrations - generated with the kriging routine - to produce the final pollution estimates.

### **5.1.1 Generating Pollution Concentrations for a Set of Pre-Defined Variables**

To help establish the pollution concentrations for a pre-defined set of conditions, 5 of the input variables were entered as constants:

emission factor	1.62 g/km
traffic volume	4000 vehicles/hour
averaging time	1 hour
source height	0
receptor height	2 m

The co-ordinates of the road link were also kept constant, fixing the length of the road link to 200m. The emission factor of 1.62 g/km was derived from Gillam *et al* (1992) to represent an average urban mixture of light and heavy traffic.

The remaining variables - surface roughness, mixing zone, wind speed, wind direction and stability class - were assumed to vary spatially and temporarily across the study area. A simplified set of conditions, chosen to reflect the full range of possible values available (Benson, 1979), were therefore established for these variables. Reference values were chosen to reflect the simplified conditions and code numbers were assigned to the reference values. In the case of the Pasquill Stability Scheme, the values range from very unstable (A) through to very stable (F), where D is considered to be the most neutral value. The code numbers and reference values for surface roughness, mixing zone, wind direction, wind speed and atmospheric stability are presented in Table 5.1. It can be seen from the table, that the reference values for wind direction only range from 0 to 150 degrees. This is due to the fact that pollution concentrations are a mirror image about 180 degrees and therefore the full range can be represented by half the circle. The derivation of these parameters is explained in the next section.

*Table 5.1 Codes and reference values for wind direction, wind speed, atmospheric stability, surface roughness and mixing zone.*

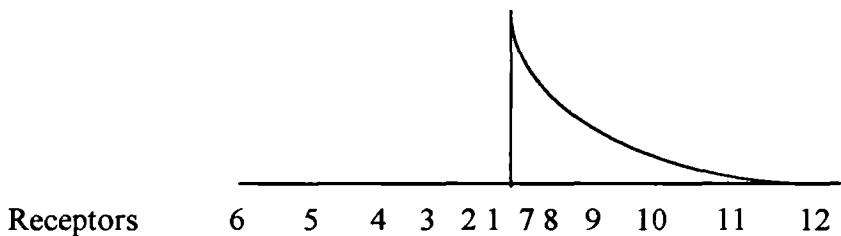
Code	Wind Direction (degrees)	Wind Speed (m/s)	Stability (Pasquill Scheme)	Surface Roughness (cm)	Mixing Zone (m)
1	0	1	A	5	28
2	30	3	B	100	21
3	60	5	C	250	13.5
4	90	8	D	300	11.5
5	120	15	E	350	
6	150		F		

Pollution concentrations were calculated for receptors at distances 5, 20, 40, 65, 100 and 150m perpendicular to the road links. Output from CALINE3 was presented as a series of tables in ASCII format. The FORTRAN program *concsext* (Appendix 3), written by the author, was applied to extract the concentrations and write them to a new text file, in a format compatible with INFO. The information was then transferred to a defined table in INFO (*concs.dat*), a sample of which can be seen in table 5.2.

*Table 5.2 Sample data from the concs.dat data file.*

CONCCODE	RP1	RP2	RP3	RP4	RP5	RP6	RP7	RP8	RP9	RP10	RP11	RP12
11111	0	0	0	0	0	0	152	94	62	41	26	17
11112	0	0	0	0	0	0	72	43	28	18	11	7
11113	0	0	0	0	0	0	47	28	18	12	7	5
~	~	~	~	~	~	~	~	~	~	~	~	~
23413	112	36	11	2	0	0	112	36	11	2	0	0
23423	49	15	5	1	0	0	49	15	5	1	0	0
~	~	~	~	~	~	~	~	~	~	~	~	~
45656	19	12	9	7	5	4	0	0	0	0	0	0

CONCCODE is a unique code number. The five digits are formed from a combination of the codes for mixing zone, surface roughness, wind direction, wind speed and atmospheric stability class. RP# is the receptor location; receptors 1 to 6 are located on the opposite side of the road to receptors 7 to 12. Receptors 1 and 7 are located nearest to the roads, as in Figure 5.1.



*Figure 5.1 Location of receptors.*

### 5.1.2 Description of the TRAFFPOL Routine

Near-source pollution concentrations were generated in ARC/INFO at the arc prompt with the command *traffpol* (Appendix 4), written by the author using a combination of AML and compiled FORTRAN programs. The *traffpol* routine requires three sources of input: a line coverage of roads in ARC/INFO format, with road type code and traffic volume attributes; a polygon coverage for surface roughness in ARC/INFO format; an

ASCII text file containing the weather data (wind direction, wind speed and atmospheric stability class). The form of the command line is (Collins, 1997):

```
&run traffpol <cover> <outgrid> <roadtype_item> <traffvol_item>
<surface_rough> <weather_data>
```

arguments:

<cover>	line coverage of roads to be analysed
<outgrid>	output grid of near-source air pollution
<roadtype_item>	item containing the code for road type
<traffvol_item>	item containing the traffic volumes in vehicles/hour
<surface_rough>	polygon coverage of classified surface roughness (with a surface roughness item named SUR_CODE)
<weather_data>	text file containing the weather data (with extension)

Within *traffpol*, the line coverage was converted to co-ordinates with the *ungegenerate* command in ARC. The compiled program *arcirection*, generated from the *adirect* program (Appendix 5), written by the author, scans the arcs and calculates the bearing between successive vertices (pairs of co-ordinates). The maximum arc ID is identified and the individual sectors between vertices are classified according to direction. The arcs were scanned again and split for changes of class (i.e. change of direction), assigning new IDs to new arcs. The segmented arcs were then written to a new file in arc ungenerated format and the arc IDs and direction classes written to a further text file for import into INFO.

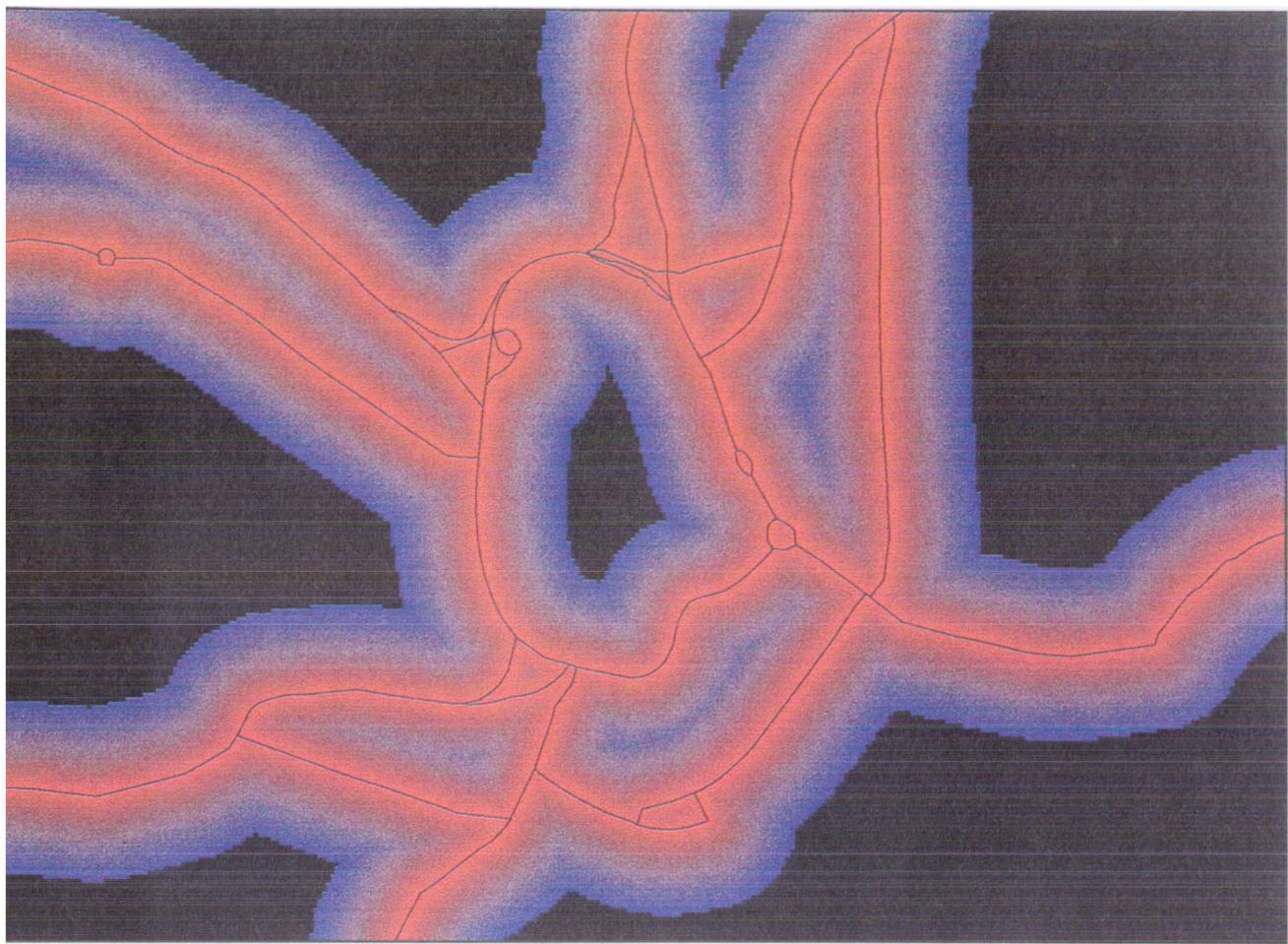
The arcs were then transformed back to ARC/INFO format with the *generate* command. The direction class is attached to the new arc coverage with the *joinitem* command. Three grids were then created with the *linegrid* command. The grids were

created on items stored in the arc attribute table, namely: arc direction, road type and traffic volume. In GRID, for all cells within the near-source area (i.e. less than 200m from the roads), the commands *eucdistance* and *eucallocation* are applied to calculate:

- the distance to the nearest road (Figure 5.2)
- the value of the nearest road by type, i.e. motorway, A-road (Figure 5.3)
- the orientation of the nearest road, between 0 and 180°, with respect to North (Figure 5.4)
- the perpendicular direction to the nearest road (Figure 5.5)
- the traffic volume on the nearest road (Figure 5.6)

Distances from the nearest road were then reclassified into six zones, with the central value of each zone corresponding to the location of the receptors (section 5.1.1). The orientation of the nearest road and the six wind direction classes were used to create grids of relative road orientation, where relative orientation is the difference between the orientation of the nearest road and the wind direction. The wind direction, relative road orientation and perpendicular direction to the nearest road were then used to identify active cells, i.e. those cells that will be in receipt of pollution for a particular wind direction.

The compiled program *weatherweight*, generated from the FORTRAN program *weather* (Appendix 6), written by the author, was then applied to the weather data. In the first instance, the program classifies the wind directions and wind speeds into 12 and 5 classes respectively, as shown in Tables 5.3 and 5.4. The atmospheric stability class is based upon the Pasquill stability scheme and the classes A to F are reclassed 1 to 6. The program then calculates the frequency of each unique combination of wind direction class, wind speed class and atmospheric stability class, referred to as a 'weather period' (Collins *et al.*, 1995).



## LEGEND

Distance



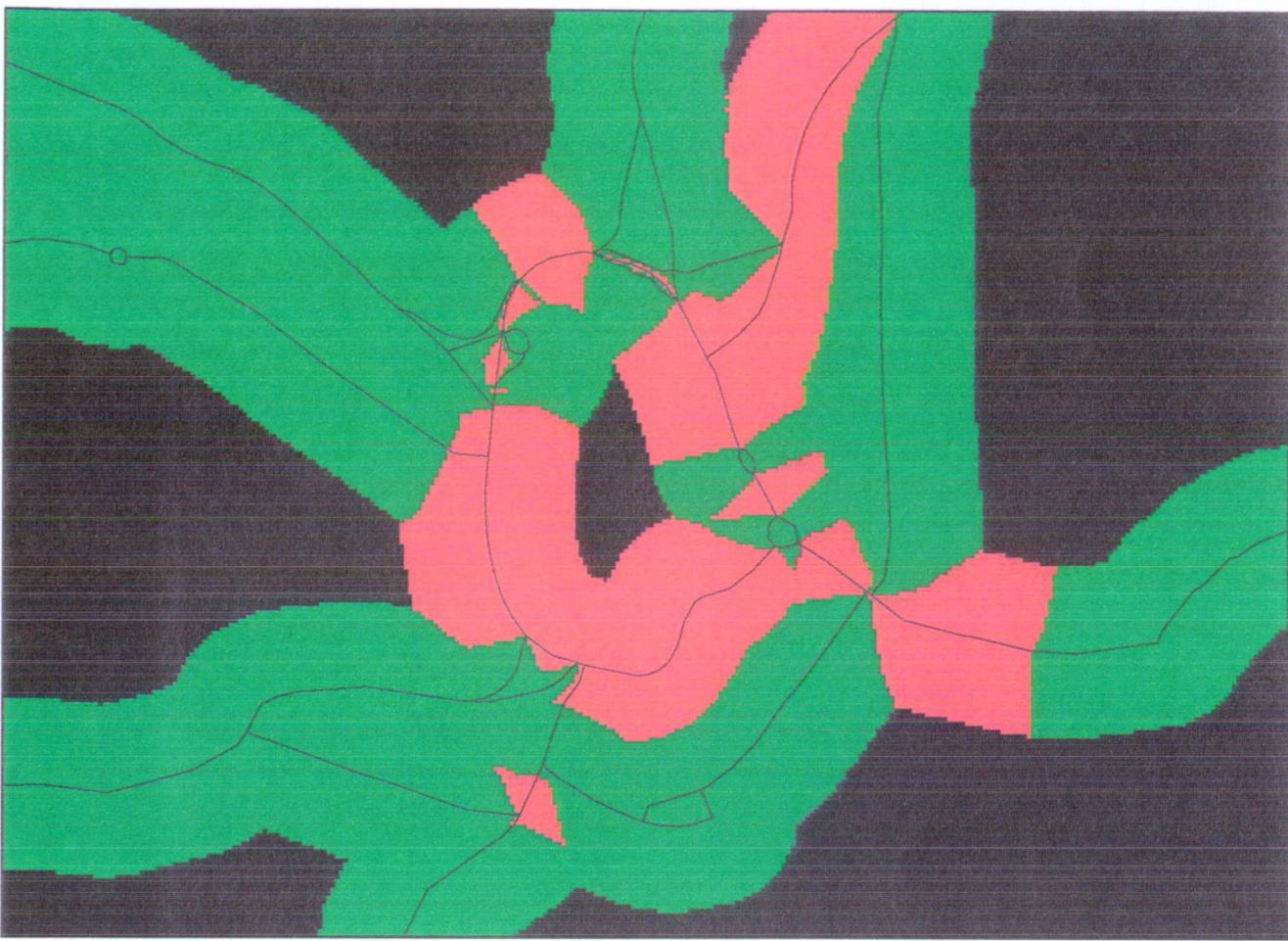
0 m

200 m



0.5 km

*Figure 5.2 Distance to the nearest road.*



## LEGEND

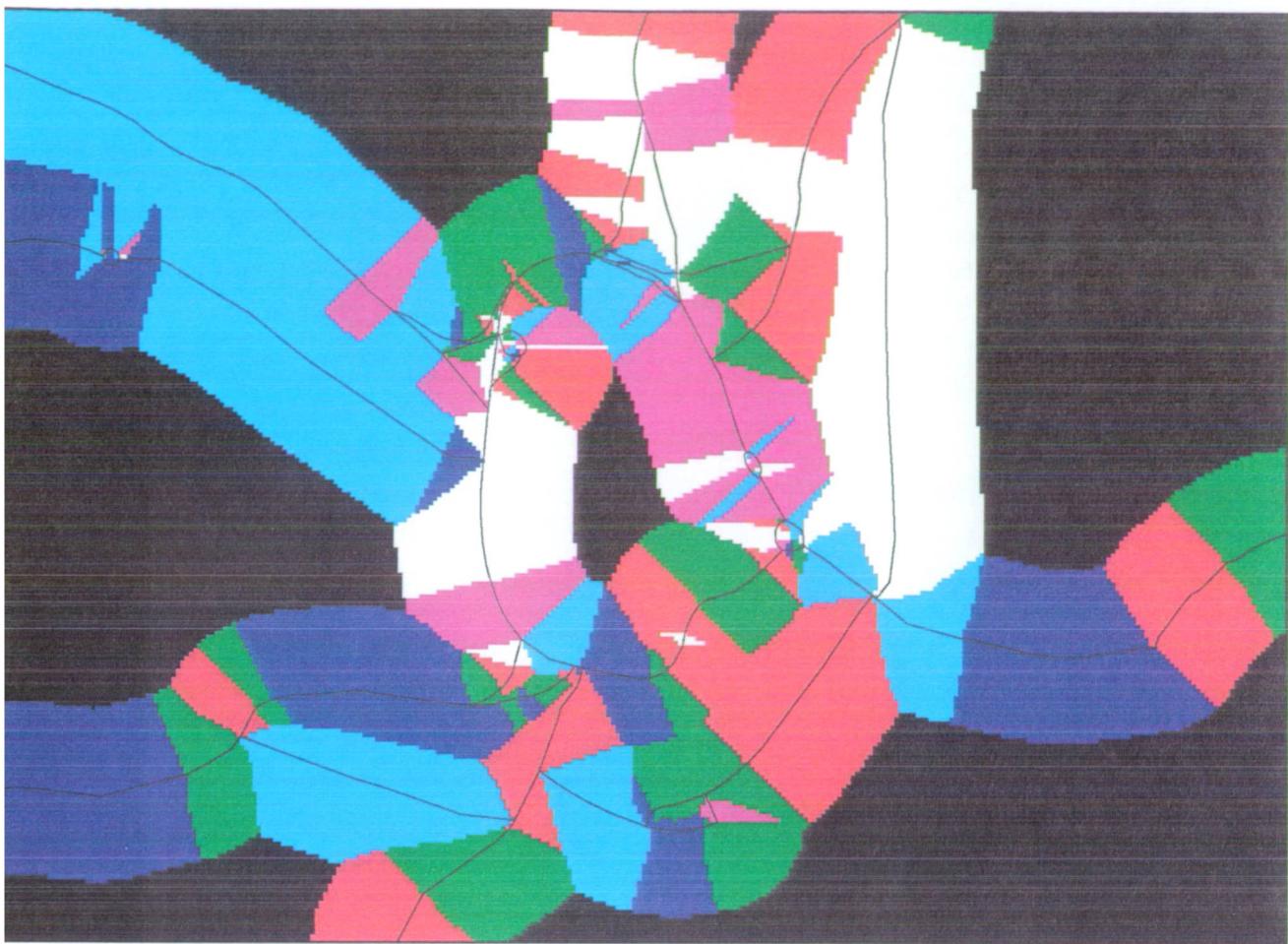
### Type of road

- Motorway
- Trunk road
- Main and secondary roads
- Other roads



0.5 km

*Figure 5.3 Value of the nearest road by type of road.*



## LEGEND

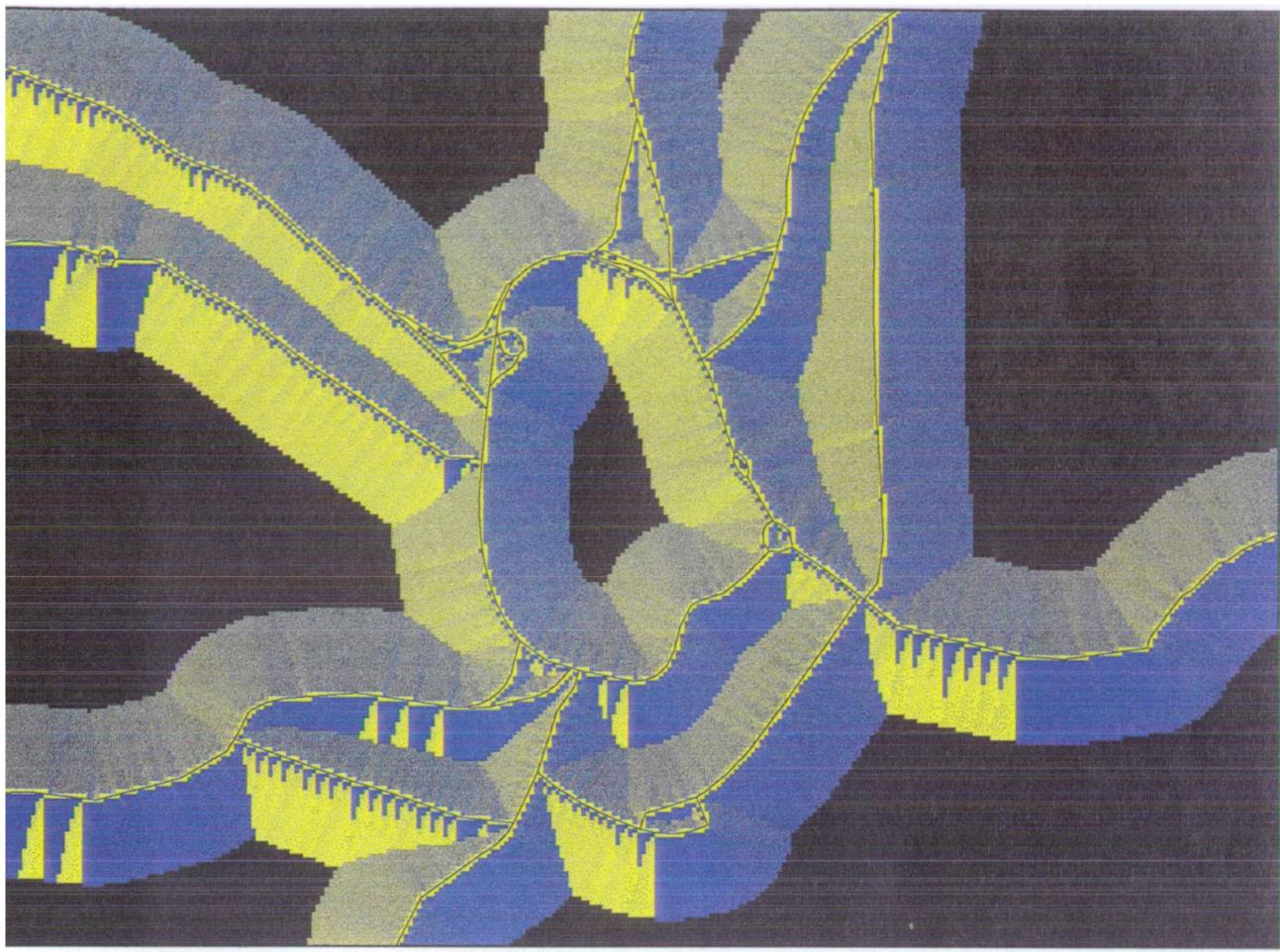
### Road orientation (degrees)

- 345 - 15 / 165 - 195
- 15 - 45 / 195 - 225
- 45 - 75 / 225 - 255
- 75 - 105 / 255 - 285
- 105 - 135 / 285 - 315
- 135 - 165 / 315 - 345



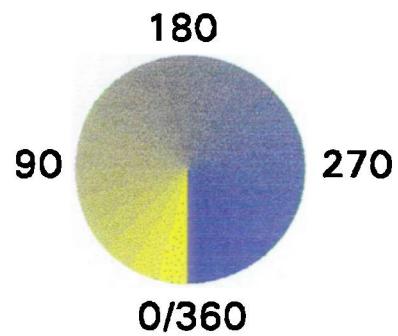
0.5 km

*Figure 5.4 Orientation of the nearest road with respect to North.*



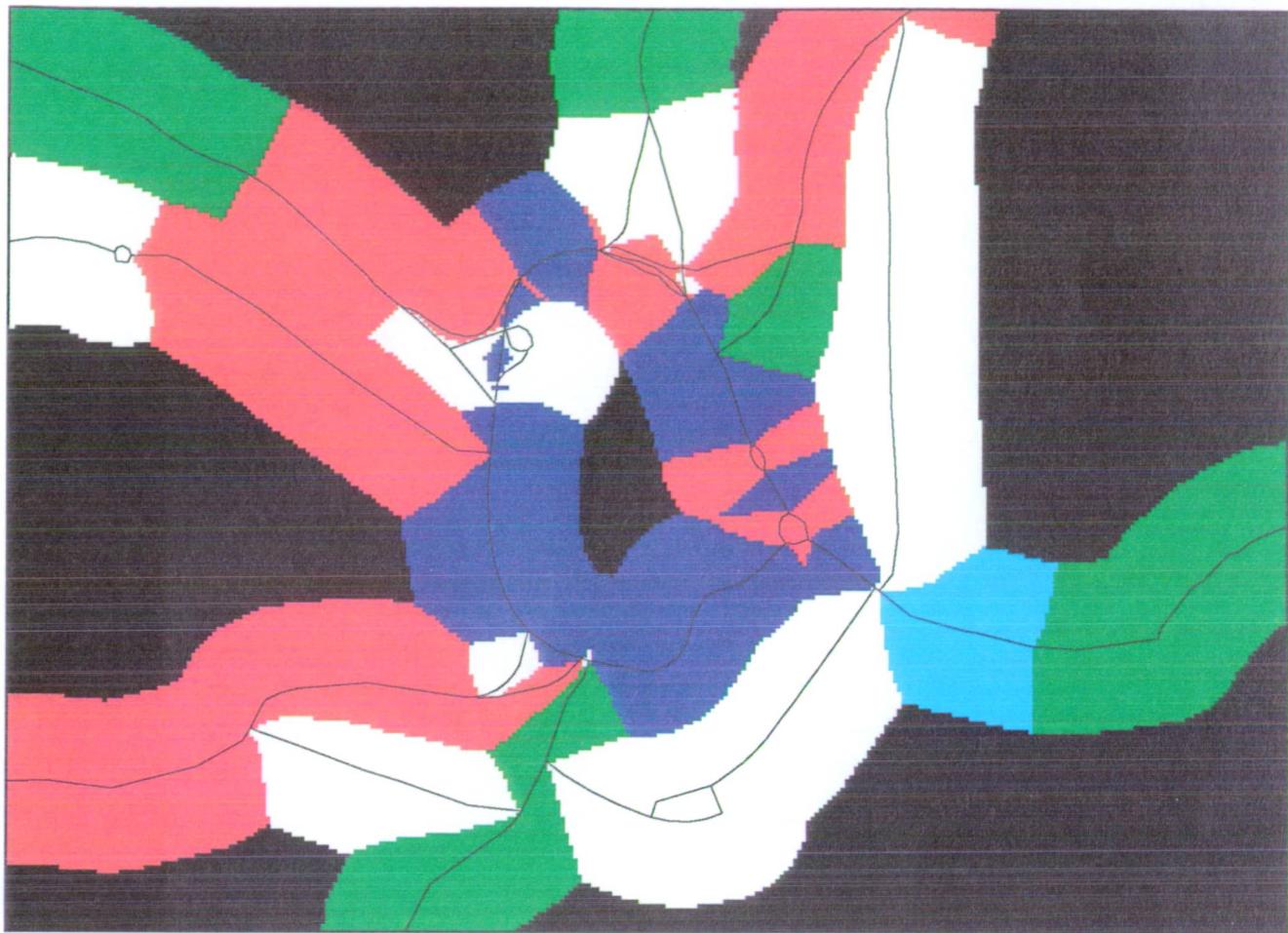
## LEGEND

Perpendicular direction (degrees)



0.5 km

*Figure 5.5 Perpendicular direction to the nearest road.*



## LEGEND

Day-time traffic volume (vehicles/hour)

- 500 - 1000
- 1000 - 1500
- 1500 - 2000
- 2000 - 2500
- 2500 - 3000



0.5 km

*Figure 5.6 Day time traffic volume on the nearest road.*

*Table 5.3 Wind direction classification scheme for the weather data.*

Code	Wind Direction (degrees)
11	345 - 15
12	15 - 45
13	45 - 75
14	75 - 105
15	105 - 135
16	135 - 165
21	165 - 195
22	195 - 225
23	225 - 255
24	255 - 285
25	285 - 315
26	315 - 345

*Table 5.4 Wind speed classification scheme for the weather data.*

Code	Wind Speed (m/s)
1	0 - 2
2	2 - 4
3	4 - 6
4	6 - 10
5	>10

Within the program, a new text file is generated containing the weather period and a weight for that weather period (based upon its frequency in relation to the total number of records in the original weather file). The file is then converted into an INFO data file, where each record contains a unique weather period and its respective weight. The *cursor* functionality in ARC/INFO is employed to select one record of the data file at a time. For each record the grid of active cells and the grid of relative road orientation, for that weather period, is selected.

The selected grids are used in combination with the grids of nearest road type, surface roughness, classified distance from roads, wind speed code and atmospheric stability class to give a seven digit code for each cell. The seven digit code is then used to extract the concentrations from the *concs.exp* table. The first two digits correspond to the receptor number (RP#) and the last five digits to the unique code CONCODE (see Table 5.2). The concentrations are then weighted for that weather period and the active cells updated to include the weighted concentration, using equation 5.1.

$$C_{ij} = \sum_{k=1}^n (c_{ijk} \cdot w_k \cdot v_{ij}) \quad [\text{Equation 5.1}]$$

where  $C_{ij}$  = the estimated concentration at cell location row  $i$  column  $j$

$c_{ijk}$  = the modelled concentration for weather period  $k$  at cell location row  $i$  column  $j$

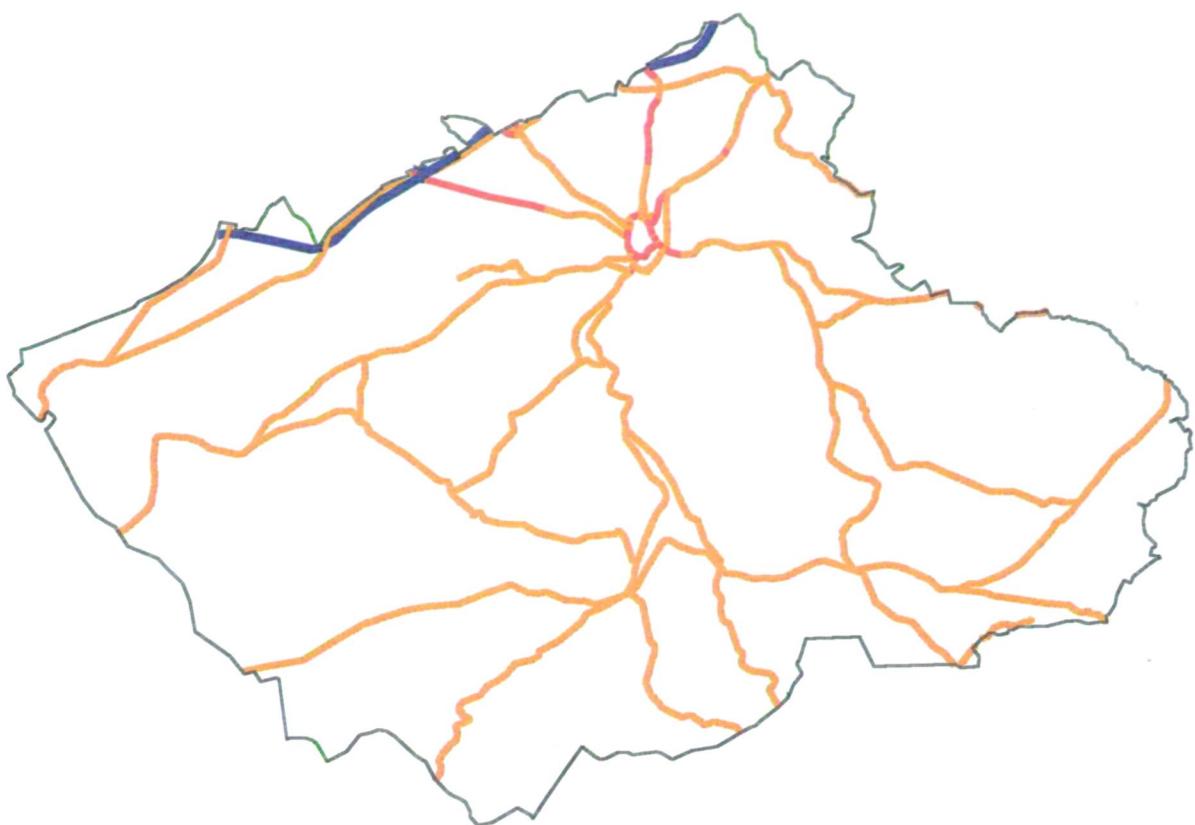
$w_k$  = the weighting factor for weather period  $k$

$v_{ij}$  = traffic volume at cell location row  $i$  column  $j$  (adjusted for the reference traffic volume)

The next record is then selected and the whole procedure is repeated. The new weighted concentrations are added to the previously calculated concentrations. The process is repeated until all the records have been selected. The final concentrations are then adjusted for hourly day-time traffic volume by dividing by 4000.

### 5.1.3 Applying the TRAFFPOL Routine to the SAVIAH Data

The *traffpol* command was applied to the SAVIAH data to estimate annual near-source pollution concentrations. Using the road network, described in section 3.3.3, major emission sources were identified (i.e. roads with day-time traffic flows > 250 vehicles/hour (Figure 5.7)) and classified according to road type, as shown in Table 5.5, using the simplified road widths and codes identified in Table 5.1. A new line coverage was created for the major emission sources (MAIN\_ROADS) with items for



## LEGEND

### Road type

-  Motorways
-  Trunk road
-  Main and secondary roads
-  Other roads



5 km

*Figure 5.7 The main emission sources.*

the classified road type (ROAD\_TYPE) and day-time hourly traffic volume (DH\_TVOL).

*Table 5.5 Road type classification scheme.*

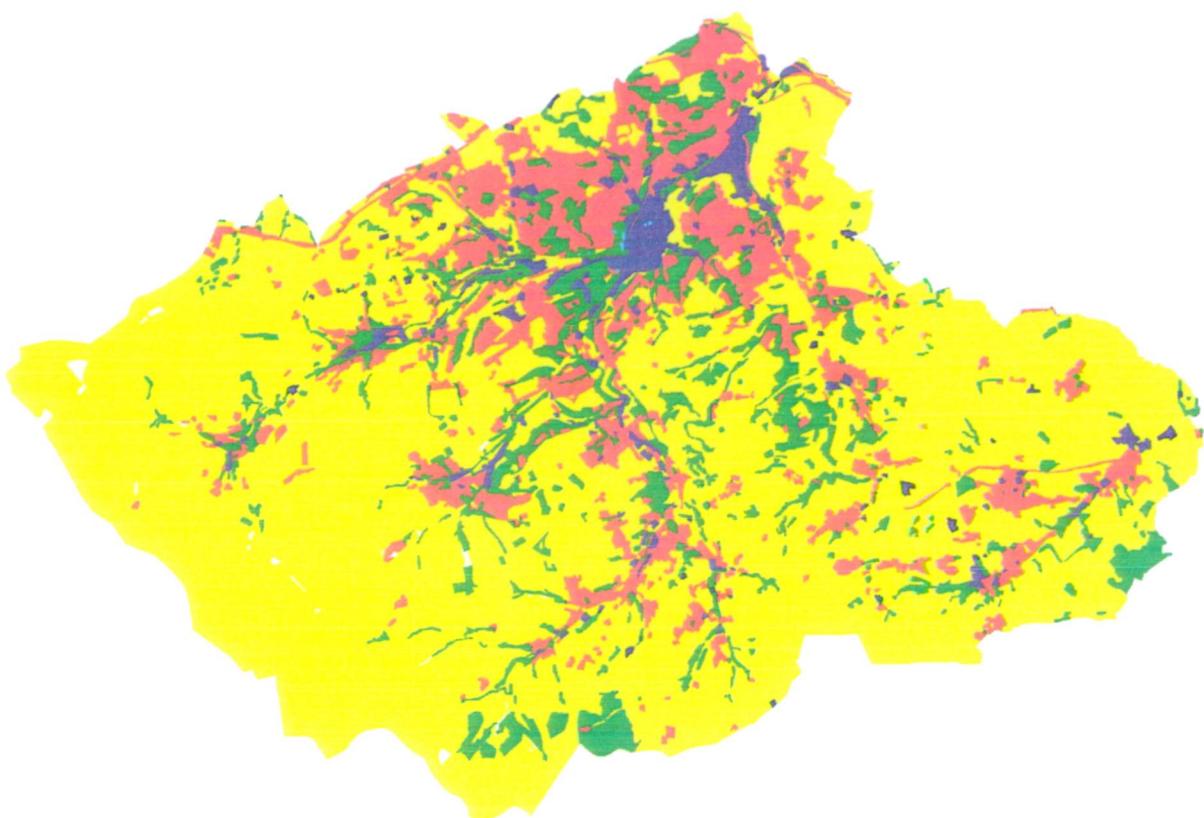
Code	Road Type
1	Motorway
2	Dual-carriageway
3	A and B Roads
4	Minor Roads

Using the land cover data, described in section 3.3.4, a surface roughness coverage (SURF\_ROUGH) was generated by applying the classification scheme outlined in Table 5.6; results can be seen in Figure 5.8. The land cover classification is based upon the different surface types and associated surface roughness (cm) presented in Benson (1979).

*Table 5.6 Surface roughness classification scheme.*

Code	Land Cover
1	Pasture, rough grass, reservoir, arable land, moor grass, peat sedge
2	Recreation, urban green space, very low and low density housing, quarries, disused and sequestered land
3	All tree types, public institutions, high density housing
4	Very high density housing
5	High density commercial, industry

Data on meteorology were obtained from the Meteorological Office at Leeds. The data were collected on an hourly basis for the duration of the study period, from June 1993 to June 1994. Daily averages for wind direction, wind speed and atmospheric stability were then calculated and written to an ASCII text file (an\_weather.txt). Thus,



## LEGEND

### Surface roughness class

- █ 1
- █ 2
- █ 3
- █ 4
- █ 5



5 km

*Figure 5.8 The surface roughness class.*

each record in the text file corresponded to one day of the study period. Three columns of data represented wind direction, wind speed and atmospheric stability.

The *traffpol* routine was run for the Huddersfield study area with the following command line:

```
Arc: &run traffpol main_roads near_source road_type dh_tvol surf_rough  
an_weather.txt
```

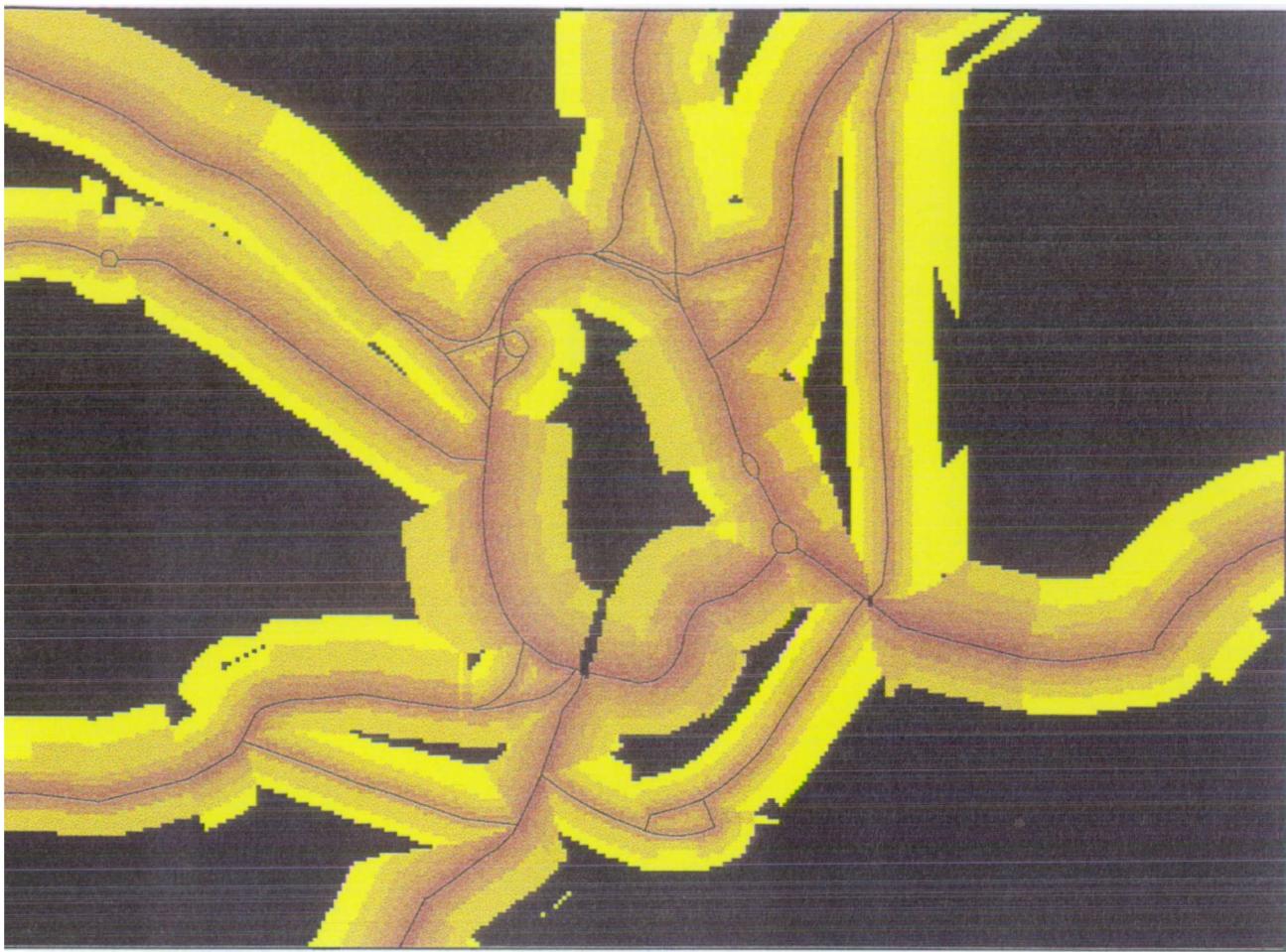
The resultant 10m grid (NEAR\_SOURCE) contains pollution concentrations for those cells within the 200m near-source band, Figure 5.9. The resultant grid was then added to the kriged background pollution.

## 5.2 Background Pollution

The background pollution surface was generated by applying ordinary kriging in ARC/INFO to the adjusted annual mean for the permanent monitoring sites that fall in the background areas (i.e. beyond the 200m near-source band). In the study, 24 of the 80 permanent sites were found to fall in these areas (Figure 5.10). A number of variograms were then created using different models and lags (see Section 4.1.2). The set of parameters which visually provided the best-fit and with the least variance were selected. This was found to be a spherical model with the following parameters:

Nugget	0.00
Range	4442
Sill	26.64
Lag	575

The kriged map and the map of variance can be seen in Figures 5.11 and 5.12. The near-source component was then added to the kriged map to produce a final pollution



#### LEGEND

$\mu\text{g}/\text{m}^3$



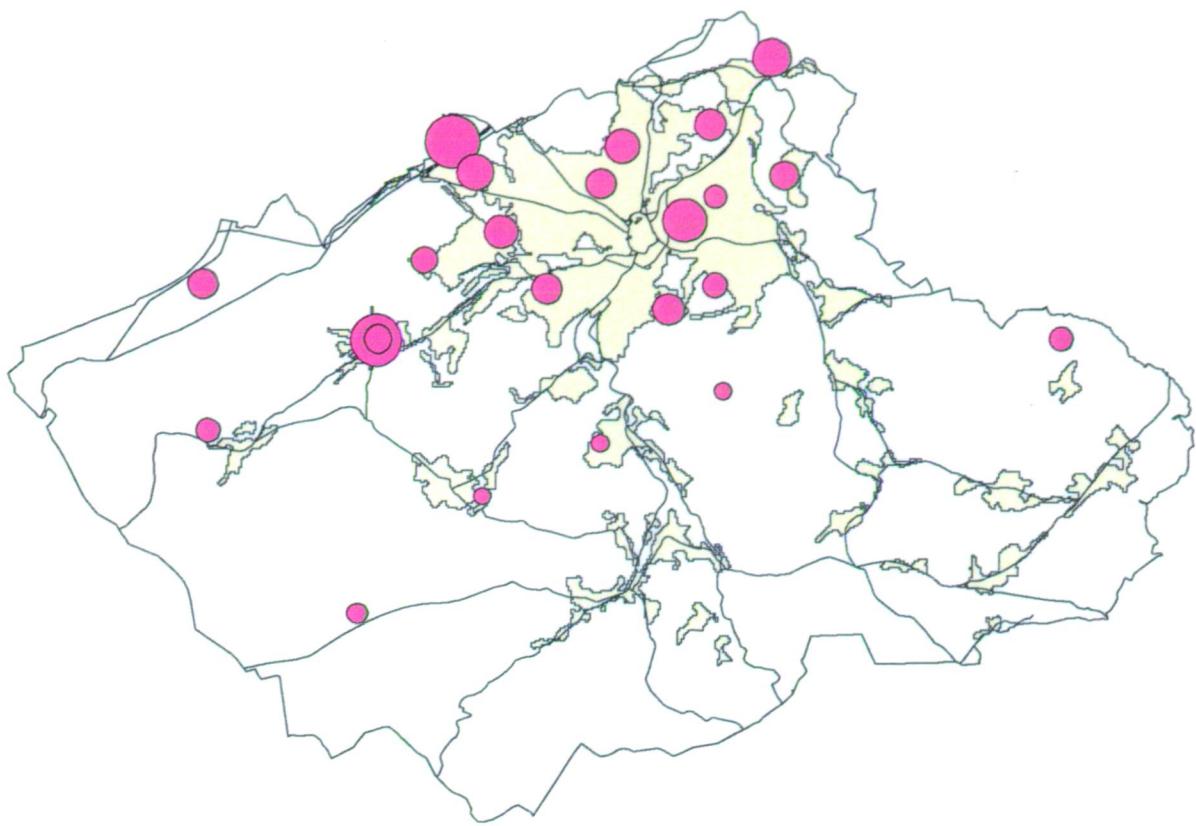
1

97



0.5 km

*Figure 5.9 Near-source  $\text{NO}_2$  concentrations.*



## LEGEND

Urban areas

Main roads

$\mu\text{g}/\text{m}^3$

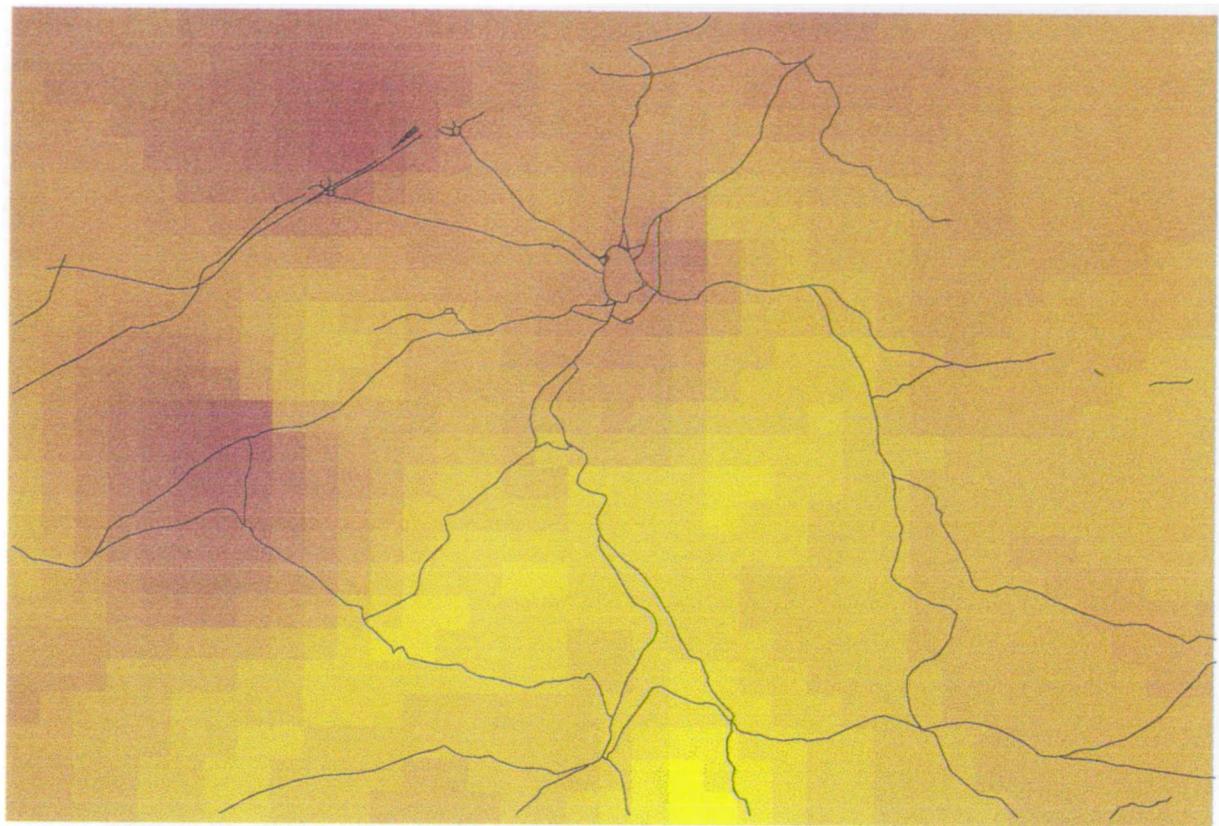
• 17.8

42.0



5 km

*Figure 5.10 Annual mean NO<sub>2</sub> at the background monitoring sites.*



## LEGEND

ug/m<sup>3</sup>



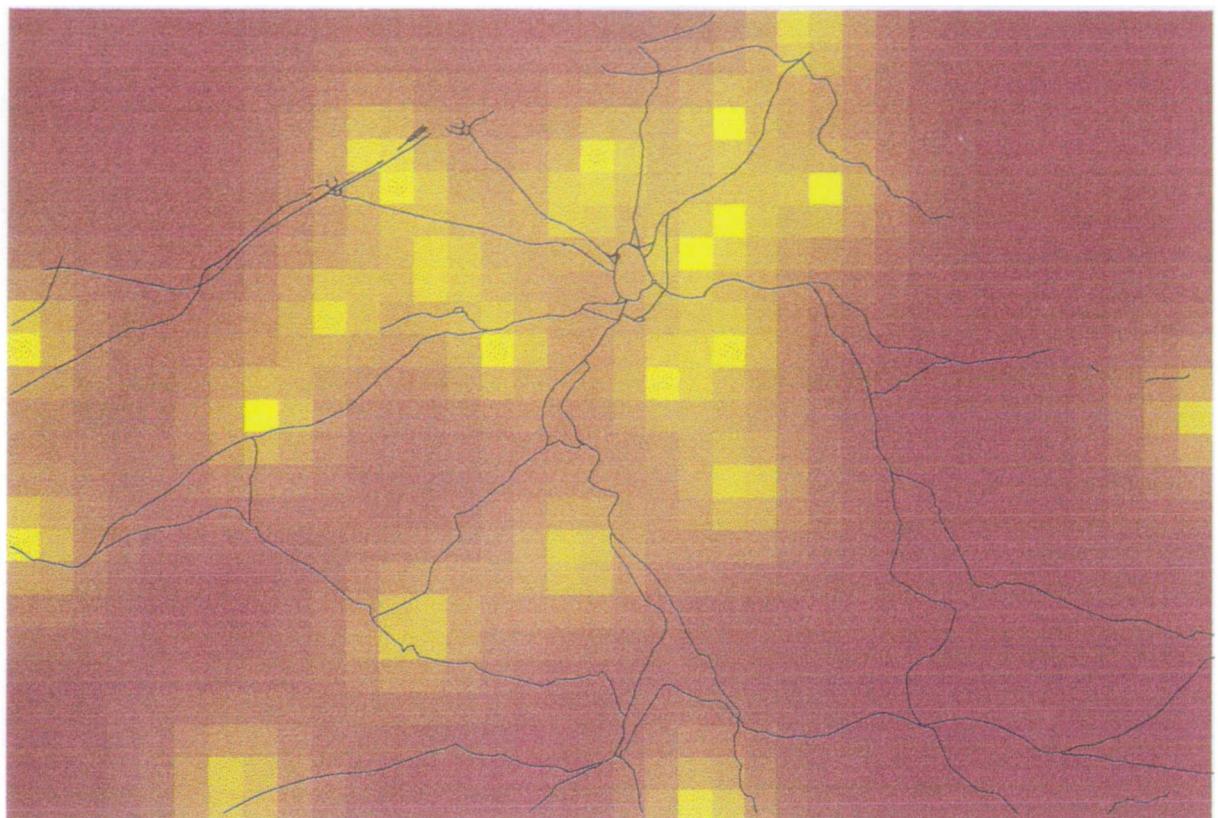
21

42



5 km

. *Figure 5.11 Map of background NO<sub>2</sub> by kriging.*



## LEGEND

### Variance



1

30



5 km

*Figure 5.12 Variance for the kriged estimates of NO<sub>2</sub>.*

map. The pollution map for the Huddersfield study area can be seen in Figure 5.13. The structure of the AMTRAP model is shown in Figure 5.14.

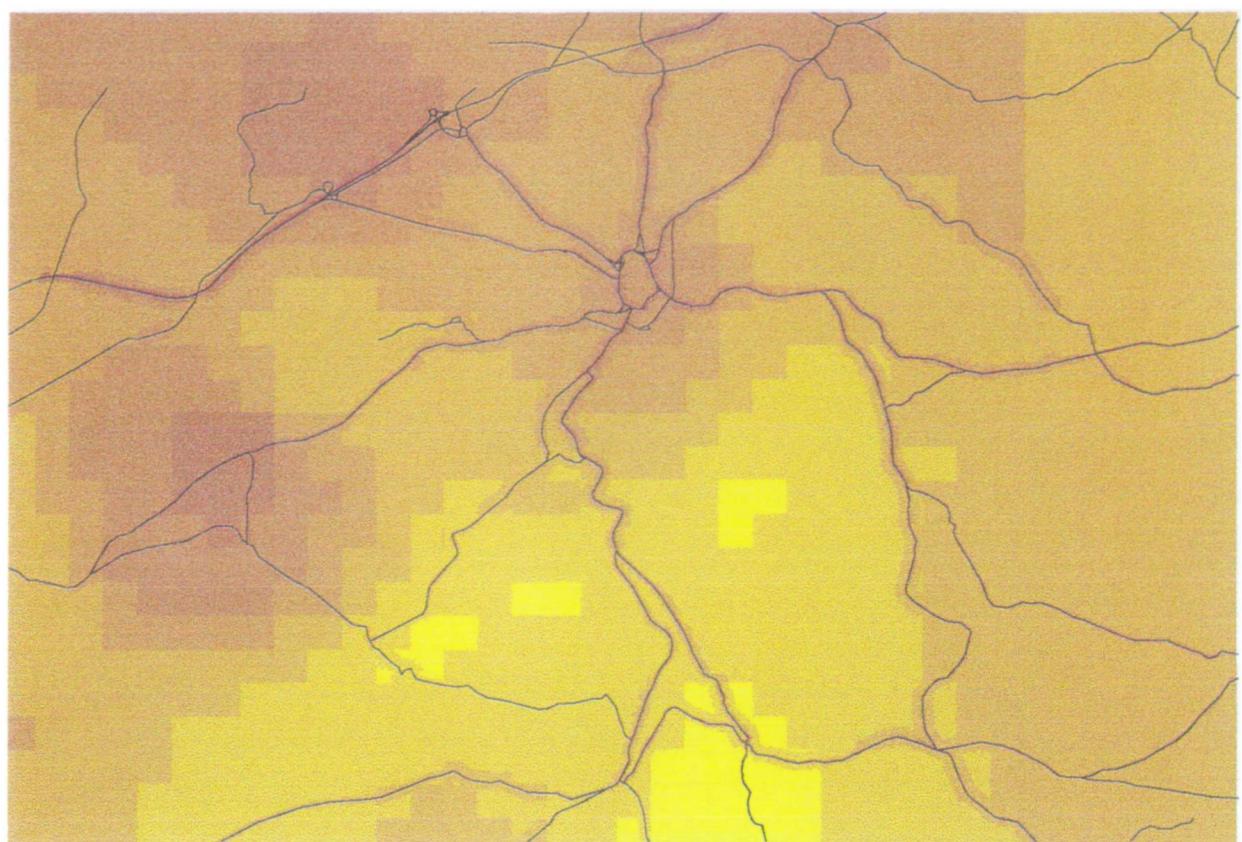
### 5.3 Validation of the Method

The final pollution map represents predictions of pollution concentrations at unsampled locations. How accurately the hybrid approach predicts pollution concentrations in these locations was tested by comparing two sets of monitored data - which had not been used to develop the model - with predicted values. In the first instance, predictions were derived for the 8 consecutive monitoring sites (section 3.3.1). The annual mean measured values and the predicted values for the 8 sites can be seen in Table 5.7 and the scatter plot in Figure 5.15. The adjusted  $r^2$  value between the annual mean and the AMTRAP predictions was found to be 0.628 ( $se = 5.25 \text{ ug/m}^3$ ). The graph shows that there is a good linear relationship between the measured and predicted values; however, the site with the highest prediction does appear to be an outlier, overestimating  $\text{NO}_2$ , influencing the slope of the regression line.

*Table 5.7 Annual mean  $\text{NO}_2$  values and AMTRAP predictions for the 8 consecutive monitoring sites.*

Site-id	Annual Mean $\text{NO}_2$ ( $\text{ug/m}^3$ )	AMTRAP Prediction	Difference
103	31.3	23.4	7.9
104	28.6	26.1	2.5
105	34.3	27.5	6.8
106	19.9	20.8	-0.9
107	21.7	25.1	-3.4
108	44.6	58.4	-13.8
109	31.0	31.8	-0.8
110	41.5	37.0	4.5

Predictions were also derived for the 56 monitoring sites in the near-source areas (which again had not been used to develop the model). Out of the 56 near-sources



#### LEGEND

ug/m<sup>3</sup>



18

125



5 km

Figure 5.13 Map of annual mean NO<sub>2</sub> by the AMTRAP method.

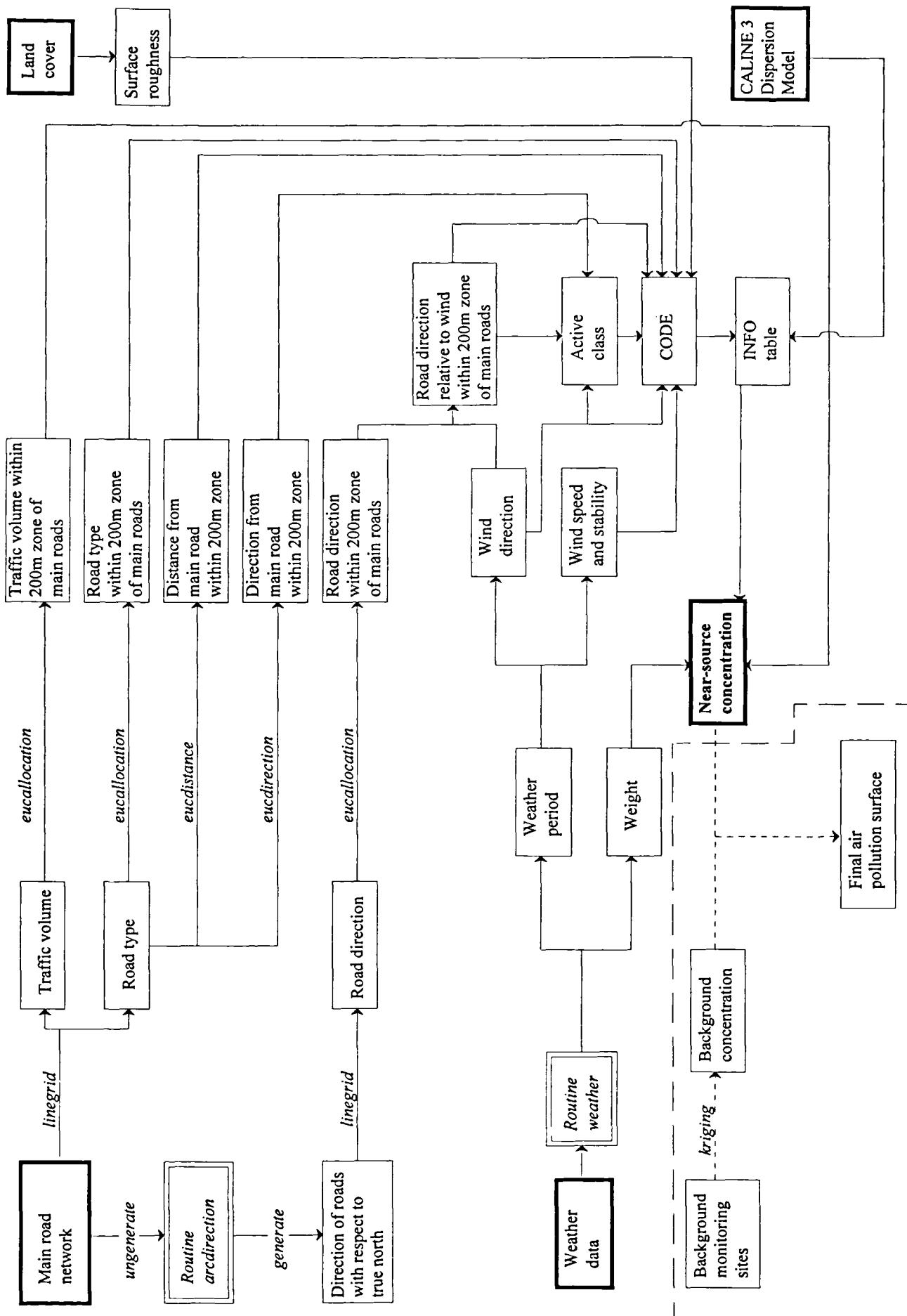
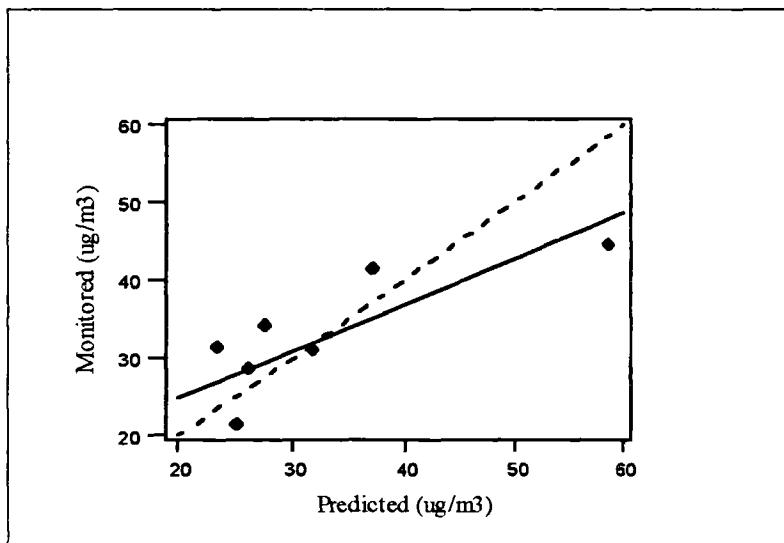


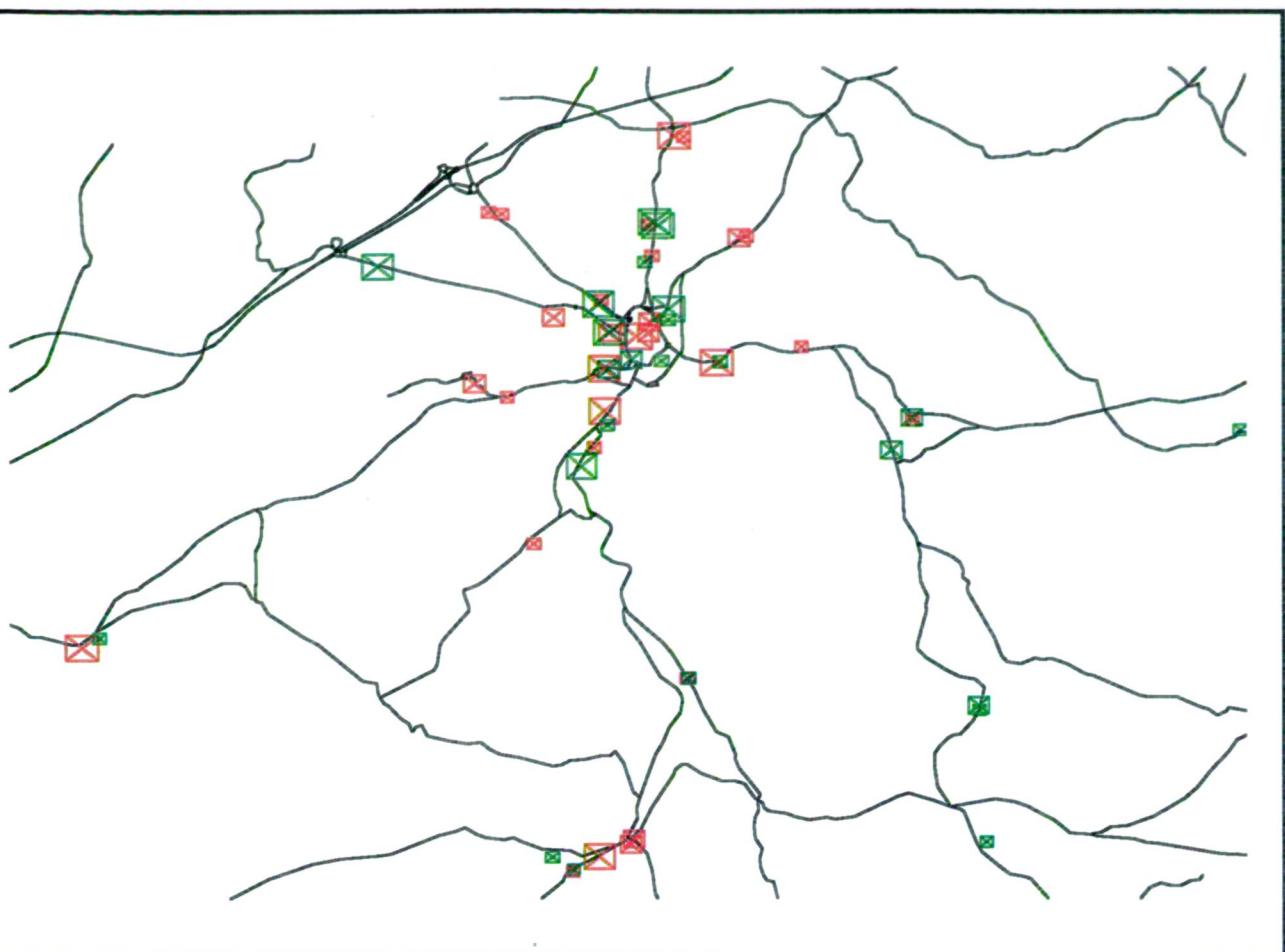
Figure 5.14 The structure of the AMTRAP model.

sites, 2 fell beyond the extents of the kriged map and were therefore not used in the evaluation. Monitored data were compared with the modelled estimates at 54 sites and the adjusted  $r^2$  was found to be 0.309 ( $se = 7.59 \text{ ug/m}^3$ ). There is a weak linear correlation between the predicted values and the monitored values, as Figure 5.16 demonstrates.



*Figure 5.15 Monitored annual mean  $\text{NO}_2$  against AMTRAP predictions for the 8 consecutive monitoring sites.*

The residuals between the measured annual means and the values predicted using the full set of meteorology at the 54 sites were calculated and mapped (Figure 5.17). Positive residuals indicate under-estimates of the pollution concentration and negative residuals indicate over-estimates. The residuals suggest that the model did not accurately predict pollution in areas where there were multiple sources, for example, at road junctions or where roads ran parallel and very close to each other. This reflects the structure of the model: whereas pollution levels are a function of all the emission sources in the near vicinity, the AMTRAP model is only based upon the characteristics of the nearest road.

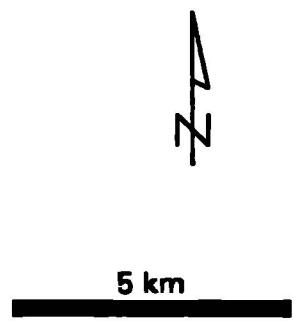


## LEGEND

■ Main roads

Residuals

- > +10
- +6 to +10
- +1 to +5
- < -10
- -6 to -10
- -1 to -5



. *Figure 5.17 Distribution of the residuals at the 54 near-source monitoring sites.*

The problem of multiple sources would be difficult to implement within the AMTRAP model. The AMTRAP model was not, therefore, developed further to overcome this problem; this is discussed more fully in Chapter 8.

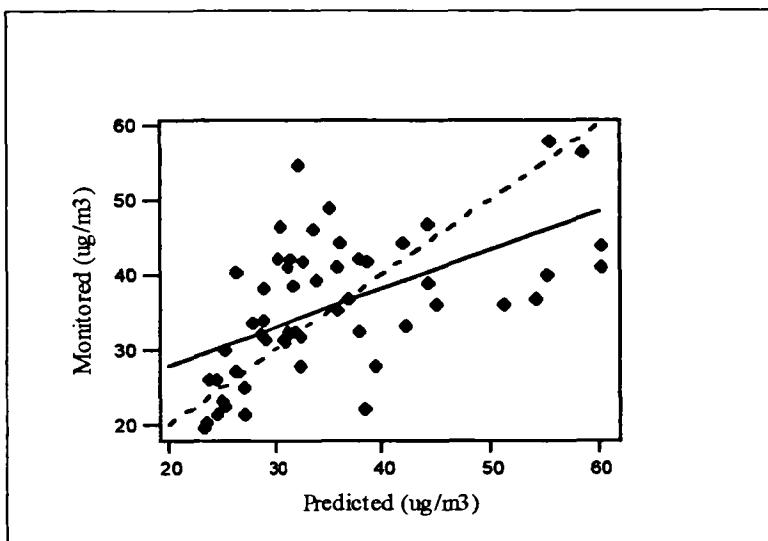


Figure 5.16 Monitored annual mean NO<sub>2</sub> against AMTRAP predictions at the 54 near-source monitoring sites.

#### 5.4 The Application of the Meteorological Data

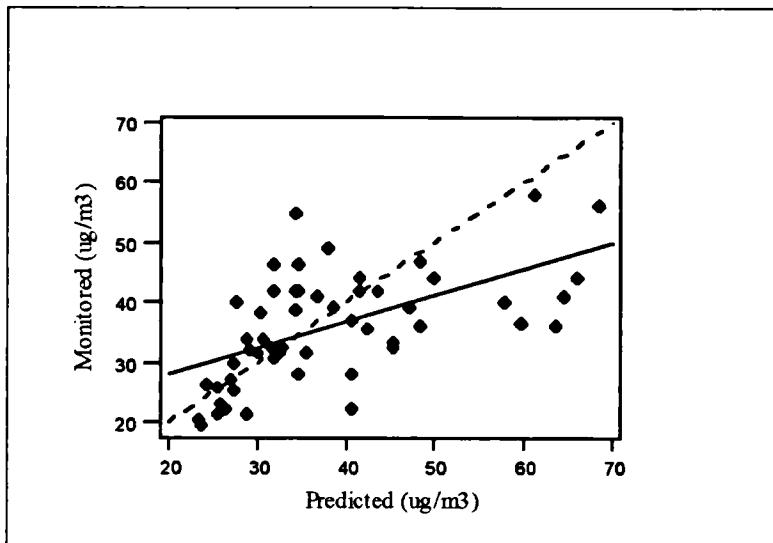
The AMTRAP approach is quite data intensive and computationally expensive, primarily due to the extensive amount of meteorology data used in the near-source component of the model. In long-term (annual) studies, this degree of complexity may, in fact, be unnecessary; it is possible that the many different weather conditions that occur over the course of a year effectively cancel each other out, making the use of short-term (e.g. daily) meteorological data redundant. Similarly, the effect of roadside buildings may have only a marginal effect in relation to estimates of mean annual concentrations. Eerens *et al* (1993) quote Tonkelaar and Hout (1980) as saying:

'although short-term (hourly) concentration values are strongly affected by the buildings in the streets, the long-term (yearly) average concentration pattern is much less sensitive to these than it is often assumed. Further, it was found that the ratio between the annual average values and the high percentiles of the frequency distribution of concentrations did not vary much from street to street. This is mainly due to the fairly constant diurnal emission patterns and the linear source configuration, which strongly reduce the variability of the concentration as a function of wind direction'.

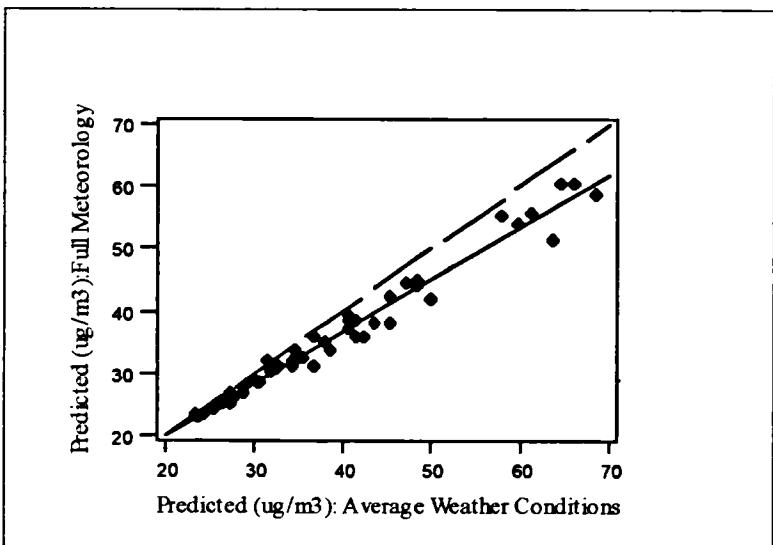
Over long periods of time, therefore, the effects due to meteorology and local buildings may become negligible. To test this assumption, the model was then run with an annual average wind rose for wind direction and annual mean values for wind speed and atmospheric stability. Predictions were derived for the 54 near-source monitoring sites. The monitored data were compared with the predicted values and the adjusted  $r^2$  was found to be 0.322 ( $se = 7.52 \text{ ug/m}^3$ ). The scatter plot of monitored data against predicted values (with average meteorology conditions) can be seen in Figure 5.18. When average conditions are applied, the model appears to provide similar predictions to the full meteorology model. The only noticeable difference, is that the average conditions model overestimates high concentrations to a greater extent than the full meteorology model.

The predictions with mean annual meteorology data were then compared with predictions at the same sites using the full set of meteorological data (section 5.3). Figure 5.19 shows that there is a very good correlation between the predicted values with full meteorology and the predicted values with average meteorology. The adjusted  $r^2$  value was found to be 0.975 ( $se = 1.59 \text{ ug/m}^3$ ).

These results suggest that, when measuring the long-term (annual) pollution concentrations, additional data relating to meteorology do not significantly improve the accuracy of the model. Furthermore, using annual mean meteorological data makes the application of the model computationally more efficient and less data intensive, while still providing the same degree of accuracy.



*Figure 5.18 Monitored annual mean NO<sub>2</sub> against AMTRAP predictions (with average meteorology conditions) at the 54 near-source monitoring sites.*



*Figure 5.19 Predicted values with full meteorology against predicted values with average meteorology conditions.*

## **5.5 Applying the AMTRAP Model to Traffic Related Pollution**

The AMTRAP model has shown how it is possible to link closely line dispersion models to GIS and that the result of the modelling is moderately good. The performance of the model, and how it compares to the other approaches, is discussed in detail in Chapter 8.

## CHAPTER 6 THE MOVING WINDOW APPROACH

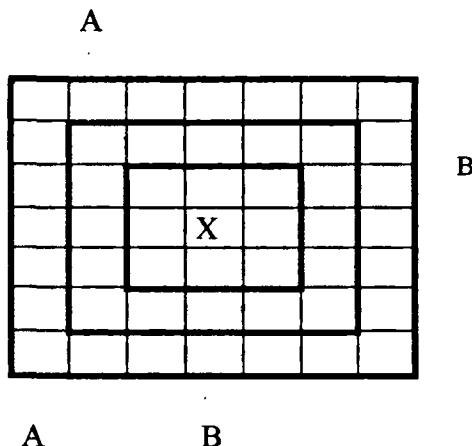
Moving windows is a technique that has not been widely used for air pollution mapping, but can effectively be applied within a GIS environment using techniques readily available for analysing spatial data. The moving window is applied to data in a regular grid on a cell by cell basis. As discussed in section 2.5, at every location in the grid a value is computed based upon the values in neighbouring cells. The number of neighbouring cells included in the calculation is defined by the size of the window, which equates to an area or zone surrounding the central cell. The computed value at the centre of the window is therefore a function of the local neighbourhood.

The moving window approach developed here uses a combination of spatial analysis techniques and spatial interpolation. The spatial analysis component of the approach (i.e. the moving window) is applied to predict near-source pollution at any location. Spatial interpolation (i.e. kriging) is then applied to pollution data measured at background pollution sites. As with the AMTRAP approach, the two separate components are added together to produce the final pollution map.

The approach is based upon the principle that pollution at any location is a function of all the emission sources in the near vicinity and that nearer sources contribute more pollution to a location than distant sources. The approach therefore assumes that the effect of traffic volume on a location is a function of the distance to roads and traffic volume on those roads. For example, roads the same distance away and carrying the same volume of traffic will have equal effect on a location, a road the same distance away with greater traffic volume will have more of an effect and a road further away with the same traffic volume will have less of an effect.

Figure 6.1 is a simple example. The central location of a sample grid is marked with an X and two roads (A-A and B-B) are indicated close by. Assuming that both roads

carry the same traffic volume, then the section of road B that falls in the first ring of cells around X will have a greater effect on X than the sections of road B and A that fall in the second ring of cells, and so on. If the decay of concentration away from the source, over distance, is known, then a weight can be applied to each ring of cells to calculate pollution at X.



*Figure 6.1. An example of a grid showing the location of two roads, A and B.*

### 6.1. Moving Windows

Using moving windows involves passing a template (or kernel) over a grid, one cell at a time. Instead of a calculation being applied to the whole grid - for example, the sum or mean of all cells in the grid - the calculation is applied only to those values that fall under the kernel. The result is calculated for the central cell of the template and placed in a new grid in the location of the central cell. A new grid is created so that the values in the original grid are not altered as the window passes over the data. Since the result is calculated for the central cell, templates tend to have an odd number of cells - for example, 3x3 window, 5x5 window or 9x9 window - to include the surrounding ring(s) of cells and the middle cell.

In Figure 6.2 a 3x3 window is passed over a grid. In Figure 6.2.a the mean of all 9 values under the template is calculated and the result placed in a new grid in a cell

location equivalent to the location of the central cell (Figure 6.2.c). The window then moves to the next cell (Figure 6.2.b) and calculates the mean again and places the result in the central value of the template in the new grid (Figure 6.2.d). Once the window reaches the end of the first row, it moves down to the start of the next row and continues to calculate the mean.

4	2	1	1	3	5	7
4	3	2	1	4	5	4
3	1	2	3	4	4	5
4	2	3	3	5	6	6
2	2	4	5	4	7	6

a)

4	2	1	1	3	5	7
4	3	2	1	4	5	4
3	1	2	3	4	4	5
4	2	3	3	5	6	6
2	2	4	5	4	7	6

b)

	2.4						

c)

	2.4	1.8					

d)

*Figure 6.2 A 3x3 window passes over a grid a) and b) calculating the mean value and placing the result in a new grid c) and d).*

As can be seen from Figure 6.2, however, one of the problems with moving windows is how to deal with the boundary problem (i.e. the outermost ring of the grid - or rings, depending on the size of window), where some of the values in the template fall outside the grid (Figure 6.3). One solution is to ignore these locations and only apply the template where the data is complete. In this case the output grid will be smaller than the input grid. Another solution is to adapt the calculation to the number of cells that contain values; for example, in Figure 6.3 the new value in the cell in the upper-left-hand corner will be the mean of 4 values.

4	2	1	1	3	5	7	
4	3	2	1	4	5	4	
3	1	2	3	4	4	5	
4	2	3	3	5	6	6	
2	2	4	5	4	7	6	

*Figure 6.3 Moving window in the outermost ring of the grid.*

In order to simulate the effects of distance from source on pollution levels, weights can be attached to the template and values in the grid multiplied by the weights in the template before the calculation is applied. Figure 6.4. shows a possible weighted template that could be applied to the data in Figure 6.2.a.

0.1	0.1	0.1
0.1	2.0	0.1
0.1	0.1	0.1

*Figure 6.4 An example of a weighted template.*

In Figure 6.4 the central value has been given a higher weight than the neighbouring values. Thus the central value has a greater influence over the sum total. The total for the output cell would therefore be:

$$(0.1 \times 4) + (0.1 \times 2) + (0.1 \times 1) + (0.1 \times 4) + (2.0 \times 3) + (0.1 \times 2) + (0.1 \times 3) + (0.1 \times 1) + (0.1 \times 2) = 7.9$$

## **6.2 Modelling Near-Source Pollution**

The contribution of any source to the pollution concentration at a location is a function of the decay of concentration over distance from source: where concentration declines in a curvilinear manner with increasing distance from the roadside. In order to model the contribution of different road sources to pollution levels at the centre of the template it is necessary to derive weights which represent this distance effect. Thus, if the weights in the template represent the decay of concentration away from the source over distance then the total pollution at a location can be calculated.

### **6.2.1 Defining the Air Pollution Template**

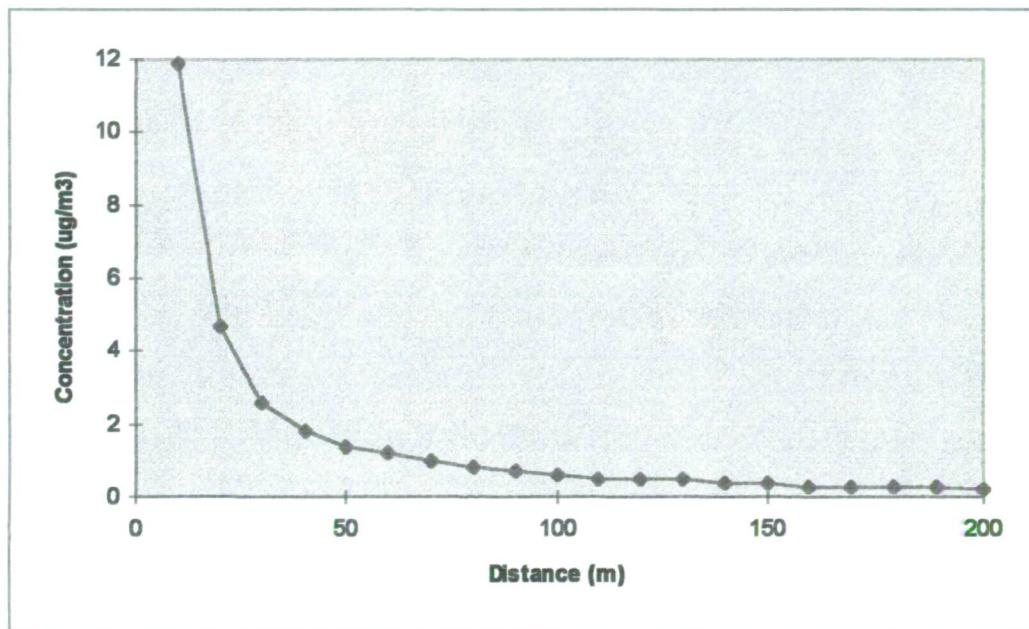
If near-source pollution is presumed to be a 200m band either side of the road, and the road network is represented as a 10m grid, then a 41 x 41 template (410m diameter) is necessary to reflect the near-source pollution.

The template consists of 41 rows and 41 columns of values and is stored as a text file. The values in the template are weights which are applied to the data in the grid. The weights are a function of the cell's distance to the central location of the template. The value of the weight is derived from the curve which is a function of the decay of concentration with distance from the source. The weights are therefore defined by  $f(D_t C_d)$ , where  $D_t$  is the distance from the centre of the template and  $C_d$  is the decay of concentration over  $D_t$ . The following steps were applied to derive the weights.

#### ***Step 1. Define the decay of concentration over distance curve.***

CALINE3 was run for eight wind directions at 45° intervals starting from 0°, with average wind speed and atmospheric stability for the study period. A constant traffic volume of 4000 vehicles/hour and emission rate of 1.62 g/km were entered into the

model. The co-ordinates for 20 receptors were also entered into the model. The receptors were positioned in the model at 10m intervals perpendicular to a 200m road link. Results for the receptors were produced for the eight wind directions and the average concentration was then calculated for each receptor. The average concentrations were then plotted on a graph against distance (Figure 6.5).



*Figure 6.5 The decay of concentration over distance.*

**Step 2. Establish weights for the template.**

The distances between the central point of the cells in the template and centre of the template was calculated. The distance is a simple multiple of 10 for adjacent cells, but more complex for non-adjacent cells, as the 5 x 5 template in Figure 6.6 demonstrates. The weights for the template were then interpolated from the concentration against distance graph (Figure 6.5) based upon the calculated distance between the centre of the cell and the centre of the template.

28.3m	22.4m	20m	22.4m	28.3m
22.4m	14.1m	10m	14.1m	22.4m
20m	10m		10m	20m
22.4m	14.1m	10m	14.1m	22.4m
28.3m	22.4m	20m	22.4m	28.3m

Figure 6.5 Distance between the central point of the cell and the centre of the template.

### Step 3. Create a text file of weights.

The weights were then entered into a text file with 41 columns and 41 rows (Figure 6.6). Although the text file consists of 41 columns and 41 rows, the template is in fact a circular template; since concentrations were only established for distances 200m or less, any cell with a distance between its centre and the centre of the template greater than 200m has a value of zero.

#### 6.2.2 Applying the Air Pollution Template

A 10m grid of the main emission sources (section 5.1.3) was created from the road network and hourly day-time traffic volumes assigned as the values of the cells. Cells containing data existed where roads existed and the value of each cell corresponded to the mean day-time-hourly traffic volume at that location. The template was then applied to the grid on a cell by cell basis with the command *focalsum* - with the WEIGHT option - in the GRID module of ARC/INFO. The syntax for the command is as follows:

```
ns_grid = focalsum (tvol_grid,WEIGHT,template.txt)
```

where `ns_grid` = the output grid of near-source pollution  
`tvol_grid` = the input grid of hourly day-time traffic volume for the main emission sources  
`template.txt` = text file containing the values for the weights

*Figure 6.6 The weighted template.*

At every location of the input grid the weights in the template are multiplied by the values in the grid. Where a road exists then the weight in the template is multiplied by

day-time traffic volume. Where a road does not exist, the value in the grid is zero and therefore the weighted value is zero. The sum of weighted values is calculated and the total placed in the output grid at the location of the central cell of the template. The contribution of air pollution to the central cell is therefore equal to the sum of all weighted values under the template. Figure 6.7 shows a sample grid and a 5 x 5 filter; traffic volumes on roads A and B are 50 and 25 day-time vehicles per hour respectively. The output value at location row 4, column 4, is therefore equal to 845, calculated as follows:

$$(1.6 * 50) + (2.0 * 25) + (3.8 * 25) + (2.0 * 50) + (6.2 * 25) + (2.4 * 50) + (3.8 * 25) + (2.0 * 50) + (2.0 * 25)$$

1					A			
2	1.6	2.0	2.4	2.0	1.6			
3	B	2.0	3.8	6.2	3.8	2.0		
4	2.4	6.2	6.2	6.2	2.4			
5	2.0	3.8	6.2	3.8	2.0	A	A	A
6	1.6	2.0	2.4	2.0	1.6			
7								
8								
	1	2	3	4	5	6	7	8

Figure 6.7 A sample grid and template.

The resultant concentrations in the output grid were then adjusted to account for the reference traffic volume of 4000 vehicles per hour. In addition, in CALINE3, concentrations at the individual receptors were calculated for a 200m road link. The moving window method, however, assumes a 10m road link; therefore, all the total concentrations were divided by 20 to provide a comparable value.

### **6.3 Background Pollution**

The kriged coverage created for the hybrid approach (section 5.2) was used for the background pollution. This was added to the near-source pollution to produce a final pollution map (Figure 6.8).

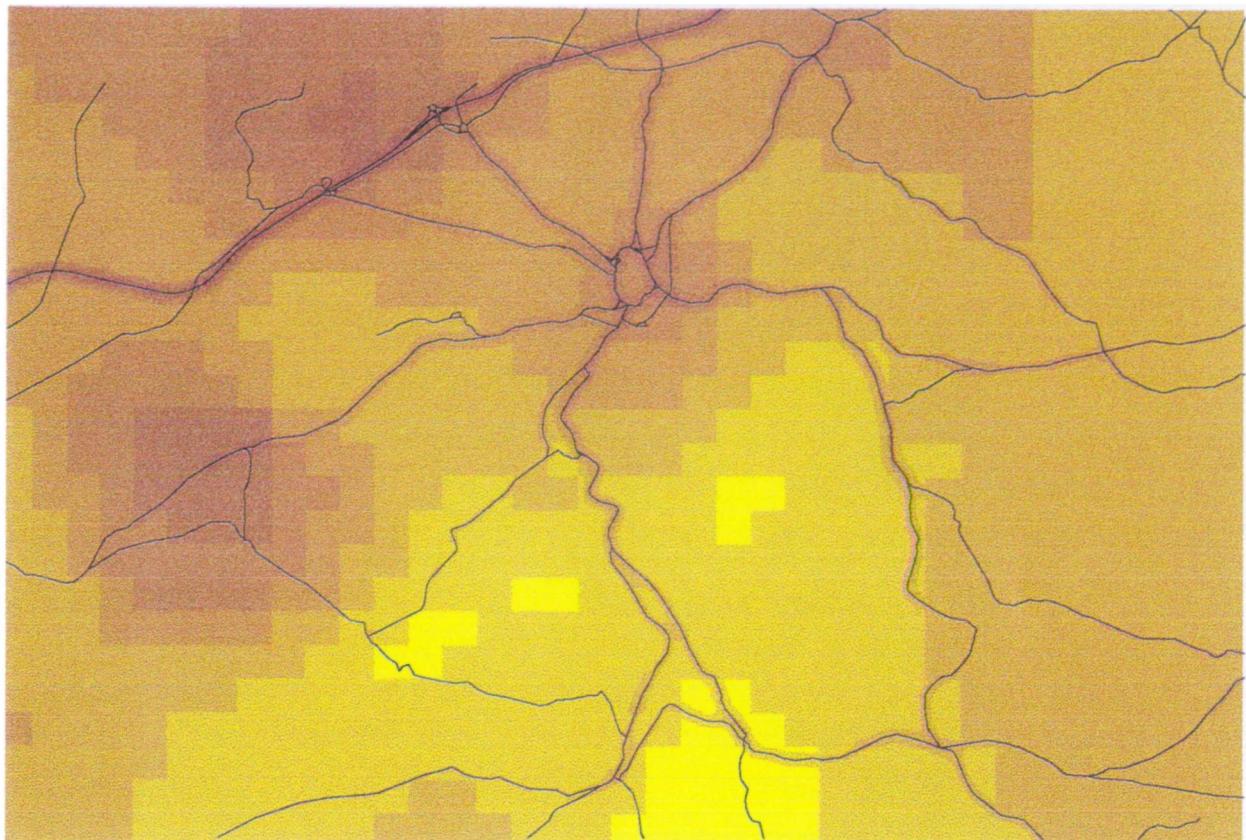
### **6.4 Validation of the Method**

The final pollution map represents predictions of pollution concentrations at unsampled locations. The accuracy of the moving window approach was examined by comparing two sets of predicted pollution concentrations in these locations with monitored data; concentrations at the 8 consecutive monitoring sites and at the 54 near-source sites.

The 8 consecutive monitoring sites (section 3.3.1) were used to test the accuracy of the model - these sites were not used to develop the model. The annual mean measured values and the predicted values for the 8 sites can be seen in Table 6.1. The moving window model appears to be systematically underestimating levels of pollution at these sites (Figure 6.9). The adjusted  $r^2$  value between the annual mean values and predictions from the moving window method was found to be 0.673 ( $se = 4.93 \text{ ug/m}^3$ )

*Table 6.1 Annual mean NO<sub>2</sub> values and filtering predictions for the 8 consecutive monitoring sites.*

Site-id	Annual Mean NO <sub>2</sub> (ug/m <sup>3</sup> )	Filtering Prediction	Difference
103	31.3	23.4	7.9
104	28.6	26.1	2.5
105	34.3	27.5	6.8
106	19.9	20.8	-0.9
107	21.7	25.1	-3.4
108	44.6	40.2	4.4
109	31.0	29.0	2.0
110	41.5	30.4	11.1



#### LEGEND

ug/m<sup>3</sup>



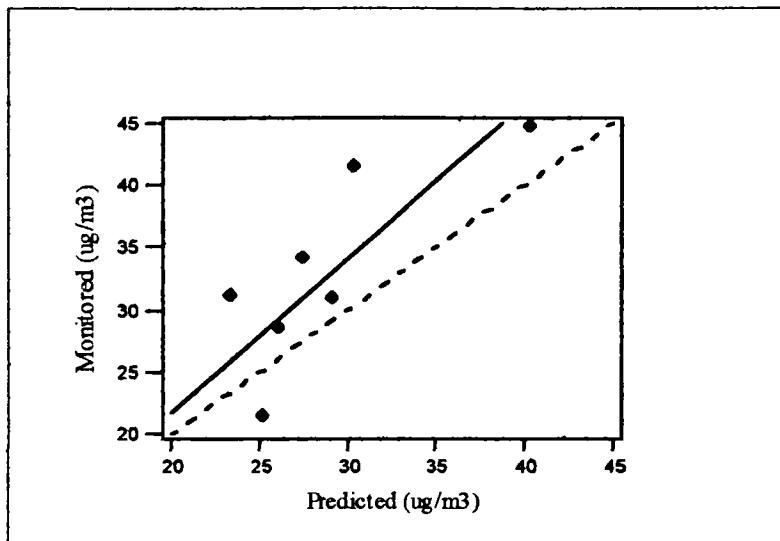
18

94



5 km

. Figure 6.8 Map of annual mean NO<sub>2</sub> by the moving window method.



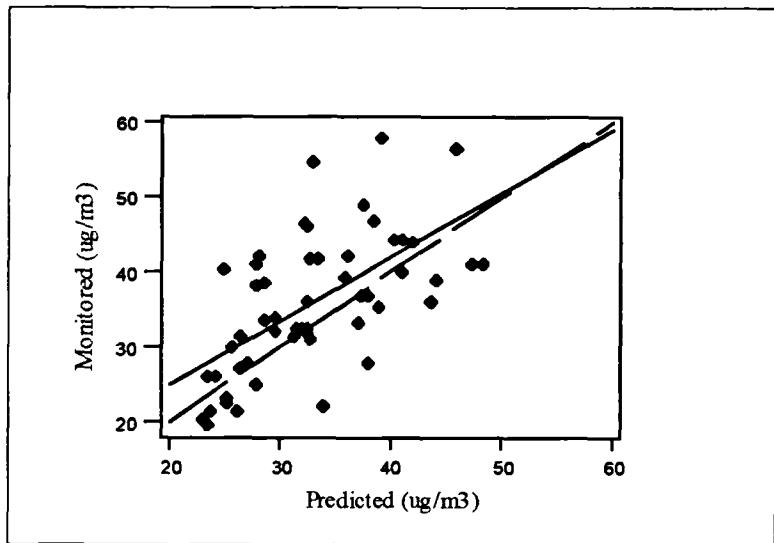
*Figure 6.9 Monitored annual mean NO<sub>2</sub> against moving window predictions for the 8 consecutive monitoring sites.*

Predictions from the moving window approach were also derived for the 54 near-source monitoring sites - not used to develop the model (described in section 5.3). The adjusted  $r^2$  value between the measured annual mean data and the moving window predictions at the 54 sites was found to be 0.360 ( $se = 7.23 \text{ ug/m}^3$ ). The measured annual means plotted against the moving window predictions are shown in Figure 6.10. The graph shows that predictions at some of the sites are outliers. These sites apart, there is a reasonably good linear relationship between the monitored and predicted values at the remaining sites.

## 6.5 Applying Moving Windows to Traffic Related Air Pollution

As this Chapter has demonstrated, it is possible to use a moving window to model traffic related pollution and it has worked reasonably well in relation to NO<sub>2</sub> concentrations in Huddersfield. The moving window has the capacity to model the contribution of pollution at a location from all sources in the immediate area. Modelling pollution as a function of the decay of concentration over distance from

source, using a readily available tool, has proved a simple and effective approach. How the moving window performs in comparison to the other models is covered in detail in Chapter 8.



*Figure 6.10 Monitored annual mean  $\text{NO}_2$  against moving window predictions at the 54 near-source monitoring sites.*

## **CHAPTER 7 THE REGRESSION APPROACH**

The final approach examined in this research was regression mapping. As outlined in section 2.4, this refers to the use of empirically derived multiple regression equations, describing the relationship between environmental factors and monitored air pollution levels, as a basis for mapping spatial patterns of pollution. It is a technique which has not been widely used in air pollution studies before, but which in many ways would seem well-suited as an application in a GIS environment for it draws upon the ability of GIS to predict conditions at each unsampled point by weighted combination of a series of map layers, each representing the independent variables in the regression model.

### **7.1 Identifying the Independent Variables**

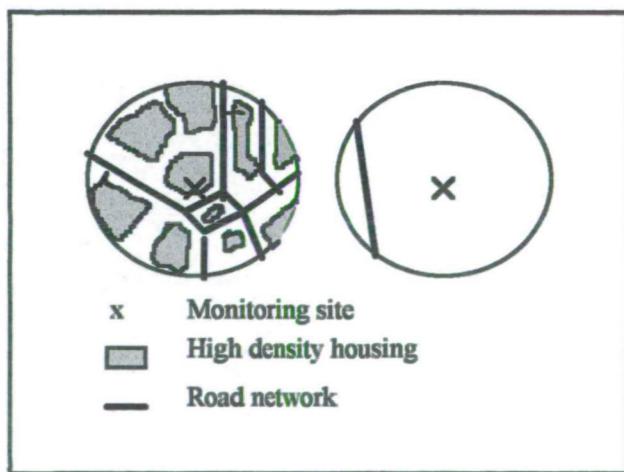
Regression mapping proceeds by first establishing a relationship between a range of predictor variables and the dependent variable of interest - in this case NO<sub>2</sub> concentration - then applying this model to unsampled locations. The choice of predictor variables for entry into the regression analysis is clearly crucial in this process, for this determines to a large extent the explanatory power of the regression model.

In the case of traffic related air pollution, three main sets of factors can readily be identified, which can be expected to account for much of the variation in air pollution levels:

- emissions from vehicles
- other diffuse sources (e.g. industrial or residential activities)
- dispersion patterns and rates (e.g. surface roughness, microclimate)

Within the GIS, the road network and associated traffic volume (section 3.3.3) could be used to characterise emissions from vehicles, land cover (section 3.3.4) to characterise other diffuse sources and land cover and altitude (section 3.3.5) to characterise dispersion patterns and rates.

As discussed in Chapter 6, any location across the study area is affected not only by emission sources and dispersion patterns at that specific location, but also by the emission sources and dispersion patterns in the surrounding area (Figure 7.1). Variables therefore need to be defined that are a measure of all significant contributing emission sources and dispersion patterns within the surrounding area. In this context, using the data stored in the GIS (road network, land cover and altitude) allows for potentially useful measures to be derived. In the case of traffic related pollution, potential indicators of vehicle emissions might include total length of road, total number of road segments or total traffic volume in the surrounding area. These local measures of vehicle emissions are likely to provide more reliable indicators than the simple one-dimensional indicators such as distance to the nearest road or traffic volume on the nearest road.



*Figure 7.1. Emission sources and dispersion patterns for two different monitoring sites.*

Although variables such as length of road, road density and traffic volume are all acceptable indicators of vehicle emissions, they are all very highly correlated. One of the variables - total traffic volume in km/hour - was therefore chosen here as the emission indicator. Traffic volume was felt to be the most reliable indicator of vehicle emissions, providing the best measure of vehicle density in urban environments.

Potential measures of emission sources and dispersion patterns can be derived by drawing a circular zone of influence, radius  $r$ , around the monitoring site and calculating variables as a function of the data within that zone. The radius can be set to reflect the area of influence around the site. However, as suggested in Chapter 6, nearby sources will tend to contribute more pollution at a location than distant sources. Similarly, nearby topographic conditions will have a greater effect on dispersion processes around a site than more distant conditions. Against this background, these independent variables were calculated for 20m bands (up to a maximum of 300m) around each monitoring site. Data from different bands were then combined, by weighted aggregation, to produce compound indicators, as will be demonstrated later in this Chapter. Least squares regression techniques were used to identify the best combination of band width and weights that explained the greatest degree of variation in NO<sub>2</sub>.

The independent variables were entered into a multiple regression analysis with the annual mean NO<sub>2</sub> as the dependent variable. The resultant regression equation was then used to estimate NO<sub>2</sub> at every location across the study area. The approach is described in further detail later in this Chapter.

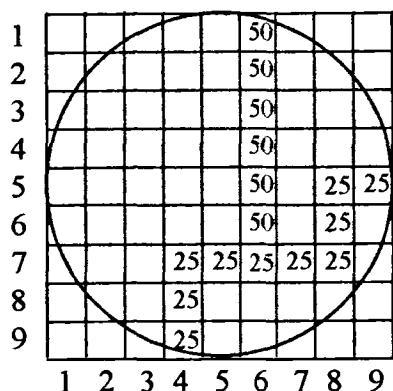
## **7.2 Calculating the Independent Variables**

### **7.2.1 Traffic Volume**

A grid, with a 10m resolution, was created for the road network with values representing day-time hourly traffic volume. A cell with a value greater than zero

indicates the presence of the road - the value of the cell reflects traffic volume on the road - and a cell with a value of zero indicates the absence of a road. In the GRID module of ARC/INFO, the command *focalsum* is applied to the traffic volume grid to calculate the total traffic volume within 15 zones of varying radii from 20m to 300m at 20m intervals (i.e. 0-20m, 0-40m, 0-60m etc.). A circular filter of radius  $r$  is passed over the grid, the total sum of all cells under the filter is calculated and the result placed in the location of the central cell. The command *latticespot* was then used to find the total traffic volumes in the different zones for all permanent monitoring sites. *Latticespot* calculates the value at a point by interpolating from the central points of the four nearest grid cells.

Figure 7.2 shows a sample grid. A 40m filter is applied to the grid cell at location row 5, column 5. *Focalsum* adds together all the values in the filter and places the total in the output grid at location row 5, column 5. In Figure 7.2 the total sum is 550. Taking into consideration the fact that the cell size is 10m, traffic volume in metres per hour can be approximated by multiplying the output value by 10 - then dividing by 1000 to give traffic volume in km per hour. In the example shown in Figure 7.2, the traffic volume in km per hour within a radius of 40m is therefore 5.5. It should be noted, however, that at the edge of the circle, only cells with more than 50% of their area within the circle are included in the calculation.



*Figure 7.2. A sample of the traffic volume grid and a 50m filter.*

In the first stage of the approach - identifying the best combination of band widths and weights - the filtering technique had to be applied many times. As the radius is increased from 20m to 300m this becomes extremely time consuming. At this stage in the approach, however, the filter only needs to be applied at the location of the monitoring sites and not to the whole grid.

To make the technique more efficient a mask was created around the monitoring sites, to ensure that the filtering was only applied at the location of the monitoring sites. This was achieved by creating a 10m grid from the monitored point data. Thus, a cell containing a value greater than zero in the resultant grid indicated the location of a monitoring site and all other cells had a value of zero. The *eucallocation* command in GRID was used to calculate the distance to the nearest monitoring site for all cells. Those cells within 350m (a radius set to just beyond the 300m maximum) of the monitoring sites were set to zero and all other cells were set to NODATA. This created a mask.

The traffic volume grid was overlaid with the mask by adding the two grids together on a cell by cell basis. The traffic volume grid represents hourly day-time traffic volumes and contains values above zero where roads exist and zero where roads do not exist. Where values in the mask contain NODATA, values in the output grid contain NODATA. Thus, traffic volume data only exists in the 350m zone around the monitored points (Figure 7.3). When the filter is passed over the data, if a NODATA value is encountered in any location of the filter, the filter is not applied to that location and moves onto the next cell. The filtering therefore only works in the small nuclei around the monitoring sites. Applying the mask in this manner dramatically speeded up the experimental stage of the regression approach.

The adjusted mean NO<sub>2</sub> concentrations and the total traffic volumes within the 15 zones for the 80 permanent monitoring sites were then imported into the statistical package SPSS. The 15 zones of traffic volume were used to calculate traffic volume for 2 or 3 bands whose widths were chosen to reflect the rapid fall in concentrations

from the source - for example, 0-20m, 20-100m, 100-300m; 0-60, 60-300m; 0-40, 40m-140m, 140-300m.

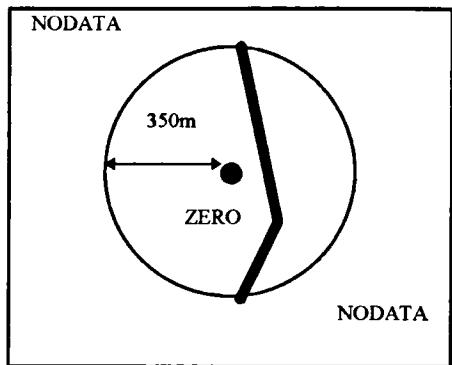


Figure 7.3. Creating a 350m mask for traffic volume around the monitoring sites.

Within SPSS, multiple regression techniques were used to establish a weighted traffic variable (TRAVOL). Traffic volumes within the different bands were entered into a multiple regression analysis against the adjusted mean NO<sub>2</sub> concentrations with the monitored data as the dependent variable and the traffic volumes as the independent variables. Different combinations of band width were entered into the regression analysis and compared in terms of their  $r^2$  values on a trial and error basis. A range of combinations were found to give broadly similar results: a combination of the 0-40m and 40-300m bands was selected on the basis of a marginally higher  $r^2$  value (adjusted  $r^2 = 0.432$ ).

The slope coefficients for the 0-40m band and the 40-300m band were used to help establish the weights for each band. The ratio between the two slope coefficients was used to define the weights. The coefficients were found to be  $75 \times 10^{-6}$  and  $5 \times 10^{-6}$  for the 0-40m and the 40-300m bands respectively, providing a ratio of approximately 15 for the 0-40 band to 1 for the 40-300m band. This resulted in the compound indicator represented by equation 7.1.

$$\text{TRAVOL} = (15 * \text{traffic volume } 0\text{-}40\text{m band}) + (1 * \text{traffic volume } 40\text{-}300\text{m band}) \quad [\text{Equation 7.1}]$$

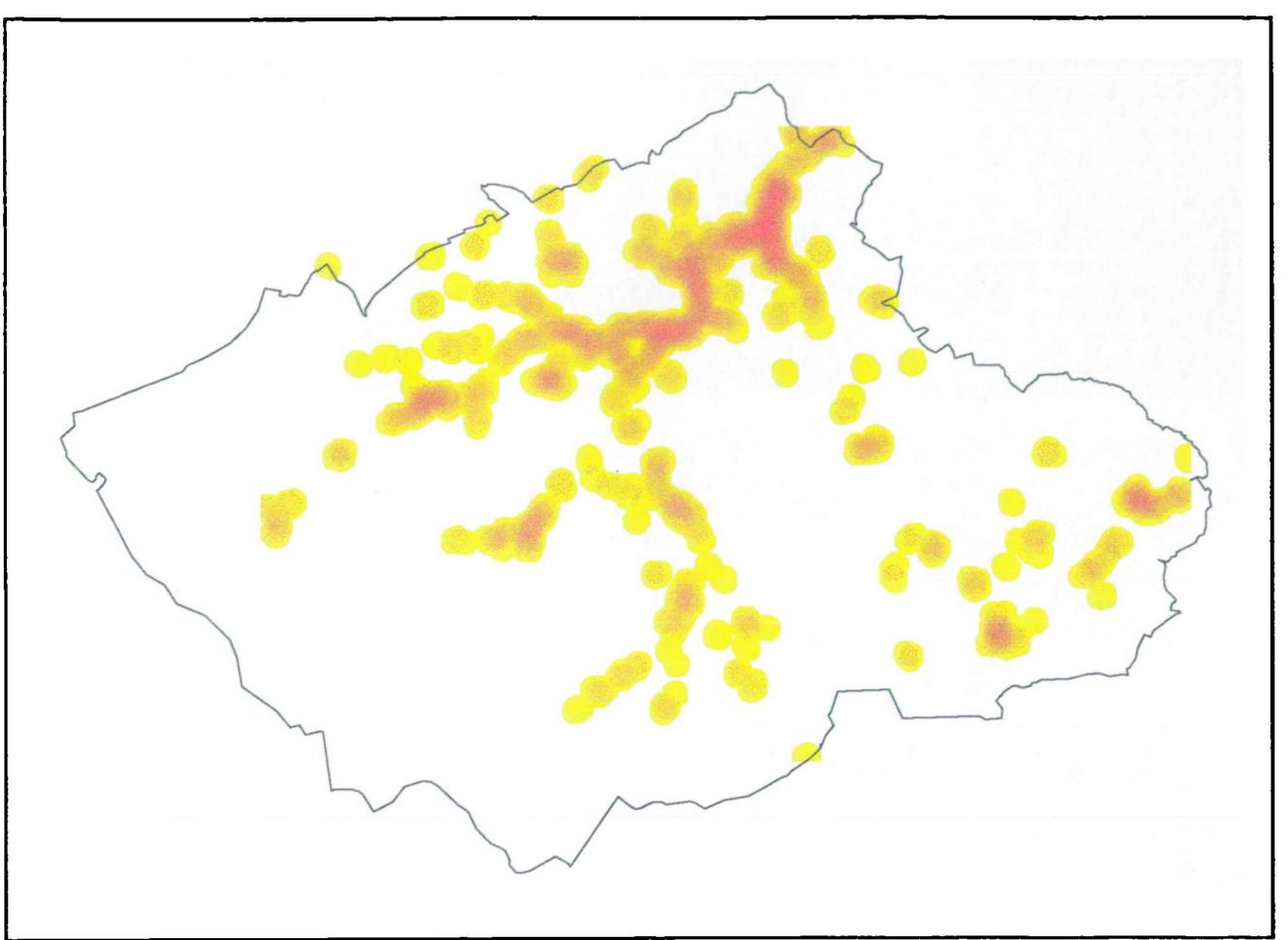
### 7.2.2 Land Cover

All land cover categories thought to reflect spatial variations in air pollutant were selected from the land cover data for analysis. These included:

- Very low density housing
- Low density housing
- High density housing
- Very high density housing
- High density commercial
- Industry
- Public institutions

Using the land cover coverage, 10m Boolean grids were created for all selected land cover categories. A cell with a value of 1 indicates the presence of the land cover and a cell with a value of zero indicates its absence. The total number of cells with a value of 1 for each category within 15 zones of varying radii (from 20m to 300m at 20m intervals) were found in GRID using the *focalsum* command. The number of cells were multiplied by 100 to give total area in square metres and then divided by 10000 to give the area in hectares. Total industrial land in the 0-300m zone for industry can be seen in Figure 7.4. The total areas, in hectares, for the 80 permanent monitoring sites were imported into SPSS. Once again, the *focalsum* command was applied in an efficient manner by overlaying the mask, described in section 7.2.1, with the Boolean grids.

Results were entered into a multiple regression analysis against the residuals from the TRAVOL analysis - to examine the extent to which land cover explained the remaining variation. A compound land cover variable (LANDCOV) was established by



## LEGEND

Industrial land (hectares)



1.0

28.2



5 km

*Figure 7.4 Total industrial land in the 0-300m zone.*

comparing the  $r^2$  values from different combinations of land cover categories and radii to find the best-fit combination. The best combination was found to be high density housing and industry in the 0-300m band (adjusted  $r^2 = 0.138$ ) and was thus selected for inclusion in the model. Examination of the slope coefficients ( $49 \times 10^{-6}$  for high density residential and  $28 \times 10^{-6}$  for industry) again gave the weighted compound indicator, presented in equation 7.2.

$$\text{LANDCOV} = (1.8 * \text{high density residential } 0\text{-}300\text{m band}) + (1 * \text{industry } 0\text{-}300\text{m band}) \quad [\text{Equation 7.2}]$$

### 7.2.3 Topography

The 50m Digital Terrain Model (DTM) stored in the GIS was also used to provide two potential measures of dispersion patterns and rates. The first one is simply the altitude at a location; the second is a measure of relative relief at a location. Relative relief (RRELIEF) was calculated from the DTM in GRID as a function of the difference between the central cell and its eight neighbouring cells. The DTM was exported as an ASCII text file and the FORTRAN program *rel\_relief* (Appendix 7), written by the author, applied to the data. Within the program the difference between the central cell and each of the eight neighbouring cells is calculated in turn and the resultant output value is the sum of the differences. The program works in a similar manner to a filter and results are placed in the location of the central cell. The output values were written to a new text file and imported into ARC/INFO as a new grid. Negative RRELIEF values indicate that the site is open and positive RRELIEF values indicate that the sites is sheltered.

*Latticespot* was also used to find the altitude and relative relief (RRELIEF) at the 80 sites from the 50m Digital Terrain Model (DTM) and grid of RRELIEF values respectively.

#### **7.2.4 Height of the Sampler and Topex**

Other independent variables, related to the monitoring sites, were also identified. These included the height of the sampler and mean angle to the visible horizon, topex (see section 3.4.5).

#### **7.4 The Regression Equation**

The variables BUILT, TRAVOL, RRELIEF and altitude (variously transformed), along with the sample height and topex, were all entered into a stepwise regression analysis against the annual mean NO<sub>2</sub> concentrations for the 80 permanent monitoring sites to give a regression equation for NO<sub>2</sub>. The annual mean NO<sub>2</sub> concentrations were entered as the dependent variable.

It was found that RRELIEF, altitude and topex did not contribute significantly to the equation and were therefore removed from the analysis. An inverse sine transformation of altitude (SIN(altitude)<sup>-1</sup>), however, was found to be significant and was used in the final analysis. Equation 7.3 represents the final regression equation.

$$\text{MeanNO}_2 = 11.83 + (0.00398 * \text{TRAVOL}) + (0.268 * \text{BUILT}) - (0.0355 * \text{SIN}(\text{altitude})^{-1}) + (6.777 * \text{sample height}) \quad [\text{Equation 7.3}]$$

All the variables included within the equation were found to be significant at the 0.005 level (i.e. < 0.005) and the t values were also found to be significant (i.e. > 2 or < -2). The regression coefficients, confidence intervals, significance values, standard errors and t values for the four variables are shown in Table 7.1.

*Table 7.1 Regression coefficients, confidence intervals, significance values, standard errors and t values for the variables in the regression equation.*

Variable	Regression coefficient	95% Confidence Interval		Significance	Standard error	t value
TRAVOL	0.00398	0.00027	0.00053	0.0000	0.00065	6.07
BUILT	0.268	0.0137	0.0398	0.0001	0.0656	4.08
SIN(altitude)-1	-0.0355	-0.598	-0.0111	0.0049	0.0122	-2.90
sample height	6.777	5.305	18.356	0.0005	1.6374	4.14

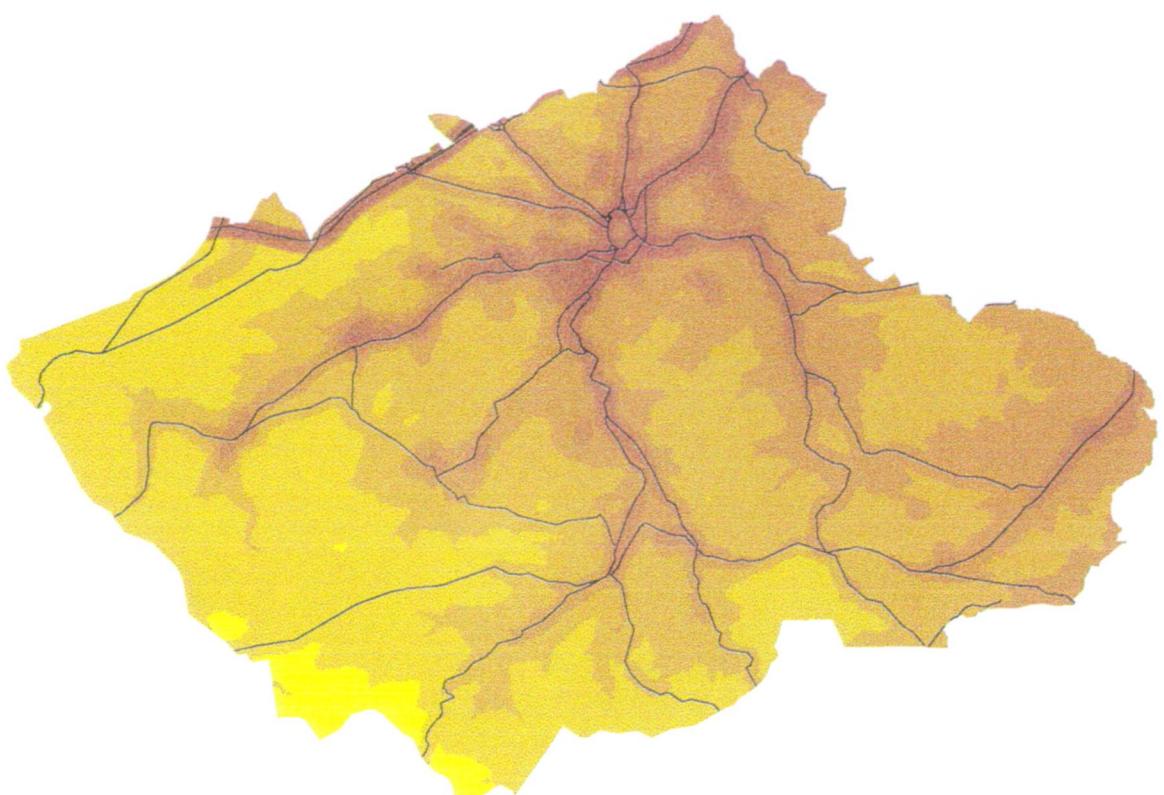
The weighted traffic volume variable was found to explain 44.0% of the NO<sub>2</sub> variation within the study area whilst the weighted measure of high density housing and industry explained a further 8.5% of the variation. Inclusion of the sample height and altitude variables increased the explanation of NO<sub>2</sub> variation within the study area to 60.7%.

#### **7.4.1 Generating the Air Pollution Map**

Within the GIS the variables TRAVOL and BUILT were calculated for all locations by applying Equations 7.3 and 7.4 to all cells in the grids. The inverse sine of altitude was calculated and a constant sample height of 2m was applied for the whole of the study area. Equation 7.5 was then applied in the GIS on a cell by cell basis to calculate area-wide estimates of annual mean NO<sub>2</sub>. The final pollution map for the estimated NO<sub>2</sub> in the Huddersfield study area is shown in Figure 7.5.

### **7.5 Results and Validation**

The measured annual means against the predicted values for the 80 permanent sites can be seen in Figure 7.6. The graph shows that the regression line fits the data well, with slight over-estimation for the lower concentrations and under-estimation for the higher concentrations. A plot of normal scores for the residuals (Figure 7.7) shows that the



#### LEGEND

$\mu\text{g}/\text{m}^3$



21

73



5 km

Figure 7.5 Map of annual mean  $\text{NO}_2$  by the regression method.

residuals appear to be acting normally, i.e. there are no visible outliers nor any significant departures from normality.

As discussed in Chapter 2, the problem with applying regression to spatial data is that the geographical location of the monitoring sites is not taken into consideration. Due to the spatial nature of the data, the residuals may show some degree of spatial autocorrelation, i.e. the residuals are more likely to resemble their neighbouring residuals than distant residuals. This would imply that spatial dependence still existed in the data and that the residuals were not random. The disadvantage of the regression technique is that it assumes that there is no spatial autocorrelation. This is where kriging has the advantage over regression, because spatial autocorrelation of the residuals is taken into account.

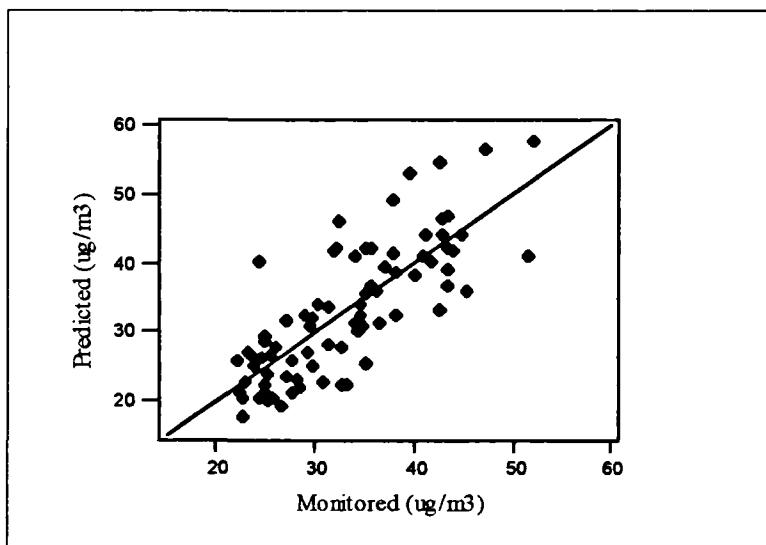


Figure 7.6 Monitored values against predicted values at the 80 permanent monitoring sites.

Spatial autocorrelation is more likely to occur when the percentage of variation explained is small and the independent variables are not significantly related to the dependent variable. Introducing more predictive variables into the regression model may help explain more of the spatial dependence in the data. As Knotters *et al* (1995)

suggest, ‘by adding additional variables to the regression model, thereby explaining a greater part of the variance, we expect that the assumption of the absence of autocorrelation of the errors will be satisfied more’. If spatial dependence remained in the data, however, one solution would be to apply kriging to the residuals.

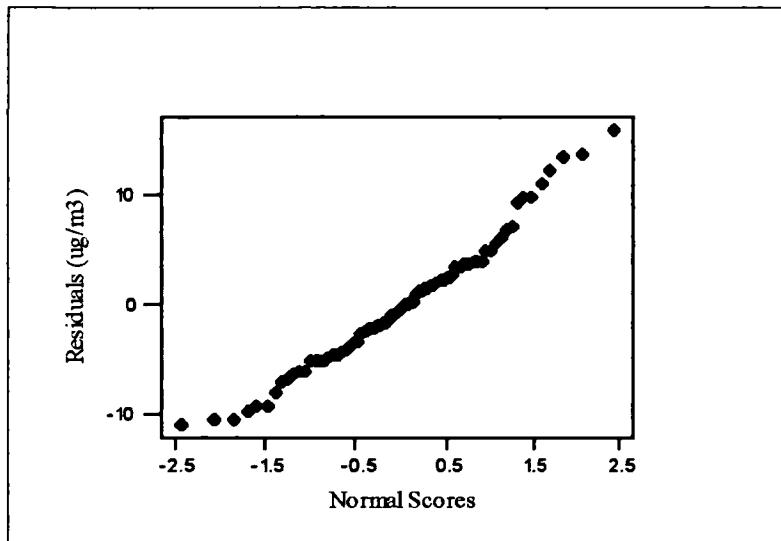


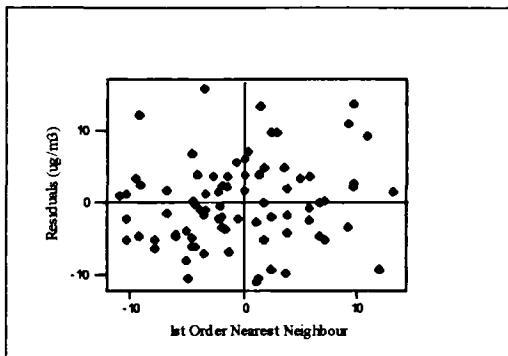
Figure 7.7 Normal scores for the residuals.

In the case of the  $\text{NO}_2$  data, analysis of the first, second and third order nearest neighbour residuals did not reveal any further spatial dependence in the data, as can be seen in Figures 7.8.a, 7.8.b and 7.8.c. Kriging was also applied to the regression residuals, but, as expected, this confirmed that there was no remaining spatial dependence in the data. This is substantiated by the strong relationship between  $\text{NO}_2$  and the explanatory variables and the high percentage of variation explained.

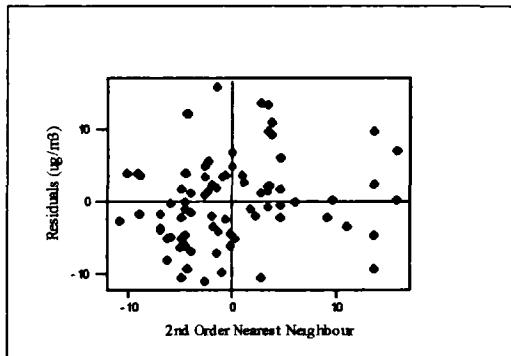
To examine how good the regression model predicts concentrations in unsampled locations, predictions were calculated for the 8 consecutive monitoring sites (section 3.3.1) and these were compared with the monitored annual means (the consecutive monitoring sites were not used to develop the regression model). The monitored annual means and the predicted values for the 8 sites can be seen in Table 7.2 and in Figure 7.9, the adjusted  $r^2$  was found to be 0.817 ( $\text{se} = 3.68 \text{ ug/m}^3$ ). There is a very

good linear relationship between the monitored and predicted values, as the table and graph show, with the slope of the regression line very close to 1.

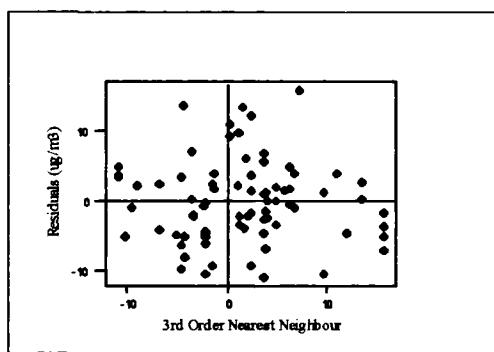
a)



b)



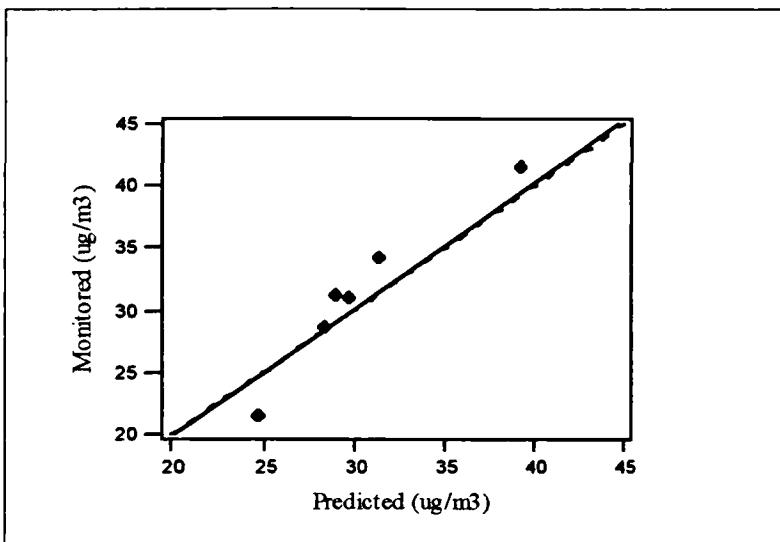
c)



*Figure 7.8 First, second and third order nearest neighbour residuals.*

*Table 7.2 Annual mean NO<sub>2</sub> values and regression predictions for the 8 consecutive monitoring sites.*

Site-id	Annual Mean NO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ )	Regression Prediction	Difference
103	31.3	28.9	2.4
104	28.6	28.3	0.3
105	34.3	31.5	2.8
106	19.9	26.4	-6.5
107	21.7	24.7	-3.0
108	44.6	48.0	-3.4
109	31.0	29.7	1.3
110	41.5	39.2	2.3



*Figure 7.9 Monitored annual mean  $NO_2$  against regression predictions for the 8 consecutive monitoring sites.*

## 7.6 Applying Regression Mapping to Traffic Related Pollution

This approach has shown that regression mapping can be applied to air pollution mapping in urban environments to help identify small area variations in traffic related air pollution. Equally, the approach has demonstrated how statistical techniques and GIS tools can be applied to provide high resolution estimates of the pollution surface. Furthermore, measures of the independent variables can be easily and effectively generated in the GIS. The technique appears to have worked moderately well.

The method can easily be transferred to other cities. As part of the SAVIAH study, the methodology was successfully applied in Amsterdam and Prague (Briggs *et al*, 1997). Due to differences in the availability of data, topography and methods of data collection in Amsterdam and Prague, compared to Huddersfield, a unique regression model was developed for each country. Development of the different models was, however, subject to the constraints that the models included terms for traffic volume, land cover and topography, and that the same approach for generating band widths (section 7.2) was applied. In each centre, the regression maps were validated against

8-10 consecutive monitoring sites. These sites were not used to establish the regression equations. The adjusted  $r^2$  values were found to be 0.79 ( $se = 4.45 \text{ ug/m}^3$ ) and 0.87 ( $se = 4.67 \text{ ug/m}^3$ ) in Amsterdam and Prague respectively. Further details regarding the development of the regression models and the resultant regression equations in the two cities can be found in Briggs *et al* (1997).

As part of the SAVIAH study, pollution data were also collected for 20 monitoring sites in Huddersfield in the year following the four surveys described in section 3.3. The measured concentrations were compared against the pollution map produced from the regression approach outlined in Chapter 7 and a good correlation was found between the two, with an adjusted  $r^2$  value of 0.59 (Briggs *et al*, 1997). This indicates that the air pollution map derived from the regression approach could be used in the same area for other years.

Recently, the regression model has been applied, unaltered, in other areas of the UK, namely Hammersmith, Ealing (Wills, 1998) and Sheffield (de Hoogh, 1998), to produce maps of air pollution. Strong correlations were again found between the pollution maps and measured concentrations. In each case, the adjusted  $r^2$  was found to be about 0.7 across 10 to 30 monitoring sites. These studies suggest that in order to apply the regression model in other areas, there is no need to recalculate the coefficients for the regression equation using locally derived data. It would, however, be necessary to validate, and even calibrate, the model with locally monitored pollution data.

The rationale underpinning these studies is that most of the spatial variation in traffic-related pollution is local (i.e. within 50 metres) and related to road vehicles. The regression model is thus providing an accurate picture of the pollution surface.

The performance of the model compared to the other approaches used is considered in more detail in Chapter 8.

## **CHAPTER 8 EVALUATION AND COMPARISON OF THE DIFFERENT APPROACHES**

Three different approaches have been developed to model air pollution ( $\text{NO}_2$ ) in urban environments as part of this study and four approaches have been applied, tested and validated:

- kriging - geostatistical techniques
- AMTRAP - geostatistical techniques and dispersion modelling
- moving window - geostatistical techniques and spatial analysis
- regression - statistical techniques and spatial analysis

As has been demonstrated in Chapters 4 to 7, all these approaches can be used to generate maps of air pollution. However, they all have different advantages and disadvantages related to the data requirements and the processing time. The accuracy of the methods and the degree to which they explain variations in air pollution also varies.

This Chapter evaluates and compares the different methods. The Chapter is split into three distinct sections. In the first section, the pollution maps generated by the four different approaches are compared. In the second section, the methods are further compared by reviewing their potential application in epidemiology, and, in the final section, the advantages and disadvantages of the four techniques are discussed in detail .

### **8.1 Comparison of the Pollution Maps**

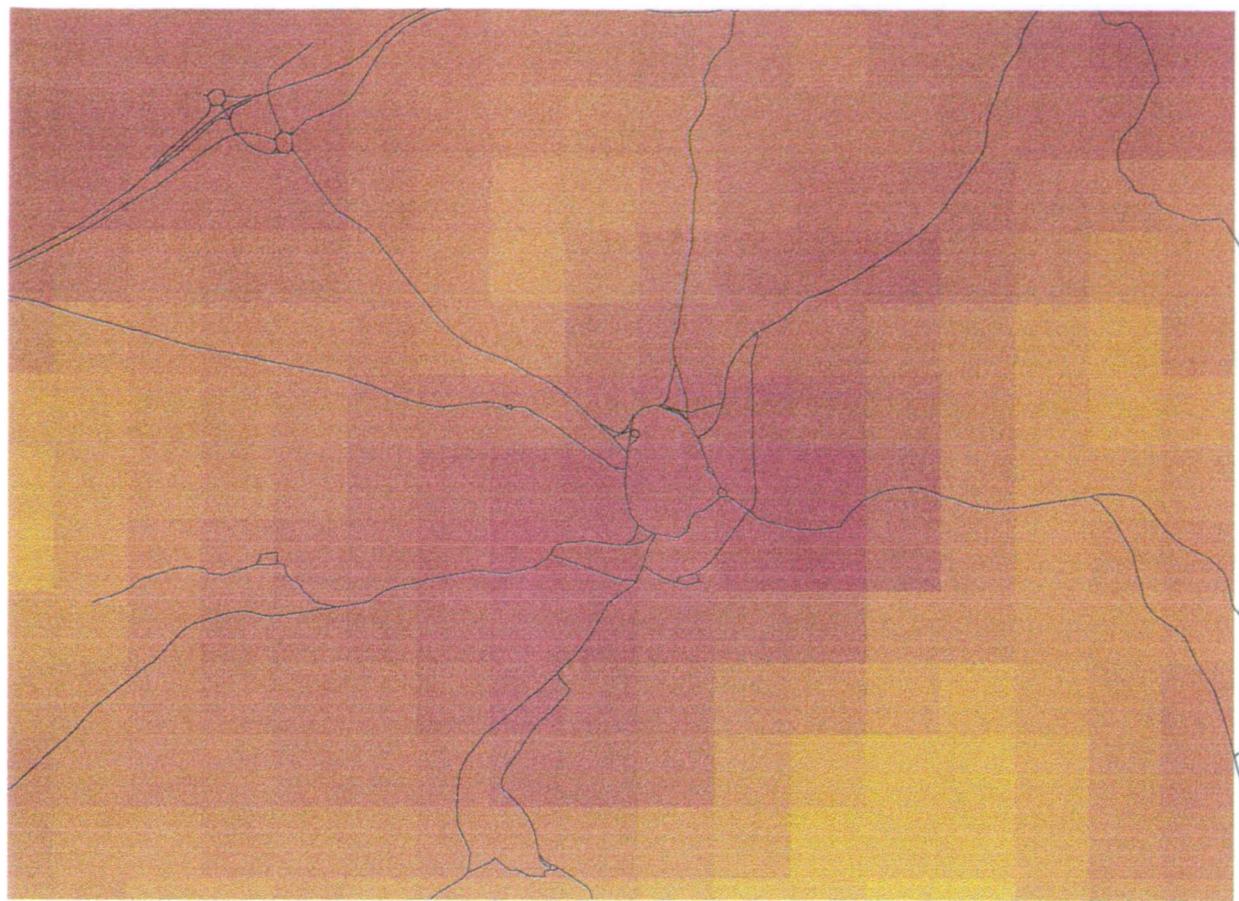
The pollution surfaces generated with the kriging, AMTRAP, moving window and regression approaches for the residential area of Huddersfield are shown in Figures

8.1, 8.2, 8.3 and 8.4 respectively. It can be seen that the pollution surface generated with the kriging technique is markedly smoother than the other three surfaces and not as detailed - with a comparatively small range between the minimum and maximum pollution values. The map shows gradual changes from low areas of pollution to high areas. The overall trend in levels of pollution across the residential area is quite distinctive, with high levels of pollution over the centre of the town, and a general decline towards peripheral areas.

The pollution surfaces generated with the AMTRAP and moving window models are very similar. The effects of meteorology on pollution levels is visible, however, in the AMTRAP map, with different levels of pollution at different sides of the roads. Both the AMTRAP and moving window approaches display a wide range of pollution values compared to kriging, with smooth low background levels in pollution, punctuated by the marked peaks of high levels that follow the pattern of the road network. At the apex of the peaks, levels in pollution can be nearly three times higher than background levels. The change between background levels and near-source levels is very sudden and occurs over a very short distance.

The pollution surface generated with the regression approach maintains the high peaks where the roads are displayed in the AMTRAP and moving window approaches, but overall changes in levels of pollution are much more gradual. The maximum pollution values are not as high as the AMTRAP and moving window approaches, but variation in levels of pollution is greater than for the kriging technique. High levels of pollution can also be found in the industrial areas and areas of high density residential housing. The background levels of pollution also reflect variations in topography. Overall, the regression map displays far greater variation in levels of pollution across the residential area than the AMTRAP and moving window approaches and a distinctive pattern across the city is visible.

The performance of the different methods was further compared by examination of the predicted annual mean NO<sub>2</sub> values for the 8 consecutive monitoring sites (section 3.3.1). Graphs of annual mean NO<sub>2</sub> against predicted NO<sub>2</sub> for kriging, AMTRAP,



## LEGEND

$\mu\text{g}/\text{m}^3$



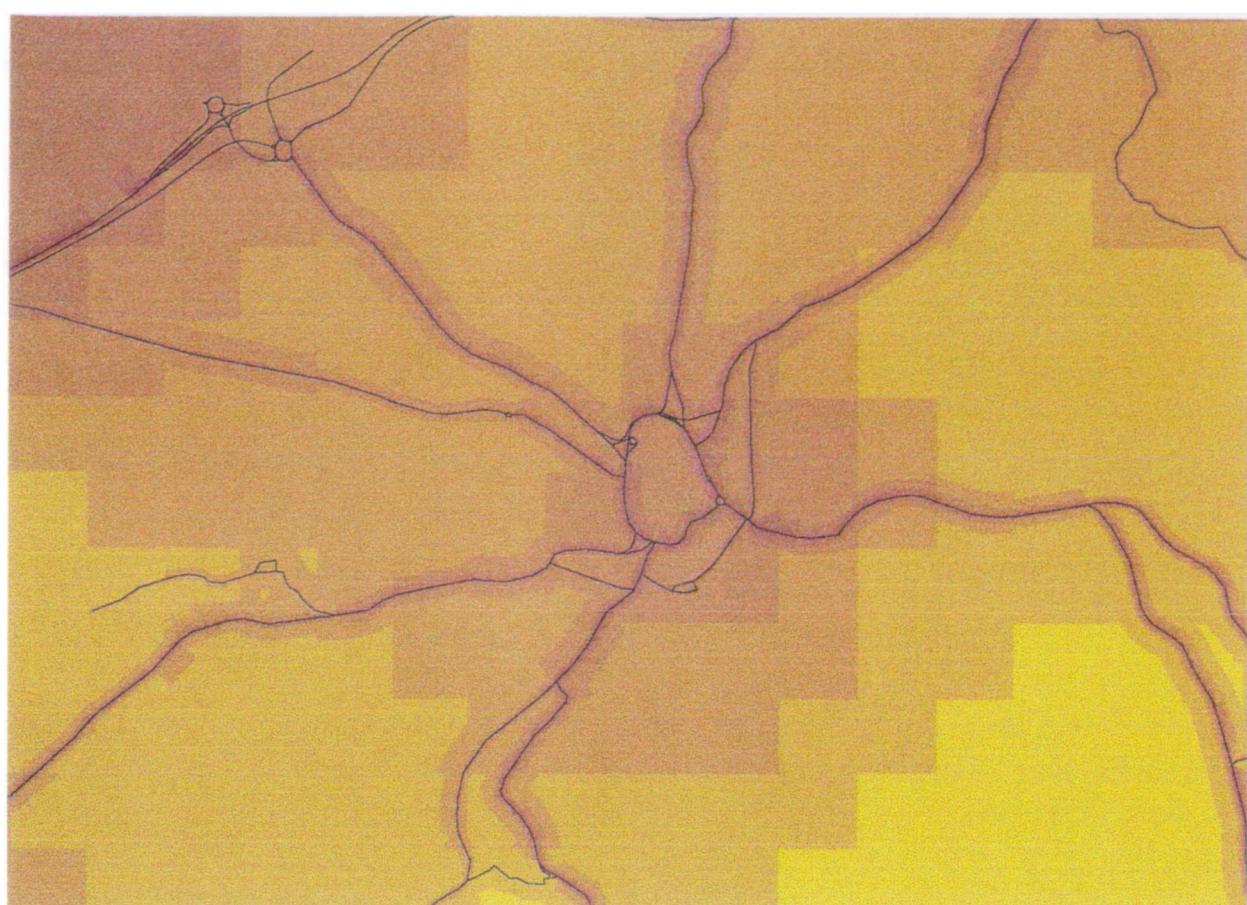
21

42



2 km

Figure 8.1 Annual mean  $\text{NO}_2$  in the residential area of Huddersfield by kriging.



## LEGEND

$\mu\text{g}/\text{m}^3$



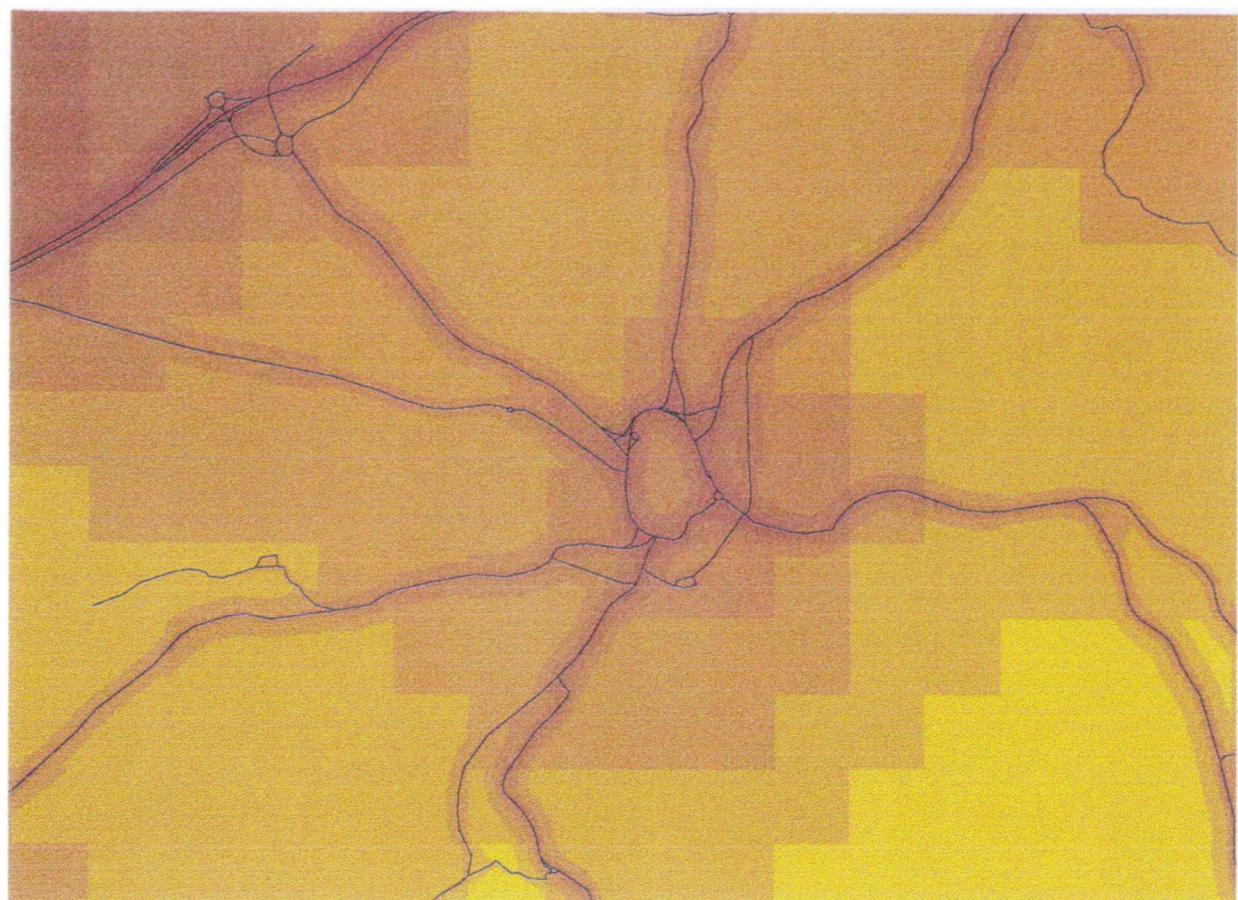
18

125



2 km

Figure 8.2 Annual mean  $\text{NO}_2$  in the residential area of Huddersfield by the AMTRAP approach.



#### LEGEND

ug/m<sup>3</sup>



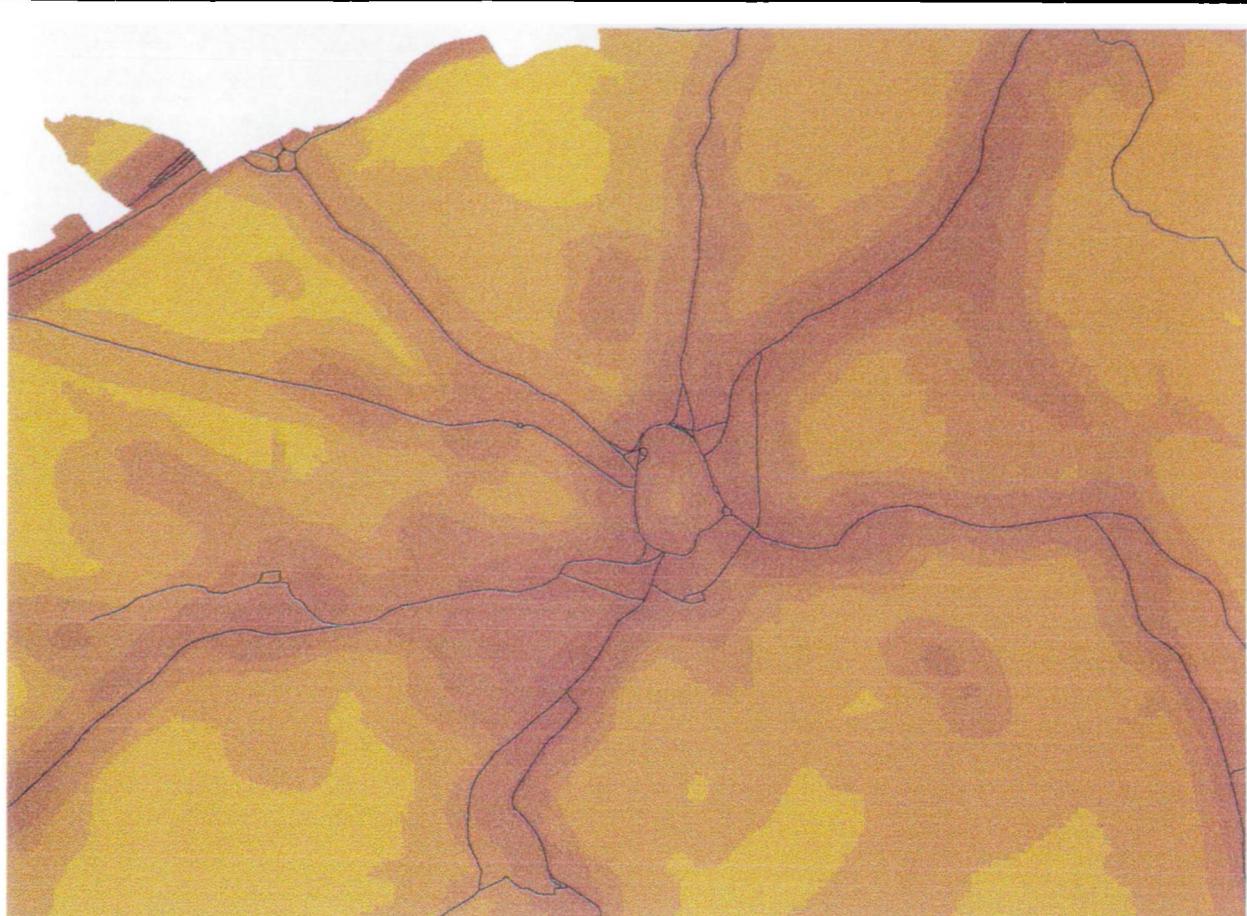
18

94



2 km

, *Figure 8.3 Annual mean NO<sub>2</sub> in the residential area of Huddersfield by the moving window approach.*



#### LEGEND

ug/m<sup>3</sup>



21

73



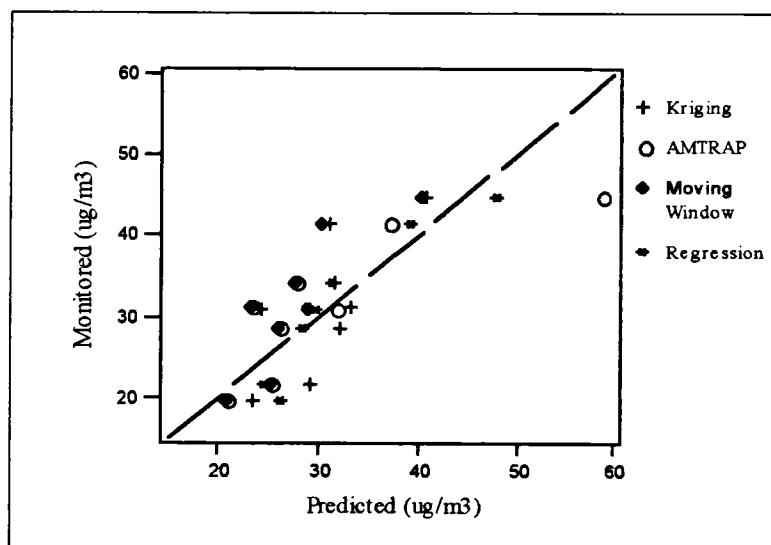
2 km

*Figure 8.4 Annual mean NO<sub>2</sub> in the residential area of Huddersfield by the regression approach.*

moving window and regression approaches can be seen in Figure 8.5. The adjusted  $r^2$  values, constant, slope coefficients, standard errors and range for the four approaches can be seen in Table 8.1. In the AMTRAP approach, one of the monitoring sites was identified as an outlier; in the second AMTRAP run (AMTRAP\*) the outlier has been removed.

*Table 8.1. Adjusted  $r^2$ , constant, slope coefficients, standard error and range for the kriging, AMTRAP, moving window and regression approaches.*

Approach	Adjusted $r^2$	Constant	Slope coefficient	Standard error of the estimate	Range (min-max)
Kriging	43.9%	-3.44	1.14	6.45	21.6-41.7
AMTRAP	62.8%	13.21	0.59	5.25	18.1-125.5
AMTRAP*	60.6%	-1.07	1.11	4.63	18.1-125.5
Moving window	67.3%	-2.97	1.24	4.92	18.1-94.5
Regression	81.7%	-0.60	1.02	3.68	21.3-73.7



*Figure 8.5 Monitored annual mean  $\text{NO}_2$  against kriging, AMTRAP, moving windows and regression predictions at the 8 consecutive monitoring sites.*

Kriging has the largest standard error and the  $r^2$  value is low compared to the other approaches. The relationship between the monitored and predicted values is relatively

poor with a high degree of under- and over-predicting. This reflects the smoothing produced by the technique and the small range in variation between the minimum and maximum pollution values. The AMTRAP approach has a lower standard error and, apart from the one outlier where the AMTRAP model has overestimated the pollution concentration by 25%, the predicted and monitored values display a good linear relationship. The outlier reflects the very high maximum pollution values at the side of the roads generated by the AMTRAP model compared to the other approaches. The standard error for the moving window approach is similar to that of the AMTRAP method, but the model appears to be systematically under-estimating the pollution concentrations. The standard error for the regression method is the lowest of all the approaches and the predicted and monitored values display a good linear relationship, with a regression slope of almost 1. There is, however, a tendency for the model to over-estimate at low concentrations and under-estimate at high concentrations, but this effect is small compared to the other methods.

The regression approach, therefore, appears to be the best predictor of levels of pollution for the 8 consecutive monitoring sites. The adjusted  $r^2$  value for the regression approach is nearly twice that of the kriging technique, with the AMTRAP and moving window approaches falling approximately half way between the two. This indicates that the additional near-source variation described by the AMTRAP and moving window approaches helps to provide better predictions of levels of pollution compared to kriging. However, even more accurate predictions can be attained by establishing a relationship between monitored data and predictive variables - as in the case of the regression approach. Furthermore, at 81.7%, the  $r^2$  value for the regression approach is exceptionally good, indicating that the approach is accounting for almost all the spatial variation in the monitored data.

## **8.2 Estimating the Population at Risk**

One of the main purposes of developing air pollution maps is to estimate and identify the population at risk, for example by overlaying the pollution surface on maps of

population distribution. How well this procedure works depends upon the accuracy of the pollution maps and thus upon the methodology used for air pollution mapping.

As Figures 8.1 - 8.4 demonstrate, the maps created by the four different approaches show very different patterns of pollution in the residential areas and therefore the population at risk will vary for the different approaches. The extent of this problem can easily be demonstrated using one of the available data sets collected for the SAVIAH study: the place of residence of the children

The 100m grid references relating to the postcodes of the children are stored in the GIS (section 3.3.6). The co-ordinates can be overlaid with a pollution map and an individual pollution score attached to the place of residence. This is just one method of estimating individual exposure; the appropriateness of this approach and alternative methods are discussed in Chapter 9. It should be noted, however, that the 100m grid reference represents the south-west corner of the grid square within which the first address in the unit postcode lies. Since the first address often lies at the edge of the postcode, this process may result in large errors between the actual location of the address and the 100m grid reference. To help reduce this error, Gatrell (1989) suggests adding 50m to the 100m grid references, shifting them North and East to the centre of the 100m grid squares. Following Gatrell's (1989) suggestion, therefore, 50m were added to the place of residence co-ordinates.

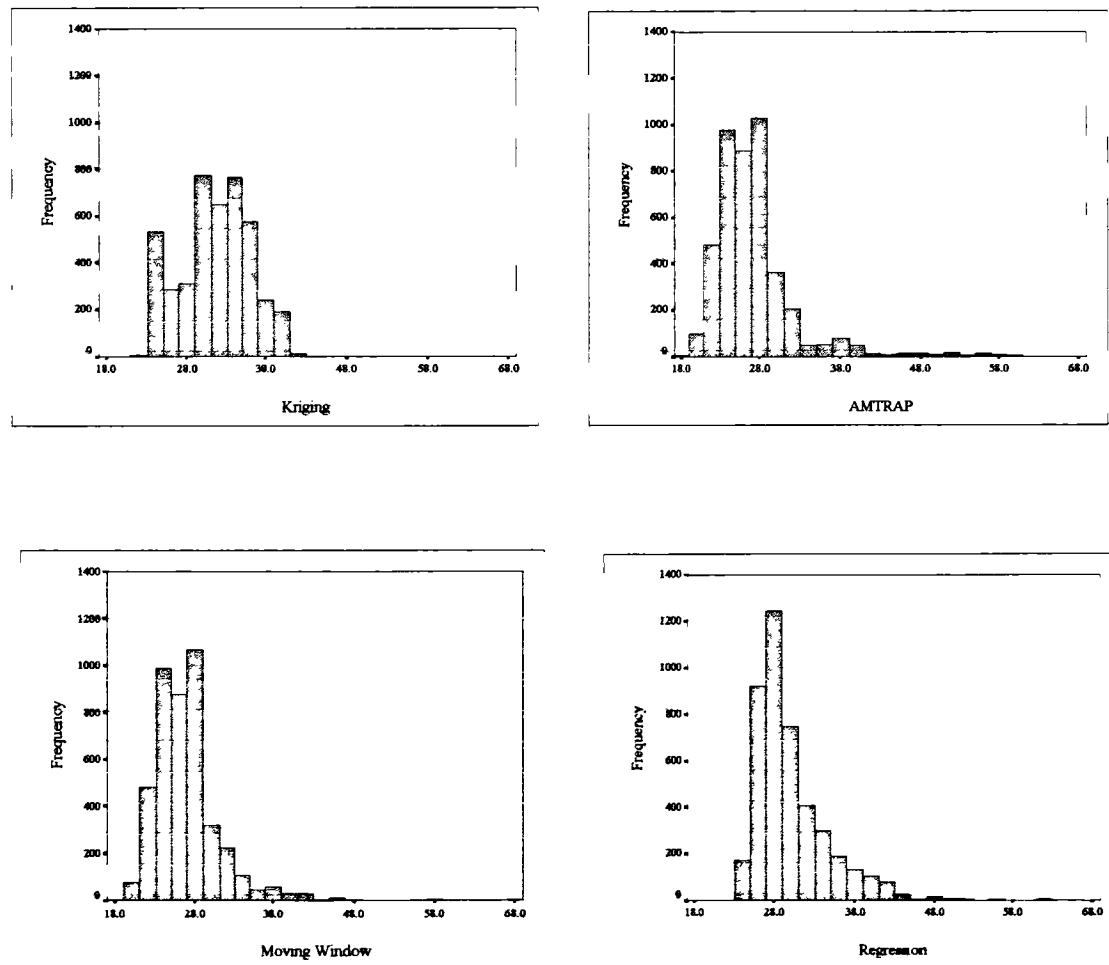
The co-ordinates for all 5027 children were overlaid with the four different pollution maps and four different pollution scores were established for each location. The kriged map, however, does not cover the full extent of the Huddersfield study area and therefore not all children had pollution scores for all methods. Only those children with four pollution scores (a total of 4357) were selected to help demonstrate the differences in the pollution maps.

The extent to which the pollution maps provide different pollution scores for the same locations was examined by calculating the minimum, maximum, interquartile range, mean and standard deviations for all scores for the four pollution maps. The results

can be seen in Table 8.2. Histograms (Figure 8.6) of the distribution of the pollution scores for the different pollution maps were also used to compare the approaches.

*Table 8.2 Minimum, maximum, interquartile range, mean and standard deviation pollution scores for the kriging, AMTRAP, moving window and regression approaches.*

Approach	min ug/m <sup>3</sup>	25% ug/m <sup>3</sup>	50% ug/m <sup>3</sup>	75% ug/m <sup>3</sup>	max ug/m <sup>3</sup>	mean ug/m <sup>3</sup>	SD ug/m <sup>3</sup>
Kriging	22.11	28.85	31.66	34.83	41.51	31.57	4.415
AMTRAP	18.15	23.98	26.67	28.53	63.25	27.09	4.983
Moving window	18.15	23.99	26.71	28.52	67.06	26.93	4.390
Regression	23.44	26.99	28.62	31.77	61.72	30.00	4.471



*Figure 8.6 Histograms of pollution scores for the kriging, AMTRAP, moving window and regression approaches.*

The histograms in Figure 8.8 and the values in Table 8.4 show how the distribution of pollution scores differs for the different approaches. Pollution scores from the kriging map represent an almost normal distribution, with a relatively small range between the minimum and maximum values and a relatively large range between the 25% and 75% quartile values compared to the other approaches.

The distributions of the pollution scores for the AMTRAP and moving window approaches are very similar. The histograms are negatively skewed and the range between the 25% and 75% quartile values is smaller than for any of the other approaches; they also have the lowest pollution scores falling between these values. The regression approach falls somewhere between kriging and the AMTRAP and moving window approaches. The histogram is also negatively skewed, but the spread of pollution scores falling between the minimum and maximum values displays a smoother distribution than for the AMTRAP and moving window approaches. The range between the 25% and 75% quantile values is slightly larger than the AMTRAP and moving window approaches and the pollution scores fall approximately half way between these and the kriging approach. The mean pollution scores for all four pollution maps show a 5  $\text{ug}/\text{m}^3$  range.

These results reflect the way in which the kriging technique has smoothed the pollution surface and fails to detect the linear variations in the pollution surface. In the case of the AMTRAP and moving window approaches most of the pollution scores are very low compared to the kriging and regression approaches, with the very high pollution scores representing locations very close to the emission sources. This can be further illustrated by calculating the percentage of children living in locations above the UK National Air Quality Strategy standard annual mean for  $\text{NO}_2$  of 37.6  $\text{ug}/\text{m}^3$  for the four pollution maps. In addition, the annual average  $\text{NO}_2$  monitored value was calculated for the study area from the 80 permanent monitoring sites and the percentage of children with scores above and below this value calculated to help demonstrate variations in the pollution surfaces. This was found to be 33.1  $\text{ug}/\text{m}^3$ . The results are shown in Table 8.3.

*Table 8.3. Percentage of children with a pollution score greater than the National Air Quality Strategy standard and the study average monitored value for the kriging, AMTRAP, moving window and regression approaches.*

Approach	National Air Quality Strategy standard: pollution scores > 37.6 ug/m <sup>3</sup>	Study monitored average: pollution scores > 33.1 ug/m <sup>3</sup>
Kriging	8.0 %	40.2 %
AMTRAP	4.3 %	7.2 %
Moving window	2.9 %	7.0 %
Regression	7.5 %	19.6 %

As the table demonstrates, there is nearly a three-fold difference in the percentage of children with pollution scores above the National Air Quality Strategy standard for NO<sub>2</sub> calculated from the different methods. The kriging approach suggests that nearly three times as many children live in locations above the standard compared to the moving window approach. In the case of the study area average NO<sub>2</sub> value, at one extreme, kriging suggests that nearly half the children live in locations where the pollution level is above the monitored average, while at the other extreme the AMTRAP and moving window approaches indicate that less than one tenth of children fall in this category. These figures demonstrate how the different mapping approaches applied to the data provide different estimates of exposure, and thus affect epidemiological inferences.

It is, however, also useful to examine the health risk associated with the high pollution areas identified by the four different mapping approaches. In order to achieve this, the response to two questions on the health questionnaire - life-time prevalence of wheezing or whistling and prevalence of wheezing or whistling in the last twelve months (section 3.4.6) - were compared against the four different pollution scores for the 4357 children. All children with positive responses to these questions (i.e. either a 1 for 'no' or a 2 for 'yes') were selected for inclusion in the analysis. Parents were only asked to respond to the second question if the response to the first was a 'yes'. This resulted in 4202 children for life-time prevalence of wheezing or whistling and 1357 children for wheezing or whistling in the last twelve months. Logistic regression

was applied to the data to measure the risk associated with raised levels of pollution. Odds ratios and 95% confidence intervals for the two health related questions against the four pollution approaches can be seen in Table 8.4.

*Table 8.4 Odds ratios and 95% confidence intervals for the four pollution approaches against health outcome.*

		Kriging	AMTRAP	Moving Window	Regression
Life-time prevalence of wheezing or whistling	Odds Ratio	0.98744	1.00653	1.00419	0.97795
	L 95% CI	0.97267	0.99332	0.98929	0.96295
	U 95% CI	1.00244	1.01992	1.01930	0.99321
Prevalence of wheezing or whistling in the last twelve months	Odds Ratio	0.99623	1.00727	1.00460	1.00069
	L 95% CI	0.97118	0.98656	0.98140	0.97686
	U 95% CI	1.02191	1.02840	1.02835	1.02510

These results show that there is no evidence of an increased risk to health in areas of high levels of pollution. This in turn suggests that there is no effect of exposure on the chronic respiratory health of children at these pollution levels. While this may indeed be a true interpretation of the results, it should, however, be taken into consideration that the lack of a relationship between levels of pollution and respiratory health in children may be due to other factors, such as inaccuracies in the data and the chosen measure of exposure.

In particular, this could be related to limitations in the health data, where there may be inconsistencies in the doctors diagnosis of asthma and respiratory disorders. In addition, parents coming from different social backgrounds may have different levels of knowledge and awareness of the symptoms of asthma, which may result in geographical variations in the response to the questionnaire. Furthermore, the logistic regression was applied to health data which had not been controlled for other risk factors (for example, house dust mite and pollen), nor the effects of cumulative exposure over time, which had not been modelled in the SAVIAH study.

With regard to the chosen measure of exposure, as discussed earlier in this section, individual exposure scores were established for the children by overlaying 100m grid

references with the pollution maps. However, changes in traffic related pollution have been shown to occur over very small distances, and indeed, the AMTRAP, moving window and regression maps have all been generated at a 10m resolution to reflect this. It is unlikely, therefore, that the 100m grid references are accurate enough to provide a true estimate of exposure at an address. The Ordnance Survey's AddressPoint data would, for example, provide a more accurate reference point for the postal addresses. The co-ordinates are referenced to a point that falls inside the permanent structure of an address to a resolution of 0.1m.

Equally, exposure at the address of residence alone may be a poor measure of exposure for the children. As discussed in Chapter 1, people are very mobile and individuals are exposed to different levels and combinations of pollutants as they go about their daily routines. The degree of exposure is often dependent upon other factors such as the age and health of the individual, their activities during exposure and the duration of exposure. Children, for example, spend a great deal of time at school and travelling to and from school. Furthermore, this thesis concentrates on outdoor exposure, however, sources of pollution can also be found inside the home (section 1.1.1) and a measure of indoor pollution may also help to provide more reliable estimates of exposure. It is possible, therefore, that the pollution scores defined here may only represent a measure of pollution and not a measure of exposure.

### **8.3 Advantages and disadvantages**

The different approaches vary not only in terms of their ability to model the variations in air pollution in urban areas, but also in terms of their data requirements, their complexity and the time it takes to apply the approaches. Before the approaches are discussed in detail their data requirements and processing times are presented in Table 8.5.

It should be noted that, apart from kriging, these processing times are based upon 10m resolution grids for the Huddersfield study area - which is approximately 30km x

20km. They also represent relative processing times, in that they are clearly dependent upon the speed and age of the machine on which they were run.

*Table 8.5 Processing times and data requirements.*

	Kriging	AMTRAP	AMTRAP no meteorology	Moving window	Regression
Processing times	2 minutes	10 hours	2 hours	30 minutes	24 hours
Data sets					
Monitored data	dense network	background	background	background	semi-dense network
Geographical data		road network land cover	road network land cover	road network	road network land cover topography
Meteorology		hourly or daily	study period average		
Field data					sample height

### 8.3.1 Kriging

The geostatistical technique - kriging - is used to interpolate from a set of monitored data to a continuous surface. The main advantage of using kriging is that the only data that needs to be collected is the monitoring data. In the case of air pollution, data is collected for a set of monitored sites. Depending on the scale of the study, this could be data collected from automatic monitoring stations, fixed site samplers or diffusion tubes. At unsampled locations kriging also calculates the variance, which is a measure of the confidence of the estimates. The only constraints on the collection of the data are that - a) air pollution is monitored at enough sites to ensure that a minimum number of pairs of points fall into each class of the semi-variogram and - b) there is an even distribution of sites to insure against large estimation errors due to gaps in the sampling network. In some cases the sampling may need to be stratified in order to ensure that particular pollution environments are taken into account. One advantage of using the monitored data is that the resultant pollution surface reflects the measured

values and the time period for which the measurements were taken. Conversely, the technique cannot readily be used to predict air pollution beyond the extremities of the sampling network or the study period. A disadvantage of only using monitored data is that random errors in the collection and measurement of the pollutant are difficult to detect and easily transferred to the pollution surface.

Kriging has on occasions been used to model background or rural levels of pollution (Campbell *et al*, 1994; Atkins and Lee, 1995). For urban areas, however, where marked variations in the level of pollution occur, an unmanageable density of monitoring sites would be necessary in order to pick up the peaks in levels of pollution that correspond to the roads. Due to lack of monitoring sites, it is likely that kriging will smooth the peaks that are present in urban environments.

It is possible that other kriging techniques, such as co-kriging using co-variates, may produce better estimates. Vauclin *et al* (1983), Knotters *et al* (1995) and Liu *et al* (1995) all compare kriging with co-kriging and conclude that co-kriging provides better estimates than kriging. As discussed in Chapter 2, co-kriging uses the correlation with other variables sampled at the same location, but which have also been sampled or measured at other locations, to help calculate estimates at these other locations. The geographical information stored in the GIS, for example land cover and total traffic volume in the surrounding area, could be sampled at the monitoring sites and also at locations on a regular grid and then used in the co-kriging calculations. However, since the regression approach explains so much of the variation in monitored NO<sub>2</sub> (81.7%), it is likely the variation would be adequately explained by the co-variates and that the kriging element of co-kriging would add little in terms of accuracy. This is, in fact, implied by the analysis of the results from the regression mapping method; as noted in section 7.5, no spatial dependence remained in the residuals from the regression analysis, showing that the covariates had explained all the spatially-dependent variation in the data.

### **8.3.2 AMTRAP**

With the AMTRAP approach, kriging is used to interpolate from the background monitoring sites to a continuous surface and dispersion modelling is used to model pollution from road traffic. The two components were additively combined. The main advantage of the AMTRAP approach is that the peaks in levels of pollution, corresponding to the road network, are identified and mapped. The AMTRAP approach requires only a limited number of sites in background locations to create a background surface - considerations regarding the number and distribution of sites, as with the kriging approach, still apply.

The approach uses data related to the road network (including traffic volume and road type), information about the land cover (to help establish surface roughness) and meteorology. The model can easily be altered to reflect changes to the road network or surface roughness. This is useful for modelling scenarios such as increases in traffic volume on the roads or the effect of building a new road. The model could also be adapted to reflect daily (rush-hour traffic) or seasonal variations in levels of pollution.

Compared to the other approaches, AMTRAP has the greatest demands on data and is the most complex of all the methods. This is primarily due to the application of the meteorological data in AMTRAP. There are a number of reasons why including meteorology causes so many problems. In the first instance, there is only a sparse network of meteorological sites and the data therefore are not always readily available. Even where data are available it is likely that there will only be one site, or at most a few sites, in the study area and consequently the meteorology has to be assumed to be constant across the whole of the study area. Secondly, because different temporal meteorological conditions have to be taken into account, levels of pollution have to be established for each unique set of conditions and this is very time consuming. The longer the study period the more difficult it becomes to model all the different meteorological conditions.

AMTRAP is the only one of the four methods to use daily meteorological data. However, meteorology was found to have little effect on levels of pollution over long periods of time (section 5.4). Taking into consideration the fact that the methods were developed to model long-term pollution, short-term changes in meteorology can therefore be disregarded and average values for meteorology can be used instead. The other two methods developed use no meteorology and provide similar or better estimates of pollution at unsampled locations.

One of the disadvantages of the AMTRAP approach is that the near-source pollution concentrations are only based upon the traffic volume on the nearest main road. Where two main roads run very close together this is not always an appropriate measure - the pollution concentration at these locations is likely to be a product of both roads. However, since the regression method produced better estimates than the AMTRAP method, the model was not developed further to accommodate this problem.

As discussed in Chapter 5, the CALINE3 dispersion model used to develop the AMTRAP approach only predicts concentrations of CO and other inert gases, and therefore NO<sub>2</sub> had to be treated as an inert gas, which may have introduced some errors into the measurements. Other line dispersion models, such as CALINE4 and ADMS (discussed in detail in section 2.2.2), do have the capacity to model NO<sub>2</sub> and may provide more accurate estimates of pollution. However, these more sophisticated line dispersion models, do have far greater data requirements, and for a large study area, such as Huddersfield, it would have been extremely difficult and very time consuming to collect all the relevant data.

In addition, the dispersion model - adapted to work in the GIS - is designed to model pollution for detailed, short-term, small-scale situations and is sensitive to variations in the input data. Small variations in traffic volume, for example, result in noticeable changes to the pollution concentrations. The Huddersfield study is a large-scale, long-term study, and consequently, to develop an automatic approach in the GIS, many of

the input variables had to be generalised. This may have introduced further errors into the AMTRAP model.

Where short-term pollution levels need to be estimated - for example, for specific pollution events - there are likely to be significant advantages in including meteorology in the analysis; in these cases, the AMTRAP approach may out-perform the other methods used here. Nevertheless, it should also be noted that the AMTRAP approach was applied in this case only with estimates for the mean daily traffic flow. If shorter-term modelling is to be conducted, more detailed (e.g. hourly) traffic flow data will also be required. This greatly adds to the data demands and computational burden of the approach.

### **8.3.3 Moving Window**

The moving window approach is similar to the AMTRAP approach, in so far as it is a combination of two separate components. Once again kriging is used to interpolate from the background monitoring sites to a continuous surface, but, with the moving window approach, a weighted template is passed over the road network to model the pollution from the road traffic. The two different components were again additively combined. As with the AMTRAP approach, only a limited number of sites in background locations are required to create a background surface, so long as the number and distribution of the sites satisfy the kriging requirements.

One of the advantages of the moving window approach is that, apart from the monitored data, once the template has been defined - and this only needs to be done once - it requires very little additional data. The approach is applied to the road network and the only information that is needed is the traffic volume. At any one location, all sources of pollution from motor vehicles in the surrounding area were identified and estimates of the pollution concentration calculated to reflect this. Thus the peaks in levels of pollution were identified and modelled. A further advantage is that the approach can be applied using techniques readily available in the GIS.

The approach is relatively quick to apply because the moving window only has to be applied to the data once. The model can easily be altered to reflect changes to the road network and traffic volume and the moving window simply applied again. Different measures of traffic volume - for example, representing seasonal or daily variations - could be stored in the database and the moving window applied to any of these items. To derive the weights for the template, the approach assumes that meteorology is constant with average conditions. Consequently, the approach is likely to be more applicable to long term studies of air pollution.

Like the AMTRAP approach, the moving window approach also used estimated data from the dispersion model - this time to help establish weights for the template. Alternatively, weights might be derived from simple ‘first principles’ - for example, by assuming that the contribution of any source varies with the square of the distance. Again, generalisation of the input variables may introduce errors into the approach. However, due to the aggregation of data under the template, it is unlikely to be as sensitive to variations in the input data as the AMTRAP approach.

#### **8.3.4 Regression**

The regression approach uses spatial analysis techniques to establish variables which are a measure of emission sources and dispersion patterns in the area surrounding the monitoring sites. The approach then uses statistical techniques to produce a regression equation for NO<sub>2</sub> from the monitored data and the measures of emission sources and dispersion patterns. The regression equation is then used to predict levels of NO<sub>2</sub> at locations where NO<sub>2</sub> has not been monitored. The approach could easily be applied to estimate levels of other pollutants.

The regression approach differs from the other three approaches in that it does not use spatial interpolation techniques. As previously mentioned, there are problems associated with spatial interpolation, primarily related to the distribution and number of

sites. The advantage of the regression approach, therefore, is that the distribution and number of monitoring sites are not as important as for the other three approaches - so long as they provide a representative sample of locations across the study area. Hence, the technique can be used to predict pollution concentrations beyond the geographical extents of the sampling network. Enough monitoring sites, however, are needed to build a robust regression model. In the case of the SAVIAH study, 80 monitoring sites were thought to be adequate. In practice, fewer sites, strategically located, would probably have been sufficient to establish a robust model.

The regression approach requires a number of geographical data sets, including the road network (with traffic volumes), land cover and topography. However, in different study areas, and for different pollutants, other variables may be more appropriate and significant. One of the disadvantages of the regression approach is that the experimental stage of the approach is relatively time consuming. Nevertheless, once the predictive variables have been identified and defined, applying the equation only takes a few minutes. A further disadvantage is that the variables have to be exported out of the GIS and the statistical analysis undertaken elsewhere - often on another platform.

One of the major advantages of the regression approach is that detailed knowledge about emission sources and dispersion patterns which reflect changes in levels of air pollution between the monitoring points is applied and modelled. Using so much predictive information is likely to provide more reliable estimates of pollution concentrations.

### **8.3.5 Summary**

The advantages and disadvantages of the four different approaches are summarised in Table 8.6.

*Table 8.6 Advantages and disadvantages*

Approach	Advantages	Disadvantages
Kriging	<ul style="list-style-type: none"> <li>- requires very little data</li> <li>- quick to apply</li> <li>- provides a measure of the confidence of the estimates</li> </ul>	<ul style="list-style-type: none"> <li>- dependent on the distribution and number of sites</li> <li>- ideally requires a dense network of monitoring sites</li> <li>- smoothes the pollution surface</li> <li>- errors in the measured data are difficult to detect</li> </ul>
AMTRAP	<ul style="list-style-type: none"> <li>- links dispersion models to GIS</li> <li>- models the peaks of pollution related to road traffic</li> <li>- easily altered to reflect changes in the road network or surface roughness</li> <li>- can be applied to predict future and past levels of pollution</li> <li>- can be applied to model short-term levels of pollution</li> </ul>	<ul style="list-style-type: none"> <li>- background concentrations are dependent upon the distribution and number of monitoring sites</li> <li>- computationally complex model</li> <li>- long processing time</li> <li>- requires meteorological data</li> <li>- estimates are based upon traffic volume on the nearest road</li> </ul>
Moving window	<ul style="list-style-type: none"> <li>- computationally efficient</li> <li>- quick to apply</li> <li>- uses readily available tools in GIS</li> <li>- models the peaks of pollution related to road traffic</li> <li>- easily altered to reflect changes in the road network or surface roughness</li> <li>- can be applied to predict future and past levels of pollution</li> </ul>	<ul style="list-style-type: none"> <li>- background concentrations are dependent upon the distribution and number of monitoring sites</li> </ul>
Regression	<ul style="list-style-type: none"> <li>- can predict levels of pollution beyond the extents of the sampling network</li> <li>- can be adapted to suit a particular study area or available data</li> <li>- applies well known and established statistical techniques</li> </ul>	<ul style="list-style-type: none"> <li>- experimental stage of the approach is time consuming</li> <li>- statistical analysis has to be applied outside the GIS</li> <li>- assumes no spatial auto-correlation</li> </ul>

#### 8.4 Discussion

The four approaches have been shown to differ in terms of their data requirements and efficiency, their ability to predict in unsampled areas and the final pollution maps that

are created. It should be taken into consideration, however, that all the models are sensitive to, and only as accurate as, the input data. As previously discussed, the AMTRAP, moving window and regression approaches all model variations in levels of pollution on a 10m grid, it is therefore essential that the input data, such as the location of the roads and the place of residence of the children, are at least as accurate, if not more so. The attribute data also needs to be accurate, changes in traffic volume and measured pollution concentrations, for example, will result in changes in the pollution surface. Applying the most accurate data possible will help to ensure that errors are not introduced into the models.

The final pollution maps produced by the different methods are markedly different. Choosing an appropriate method for pollution mapping is therefore very important and largely depends upon how the map is going to be used. In the case of traffic related pollution, areas of high levels of pollution are very localised and close to the sources of pollution, it is therefore extremely important that the pollution surface has been accurately generated. Where additional data relating to emission sources and dispersion patterns are available, then the use of exogenous information clearly helps to provide more reliable estimates of the pollution surface.

Although the different methods have been applied to estimate levels of NO<sub>2</sub>, all the methods could be adapted to estimate levels of other traffic related pollutants. If the main pollution sources could be identified, then the regression and moving window approaches would work just as well for CO and possibly PM<sub>10</sub>. The AMTRAP approach could be applied to any pollutant that can be modelled in an appropriate dispersion model. The dispersion model would then be linked to the GIS in a similar manner as the CALINE model. In the case of CO and PM<sub>10</sub>, it would be possible to apply the CALINE model again. For industry based pollutants, such as SO<sub>2</sub>, which have long ranges of dispersion compared to traffic related pollutants, other, point dispersion models, for example the Industrial Source Complex Short Term (ISCST) model, could also be linked to the GIS.

The regression approach proved to provide the most reliable estimates and, as discussed in Chapter 7, the approach can easily be transferred to other cities. Furthermore, so long as the model is validated and calibrated, the regression equation described in Chapter 7 can be applied in the UK without recalculating the coefficients.

## CHAPTER 9 CONCLUSION

In urban environments, spatial variations in air pollution have been shown to occur over very small distances. In the SAVIAH study, air pollution data collected at monitoring sites showed two-fold variations in levels of NO<sub>2</sub> at distances of less than 100m. Variations in air pollution reflect the complex pattern of emission sources in urban environments, related to road traffic, industrial and domestic activities, and characteristics of the built environment, such as housing density and street canyons, which influence dispersion rates and patterns.

As discussed in Chapter 3, there are only a small number of automatic monitoring stations regularly measuring air pollution in the UK. Consequently, levels of pollution in an urban environment are usually measured at only a single monitoring site. Since variations occur over such small distances, one site will not normally be representative of city-wide levels of air pollution. Furthermore, interpolation of the air pollution surface from the limited network of monitoring stations is unlikely to provide a meaningful measure of spatial variations in air pollution.

Chapter 1 emphasised how growth in the number of vehicles on the roads will cause major traffic congestion in many of the cities and towns. The problem will be further augmented by changes in lifestyle, with more and more people using cars for journeys of less than 5 km. In the future, with concerns about traffic congestion and the rise in emissions and its effect on public health, it will become increasingly more important to monitor and manage traffic volume and air pollution in urban environments. Providing meaningful measures of air pollution in cities, therefore, has implications for policy makers, planners and epidemiologists.

In major cities it will be necessary to control and reduce levels of air pollution and this can only realistically be achieved at the local level (i.e. at the street level). A major

part of transport policy is directed towards maintaining standards and ensuring that pollution levels do not exceed these standards. Measures of air pollution need to be derived at the local level if areas where pollution levels exceed guidelines and standards are to be identified. Transport and planning strategies can then be targeted to the areas where pollution levels are high and areas where the public may be at risk to ill-health effects.

In epidemiological research, it is important that areas where pollution levels are high can be easily identified. Overlaying the population with pollution surfaces can help to provide estimates of the population at risk and also help to establish relationships between health outcomes and pollution.

As previously mentioned, using information from automatic monitoring sites to estimate spatial variations in pollution tends to result in a poor representation of the pollution surface. The need is thus for methods which can generate more detailed and accurate pollution surfaces, across relatively wide urban areas. The application of diffusion tubes for low-cost monitoring of airborne dioxides, such as NO<sub>2</sub> and SO<sub>2</sub>, which are simple and cheap to operate, provides a higher density of sample sites, to underpin this approach. Mapping methods are then required which can use these data to generate a pollution map.

One such method is kriging. This was used here to generate pollution maps on the basis of a network of 80 passive sampler sites. One of the advantages of kriging is that a map of the error variance is produced; to help identify locations where additional monitoring sites could be placed (i.e. in locations where the estimates are not very accurate). Nevertheless, in the case of large population or city-wide studies, where estimates are required for many unsampled locations, interpolating from monitoring sites is not always a viable option. Spatial variations in air pollution from vehicle emissions occur over such small distances that it would require an unrealistic density of monitoring sites to pick up all the local variation that occurs.

The research presented in this thesis demonstrates how GIS can be applied to help provide more reliable estimates of the pollution surface by identifying and mapping the linear patterns associated with transport emissions. The AMTRAP model used a line-source dispersion model, in association with kriging; the moving window approach linked a spatial analysis technique developed in the GIS with kriging; the regression mapping method predicted variations in air pollution on the basis of a small number of covariates.

In many cases, the choice between these various approaches is largely dependent upon the application and the available resources. The three approaches developed here require far more data and have greater processing times than the kriging technique and the generalised surface generated by kriging may be adequate for some applications. However, in urban environments where detailed city-wide information about pollution levels is required, for control strategies and transport policies, or where associations with health need to be examined, then more detailed estimates are likely to be required. In terms of exposure assessment, for example, kriging could result in significant misrepresentation of the data, which is likely to dilute any association between air pollution and health. Furthermore, a map produced from monitored data will only reflect the measured values and the time period for which the measurements were taken. With the introduction of exogenous information, it is possible to predict variations in pollution, including extreme values (i.e. peaks and troughs), and introduce daily and seasonal variations into the model, without additional monitoring. Equally, it is possible to estimate past pollution levels in order to assess historic exposure patterns.

The capability to use these methods to analyse pollution levels beyond (either before or after) the period and area of available monitoring is also extremely useful for policy makers and planners. For example, models can be used to assess or compare the impact of new schemes, such as building a new road or giving restricted access to an existing road before implementation, and monitor the effect on levels of pollution after implementation. The models can also be used to predict future levels of air pollution

using projected estimates of traffic volumes and help inform the design and implementation of control schemes.

The performance of the three methods was examined by comparison with each other and also with the geostatistical technique kriging. All four methods were validated against independent, monitored data. Comparison of the methods showed that, by incorporating additional information in this way, it is possible to develop more accurate pollution maps than is possible by using kriging alone. In general, the regression mapping method was found to give the most accurate results, AMTRAP and the moving window methods were intermediate, and kriging the least accurate.

As these results imply, the choice of mapping method can, in some circumstances, have a major effect on the accuracy and utility of the results. When the various methods were used to estimate the percentage of children above the National Air Quality Strategy standard, for example, three-fold differences were found across the four methods. Potentially, these differences might significantly affect correlations between pollution levels and health outcome and any estimates of health risk. In the case of the SAVIAH study, however, it was seen that none of the methods showed any relationship between pollution levels and respiratory health in children. The apparent reliability of the regression mapping method, at least, in modelling pollution levels suggests that this lack of relationship was not due to errors in pollution modelling. As discussed in Chapter 8, however, it might be due to a number of other factors, including:

- poor estimates of exposure (e.g. due to the use of postcode centroids on a 100m grid instead of actual address location to define place of residence, and failure to take account of children's mobility)
- failure to consider other sources of exposure (e.g. indoor sources, other pollutants)
- inadequacies in the health data
- lack of control of confounding

- the lack of any effect of exposures on chronic respiratory health of children, at the pollution levels studied

The link between vehicle emissions and health is extremely complex, the associations with health are difficult to define and there is often an inconsistent diagnosis across the study population. In order to understand the association between air quality and health, therefore, there is a need to understand and model not only the spatial variations in pollution, but also the activities and mobility of the population. Against this background, the pollution maps can only reliably be used to give a general and relative exposure score, rather than to quantify actual, total exposure.

### **9.1 Future Research Issues**

In view of these conclusions, one of the most important areas for further research will be to try and provide better estimates of exposure. This would involve mapping the activities of the sample populations, identifying other locations - apart from their place of residence - where they spend extended periods of time, and also taking into account their exposure to indoor pollution.

The SAVIAH study concentrated on long-term exposure to pollution and the estimated pollution concentrations are based upon mean annual measurements. There is, however, a distinctive pattern of temporal variation in levels of pollution and this also varies from place to place. The next step in pollution modelling would be to incorporate the temporal element and map changes in pollution over time and space with a time-space pollution model. This is likely to provide even more reliable estimates of pollution and will not only help inform studies looking at the chronic health effects of pollution, but may also help inform studies examining the acute health effects.

The methods presented in this thesis were developed and tested in Huddersfield, UK, and although they work well in this study area, it would be useful to test the models in

other cities. Indeed, as mentioned in Chapter 7, the regression model has been successfully applied in other cities, the results of which are reported in Briggs *et al* (1997).

Furthermore, the methods presented here have only been applied to model NO<sub>2</sub>, but there are other important traffic related pollutants, for example, PM<sub>10</sub> and PM<sub>2.5</sub>, that are associated with ill-health effects. The development, application and testing of these models for other pollutants is clearly an important area for future research.

Other techniques may also prove useful in the application of traffic related pollution mapping. One method that is likely to be beneficial, and worthy of further examination, is the application of co-kriging.

At the present time, traffic related pollution and its effect on the health of the public is one of the most prominent issues in environmental health. To be able to produce accurate and reliable estimates of levels of pollution is therefore extremely important; without doubt, pollution mapping and exposure assessment are two topics that demand continued and further research.

## REFERENCES

- Advisory Group on the Medical Aspects of Air Pollution Episodes. 1993. *Oxides of Nitrogen*. Third Report, Department of Health.
- Alexopoulos, A., Assimacopoulos, D. and Mitsoulis, E., 1993. Model for Traffic Emissions Estimation. *Atmospheric Environment*, 27B(4), 435-446.
- Anderson, H.R. 1989. Increase in Hospital Admissions for Childhood Asthma: Trends in Referral, Severity, and Readmissions from 1980 to 1985 in a Health region of the United Kingdom. *Thorax*, 44, 614-619.
- Anderson, H.R., Butland, B.K. and Strachan, D.P. 1994. Trends in the prevalence and severity of childhood asthma. *British Medical Journal*, 308, 1600-1604.
- Apeldoorn, van. R.C., Celada, C. and Nieuwenhuizen, W. 1994. Distribution and Dynamics of the Red Squirrel (*Sciurus Vulgaris L.*) in a Landscape with Fragmented Habitat. *Landscape Ecology*, 9(3), 227-235.
- Asher, M.I., Keil, U., Anderson, H.R., Beasley, R., Crane, J., Martinez, F., Mitchell, E.A., Pearse, N., Sibbald, B., Stewart, A.W., Strachan, D., Weiland, S.K. and William, H.C. 1995. International Study of Asthma and Allergies in Childhood (ISARC): Rationale and Methods. *European Respiratory Journal*, 8, 484-491.
- Atkins, D.H.F. and Lee, D.S. 1995. Spatial and Temporal Variation of Rural Nitrogen Dioxide Concentrations across the United Kingdom. *Atmospheric Environment*, 29 (2), 223-239.
- Atkins, D.H.F., Sandalls, J., Law, D.V., Hough, A.M. and Stevenson, K. 1986. *The Measurement of Nitrogen Dioxide in the Outdoor Environment Using Passive Diffusion Tube Samplers*. United Kingdom Atomic Energy Authority Report, AERE R-12133, Harwell Laboratory, Oxfordshire.
- Bailey, T.E. and Gatrell, A.C. 1995. *Interactive Spatial Data Analysis*. Longman Scientific and Technical.
- Benson, P.E. 1979. *CALINE3 - A Versatile Dispersion Model for Predicting Air Pollution Levels Near Highways and Arterial Streets*. Abridgement of Report No. FHWA/CA/TL-79/23, Nov 1979. California Department of Transportation.
- Benson, P.E., 1992. A review of the development and application of the CALINE3 and CALINE4 Models. *Atmospheric Environment*, 26B(3), 379-90.
- Boleij, J.S.M., Lebret, E., Hoek, F., Noy, D. and Brunekreef, B. 1986. The Use of Palmes Tubes for Measuring NO<sub>2</sub> in Homes. *Atmospheric Environment*, 20 (3), 596-600.
- Bower, J.S., Broughton, G.F.J., Dando, M.T., Lees, A.J., Stevenson, K.J., Lampert, J.E., Sweeney, B.P., Parker, V.J., Driver, G.S., Waddon, C.J. and Wood, A.J. 1991.

Urban NO<sub>2</sub> Concentrations in the UK in 1987. *Atmospheric Environment*, 25B (2), 267-283.

Bower, J.S. and Vallance-Plews, J. 1995. *The UK National Air Monitoring Networks*. Paper prepared for WHO seminar. National Environmental Technology Centre.

Briggs, D.J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., Pryl, K., van Reeuwijk, H., Smallbone, K. and van der Veen, A. 1997. Mapping Urban Air Pollution Using GIS: A Regression-Based Approach. *International Journal of Geographical Information Science*. 11(7), 677-718.

Briggs, D.J., van der Veen, A.A., Pyl, K. and Collins, S. 1995. Air Pollution Mapping in the SAVIAH study. (Abstract). *Epidemiology*, 6(4), S32.

Brunekreef, B., Janssen, N.A.H., de Hartog, J., Harssema, H., Knape, M. and van Vliet, P. 1997. Air Pollution from Truck Traffic and Lung Function in Children living Near Motorways. *Epidemiology*, 8 (3), 298-303.

Buchdahl, R., Parker, A., Stebbings, T. and Babiker, A. 1996. Association between air pollution and acute childhood wheezy episodes: prospective observational study. *British Medical Journal*, 312, 661-5.

Burr, M.L. 1987. Is Asthma Increasing? *Journal of Epidemiology and Community Health*, 41, 185-189.

Burrough, P.A. 1986. *Principles of Geographical Information Systems for Land Resources Assessment*. Monographs on Soil and Resources Survey, Oxford: Clarendon Press.

Campbell, G.W., Stedman, J.R. and Stevenson, K. 1994. A Survey of Nitrogen Dioxide Concentrations in the United Kingdom Using Diffusion Tubes, July-December 1991. *Atmospheric Environment*, 28 (3), 477-486.

Cassiani, G. and Medina, M.A. 1997. Incorporating Auxiliary Geophysical Data into Ground-Water Flow Parameter Estimation. *Ground Water*, 35 (1), 79-91.

Collins, S. 1997. *Modelling Urban Air Pollution Using GIS*. In Geographic Information Research: Bridging the Atlantic. Edited by Craglia, M. and Couclelis, H. Taylor and Francis, 427-440.

Collins, S. 1998. *Modelling Spatial Variations in Air Quality Using GIS*. In GIS and Health in Europe. Edited by Gatrell, A.C. and Löytönen, M. Taylor and Francis (Forthcoming).

Collins, S., Smallbone, K. and Briggs, D. 1995. *A GIS Approach to Modelling Small Area Variations in Air Pollution within a Complex Urban Environment*. In Innovations in GIS 2. Edited by Fisher, P. Taylor and Francis, 245-253.

Commission of the European Communities. 1992. *Towards Sustainability: A program of action on the European Environment*. Commission of the European Communities, Brussels.

Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society. 1996. State of the Art: Heath Effects of Outdoor Air Pollution. *American Journal of Respiratory Critical Care Medicine*, 153, 3-50.

Committee on the Medical Effects of Air Pollution. 1995. *Asthma and Outdoor Air Pollution*. Department of Health. HMSO

Commission of the European Communities. 1992. *Towards Sustainability: a Program of Action on the European Environment*. Brussels: Commission of the European Communities.

Cressie, N.A.C. 1993. *Statistics for Spatial Data*. John Wiley and Sons, Inc.

Dabberdt, W.F., Ludwig, F.L. and Johnson, W.B. Jr. 1973. Validation and Applications of an Urban Diffusion Model for Vehicular Pollutants. *Atmospheric Environment*, 7, 603-618.

Dekkers, A. 1995. *Kriging: a Geostatistical Modelling Technique*. Background Paper to the SAVIAH Project with Methodological Description and Application to SO<sub>2</sub> Concentrations in Poznan, Poland. CWM-MEMO 002/95. National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands.

Department of the Environment. 1989. *National Road Traffic Forecasts (Great Britain)*. HMSO.

Department of the Environment. 1996. *Digest of Environmental Statistics No. 18 1996*. HMSO.

Department of Transport. 1996a. *National Travel Survey 1993/95*. Transport Statistics Report. HMSO.

Department of Transport. 1996b. *Transport Statistics Great Britain: 1996 Edition*. HMSO.

Design Manual for Roads and Bridges. 1994. *Air Quality*. Volume 11, Section 3, Part 1.

Deutsch, C.V. and Journel, A.G. 1992. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press.

Dockery, D.W., Pope, C.A.III, Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G.Jr. and Speizer, F.E. 1993. An Association Between Air Pollution and Mortality in Six U.S. Cities. *Journal of Medicine*, 329(24), 1753-1759.

Dubrule, O. 1984. Comparing Splines and Kriging. *Computers and Geosciences*, 10 (2-3), 327-338.

Edwards, J., Walters, S. and Griffiths, R.K. 1994. Hospital Admissions for Asthma in Preschool Children: Relationship to Major Roads in Birmingham, United Kingdom. *Archives of Environmental Health*, 49 (4), 223-227.

Eerens, H., Sliggers, C. and van der Hout, K. 1993. The CAR model: the Dutch method to determine city street air quality. *Atmospheric Environment*, 27B(4), 389-99.

Eggleson, S., Hackman, M.P., Heyes, C.A., Irwin, J.G., Timmis, R.J. and Williams, M.L. 1992. Trends in Urban Air Pollution in the United Kingdom During recent Decades. *Atmospheric Environment*, 26B(2), 227-239.

El Abbass, T., Jallouli, C., Albony, Y. and Diamant, M. 1990. A Comparison of Surface Fitting Algorithms for Geophysical Data. *Terra Research*, 2, 467-475

Elliott, P., Briggs, D., Lebret, E., Gorynski, P. and Kriz, B. 1995. Small Area Variations in Air Quality and Health, The SAVIAH study: Design and methods. (Abstract). *Epidemiology*, 6(4), S32.

European Commission. 1995. *A Review of Research in Environmental Health and Chemical Safety in the STEP (1989-92) and Environment (1991/94) Research Programmes*. Ecosystem Research Report 15, Report number EUR 15909 EN, edited by C. Nolan, European Commissions, Luxembourg.

Fischer, P., Brunekreef, B., Biersteker, K., Boleij, J.S.M., Lende, R. vd., Schouten, J. et al. 1989. Effects of Indoor Exposure to Nitrogen Dioxide on Pulmonary Function of Women Living in Urban and Rural Areas. *Environmental International*, 15, 375-381.

Gambolati, G. and Volpi, G. 1979. Groundwater Contour Mapping in Venice by Stochastic Interpolators. *Water Resources Research*, 15 (2), 281-290.

Gatrell, A.C. 1989. On the Spatial Representation and Accuracy of Address-Based Data in the United Kingdom. *International Journal of Geographical Information Systems*, 3, 335-348.

George, D.G. 1997. The Airborne Remote Sensing of Phytoplankton Chlorophyll in the Lakes and Tarns of the English Lake District. *International Journal of Remote Sensing*, 18(9), 1961-1975.

Girman, J.R., Hodgson, A.T., Robinson, B.K. and Traynor, G.W. 1983. Laboratory Studies of the Temperature Dependence of the Palms NO<sub>2</sub> Passive Sampler. Lawrence, Kirkeley. Laboratory report LBL 16302 Berkeley.

Godlee, F. 1991 Air Pollution: II - Road Traffic and Modern Industry. *British Medical Journal*, 303, 1539-1543.

Guo, J. and O'Leary, S.M. 1997. Estimation of Suspended Solids from Aerial Photographs in GIS. *International Journal of Remote Sensing*, 18(10), 2073-2086.

Haining, R. 1990. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge.

Heal, M.R. and Cape, J.N. 1997. A Numerical Evaluation of Chemical Interferences in the Measurement of Ambient Nitrogen Dioxide by Passive Diffusion Samplers. *Atmospheric Environment*, 31(13), 1911-1923.

Her Majesty's Inspectorate of Pollution. 1993. *An Assessment of the Effects of Industrial Releases of Nitrogen Oxides in the East Thames Corridor*. HMSO.

Hewitt, C.N., 1991. Spatial variation in nitrogen dioxide concentrations in an urban area. *Atmospheric Environment*, 25B(3), 429-34.

de Hoogh, K. 1998. Personal Communication.

Ishizaki, T., Koizumi, K., Ikemori, R., Ishiyama, Y., and Kushibiki, E. 1987. Studies of prevalence of Japanese cedar pollinosis among the residents in a densely cultivated area. *Annals of Allergy*, 58, 265-70.

Iverson, L.R., Dale, M.E., Scott, C.T. and Prasad, A. 1997. A GIS-Derived Integrated Moisture Index to Predict Forest Composition and Productivity of Ohio Forests (USA). *Landscape Ecology*, 12, 331-348.

Johnson, W.B., Ludwig, F.L., Dabberdt, W.F. and Allen, J. 1973. An Urban Diffusion Simulation Model for Carbon Monoxide. *Journal of the Air Pollution Control Association*, 23 (6), 490-498.

Journel, A.G. and Huijbregts, C.J. 1978. *Mining Geostatistics*. Academic Press, London.

Kirklees Traffic Monitoring Report. 1992. Highways Engineering and Technical Services, Kirklees Metropolitan Council.

Knotters, M., Brus, D.J. and Oude Voshaar, J.H. 1995. A Comparison of Kriging, Co-kriging and Kriging Combined with Regression for Spatial Interpolation of Horizon Depth with Censored Observations. *Geoderma*, 67, 227-246.

Kriz, B., Bobak, M., Martuzzi, M., Briggs, D., Livesley, E., Lebret, E., Fisher, P., Wojtyniak, B., Gorynski, P. and Elliott, P. 1995. Respiratory Health in the SAVIAH study. (Abstract). *Epidemiology*, 6(4), S31.

Lam, N.S-N. 1983. Spatial Interpolation Methods: A Review. *The American Cartographer*, 10 (2), 129-149.

Larssen, S., Tonnesen, D., Clench-Aas, J., Aarnes, M.J. and Arnesen, K. 1993. A Model for Car Exhaust Exposure Calculations to Investigate Health Effects of Air Pollution. *The Science of the Total Environment*, 134, 51-60.

Laslett, G.M, McBratney, A.B., Pahl, P.J. and Hutchinsomn, M.F. 1987. *Journal of Soil Science*, 38, 3250-3273.

Laxen, D.P.H. and Noordally, E. 1987. Nitrogen Dioxide Distribution in Street Canyons. *Atmospheric Environment*, 21 (9), 1899-1903.

Lebret, E., Briggs, D., Collins, S., Reeuwijk, H. van. and Fisher, P.H. 1995. Small Area Variation in Exposure to NO<sub>2</sub>. (Abstract). *Epidemiology*, 6(4), S31.

Lebret, E., Briggs, S., Smallbone, K., van Reeuwijk, H., Fischer, P., Harssema, H., Kriz, B. and Gorynski, P. 1997. Small Area Variations in Ambient NO<sub>2</sub> Exposures in Four European Areas. Accepted for publication in *Atmospheric Environment*.

Liu, L.J.S, Rossini, A. and Koutrakis. P. 1995. Development of Cokriging Models to Predict 1- and 12-hour Ozone Concentrations in Toronto. (Abstract). *Epidemiology*, 6(4), S69.

Mattson, M.D. and Godfrey, P.J. 1994. Identification of Road Salt Contamination Using Multiple Regression and GIS. *Environmental Management*, 18 (5), 767-773.

McBratney, A.B. and Webster, R. 1986. Choosing Functions for Semi-Variograms of Soil Properties and Fitting them to Sampling Estimates. *Journal of Soil Science*, 37, 617-639.

Miller, D.P. 1988. Low-level Determination of Nitrogen Dioxide in Ambient Air Using the Palmes Tube. *Atmospheric Environment*, vol 22 (5), 945-947.

Murakami, M., Ono, M. and Tamura, K. 1990. Health Problems of Residents Along Heavy-Traffic Roads. *Journal of Human Ergology*, 19, 101-106.

Nitta, H., Sato, T. Naki, S., Maeda, K., Aoki, S. and Ono, M. 1993. Respiratory health associated with exposure to automobile exhaust: I. Results of cross-sectional study in 1979, 1982 and 1983. *Archives of Environmental Health*, 48, 53-8.

Nystad, W., Magnus, P., Gulsvik, A., Skarpaas, I.J.K and Carlsen, K-H. 1997. Changing Prevalence of Asthma in School Children: Evidence for Diagnostic Changes in Asthma in Two Surveys 13 Years Apart. *European Respiratory Journal*, 10, 1046-1051.

Oliver, M. A. and Webster, R. 1990. Kriging: a Method of Interpolation for Geographical Information Systems. *International Journal of Geographical Information Systems*, 4 (3), 313-332.

Oosterlee, A., Drijver, M., Lebret, E. and Brunekreef, B. 1996. Chronic Respiratory Symptoms in Children and Adults Living Along Streets with High Traffic Density. *Occupational and Environmental Medicine*, 53, 241-247.

Palmes, E.D., Gunnison, A.F., DiMattio, J. and Tomczyk, C. 1976. Personal Sampler for Nitrogen Dioxide. *Am. Ind. Hyg. Ass. J.* 37, 570-577.

Pannatier, Y. 1994. VARIOWIN 2.0. Product of a completed PhD.

Peat, J.K., Haby, M., Spijker, J., Berry, G. and Woolcock, A.J. 1992. Prevalence of Asthma in Adults in Busselton, Western Australia. *British Medical Journal*, 305, 1326-1329.

Peters, A., Dockery, D.W., Heinrich, J. and Wichmann, H.E. 1997. Short-Term Effects of Particulate Air Pollution on Respiratory Morbidity in Asthmatic Children. *European Respiratory Journal*, 10, 872-879.

Pikhart, H., Bobak, M., Kriz, B., Danova, J., Celko, M., Prikazsky, V., Pryl, K., Briggs, D. and Elliott, P. 1997. Outdoor Air Concentrations of Nitrogen Dioxide and Sulphur Dioxide and Prevalence of Wheezing in School Children in Two Districts of Prague. Submitted for publication.

Photochemical Oxidants Review Group. 1993. *Ozone in the United Kingdom 1993*. The Third Report of the Photochemical Oxidants Review Group, AEA Technology, Harwell Laboratory, Oxfordshire.

Ponce de Leon, A., Anderson, H.R., Bland, J.M., Strachan, D.P. and Bower, J. 1996. Effects of Air Pollution on Daily Hospital Admissions for Respiratory Disease in London between 1987-88 and 1991-92. *Journal of Epidemiology and Community Health*, 33(Suppl 1), S63-S70.

Pönkä, A. and Virtanen, M. 1996. Asthma and Ambient Air Pollution in Helsinki. *Journal of Epidemiology and Community Health*, 50 (Supp 1), S59-S62.

Pope, C.A.III, Thun, M.J., Namboodiri, M.M., Dockery, D.W., Evans, J.S., Speizer, F.E. and Heath, C.W.Jr. 1995a. Particulate Air Pollution as a Predictor of Mortality in a Prospective Study of U.S. Adults. *American Journal of Respiratory and Critical Care Medicine*, 151, 669-674.

Pope, C.A.III, Bates, D.V. and Raizenne, M.E. 1995b. Health Effects of Particulate Air Pollution: Time for Reassessment? *Environmental Health Perspectives*, 103(5), 472-480.

Pratt, N.D., Bird, A.C., Taylor, J.C. and Carter, R.C. 1997. Estimating Areas of Land Under Small-Scale Irrigation Using Satellite Imagery and Ground Data for a Study Area in N.E. Nigeria. *International Journal of Remote Sensing*, 163(1), 65-77.

Quality of Urban Atmospheric Review Group. 1993. *Urban Air Quality in the United Kingdom. First Report of the Quality of Urban Atmosphere Review Group*, Department of the Environment, Bradford.

Rasmassen, M.S. 1997. Operational Yield Forecast Using AVHRR NDVI data: Reduction of Environmental and Inter-Annual Variability. *International Journal of Remote Sensing*, 18(5), 1059-1077.

van Reeuwijk, H., Lebret, E., Fisher, P.H., Smallbone, K., Celko, M. and Harssema, H. 1995. Performance of NO<sub>2</sub> Passive Samplers in Ambient Air in Dense Networks. (Abstract). *Epidemiology*, 6 (4), S60.

Reeuwijk van, H., Fischer, P.H., Harssema, H., Briggs, D.J., Smallbone, K. and Lebret, E. 1997. Field Comparison of Two NO<sub>2</sub> Passive Samplers to Assess Spatial Variation. Accepted for publication in Environmental Monitoring and Assessment.

Rodden, J. B., Green, N.J., Messina, A.D. and Bullin, J.A. 1982. *Journal of the Air Pollution Control Association*, 32 (12), 1226-1228.

Royal Commission of Environmental Pollution. 1994. *Transport and the Environment*. HMSO.

Russell, A.G. 1988. *Mathematical Modelling of the Effect of Emission Sources on Atmospheric Pollutant Concentrations*. In Air Pollution, the Automobile and Public Health. Edited by Kennedy, D. Health effects Institute, National Academy Press, Washington D.C., 161-205.

Schaug, J., Iversen, T. and Pedersen, U. 1993. Comparison of Measurements and Model Results for Airborne Sulphur and Nitrogen Components with Kriging. *Atmospheric Environment*, 27A (6), 831-844.

Schwartz, J. 1993. Particulate Air Pollution and Chronic Respiratory Health. *Environmental Research*, 62, 7-13.

Simpson, D., Perrin, D.A., Varey, J.E. and Williams, M.L. 1990. Dispersion Modelling of Nitrogen Oxides in the United Kingdom. *Atmospheric Environment*, 24A (7), 1713-1733.

Smallbone, K. 1998. Mapping Ambient Urban Air Pollution at the Small Area Scale: A GIS Approach. Unpublished PhD Thesis. University of Huddersfield.

Stedman, J.R. 1995. Estimated High Resolution Maps of the United Kingdom Air Pollution Climate. Harwell: AEA Technology report AEA/CS/16419035/001.

The United Kingdom National Air Quality Strategy. 1997. Department of the Environment. Stationery Office Ltd.

Trotter, C.M., Dymond, J.R. and Goulding, C.J. 1997. Estimation of Timber Volume using Landsat TM. *International Journal of Remote Sensing*, 18(10), 2209-2223.

Tyler, A.L., MacMillan, D.C. and Dutch, J. 1996. Models to Predict the General Yield Class of Douglas Fir, Japanese Larch and Scots Pine on Better Quality Land in Scotland. *Forestry*, 69(1).

Vauclin, M., Vieira, R., Vachaud, G. and Nielsen, D.R. 1983. The Use of Cokriging with Limited Field Soil Observations. *Soil Science Society American Journal*, 47 (2), 175-184.

Venables, W.N. and Ripley, B.D. 1994. Modern Applied Statistics with S-Plus. Springer-Verlag, New York.

Wardlaw, A.J. 1993. The Role of Air Pollution in Asthma. *Clinical and Experimental Allergy*, 23(2), 81-96.

Webster, R. and Oliver, M.A. 1990. *Statistical Methods in Soil and Land Resource Survey*. Oxford University Press.

Wichmann, H.E., Hubner, H.R., Malin, E., Köhler, B., Hippke, G., Fischer, D., Bontemps, M., Huenges, R., Rebmann, H., Walzer, H., Wolf, U., Ludwig, H., Pizard-Weyrich, M., Gruner, K., Döller, G. and Herrmann, S. 1989. The relevance of health risks by ambient air pollution, demonstrated by a cross-sectional study of Croup syndrome in Baden-Wurttemburg. *Off Gesundheitwes*, 51, 414-20.

Wieland, S.K., Mundt, K.A., Ruckmann, A., and Keil, U. 1994. Reported wheezing and allergic rhinitis in children and traffic density on streets of residents. *AIR*, 4, 79-84.

Willems, J.J.H. and Hofschreuder, P. 1991. *A Passive Monitor for Measuring Ammonia*. In Field Intercomparison Exercise on Ammonia and Ammonium Measurements. Air Pollution Research Report 37. CNR, Commission of the European Communities.

Wills, J. 1998. Personal Communication.

Wijst, M., Reitmeir, P., Dold, S., Wulff, A., Nicolai, T., von Leoffelholz-Colburg, E., and von Mutius, E. 1993. Road traffic and adverse effects on respiratory health in children. *British Medical Journal*, 307, 596-60.

Wood, G., Oliver, M.A. and Webster, R. 1990. Estimating Soil Salinity by Disjunctive Kriging. *Soil Use and Management*, 6 (3), 97-104.

Wordley, J., Walters, S. and Ayers, J.G. 1997. Short Term Variations in Hospital Admissions and Mortality and Particulate Air Pollution. *Occupational and Environmental Medicine*, 54, 108-116.

World Health Organisation. 1987. Air Quality Guidelines for Europe. WHO Regional Publications, European Series, No 23, Copenhagen.

Yamartino, R.J. and Weigand, G. 1986. Development and evaluation of simple models for the flow, turbulence and pollutant concentration fields within an urban street canyon. *Atmospheric Environment*, 20(11), 2137-56.

## **APPENDIX 1**

### **The *hydro* program**

## HYDRO.F

```
program hydro
```

```
*****
```

```
* Program to transform hydrological DTM data to a format  
* compatible with Idrisi.  
*  
* Author: Susan Collins  
* Last updated: 26/10/92  
*****
```

```
* Parameter statements
```

```
parameter (size=8692)
```

```
* Dimension the arrays
```

```
integer value(1:size,1:10)  
integer totalc(1:50),a  
integer column(1:50),gap,endgap  
integer zero(1:8000)
```

```
* Open the files for input and output
```

```
open (unit=12,file='kirk.inf',status='old')  
open (unit=16,file='inf2.dat',status='unknown')
```

```
* Format statements
```

```
10  format (17x,i6,6x,i4)  
20  format (10i8)  
30  format (i8)
```

```
* Loop to set the zero values to 0
```

```
do 100 I=1,8000  
    zero(i)=0  
100 continue
```

```
* Set the minimum and maximum values
```

```
min=999  
max=0
```

```
* Loop to read the easting and northing of the south west  
* corners of the 1km squared data blocks and find the minimum  
* and maximum easting values
```

```

do 120 i=1,size,41
  read (12,10) value(i,1),value(i,2)
  if (value(i,1).le.min) min=value(i,1)
  if (value(i,1).ge.max) max=value(i,1)

* Loop to read the hydrological data for the 1km squared data
* block

  do 110 k=i+1,i+40
    read (12,20) (value(k,j), j=1,10)
110  continue
120  continue

* Set kmeast to the northing of the first km block and the count
* to zero, to find the number of km blocks in an east direction
* for each northing value

  kmeast=value(1,2)
  count=0

  do 130 i=1,size,41
    if (value(i+41,2).eq.kmeast) then
      count=count+1
    else
      totalc(kmeast)=count+1
      count=0
      kmeast=value(i+41,2)
    endif
130  continue

* Set the width to the maximum number of km blocks in an
* east direction, kmeast to the first northing and the count
* to one

  width=max-min
  kmeast=value(1,2)
  count=1

* For all the east km blocks with a constant northing

  do 180 i=1,size,41
    if (i.eq.count) then
      gap=value(i,1)
      column(gap)=value(i,1)-min
      endgap=width-column(gap)-totalc(kmeast)+1

      do 170 a=1,39,2
        do 140 l=1,column(gap)*20
          write (16,30) zero(l)

```

```
140      continue

        do 150 n=i,i+(totalc(kmeast)*41)-1,41
              write (16,30) (value(n+a,j), j=1,10)
              write (16,30) (value(n+a+1,j), j=1,10)
150      continue

        do 160 l=1,endgap*20
              write (16,30) zero(l)
160      continue
170      continue

        count=count+(totalc(kmeast)*41)
        kmeast=kmeast+1

      endif

180      continue

* Close I/O

      close (unit=12)
      close (unit=16)

    end
```

**APPENDIX 2**

**The Health Questionnaire**

**QUESTIONNAIRE SURVEY OF CHEST PROBLEMS IN SCHOOL CHILDREN  
AGED 7-9 IN THE HUDDERSFIELD AREA**

INFORMATION IN THIS QUESTIONNAIRE WILL BE TREATED AS STRICTLY CONFIDENTIAL AND WILL BE USED FOR STATISTICAL PURPOSES ONLY. PLEASE ANSWER THE QUESTIONS BY TICKING ONE OF THE BOXES OR WRITING IN THE SPACES PROVIDED. IF YOU DO NOT KNOW THE ANSWER TO A QUESTION, PLEASE LEAVE THE BOXES BLANK.

What is the name of the child who brought this questionnaire home from school?

First name \_\_\_\_\_ Surname \_\_\_\_\_

What is the post code of the child's address? \_\_\_\_\_

1-9

Which school does he or she attend? \_\_\_\_\_

10-11

1. Is this child a boy or a girl?

Boy

1

Girl

2 12

2. What is the child's date of birth? \_\_\_\_\_

\_\_\_\_ / \_\_\_\_ / \_\_\_\_  
day month year

13-18

3. Where was the child born?

In the Kirklees area

1

Outside the Kirklees area

2 19

**ABOUT THE HEALTH OF THE CHILD WHO BROUGHT THIS QUESTIONNAIRE HOME FROM SCHOOL:**

4. Has the child *ever* had wheezing or whistling in the chest *at any time* in the past?

Yes

1

No

2 20

If NO, skip to question 10 over the page.

If YES, then please answer questions 5-9.

5. At what age did the wheezing first occur?

Before age 1

1

Age 1 to 3

2

Age 4 to 6

3

Age 7 to 9

4 21

6. Has this child *in the last 12 months* had wheezing or whistling in the chest?

Yes

1

No

2 22

7. How many attacks of wheezing *in the last 12 months* has he or she had?

None

1

1 to 3

2

4 to 12

3

More than 12

4 23

8. Has wheezing *in the last 12 months* ever been severe enough to limit the child's speech to only one or two words at a time between breaths?

Yes

1

No

2 24

9. Has the child's chest *in the last 12 months* sounded wheezy during or after exercise?

Yes

1

No

2 25

## FOR EVERYONE:

10. How many days *in the last 12 months* has the child been absent from school because of illness?
- |              |                          |      |
|--------------|--------------------------|------|
| None         | <input type="checkbox"/> | 1    |
| 1-5          | <input type="checkbox"/> | 2    |
| 6-10         | <input type="checkbox"/> | 3    |
| 11-15        | <input type="checkbox"/> | 4    |
| More than 15 | <input type="checkbox"/> | 5 26 |
11. Has this child been bothered *in the last 12 months* by a wheezy chest, *apart from colds*?
- |     |                          |      |
|-----|--------------------------|------|
| Yes | <input type="checkbox"/> | 1    |
| No  | <input type="checkbox"/> | 2 27 |
12. Has this child *ever* been bothered by attacks of shortness of breath with wheezing?
- |     |                          |      |
|-----|--------------------------|------|
| Yes | <input type="checkbox"/> | 1    |
| No  | <input type="checkbox"/> | 2 28 |
13. Has this child had a dry cough at night *in the last 12 months* apart from a cough associated with a cold or chest infection?
- |     |                          |      |
|-----|--------------------------|------|
| Yes | <input type="checkbox"/> | 1    |
| No  | <input type="checkbox"/> | 2 29 |
14. Has a *doctor ever* said this child had asthma?
- |     |                          |      |
|-----|--------------------------|------|
| Yes | <input type="checkbox"/> | 1    |
| No  | <input type="checkbox"/> | 2 30 |
- If NO, skip to question 16.*  
*If YES, then please answer question 15.*
15. Has this child been treated for asthma by a doctor *in the last 12 months*?
- |     |                          |      |
|-----|--------------------------|------|
| Yes | <input type="checkbox"/> | 1    |
| No  | <input type="checkbox"/> | 2 31 |
6. Did your child suffer from a serious chest illness or severe cold, going to the chest, *in the first two years of life*?
- |     |                          |      |
|-----|--------------------------|------|
| Yes | <input type="checkbox"/> | 1    |
| No  | <input type="checkbox"/> | 2 32 |
7. Apart from after vigorous exercise, has this child *ever* been too breathless to talk?
- |     |                          |      |
|-----|--------------------------|------|
| Yes | <input type="checkbox"/> | 1    |
| No  | <input type="checkbox"/> | 2 33 |
8. Has a *doctor ever* said this child had bronchitis?
- |     |                          |      |
|-----|--------------------------|------|
| Yes | <input type="checkbox"/> | 1    |
| No  | <input type="checkbox"/> | 2 34 |
9. Does the child *usually* cough in the morning in the autumn-winter season?
- |     |                          |      |
|-----|--------------------------|------|
| Yes | <input type="checkbox"/> | 1    |
| No  | <input type="checkbox"/> | 2 35 |
0. How long has the child lived at this address?
- |                    |                          |      |
|--------------------|--------------------------|------|
| Less than one year | <input type="checkbox"/> | 1    |
| 1 to 3 years       | <input type="checkbox"/> | 2    |
| 4 to 6 years       | <input type="checkbox"/> | 3    |
| 7 to 9 years       | <input type="checkbox"/> | 4 36 |
1. If less than 3 years please give previous address and post code:
-

**ABOUT THE HOME:**

22. Do you cook by:  
**TICK ALL BOXES THAT APPLY**
- |                               |                          |         |
|-------------------------------|--------------------------|---------|
| Gas                           | <input type="checkbox"/> | 1       |
| Electricity                   | <input type="checkbox"/> | 2       |
| Other - please specify: _____ |                          | 3 44-46 |
23. Does the home have central heating, or district (communal) heating?  
Yes  1  
No  2 47
24. Do you have other forms of heating in the home?  
**TICK ALL BOXES THAT APPLY**
- |                               |                          |         |
|-------------------------------|--------------------------|---------|
| No other heating              | <input type="checkbox"/> | 1       |
| Open gas fire                 | <input type="checkbox"/> | 2       |
| Electricity                   | <input type="checkbox"/> | 3       |
| Open coal fire                | <input type="checkbox"/> | 4       |
| Gas wall heater               | <input type="checkbox"/> | 5       |
| Other - please specify: _____ |                          | 6 48-53 |
25. Is the child's bedroom heated by:  
**TICK ALL BOXES THAT APPLY**
- |                               |                          |         |
|-------------------------------|--------------------------|---------|
| No heating                    | <input type="checkbox"/> | 1       |
| Open gas fire                 | <input type="checkbox"/> | 2       |
| Electricity                   | <input type="checkbox"/> | 3       |
| Central heating radiator      | <input type="checkbox"/> | 4       |
| Other - please specify: _____ |                          | 5 54-58 |
26. Have you noticed any damp spots or mould on the walls of your home during the past two years?  
Yes  1  
No  2 59
27. Are you disturbed by traffic noise and/or traffic fumes in your home?  
Yes  1  
No  2 60
28. Do or did you have pets *with fur or feathers* in the home?  
Yes, since 19 \_\_\_\_\_  
year  
No, not since 19 \_\_\_\_\_  
year  
No, never  3 61-62 63-64 65

**ABOUT CIGARETTE SMOKING IN THE HOME:**

29. Does the mother of this child smoke at present?  
Yes, started smoking in 19 \_\_\_\_\_  
year  
No, stopped smoking in 19 \_\_\_\_\_  
year  
No, has never smoked  70
- If YES, then:  
How many cigarettes does the mother smoke *on average per day inside the home?* \_\_\_\_\_ 71-72
30. How many other members of the household smoke *inside the home?* 73-74
31. In total about how many cigarettes are smoked *on average per day inside the home?* 75-77

## ABOUT THE CHILD:

32. Is the child:	White	<input type="checkbox"/>	1		
	Asian (Indian, Pakistani, Bangladeshi)	<input type="checkbox"/>	2		
	Chinese, Malaysian	<input type="checkbox"/>	3		
	Afro-Caribbean	<input type="checkbox"/>	4		
	Mixed race	<input type="checkbox"/>	5		
	Other	<input type="checkbox"/>	6 78		
33. How many <i>older</i> brothers and sisters does the child have?	— —		79-80		
34. How many <i>younger</i> brothers and sisters does the child have?	— —		81-82		
35. How does the child <i>normally</i> travel to and from school? <b>TICK ALL BOXES THAT APPLY</b>	Walks	<input type="checkbox"/>	1		
	Cycles	<input type="checkbox"/>	2		
	Car	<input type="checkbox"/>	3		
	Bus/train	<input type="checkbox"/>	4		
	Other - please specify:	_____	5 83-87		
6. Has the mother or the father of the child had any of the following conditions, <i>now or in the past</i> ?					
<b>TICK ALL BOXES THAT APPLY</b>	<b>MOTHER</b>		<b>FATHER</b>		
	Asthma	<input type="checkbox"/>	Asthma	<input type="checkbox"/>	1
	Hay fever	<input type="checkbox"/>	Hay fever	<input type="checkbox"/>	2
	Atopic eczema	<input type="checkbox"/>	Atopic eczema	<input type="checkbox"/>	3
	None of the above	<input type="checkbox"/>	None of the above	<input type="checkbox"/>	4 88-95

## ABOUT YOU:

7. What is your relationship to the child?	Mother of the child	<input type="checkbox"/>	1
	Father of the child	<input type="checkbox"/>	2
	Other - please specify:	_____	3 96
8. What type of school (or course of education) did <i>you</i> last attend?	None	<input type="checkbox"/>	1
	Primary	<input type="checkbox"/>	2
	Secondary	<input type="checkbox"/>	3
	Further education	<input type="checkbox"/>	4
	University	<input type="checkbox"/>	5 97
9. About how many years of schooling or formal education have <i>you</i> completed? _____			98-99

We would like to be able to contact you, if we need to, for further information. Could you please provide your name and address. PLEASE PRINT CLEARLY.

our name \_\_\_\_\_

our address \_\_\_\_\_

Post code \_\_\_\_\_

Thank you for taking the time to answer these questions. If you wish to add anything please use a separate sheet of paper. Please check carefully that you have not missed anything, then return the questionnaire promptly to your child's school in the envelope provided.

## **APPENDIX 3**

### **The *consect* Program**

## **CONSECT.F**

program consectetur

```
*****  
* Program to reduce the results from the CALINE3 dispersion  
* model, extract the concentrations, then give each scenario a  
* unique code number and finally export the data to INFO  
* format. Storing with each record, the code and concentrations  
* at all receptors for that scenario.  
*  
* Author: Susan Collins  
* Last Update: 03/02/94  
*****
```

```
* Call the 'reformat' subroutine to remove unwanted lines from  
* the input file, leaving the wind speed and direction, stability and  
* the concentrations
```

### **CALL REFORMAT**

```
* Call the 'dataret' subroutine to create a unique code and transform  
* the data to INFO format.
```

### **CALL DATARET**

end

```
* Reformat subroutine
```

### **SUBROUTINE REFORMAT**

```
* Parameter statements
```

parameter (n=7570)

```
* Dimension the arrays
```

character cdata(n)\*70

```
* Open the files for input and output
```

```
open (unit=12,file='mway1.lst',status='old')  
open (unit=14,file='mway2.lst',status='unknown')
```

```
* Format statements
```

10 format (a70)

```

* Read input data

do 100 i=1,n
  read (12,10) cdata(i)
100 continue

* Main loop - start at line 9 and then loop every 42 lines

do 400 i=9,n,42

* For every run of the main loop write lines i+10 and i+11
* to the new file

do 200 k=i+10,i+11
  write (14,10) cdata(k)
200 continue

* For every run of the main loop write line i+30 to i+41
* to the new file

do 300 k=i+30,i+41
  write (14,10) cdata(k)
300 continue

400 continue

* Close I/O

close (unit=12)
close (unit=14)

end

* Dataret subroutine

SUBROUTINE DATARET

* Parameter statements

parameter (n=2520)

* Dimension the arrays

dimension speed(n)
integer stab(n)
dimension wind(n)
dimension conc(1:n,1:12)
integer code,rd

```

\* Open the input and output files

```
open (unit=12,file='mway2.lst',status='old')
open (unit=14,file='mway3.lst',status='unknown')
```

\* Format Statements

```
10  format (11x,f5.2,24x,i1)
20  format (11x,f5.1)
30  format (62x,f5.1)
40  format (i1,i1,i3,',',12(f5.2,',',))
```

\* Enter the road type and land class in coded form

```
write (*,4)
4   format ('Enter the code for road type: '$)
      read (*,*) rd
6   format ('Enter the code for land class: '$)
      read (*,*) lc
```

\* Main loop

```
do 200 i=1,n,14
```

\* Read in the wind speed and stability

```
read (12,10) speed(i),stab(i)
```

\* Code the wind speed (stability is already in code)

```
if (speed(i).eq.1) speed(i)=10
if (speed(i).eq.3) speed(i)=20
if (speed(i).eq.5) speed(i)=30
if (speed(i).eq.8) speed(i)=40
if (speed(i).eq.15) speed(i)=50
```

\* Read in the wind direction

```
read (12,20) wind(i+1)
```

\* Code the wind direction

```
if (wind(i+1).eq.0) wind(i+1)=100
if (wind(i+1).eq.30) wind(i+1)=200
if (wind(i+1).eq.60) wind(i+1)=300
if (wind(i+1).eq.90) wind(i+1)=400
if (wind(i+1).eq.120) wind(i+1)=500
if (wind(i+1).eq.150) wind(i+1)=600
```

\* Define the last three digits of the unique code

```
code = wing(i+1)+speed(i)+stab(i)
```

\* Loop to read the concentrations for the 12 receptors and then

\* write the full unique code, followed by all the concentrations

\* for that code, on one line in INFO format to the new file

```
do 100 k=i+2  
    read (12,30) (conc(k,j), j=1,12)  
    write (14,40) rd,lc,code,(conc(k,j), j=1,12)  
100      continue
```

```
200  continue
```

\* Close I/O

```
close (unit=12)  
close (unit=14)
```

```
end
```

**APPENDIX 4**

**The *TRAFFPOL* Program**

## TRAFFPOL.AML

```
/* ****
/* AML to calculate the near-source pollution concentrations.
/*
/* Author: Susan Collins
/* Last update: 15/5/95
*****
```

/\* Set the arguments

&args cover roadtype traffvolume surfacerough weatherfile

/\* Check that all arguments present

&if [null %cover%] &then &call show\_usage  
&if [null %roadtype%] &then &call show\_usage  
&if [null %traffvolume%] &then &call show\_usage  
&if [null %surfacerough%] &then &call show\_usage  
&if [null %weatherfile%] &then  
 &return &inform Invalid number of arguments

&s closestat = [close -all]

/\* Check that all coverages, items and files exist

&if not [exists %cover% -coverage] &then  
 &return The coverage [translate %cover%] does not exist

&if not [iteminfo %cover% -arc %roadtype% -exists] &then  
 &return The item [translate %roadtype%] does not exist

&if not [iteminfo %cover% -arc %traffvolume% -exists] &then  
 &return The item [translate %traffvolume%] does not exist

&if not [exists %surfacerough% -coverage] &then  
 &return The coverage [translate %surfacerough%] does not exist

&if not [exists %weatherfile%] &then  
 &return The file %weatherfile% does not exist

&if [exists %cover%.txt] &then  
 &return The file %cover%.txt already exists

/\* Open the file to hold the name of the ungenerated line file

&s fileunit = [open roads.save openstat -write]

&if %openstat% ne 0 &then

```

&return Could not open file

/* Ungenerate the line coverage of roads and write results to a new file

ungenerate line %cover% %cover%.txt

&s writestat = [write %fileunit% %cover%.txt]

&s closestat = [close -all]

/* Run the compiled arcdirection program to establish the direction of the
/* lines

arcdirection

&sv ok = [delete holdall2.txt -file]
&sv ok = [delete holdall3.txt -file]
&sv ok = [delete %cover%.txt -file]

/* Convert the ungenerated data back into a new line coverage and import
/* associated line directions into an INFO table

&type
&if [exists %cover%6 -coverage] &then
  &return The coverage [translate %cover%6] already exists

generate %cover%6;input outarcs.txt;lines;q
build %cover%6 lines

&sv ok = [delete outarcs.txt -file]

&type
&if [exists %cover%6.exp -info] &then
  &return The info data file [translate %cover%6.exp] already exists

&data arc tables
DEFINE [translate %cover%]6.EXP
[translate %cover%]6-ID,4,5,B
DIRECTION,1,2,I
~
ADD FROM [translate outarcs].REL
Q STOP
&end

&sv ok = [delete outarcs.rel -file]

/* Join the INFO table to the new line coverage

joinitem %cover%6.aat %cover%6.exp %cover%6.aat %cover%6-id %cover%6-id

```

```

/* Delete the INFO table
&sv ok = [delete %cover%6.exp -info]

/* Generate the grids of arc direction, road type, traffic volume and surface
/* roughness

linegrid %cover%6 %cover%6grid direction;10;y;NODATA

kill %troads%6

linegrid %cover% rdtypeg %roadtype%;10;y;NODATA

linegrid %cover% traffgrid %traffvolume%;10;y;NODATA

polygrid %surferough% roughgrid surf_code;10;y;NODATA

/* Start grid

grid

/* Run the euclidean distance and euclidean allocation commands on the line grids
/* in the 200m near-source zone

%cover%6dist = euclidean (%cover%6grid,%cover%6dir,%cover%6allo,200)
rdtypeallo = euclidean (rdtypeg,#,#,200)

traffallo = euclidean (traffgrid,#,#,200)

kill %cover%6grid
kill rdtypeg

/* Reclassify the grids

%troads%dist2 = reclass(%troads%6dist,remapdist34.table)

kill %troads%6dist

%troads%6dir2 = reclass (%troads%6dir,remapdir.table)
kill %troads%6dir

/* Establish which side of the road the cell is on (down-wind or up-wind)

docell;abwa := %troads%6dir2 - 0;if (abwa lt 0) abwa := abwa + 360;~
aballo1 = %troads%6allo;&sv .var = 1;&call section;end
docell;abwa := %troads%6dir2 - 30;if (abwa lt 0) abwa := abwa + 360;~
aballo2 = %troads%6allo - 1;if (aballo2 le 0) aballo2 = aballo2 + 6;~
&sv .var = 2;&call section;end
docell;abwa := %troads%6dir2 - 60;if (abwa lt 0) abwa := abwa + 360;~

```

```

aballo3 = %troads%6allo - 2;if (aballo3 le 0) aballo3 = aballo3 + 6;~
&sv .var = 3;&call section;end
docell;abwa := %troads%6dir2 - 90;if (abwa lt 0) abwa := abwa + 360;~
aballo4 = %troads%6allo - 3;if (aballo4 le 0) aballo4 = aballo4 + 6;~
&sv .var = 4;&call section;end
docell;abwa := %troads%6dir2 - 120;if (abwa lt 0) abwa := abwa + 360;~
aballo5 = %troads%6allo - 4;if (aballo5 le 0) aballo5 = aballo5 + 6;~
&sv .var = 5;&call section;end
docell;abwa := %troads%6dir2 - 150;if (abwa lt 0) abwa := abwa + 360;~
aballo6 = %troads%6allo - 5;if (aballo6 le 0) aballo6 = aballo6 + 6;~
&sv .var = 6;&call section;end

kill %troads%6dir2
kill %troads%6allo

/* Open the file to hold the name of the file which contains the weather data
/* and write the file name to it

&s fileunit = [open weather.save openstat -write]

&s writestat = [write %fileunit% %weatherfile%]

&s closestat = [close -all]

/* Run the compiled program wp to establish weights for unique
/* weather periods

&data arc
wp
q
&end

/* Transfer the results to an INFO table

&data arc tables
DEFINE WEATHER.EXP
HALF,1,2,I
WINDDIR,1,2,I
SPEEDSTAB,2,3,I
WEIGHT,5,6,N,3
~

ADD FROM WEATHER.WGTS
Q STOP
&end

/* Pull out the first record of the INFO file with the weighted weather data

cursor wp declare weather.exp info
cursor wp open

```

```

&sv .hf = %:wp.half%
&sv .wd = %:wp.winddir%
&sv ss = %:wp.speedstab%
&sv wg = %:wp.weight%

/* Calculate the five digit CONCODE for all cells

surface2 = (rdtypeallo * 10000) + %roughgrid% + (aballo%.wd% * 100) + %ss% 

/* Establish which side of the road will be in receipt of pollution

&if %.hf% = 1 &then
  leftright = ablr%.wd%
&else &if %.hf% = 2 &then
  &do
    if (ablr%.wd% == 1) leftright = 2
    else if (ablr%.wd% == 2) leftright = 1
    endif
  &end

/* Combine the CONCODE grid with the grid that identifies cells
/* that will be in receipt of pollution and cells that will not

surface3 = combine (surface2,leftright,%troads%dist2)

/* Add a new item to the grid to store the pollution concentrations

&data arc
additem surface3.vat surface3.vat conc 5 6 n 1
&sv ok = [delete weather.wgts -file]
q
&end

calculate surface3.vat info conc = 0.0

/* Link the grid to the table that contains the CONCODEs and concentrations
/* for the 12 receptors

relate add
concsurf
concs.exp
info
surface2
surface2
ordered
ro
~

/* Calculate the weighted concentrations

```

```

&sv numb = 1

&do i = 1 &to 2
  &do j = 1 &to 6

    reselect surface3.vat info leftright = %i% and %troads%dist2 = %j%
    calculate surface3.vat info conc = conc + ( concsurf/rcpt%numb% * %wg% )
    aselect surface3.vat info
    &sv numb = %numb% + 1

  &end
&end

surface4 = surface3.conc

kill leftright
kill surface2
kill surface3

/* The whole process - from selecting a record of weather conditions - is repeated
/* in a loop until the last record is reached

/* Pull out the next record of the INFO file with the weighted weather data

cursor wp next

&do &while %:wp.AML$NEXT% = .TRUE.

  &sv .hf = %:wp.half%
  &sv .wd = %:wp.winddir%
  &sv ss = %:wp.speedstab%
  &sv wg = %:wp.weight%

  /* Calculate the five digit CONCODE for all cells

  surface2 = (rdtypeallo * 10000) + roughgrid2 + (aballo%.wd% * 100) + %ss%

  /* Establish which side of the road will be in receipt of pollution

  &if %.hf% = 1 &then
    leftright = ablr%.wd%
  &else &if %.hf% = 2 &then
    &do
      if (ablr%.wd% == 1) leftright = 2
      else if (ablr%.wd% == 2) leftright = 1
      endif
    &end

  /* Combine the CONCODE grid with the grid that identifies cells

```

```

/* that will be in receipt of pollution and cells that will not
tempsurf1 = combine (surface2,lefright,%troads%dist2)

/* Add a new item to the grid to store the pollution concentrations

&data arc
additem tempsurf1.vat tempsurf1.vat conc 5 6 n 1
q
&end

calculate tempsurf1.vat info conc = 0.0

/* Calculate the weighted concentrations

&sv numb = 1

&do i = 1 &to 2
  &do j = 1 &to 6

    aselect tempsurf1.vat info
    reselect tempsurf1.vat info lefright = %i% and %troads%dist2 = %j%
    calculate tempsurf1.vat info conc = conc + ~
      ( concsurf//rcpt%numb% * %wg% )
    &sv numb = %numb% + 1

  &end
&end

tempsurf2 = surface4 + tempsurf1.conc

kill lefright
kill surface2
kill surface4
kill tempsurf1

surface4 = tempsurf2

kill tempsurf2
cursor wp next

&end

/* traffic vol - hourly rate

surface5 = (surface4 / 4000) * traffallo

relate drop
concsurf

```

~

q

&sv ok = [delete weather.exp -info]

kill surface4  
kill %troads%dist2

&do i = 1 &to 6

kill aballo%o%  
kill ablr%o%

&end  
&return

/\* Command line

&routine show\_usage

&type

&return &inform Usage: traffpol <cover> <roadtype\_item> <traffvol\_item> ~<surface\_rough> <weather\_data>;

&type

&return

/\* Called section routine

&routine section

&sv i = %.var%

if (aballo%o% == 1 && abwa <= 180) ablr%o% = 2  
else if (aballo%o% == 1 && abwa > 180) ablr%o% = 1  
else if (aballo%o% == 2 && (abwa > 30 and abwa <= 210)) ablr%o% = 1  
else if (aballo%o% == 2 && (abwa > 210 or abwa <= 30)) ablr%o% = 2  
else if (aballo%o% == 3 && (abwa > 60 and abwa <= 240)) ablr%o% = 1  
else if (aballo%o% == 3 && (abwa > 240 or abwa <= 60)) ablr%o% = 2  
else if (aballo%o% == 4 && (abwa > 90 and abwa <= 270)) ablr%o% = 1  
else if (aballo%o% == 4 && (abwa > 270 or abwa <= 90)) ablr%o% = 2  
else if (aballo%o% == 5 && (abwa > 120 and abwa <= 300)) ablr%o% = 2  
else if (aballo%o% == 5 && (abwa > 300 or abwa <= 120)) ablr%o% = 1  
else if (aballo%o% == 6 && (abwa > 150 and abwa <= 330)) ablr%o% = 2  
else if (aballo%o% == 6 && (abwa > 330 or abwa <= 150)) ablr%o% = 1

end

&return

## **APPENDIX 5**

### **The *adirect* Program**

## **ADIRECT.F**

program adirect

```
*****  
* Program to code ungenerated arcs from ARC/INFO into six  
* classes of direction.  
*  
* Author: Susan Collins  
* Last update: 28.7.94  
*****
```

character textcov\*20

\* Open file which contains the name of ungenerate data file

open (unit=4,file='roads.save',status='old')

\* Format statement

10 format (a20)  
read (4,10) textcov

\* Call subroutine to count the number of lines in a file

CALL COUNTLINES (textcov,numb)

\* Call subroutine to re-format data from ARC/INFO ungenerate  
\* to a format that is easier to read in FORTRAN

CALL REFORMAT (textcov,numb)

\* Call subroutine to find the bearing between consecutive  
\* pairs of coordinates (i.e. road direction) and to classify  
\* them according to direction

CALL COUNTLINES ('holdall2.txt',numb2)

CALL BEARINGS (numb2)

\* Call subroutine to split arcs according to class

CALL SPLITARC (numb2)

close (unit=4)

end

\* Countlines subroutine

SUBROUTINE COUNTLINES (filename,count)

```
character filename*20
character value*20
integer count

open (unit=12,file=filename,status='old')

10 format (a20)

count = 0

do while (count.ge.0)
  count = count + 1
  read (12,10,end=100) value
end do

100 continue

  count = count - 1

close (unit=12)

end
```

\* Reformat subroutine

SUBROUTINE REFORMAT (covname,nrecs)

\* Parameter statements

parameter (lines = 100000)

\* Dimension the arrays

```
character covname*20
doubleprecision vertex
dimension id(1:lines)
dimension vertex(1:lines,1:2)
character endch*5
```

\* Open the files

```
open (unit=12,file=covname,status='old')
open (unit=14,file='holdall2.txt',status='unknown')
```

\* Format statements

```
10  format (a5,f13.6,5x,f13.6)
20  format (i6)
30  format (2(6x,f13.6))
```

\* Loop to read vertex coordinates and find maximum ID

```
do 100 i=1,nrecs
  read(12,10) endch,(vertex(i,j), j=1,2)

  if (endch.ne.'END') then

    if (vertex(i,2).eq.0) then
      id(i)=vertex(i,1)*1000000
      write(14,20) (id(i))
    else
      write(14,30) (vertex(i,j), j=1,2)
    endif

  endif
100 continue

  close (unit=12)
  close (unit=14)

end
```

\* Bearings subroutine

SUBROUTINE BEARINGS (nrecs2)

parameter (lines=100000)

\* Dimension the arrays

```
doubleprecision vertex
dimension vertex(1:lines,1:2)
dimension id(1:lines)
```

\* Open the files

```
open (unit=12,file='holdall2.txt',status='old')
open (unit=14,file='holdall3.txt',status='unknown')
```

\* Format statements

```
10  format (i6,f13.6,6x,f13.6)
```

```

20  format (2(6x,f13.6),6x,f8.4,6x,i1)

* Loop to calculate bearing between consecutive pairs
* of coordinates

do 100 i=1,nrecs2
  read(12,10) id(i),(vertex(i,j), j=1,2)
100 continue

signeast=1
signnrth=1

do 200 i=1,nrecs2
  if (id(i).eq.0.and.vertex(i+1,2).gt.0) then
    diffeast=vertex(i+1,1)-vertex(i,1)
    diffnrth=vertex(i+1,2)-vertex(i,2)

    if (diffeast.lt.0) then
      signeast=2
      diffeast=diffeast*(-1.0)
    endif

    if (diffnrth.lt.0) then
      signnrth=2
      diffnrth=diffnrth*(-1.0)
    endif

    if (signeast.eq.1.and.diffnrth.eq.0) then
      theta=90.0
    else if (signeast.eq.2.and.diffnrth.eq.0) then
      theta=270.0
    else if (diffeast.eq.0.and.signnrth.eq.1) then
      theta=0.0
    else if (diffeast.eq.0.and.signnrth.eq.2) then
      theta=180.0

    else
      theta=atan(diffeast/diffnrth)
      if (signeast.eq.1.and.signnrth.eq.2) theta=180-theta
      if (signeast.eq.2.and.signnrth.eq.2) theta=180+theta
      if (signeast.eq.2.and.signnrth.eq.1) theta=360-theta
    end if

    if (theta.ge.345.or.theta.lt.15.or.theta.ge.165.and.
+     theta.lt.195) medclass=1
    if (theta.ge.15.and.theta.lt.45.or.theta.ge.195.and.
+     theta.lt.225) medclass=2
    if (theta.ge.45.and.theta.lt.75.or.theta.ge.225.and.
+     theta.lt.255) medclass=3
  endif
end do

```

```
    if (theta.ge.75.and.theta.lt.105.or.theta.ge.255.and.  
+      theta.lt.285) medclass=4  
    if (theta.ge.105.and.theta.lt.135.or.theta.ge.285.and.  
+      theta.lt.315) medclass=5  
    if (theta.ge.135.and.theta.lt.165.or.theta.ge.315.and.  
+      theta.lt.345) medclass=6
```

```
    write(14,20) (vertex(i,j), j=1,2),theta,medclass  
  else  
    write(14,10) id(i),(vertex(i,j), j=1,2)  
  endif
```

```
signeast=1  
signnrth=1
```

200 continue

```
close (unit=12)  
close (unit=14)
```

```
end
```

\* Splitarc subroutine

SUBROUTINE SPLITARC (nrecs2)

parameter (lines=100000)

\* Dimension the arrays

```
doubleprecision vertex  
dimension vertex(1:lines,1:2)  
dimension id(1:lines)  
dimension medclass(1:lines)
```

\* Open the files

```
open (unit=12,file='holdall3.txt',status='old')  
open (unit=14,file='outarcs.txt',status='unknown')  
open (unit=16,file='outarcs.rel',status='unknown')
```

\* Format statements

```
10 format (i6,f13.6,6x,f13.6,20x,i1)  
20 format (2(6x,f13.6))  
30 format (i6)  
40 format (a3)  
50 format (i6,a1,i1)
```

\* Loop to read in data and find maximum ID

```
maxid=0

do 100 i=1,nrecs2
  read(12,10) id(i),(vertex(i,j), j=1,2),medclass(i)
  if (id(i).gt.maxid) maxid=id(i)
100 continue
```

\* Loop to generalise classes for any single changes in class

```
do 200 i=2,nrecs2
  if (medclass(i-1).eq.0.and.medclass(i).ne.medclass(i+1).and.
+    medclass(i).ne.0.and.medclass(i+1).ne.0)
+    medclass(i)=medclass(i+1)
    if (medclass(i-1).gt.0.and.medclass(i).ne.medclass(i-1).
+      and.medclass(i).ne.medclass(i+1)) medclass(i)=medclass(i-1)
200 continue
```

\* Loop to split arcs for changes in class

```
write(14,30) id(1)
write(16,50) id(1),'',medclass(2)

do 300 i=2,nrecs2-1
  if (medclass(i).eq.0.and.medclass(i+1).gt.0) then
    write(14,40) 'END'
    write(14,30) id(i)
    write(16,50) id(i),'',medclass(i+1)
  else if (medclass(i).eq.0.and.medclass(i+1).eq.0) then
    write(14,20) (vertex(i,j), j=1,2)
```

\* This section splits an arc if (and then where) the medclass  
\* changes along its length.

```
else if (medclass(i).gt.0.and.medclass(i+1).gt.0.and.
+  medclass(i).ne.medclass(i+1)) then
  write(14,20) (vertex(i,j), j=1,2)
  write(14,20) (vertex(i+1,j), j=1,2) !create end node
  write(14,40) 'END'
  maxid=maxid+1
  write(14,30) maxid
  write(16,50) maxid,'',medclass(i+1)
else
  write(14,20) (vertex(i,j), j=1,2)
endif
300 continue

write(14,20) (vertex(nrecs2,j), j=1,2),medclass(i)
```

```
write(14,40) 'END'  
write(14,40) 'END'
```

```
close (unit=12)  
close (unit=14)  
close (unit=16)
```

```
end
```

## **APPENDIX 6**

### **The *weather* Program**

## **WEATHER.F**

program weather

```
*****  
* Program to generate weather periods and assosciated weights.  
*  
* Author: Susan Collins  
* Last update: 25.05.95  
*****
```

character infile\*20

open (unit=14,file='weather.save',status='old')

\* Read in data

5 format (a20)  
read (14,5) infile

\* Call the subroutine to count the number of lines in a file

**CALL COUNTLINES (infile,numb)**

\* Call the subroutine to establish the weather periods and weights

**CALL PERIOD (infile,numb)**

close (unit=14)

end

\* Countlines subroutine

**SUBROUTINE COUNTLINES (filename,count)**

character filename\*20  
character value\*20  
integer count

open (unit=12,file=filename,status='old')

10 format (a20)

count = 0

do while (count.ge.0)  
count = count + 1

```
    read (12,10,end=100) value
end do

100 continue

count = count - 1

close (unit=12)

end
```

\* Subroutine period

**SUBROUTINE PERIOD (filename,numb)**

\* Parameter statements

```
parameter (lines = 40000)
parameter (unip = 2699)
```

\* Dimension the arrays

```
character filename*20
integer wdata(1:lines,1:4)
integer freq(1:unip)
integer unival(1:lines)
real weight(1:unip)
integer lr(1:unip)
integer wd(1:unip)
integer ss(1:unip)
```

\* Open the files for input and output

```
open (unit=12,file=filename,status='old')
open (unit=4,file='weather.wgts',status='new')
```

10 format (3x,i3,2x,i1,3x,i1)

\* Set all the variables to zero

```
do 100 m=1,unip
freq(m) = 0
lr(m) = 0
wd(m) = 0
ss(m) = 0
100 continue
```

\* Read in the wind direction, speed and stability

```
do 150 i=1,numb  
  read (12,10) (wdata(i,j), j=2,4)  
150 continue
```

\* Check that the wind direction is within the 0-360 degrees range

```
do 200 i=1,numb  
  if (wdata(i,1).gt.360.or.wdata(i,1).lt.0) then  
    write (*,*) 'DATA INPUT ERROR -- degrees'  
    goto 999  
  endif  
200 continue
```

\* Establish up-wind or down-wind code depending on wind direction

```
do 250 i=1,numb  
  if (wdata(i,2).ge.345.or.wdata(i,2).lt.165) wdata(i,1) = 1  
  if (wdata(i,2).ge.165.and.wdata(i,2).lt.345) wdata(i,1) = 2  
250 continue
```

\* Code wind direction into one of 6 classes

```
do 300 i=1,numb  
  if (wdata(i,2).ge.345.or.wdata(i,2).lt.15) wdata(i,2) = 1  
  if (wdata(i,2).ge.15.and.wdata(i,2).lt.45) wdata(i,2) = 2  
  if (wdata(i,2).ge.45.and.wdata(i,2).lt.75) wdata(i,2) = 3  
  if (wdata(i,2).ge.75.and.wdata(i,2).lt.105) wdata(i,2) = 4  
  if (wdata(i,2).ge.105.and.wdata(i,2).lt.135) wdata(i,2) = 5  
  if (wdata(i,2).ge.135.and.wdata(i,2).lt.165) wdata(i,2) = 6  
  if (wdata(i,2).ge.165.and.wdata(i,2).lt.195) wdata(i,2) = 1  
  if (wdata(i,2).ge.195.and.wdata(i,2).lt.225) wdata(i,2) = 2  
  if (wdata(i,2).ge.225.and.wdata(i,2).lt.255) wdata(i,2) = 3  
  if (wdata(i,2).ge.255.and.wdata(i,2).lt.285) wdata(i,2) = 4  
  if (wdata(i,2).ge.285.and.wdata(i,2).lt.315) wdata(i,2) = 5  
  if (wdata(i,2).ge.315.and.wdata(i,2).lt.345) wdata(i,2) = 6  
300 continue
```

\* Code wind speed into one of five classes

```
do 350 i=1,numb  
  if (wdata(i,3).ge.0.and.wdata(i,3).lt.2) wdata(i,3) = 1  
  if (wdata(i,3).ge.2.and.wdata(i,3).lt.4) wdata(i,3) = 2  
  if (wdata(i,3).ge.4.and.wdata(i,3).lt.6) wdata(i,3) = 3  
  if (wdata(i,3).ge.6.and.wdata(i,3).lt.10) wdata(i,3) = 4  
  if (wdata(i,3).ge.10) wdata(i,3) = 5  
350 continue
```

\* Calculate the individual weather period

```
do 400 i=1,numb
    unival(i) = (wdata(i,1)*1000) + (wdata(i,2)*100) +
+        (wdata(i,3)*10) + wdata(i,4)
400 continue
```

\* Calculate the frequency of individual weather periods

```
do 500 m=1,unip
    do 450 i=1,numb
        if (unival(i).eq.m) then
            freq(m) = freq(m) + 1
            lr(m) = wdata(i,1)
            wd(m) = wdata(i,2)
            ss(m) = (wdata(i,3)*10) + wdata(i,4)
        endif
450 continue
500 continue
```

```
20 format (i1,',',i1,',',i2,',',f5.3)
```

\* Calculate the weights for the weather periods based upon the frequency

```
do 550 m=1,unip
    if (freq(m).gt.0) weight(m) = real(freq(m))/real(num)
550 continue
```

\* Write the weather periods and weights to a new file

```
do 600 m=1,unip
    if (freq(m).gt.0) write (4,20) lr(m),wd(m),ss(m),weight(m)
600 continue
```

\* Close I/O

```
999 close (unit=12)
    close (unit=4)

end
```

## **APPENDIX 7**

### **The *rel\_relief* Program**

## **REL\_RELIEF.F**

```
program rel_relief
```

```
*****  
* Program to calculate the relative relief (degree of  
* exposure from the DTM for a 3x3 neighbourhood.  
*  
* Author: Susan Collins  
* Last updated: 01/08/94  
*****
```

```
* Dimension the arrays
```

```
character*20 head(6)  
integer cell (1:566,1:640)  
integer outcell (1:566,1:640)  
integer diff,topvalue
```

```
* Open the files for input and output
```

```
open (unit=12,file='dtm2.txt',status='old')  
open (unit=14,file='topex.txt',status='unknown')
```

```
* Format statements
```

```
10 format (640i4)
```

```
* Read in the data and remove header
```

```
do 100 i=1,6  
    read (12,*) head(i)  
100 continue
```

```
* Read in the data
```

```
write (*,*) ''  
write (*,*) ' Reading in the data....'
```

```
do 200 i=7,566  
    read (12,10) (cell(i,j), j=1,640)  
200 continue
```

```
* Set the output cells to zero
```

```
do 400 i=7,566  
    do 300 j=1,640  
        outcell(i,j) = 0  
300 continue
```

400 continue

\* Calculate relative value with 3x3 filter

```
write (*,*) ' Filtering the data....'

do 800 i=8,565
  do 700 j=2,639

    if (cell(i,j).ne.999) then

      topvalue = 0

      do 600 k=i-1,i+1
        do 500 l=j-1,j+1

          if (cell(k,l).ne.999) then

            if (cell(k,l).ge.800.and.cell(k,l).lt.900) then
              cell(k,l) = topvalue - 800
            endif

            diff = cell(i,j) - cell(k,l)
            topvalue = topvalue + diff

          endif

        500      continue
        600      continue

        outcell(i,j) = topvalue

      endif

    700      continue
    800      continue
```

\* Write the results to a new file

```
write (*,*) ' Writing the results to a new file...'
write (*,*) ' '
write (14,*) 'ncols           640'
write (14,*) 'nrows           560'
write (14,*) 'xllcorner     398000'
write (14,*) 'yllcorner     402000'
write (14,*) 'cellsize        50'
write (14,*) 'NODATA_value -9999'
```

```
do 900 i=7,566
    write (14,10) (outcell(i,j), j=1,640)
900 continue
```

\* Close I/O

```
close (unit=12)
close (unit=14)
```

```
end
```