The use of GIS in Exposure-Response Studies

A regional study of Air pollution and Noise in southern Sweden

Emilie Stroh

The Department of Physical Geography and Ecosystem Analysis

ISSN: 1650-397X
ISBN: 91-974997-7-3
The use of GIS in Exposure-Response Studies
A regional study of Air pollution and Noise in southern Sweden

Emilie Stroh
The Department of Physical Geography and Ecosystem Analysis
Abstract
The use of geographical information systems (GIS) has enabled the study of geographical and temporal variations in the health of various populations. The opportunity to examine key relationships between the health of a population and environmental characteristics has rapidly led GIS to become an important tool for spatial analysis in health studies, and it is now used for various analyses within the discipline of health. This dissertation focuses on the use of GIS for the analysis of disease rates in relation to exposure to environmental factors such as air pollution and noise.

Three specific studies were carried out, all of them focused on the conditions in Scania, the southernmost county of Sweden. One of these focused on accurate modelling of pollutant fields. Problems associated with this process include the choice of scale and level of aggregation of the data. The effects of scale and data aggregation were analysed by studying the effect of the choice of study area as well as different resolutions in both time and space. The results showed that the size and character of the study area had significant effects on the results. The chosen study areas in Scania showed that if a concentration level varied considerably within the area, as is the case in urban areas, coarser pollutant grids fail to reflect these variations. However, for areas with less spatial variation, for example rural areas, coarser grids can accurately reflect these concentrations. It was also found that the choice of temporal resolution of the modelled concentrations affected the suitability and choice of optimal spatial resolution.

Another study focused on exposure estimates. For these calculations it is of the utmost importance that confounding factors, such as socio-economic or environmental factors, are accounted for in order not to cause bias in the results. Therefore, the annual level of NO$_2$ for different socio-economic groups and areas was analysed to determine whether there were any relations between level of exposure and the socio-economic factors country of birth and educational level. The exposure to air pollutants of these socio-economic groups varied considerably, both within the groups and also between the areas studied. The relationships seen also differed, not only in magnitude but also in direction. The results of the study demonstrated that there was no consistency in how individuals in different socio-economic classes were exposed to air pollutants.

Possible health effects were investigated using an exposure-response study where the level of exposure was related to the health status of various study groups. This was investigated by studying the relationship between noise in residential areas arising from road traffic and its relation to the self-reported health effects, annoyance and disturbance of daily activities for the population of Scania. GIS was also used to estimate the average exposure to road traffic of the study population in Scania, and the results showed that about a third of the population was affected by noise levels of at least 55 dB(A).
Acknowledgments

Writing this part of my dissertation proved to be extremely difficult since I was afraid of missing out someone who had helped me over the years. Hopefully, I have managed to mention all of you, but if you can’t find your name here and you think it should be, please be assured of my gratitude.

First of all, I would like to dedicate this dissertation to my friend and colleague Anna Oudin, who believed in me throughout, and proved herself right when she said that I one day would publish “something” and who never thought any publication would ever be dedicated to her – this dissertation is for you.

Secondly, I would like to thank my supervisors, whose vast knowledge and never-ending energy were, at times, frightening, but mostly just extremely impressive.

Dr Petter Pilesjö – for not only being a supportive tutor, but also a great colleague who knows that a part of the trick is to have fun and enjoy going to work.

Dr Lars Harrie – As a token of my deep gratitude for all your help, your endless support, incredible knowledge, your extraordinary patience with me when I was “fighting windmills”, and your exceptional endurance of my enormous capacity and urge to write long and never-ending sentences, with far too many adjectives, this long and never-ending sentence with far too many adjectives is just for you

Dr Kristina Jakobsson – for all her incredible support and all the positive words and gestures that carried me through. Without her encouragement and interest in my work I would still have had a long way to go.

Thirdly, I would like to thank all my wonderful, helpful colleagues at the GIS Centre for making me want to go to work, even on those cold, rainy Mondays when you are convinced that anyone in their right mind would stay in bed. I would especially like to thank:

Susanna Gustafsson – not only for all the support and help she has given me, but also for putting up with a roommate who thinks she is multitasking and (at the best of times) continues to hack away on the keyboard while she is talking to you and (at the worst of times) is completely lost in thought or mentally absorbed.

Karin Larsson – without her support, as a colleague and friend, this dissertation would probably not exist (and neither would its author).

Fourthly, I would like to thank the GISEPI members at the Department of Environmental and Occupational Medicine, Lund University, who were always very supportive and helpful.

I would also like to thank:

Åke Boalt and his employees at the GIS Department at Region Skåne, who
made me realise that everyone with a huge treasure of data and knowledge doesn’t have to be a fierce dragon.

**Maria Hedberg**, who is not only the best of friends, but who also helped me with the layout of this thesis.

My wonderful **family and friends** who, at times, still believe that I am studying Geology, but are always there, prepared to support me and use justified violence in order to tear me away from my computer. Thank you for reminding me of what’s really important in life.

And finally, **Niclas**, my beloved one; I am grateful to you for a thousand reasons, but above all for your incredible patience and ability to bring sunshine and joy into my life.
**List of Papers**

This thesis is based on the following papers, which are referred to in the text by their Roman numbers:


# Table of Contents

Abstract .................................................................................................................................................. III
Acknowledgments ................................................................................................................................. v
List of Papers ......................................................................................................................................... vii

## PART 1

GIS & HEALTH ................................................................................................................................. 13

1.1 Objectives........................................................................................................................................ 13
  1.1.1 Paper I ........................................................................................................................................... 14
  1.1.2 Paper II ......................................................................................................................................... 14
  1.1.3 Paper III ....................................................................................................................................... 14
1.2 Study area ........................................................................................................................................ 14
1.3 Structure of dissertation .................................................................................................................. 15

EMISSION SOURCES & DISPERSION MODELS ........................................................................... 19

2.1 Pollutants......................................................................................................................................... 20
  2.1.1 Air pollutants ............................................................................................................................... 20
  2.1.2 Noise ............................................................................................................................................ 20
2.2 WHO Guidelines ................................................................................................................................. 21
  2.2.1 Nitrogen Dioxide .......................................................................................................................... 21
  2.2.2 Particles ......................................................................................................................................... 21
  2.2.3 Noise ............................................................................................................................................ 21
2.3 Levels of pollutants in Sweden and Scania ....................................................................................... 22
2.4 Emission sources ................................................................................................................................. 22
  2.4.1 Air pollution ................................................................................................................................. 22
  2.4.2 Noise ............................................................................................................................................ 23
  2.4.3 Sources of air pollution considered in these studies ................................................................. 23
  2.4.4 Sources of noise considered in these studies .............................................................................. 25
2.5 Dispersion Models ............................................................................................................................... 26
  2.5.1 Air pollution ................................................................................................................................. 26
  2.5.2 Noise ............................................................................................................................................ 27
  2.5.3 Dispersion model for air pollutants used in these studies .......................................................... 27
  2.5.4 Dispersion model for noise used in these studies ........................................................................ 28
2.6 A comparative study of two dispersion models ............................................................................... 28
  2.6.1 Data and dispersion models .......................................................................................................... 28
  2.6.2 Methodology ................................................................................................................................. 29
  2.6.3 Results and discussion .................................................................................................................. 29

EXPOSURE MODELLING ..................................................................................................................... 33

3.1 Census data ....................................................................................................................................... 34
  3.1.1 Data sources................................................................................................................................. 34
  3.1.2 Spatially aggregated census data ................................................................................................. 34
  3.1.3 Census data used in these studies .............................................................................................. 35
Part I

Be careful about reading health books. You may die of a misprint.
Mark Twain (1835-1910)
Chapter 1
GIS & Health

There is a connection between a population’s location and its health status. However, it is only recently that the use of geographical information system (GIS) in epidemiological studies has actualised the importance of geographical and temporal variations in populational health (Krieger, 2003). According to Rushton (2003) some authors even suggest that the implementation of GIS, and thus the opportunity to examine key relationships between the health characteristics of a population and environmental characteristics, has developed into a new discipline within spatial epidemiology. Regardless of whether this is true or not, GIS have rapidly become an important tool for spatial analysis in health studies and is now used for mapping disease rates and health information, analysing spatial clustering and health events, studying the effects on health of environmental hazards, investigating the risk and spread of infectious diseases, exploring the ecology of vector-born diseases and analysing access to and the location of health services.

One of the areas in which the use of GIS has recently increased is the analysis of disease rates in relation to exposure to environmental factors such as air pollution. The methodologies and metrics used for these studies vary considerably, and we still have much to learn before we can obtain effective and accurate exposure-response relationships. Fundamental dilemmas are how to obtain reliable and consistent estimates of exposure and exposure-response relationships, over both long and short exposure periods. Other factors that might cause difficulties are differences in the scale and level of aggregation of data which might influence each other. The results can also be biased by attributes such as social, demographic or environmental factors.

The work described in this dissertation was carried out in an effort to try to reduce some of these problems.

1.1 Objectives
GIS is a powerful and commonly used tool for exposure-response studies of environmental factors. However, there are certain limitations within the field which must be considered and thoroughly analysed in order to achieve reliable exposure estimates. The work in this dissertation deals with some of these limitations.
Chapter 1

1.1.1 Paper 1

One difficulty in exposure-response studies is the choice of study area since both the characteristics and size of the area may influence the results. Another difficulty that might arise is possible associations between the levels of air pollution exposure and socio-economic characteristics within an area. Knowledge of such associations is important in order to ensure that the results of the epidemiological studies are not biased. Additionally, if the associations observed are dependent on the size and choice of the study area, generalising the contextual effect of socio-economic factors for too large an area may lead to erroneous findings. The main aim of this study was therefore to describe associations between the levels of mean annual concentrations of air pollution (NO$_2$) and two socio-economic indices, country of birth and level of education, in Scania, the southernmost county of Sweden. A secondary aim was to investigate the possible influence of differences in size of the study area on any associations observed.

1.1.2 Paper II

The overall objective of this study was to determine the optimal resolution for a pollutant database for the county of Scania and how this depends on the interaction between different temporal and spatial resolutions. A secondary objective was to evaluate the loss of information due to the use of a coarse space resolution in the pollutant database. This loss of information must be seen in relation to the accuracy of the estimated pollutant field.

1.1.3 Paper III

The aim of this study was to investigate residential noise arising from road traffic and its relation to self-reported health disturbances of the population in Scania. The data for the study was gathered from a large public health survey conducted in southern Sweden. Data on demography, annoyance and disturbance of daily activities, general health problems regarding concentration, sleep, stress and treatment for hypertension was obtained. GIS was also used to estimate the average exposure to road traffic for the study population in Scania.

1.2 Study area

The study area chosen for this licentiate dissertation is Scania, the southernmost county of Sweden (Figure 1.1). Scania covers around 11,350 km$^2$, which is approximately 2% of the total Swedish land area. The land use is mostly agricultural but in the northern/north-eastern part of the county some areas are forested. The region is relatively densely populated, with more than 1.1 million people living in the area, which constitutes approximately 11% of the total Swedish population. Approximately 67% of the population in Scania lives close to the western coast.
The largest city in the county is Malmö, with a population of about 260,000 (the third largest city in Sweden) and an area of 66 km². The city is the residential focus of the rich agricultural area of Scania.

There are five major motorways running through the county, as well as several harbours and a considerable amount of cargo shipping and ferry transport along the coast. These factors, and the closeness to Copenhagen in Denmark, and the European continent, contribute to high concentrations of air pollutants in the region, compared with most other regions in Sweden.

1.3 Structure of dissertation

The dissertation is divided into two main parts. The first part introduces the reader to the theoretical background and presents some of our groups previous findings, while the second part consists of three papers describing the work making up this dissertation.

Part 1

This part consists of four main chapters (Chapters 2-5) corresponding to the three central components of a GIS-based exposure-response study (Figure 1.2), i.e. modelling concentrations of air pollution, exposure estimates and, finally, estimated health effects.

The first of these chapters (Chapter 2) describes the nature and sources of the pollutants: nitrogen oxides, particles and noise. It also deals with the concept of dispersion modelling, and includes a comparative study of two dispersion models. The second chapter (Chapter 3) describes the procedures used in exposure models, and the use and limitations of different sets of census data. It also presents a study regarding differences in exposure output depending on the
Chapter 1

level of aggregation of the census data. The third chapter (Chapter 4) describes the way in which the exposure estimates can be related to health outcome, and reviews the effects that might be caused by excess exposure. The last chapter (Chapter 5) contains a summary of the papers that are included in Part 2.

Part 2

Paper I
Are associations between socio-economic characteristics and exposure to air pollution a question of study area size? An example from Scania, Sweden. *International Journal of Health Geographics*, 4:30.

Paper II
A study of spatial resolution in pollution exposure modelling.
Submitted.

Paper III
Road traffic noise in Southern Sweden and its relation to disturbance of daily activities, annoyance, and health.
Accepted for publication in the *Scandinavian Journal of Work, Environment & Health*.

In Paper I, the author coordinated the study and also preformed the necessary data extraction, exposure modelling and wrote the main part of the paper (except for the section on Statistical analysis in Methods). In Paper II the author conducted part of the programming and analysis and contributed to most of the manuscript (not chapter 2). In Paper III the author performed the exposure modelling, data extraction and supported the GIS operations.
Figure 1.2: Schematic illustration of the procedure adopted in the present work: modelling pollutant values for an area from emission sources, estimating the exposure of the population and estimating the health effects on the exposed population.
Numerous models are available to calculate the dispersion of airborne pollutants in our environment. The quality of the results given by these models is not only dependent on the amount and quality of the emission data, but also on the level of generalisation of the model.

The overall aim of this chapter is to describe the various processes involved in generating pollutant values for an area using emission sources and dispersion models (Figure 2.1). Some major pollutants commonly associated with various health problems are described in Section 2.1 together with their WHO guidelines (Section 2.2) and the levels of these pollutants in the study area, Scania (Section 2.3). Thereafter, follows a section describing emission sources in general and those used in this study (Section 2.4). Section 2.5 deals with general differences between diverse dispersion models and a description of the dispersion model used in this study. The last section (Section 2.6) describes a comparative study of two dispersion models conducted within the SNAP programme ("Swedish National Air Pollution and Health Effects Programme").
Chapter 2

2.1 Pollutants

Air, together with food and water, is one of the three basic requirements for our survival; a human being inhales approximately 10-20 m$^3$ of air per day (WHO, 2000). Unfortunately, what we need for our survival also contains a vast number of toxic pollutants, which might be directly harmful to our health.

2.1.1 Air pollutants

Numerous chemical species of nitrogen oxides ($\text{NO}_x$) exist, but the air pollutant species of most interest from the point of view of human health is nitrogen dioxide ($\text{NO}_2$) (WHO, 2000). According to the WHO (2006) nitrogen dioxide is an important atmospheric trace gas, not only because of its health effects, but also because it absorbs visible solar radiation which may lead to impaired atmospheric visibility, as well as having an effect on global climate change. Under ambient conditions, nitrogen oxides are rapidly transformed into nitrogen dioxide by atmospheric oxidants such as ozone (Brunekreef and Holgate, 2002).

Particulate air pollution may be solid, liquid, or a combination of solid and liquid particles, suspended in the air (Brunekreef and Holgate, 2002). The sizes range from a few nm to tens of µm. The coarser fractions of the particles are mechanically produced by the erosion of larger particles. Small particles, i.e. < 1 µm, are largely formed by gases, from which the smallest ultra-fine particles (< 0.1 µm) are formed by nucleation resulting from condensation or chemical reactions that form new particles. Particles are usually divided into three fractions depending on their size (groups 2 and 3 being subgroups of previous groups):

1. $\text{PM}_{10}$ (also called thoracic particles) have a diameter less than 10 µm, and are able to penetrate into the lower respiratory system,
2. $\text{PM}_{2.5}$ (also called respirable particles) have a diameter less than 2.5 µm and can penetrate into the gas-exchange region of the lungs,
3. ultra-fine particles have a diameter less than 100 nm. These particles contribute little to particle mass but are plentiful in number and their small size increases the degree of lung penetration.

2.1.2 Noise

According to Rosenlund (2005, p. 10) noise can be defined as: “unwanted sound that is characterized according to its frequency, temporal occurrence (day or night) and strength”.

In order to quantify noise, measures of sound pressure levels are often used. These levels quantify the sound, not only by the pressure level, but also by the variation in level with the time and the frequency of the sound. According to King and Davis (2003) there are three commonly used measures:
1. SEL – the sound level exposure,
2. $L_{\text{A,max}}$ – the maximum sound level exposure, and
3. $L_{\text{A,eq}}$ – the average energy equivalent level.

These three measures usually involve a weighted correction. This is a frequency dependent adjustment which is applied to sound of moderate intensity to imitate the varying sensitivity of the ear to various frequencies. For example, this correction will weight a lower frequency less than mid- or high-frequency sounds.

2.2 WHO Guidelines

The WHO (World Health Organization) is the United Nations ‘specialised agency for health and the organisations’ constitutional objective is “the attainment by all peoples of the highest possible level of health” (WHO, 2006). To achieve this goal the WHO has developed a number of guidelines regarding health issues. Based on new scientific information the guidelines for nitrogen dioxides and particulate matter are currently being re-evaluated. The guidelines presented in this section are those in force at the time of writing this dissertation (July 2006).

2.2.1 Nitrogen Dioxide

The limit for exposure to nitrogen dioxide, based on human clinical data, is a 1-hour exposure to 200 µg/m$^3$. The long-term guideline for nitrogen dioxide is 40 µg/m$^3$. This value (for want of a sufficient number of long-term health studies) is based on a review conducted for the Environmental Health Criteria document on nitrogen oxides (WHO, 2000).

2.2.2 Particles

According to the WHO (2000), health effects resulting from exposure to particulate matter have been observed at levels well below 100 µg/m$^3$ (PM$_{10}$), in short-term studies, and 20 µg/m$^3$ (as PM$_{2.5}$) or 30 µg/m$^3$ (as PM$_{10}$) in long-term studies. The information currently available does not allow an estimate to be made of concentrations below which no effects would be expected. For this reason the WHO provides no guideline value for average concentrations of particulate matter.

2.2.3 Noise

The WHO report, Guidelines for Community Noise (Berglund et al., 1999), recommends a daytime $L_{\text{A,eq}}$ value lower than 50-55 dB outdoors and a level of 30-35 dB for the indoor environment, with a $L_{\text{A,max}}$ of 45 dB in bedroom facilities. Night-time values should be at least 5-10 dB lower than the daytime values.
Chapter 2

2.3 Levels of pollutants in Sweden and Scania

According to a study by Welinder et al. (2003) the background levels of NO\textsubscript{2} in most urban areas in Sweden range between 10 and 25 µg/m\textsuperscript{3}, but the levels in areas with dense traffic may range between 30 and 50 µg/m\textsuperscript{3}. According to the present work (Paper I) the median level of NO\textsubscript{2} in Scania is approximately 14 µg/m\textsuperscript{3}, with a minimum value of 3 µg/m\textsuperscript{3} and a maximum value of 23 µg/m\textsuperscript{3} for the year 2001.

Regarding particulate matter, the annual mean levels in two of the major cities in Scania (Malmö and Lund) are approximately 20-30 µg/m\textsuperscript{3} in environments with limited traffic (Welinder et al., 2003). Half of the particles in the rural areas of Sweden are considered to originate from long-distance sources (Welinder et al., 2003).

Öhrström et al. (2006) recently found that approximately 2 million persons (= 25% of the population) in Sweden were exposed to traffic noise exceeding the WHO guidelines, and nearly one million adults are disturbed by noise in their homes, such as traffic and neighbours. In the present work (Paper III) 29% of the population in Scania was found to have an average exposure to noise of 55 dB(A), or more, at their place of residence.

2.4 Emission sources

Emission sources can be natural or artificial. Large quantities of pollutants originate from artificial sources, but the vast majority of sources are natural processes taking place in our surroundings. Although this natural contribution of toxic chemicals, particles and noise is present in our environment, industrialisation has rapidly increased emissions and thus the environmental stress not only on our surroundings, but also on ourselves.

2.4.1 Air pollution

Natural processes such as geothermal and bacterial activity emit gases and chemical compounds in such quantities that they can be considered noxious to the environment. Particles generated from nearby fields, seas, arid land or deserts are also dispersed into the air in large quantities. However the proportion of anthropogenic pollution is increasing and has done so since the beginning of the industrial era.

The major source of anthropogenic emissions of NO\textsubscript{X} into the atmosphere is the combustion of fossil fuels in stationary facilities and motor vehicles (Brunekreef and Holgate, 2002). Industry and combustion used for heating and incineration also account for a large proportion of the pollutants in the atmosphere. However, many countries have begun to put pressure on their industries through regulations on emissions (Borrego et al., 2003). Instead the increasing
amount of passenger and freight transport has now become the dominant source of air pollution, and one of the major problems associated with quality of life in urban areas (Bhat and Sen, 2006; Borrego et al., 2003). Even forms of transport such as planes, ferries and trains contribute to the rising levels of air pollutants. A large amount of air pollutants in the atmosphere is also the effect of domestic combustion which, in densely populated areas, can amount to hazardous levels (Kebin et al., 2002).

2.4.2 Noise
Noise as an environmental health hazard has a diffuse origin. Noise is produced by almost everything in our surroundings, but the strength of the noise is not always related to the degree of disturbance and the environmental stress it causes. King and Davis (2003, p.123) states: “from a physical standpoint, sound and noise are the same thing. However, when the person hearing the sound defines it as a noise, it can cause negative health effects”. Consequently, the level of disturbance experienced is an individual matter, and the same sound level of noise from two different sources, e.g. the sea and a train, may generate completely opposite physiological responses.

Most disturbing noise comes from vehicles such as cars, trains, planes and ferries, but other artificial sources, such as construction and industrial sites, also contribute to noise pollution.

2.4.3 Sources of air pollution considered in these studies
There are two fundamental principles on which an emission database is built up: the top-down and the bottom-up approaches (Gustafsson et al., 2006). In the top-down approach it is necessary to know the total amount of emission in a region. The total emission is then distributed over the area, often in relation to knowledge about e.g. population density or industrial areas. The other principle is the bottom-up approach. In order to use this we need more detailed emission data, such as emission databases for traffic-generated pollution containing information about traffic flow, number of roads in the area, type of vehicles and roads, speed limits etc. These data are then used to estimate the total amount of emission in the region (Gustafsson et al., 2006).

In the studies described in this dissertation an emission database was created using the bottom-up approach (Gustafsson et al., 2006). The database contains emission sources for the region of Scania (and to some extent also Zealand, Denmark). The input data concern mainly nitrogen oxides (NO\textsubscript{X}) and particulate matter (PM\textsubscript{10}).

The database structure is primarily designed for GIS applications and therefore contains coordinated emission sources such as point, line, polygon, and grid sources. A typical point source is a chimney, while a line source could be a
Chapter 2

road. Polygon and grid sources differ in that, for polygon sources, the emission is equally dispersed over the area, while for a grid source the emission can vary in space.

The emission sources in the database can be divided into eight categories (Gustafsson et al., 2006).

1. Road traffic
Road traffic includes vehicles such as passenger cars and different types of heavy vehicles (buses and lorries without trailers, lorries with trailers, and buses driven by gas). The data are also divided into vehicles consuming different fuels (petrol and diesel). The data used in this work originate from the Swedish Road Administration and local municipalities.

2. Shipping
Due to the differences in ship construction, temporal and spatial patterns and the quantities of air pollutants emitted the emissions from shipping are divided into five categories in the database: pleasure boats, larger ships with their home harbour in Scania, cargo ships that depart from and/or tie up in harbours in Scania, ferries that depart from/arrive at harbours in Scania, and cargo ships and other ships that pass the Scanian coast without tying up at harbours. These figures were manually collected for each harbour by Gustafsson et al. (2006).

3. Aviation
The emissions from aviation in the emission database include only emissions below 3000 ft (912 m). The emission data originate from yearly environment reports produced by the Scandinavian airports during the year of 2001.

4. Railroads
Emissions from railroads are fairly small compared to other sources, mainly because most of the railroads in Sweden are electrified and therefore generate low concentrations of air pollutants. All emissions included in the database arise from diesel engines. These are mainly used for cargo transport, but also for some passenger transport on a few lines that are not electrified.
5. Industries and major energy and heat producers
The database contains information from all the large industrial facilities in the area regarding height of chimney, chemical composition of discharge, exhaust gas temperature, etc. Air pollution data were collected from the national database “EMIR” (Emission Register administrated by Sweden’s county administration).

6. Small-scale heating
Information regarding the area’s 150,000 heating installations has been collected. Of these 70,000 installations are oil-burning heaters and 13,000 are wood- or pellet-fuelled heaters. The remaining 67,000 installations are small fireplaces. Depending on information regarding how often the chimney is swept, the frequency and use of the heater can be calculated. This information was gathered from the National Rescue Agency’s chimney register (2001).

7. Construction machinery
This group contains several vehicles, tools and machines used for construction or industrial applications, e.g. caterpillars, farming machinery, lawn mowers, trucks for loading and unloading cargo etc. This group is divided into 11 smaller groups: agriculture, forestry, harbour activities, building, domestic, mining, iron and steel industry, railroads, aviation and military defence. Most information originates from a report published by IVL, ‘Swedish Environmental Research Institute’ (Persson and Kindblom, 1999).

8. Emissions from Zealand, Denmark
Since the emissions from Zealand, Denmark are quite high and westerly winds are dominant in this area, these concentrations must be included. Information regarding the emissions was gathered from an investigation by SMHI (Swedish Meteorological and Hydrological Institute) carried out for Scania’s Air Pollution Association in 2000.

2.4.4 Sources of noise considered in these studies
The emission data used to model noise-abatement zones in these studies consist mainly of road traffic data. Data were available for a total of 21,397 road segments. Information for 17,339 road segments was obtained from the Swedish Road Administration and for 4,058 segments from local municipalities. Data regarding the number of vehicles was available for 82% of the road segments. Speed limits were available for >95% of the segments. Some of the traffic data are
old, but 93% percent is from 1985 or later and 71% of the data is from 1995 or later. For road segments without traffic data mean values were assigned to each segment based on existing data for included road types (Paper III).

2.5 Dispersion Models

Dispersion models are designed to model the distribution of pollutants from emission sources. Most dispersion models are used to model air pollutants but can also be used for modelling the extension and variation of noise-abatement zones. The methods, as well as their accuracy and complexity, differ and consequently also their results.

2.5.1 Air pollution

Particles and chemical compounds dispersed in the air can travel with air masses and winds and within days be in another geographical region or even on the other side of the globe. During transportation air pollutants may interact with other chemical compounds to form new chemical compositions. Consequently, it is not only the air pollutants emitted in our immediate surroundings that we inhale but also those from distant sources, and with unfamiliar chemical compositions.

The accuracy of the dispersion model is crucial in estimating the concentration and distribution of air pollution. The objective is to model the dispersion and concentrations of predefined chemical compounds out of a number of emission sources. Today, several dispersion models are in use.

The number of emission sources used in the different models can vary from one to several thousands in complex cases (Gustafsson et al., 2006). The complexity of the models also varies. Some models are statistical, using measured and established relations between sources and the air pollutants being emitted. Others are based on emission databases with information regarding the quantity, composition and time interval of emissions, as well as meteorological parameters, street-canyon effects, the height of emission sources or the elevation of the landscape. There are also chemical dispersion models that calculate the interactions between chemical compounds or the way in which they change over time or with sun intensity, etc. And, of course, there are also many hybrid models in between these. The appropriate model depends on factors such as: the complexity of the terrain, the size of the study area, the air pollutants of interest, the temporal resolution, the complexity of the meteorological conditions in the study area, and the data available on emission sources and meteorological conditions.
2.5.2 Noise

The parameters and models used to model the dispersion of noise differ, but most of the sources of noise are the same, i.e. road traffic, aviation, railroads and industry. The dispersion models used are divided into different categories depending on the source of the noise. The main categories for these are:

1. noise from road traffic
2. noise from rail traffic
3. noise from air traffic, and
4. noise from external industrial sites

There are many models, of varying complexity, that can be used to generate noise-abatement zones. The more advanced models require large amounts of input data such as source height, the ground type (soft or hard), land use, wind direction, shielding effects, reflection, etc.

2.5.3 Dispersion model for air pollutants used in these studies

The dispersion program used in these studies is ENVIMAN, which has been developed by the company OPSIS AB (Opsis, 2006). ENVIMAN is a combination of the dispersion models AERMOD and OSPM. AERMOD is a Gaussian dispersion model provided by the U.S. Environmental Protection Agency (US EPA, 2002). OPSIS has modified the AERMOD model to fit the emission database structure and meteorology. The altered model can calculate the dispersion over an area using average climatological data, or between specific dates, using the actual meteorological data for these dates with hourly resolution. The model is a flat 2-dimensional model. This means that topography and buildings are not taken into consideration, but the height of the emission source is taken into consideration. The other model, used in ENVIMAN, is a street-canyon model called OSPM (Operational Street Pollution Model) developed by the Danish National Environmental Research Institute (DMU); this model has been adjusted to Swedish conditions and integrated into ENVIMAN (Gustafsson et al., 2006).

According to Gustafsson et al. (2006), the meteorological data used in ENVIMAN can be either climatological data or hourly time series. Climatological data provide a meteorological average over a number of years, while hourly time series consist of measured meteorological values for each hour. To use climatological data in ENVIMAN a minimum of four meteorological parameters are needed:

1. the temperature at a height of 2 m level,
2. wind speed,
3. wind direction at 10 and 24 m above ground,
4. the global solar radiation.
Modelling in ENVIMAN is performed on a grid with adjustable resolution. All emissions, obtained from the emission database, and calculated concentrations are summed to the centroid of each cell in the grid, but the model can also calculate concentrations at receptor points, along lines and over areas.

2.5.4 Dispersion model for noise used in these studies

The dispersion program used to model noise in these studies is based on the Nordic prediction method for road traffic noise (for details see Nielsen, 1996):

$$\text{LA}_{eq} = L_1 + \Delta L_2 + \Delta L_3 + \Delta L_4 + \Delta L_5$$

where

- $\text{LA}_{eq}$ = the weighted equivalent continuous sound pressure level (in dB). This is the primary descriptor of noise from road traffic
- $L_1$ = basic noise level
- $\Delta L_2$ = distance correction
- $\Delta L_3$ = ground and barrier correction
- $\Delta L_4$ = other corrections
- $\Delta L_5$ = façade corrections

The model has been somewhat simplified in that it contains reductions due to distance and ground type (soft or hard) but excludes reduction due to noise barriers (see Eq. 2.1; the terms $L_1$, $\Delta L_2$ and parts of $\Delta L_3$ were included, while the terms $\Delta L_4$ and $\Delta L_5$ were excluded).

2.6 A comparative study of two dispersion models

Within the SNAP project (SNAP, 2006) a comparative study has been conducted of two dispersion models for air pollution. This study was undertaken in order to obtain a rough indication of the range of errors that could be expected, depending on the choice of dispersion model and the number of emission sources included.

2.6.1 Data and dispersion models

The two models compared in the study are described below.

**Model 1:** ENVIMAN: a Gaussian dispersion model modified by OPSIS (OPSIS, 2006) using an emission database created at the GIS Centre, Lund University, Sweden (briefly described in previous sections 2.4.3 and 2.5.3 and in detail by Gustafsson et al., 2006).
Model 2: URBAN: an empirical statistical calculation model created by IVL using measured values of air quality (Sjöberg et al., 2004). In contrast to the Gaussian dispersion model used by Model 1 the concentrations generated by Model 2 include neither the spatial effects of urban structures, such as road networks, nor weather. Instead, the concentrations are based on values measured at different locations (urban and rural areas, streets and roof locations, etc.) and then distributed over the surrounding area by extrapolation. Model 2 can be said to use measured values from city environments which are then spatially reduced from the city border until the concentrations reach the levels of the measured background values. In cities and towns where no air pollution levels have been measured the concentrations can be estimated based on comparison between cities/towns of similar size. The model is described in detail by Sjöberg et al. (2004).

2.6.2 Methodology
In this study both models were used to estimate the annual mean NO\textsubscript{2} (µg/m\textsuperscript{3}) concentration for Scania, southern Sweden. Unfortunately, the emission databases used not only differed in number and kind of emission sources but also in years of origin (2001 for Model 1 and 1999 for Model 2) which of course hampers the possibility of comparing the two models. The spatial resolutions of the pollutant grids were identical, 1x1 km.

The comparison was made by subtracting the output from Model 1 from the output from Model 2, i.e. the value of each concentration cell from Model 1’s pollutant grid was subtracted from the corresponding value given by Model 2’s pollutant grid (Figure 2.2).

To estimate the effect of the choice of dispersion model on an exposure study, approximately 1.12 million people, located at their place of residence were assigned levels of NO\textsubscript{2} (µg/m\textsuperscript{3}) from the two models (Figure 2.3). However, this comparison does not take the location and spatial differences into account. Therefore, another comparison was performed in which the two models were compared by looking at the difference in NO\textsubscript{2} concentration for each individual (Figure 2.4).

2.6.3 Results and discussion
The comparison between the modelled concentrations showed that the largest differences in concentrations occurred in the urban areas in the densely populated parts of the west coast of Scania where Model 2 predicted lower concentrations than Model 1 (Figure 2.2). This is probably caused by the contribution of high levels of air pollutants from the road network in Model 1, a factor that is not taken into account in Model 2.
Chapter 2

The number of individuals exposed to different NO$_2$ levels in the two models showed similar trends, with a correlation coefficient of 0.84 (Figure 2.3). This distribution shows how well the two models covary according to the number of exposed individuals in the different concentration categories but it does not take the spatial differences that might occur into account. However, figure 2.4 shows that the majority of the individuals were assigned exposure levels that differed by approximately -2 to 2 µg/m$^3$, depending on which model was used. Although the differences are quite large, it is interesting to observe that the output is relatively similar and that the exposure comparison shows great similarity. It is, however, important to note that the accuracy of the models compared with actual, measured, values of NO$_2$, was not investigated.

![Figure 2.2](image_url)

*Figure 2.2: Upper left corner: Concentration of NO$_2$ (µg/m$^3$) for Scania generated by Model 1. Upper right corner: Concentration of NO$_2$ (µg/m$^3$) for Scania generated by Model 2. Lower figure: Difference between the models in modelled concentration of NO$_2$ (µg/m$^3$).*
Figure 2.3: Comparison between the two models regarding the assigned concentration of NO$_2$ (µg/m$^3$)

Figure 2.4: Difference in individually assigned concentrations of NO$_2$ (µg/m$^3$) between the two models.
Chapter 3

Exposure Modelling

Accurate modelling of the concentrations of pollutants over an area is only the first part of an exposure-response study. The second part consists of assigning this exposure to a population or group of individuals (Figure 3.1). This is important in order to correctly evaluate the health effects and thus be able to effectively allocate resources for the reduction and regulation of pollution (Omstedt and Szegö, 1990).

It is important to correctly define the location of the individuals concerned. The level of generalisation of both the population location and the methodology has a significant influence on the results.

This chapter starts with a description of census data; the sources and aggregation levels in general, as well as for those applied in this study (Section 3.1). This is followed by a description of different methodologies (Section 3.2) and thereafter a short review of spatial resolution in exposure modelling is given (Section 3.3). Finally, the chapter ends with a comparative study of different exposure estimates that was conducted within the project BHM (“Biobränsle, Hälsa, Miljö”) financed by The Swedish Energy Agency (Section 3.4).
3.1 Census data

To estimate population change, in time and space, in order to conduct exposure studies, a variety of parameters are of interest. Examples are: household density/area unit, workplace density or, for more detailed studies, individual addresses of homes, workplaces or schools. If the individual’s mode of transportation, location of leisure activities, etc. are recorded, these factors can also be included in an exposure study, enabling the creation of a time-activity model (Briggs, 2005; Krzyzanowski et al., 2005).

3.1.1 Data sources

Census data are usually administrated by statistical organisations, municipalities, national agencies or companies. The data can either be collected from official sources such as population, tax or health registers, or can be gathered from directed or undirected surveys.

Statistics Sweden (SCB) has been assigned the responsibility of collecting, organising and developing statistic material regarding the Swedish population and official statistical parameters. Due to the structure and frequent use of personal ID numbers in Sweden the possibility of combining data on individual level from different sources is substantially better than in many other countries. For example by combining the ID number with the Swedish National Land Survey’s (Lantmäteriet) property register the coordinates of, and information on, an individual’s residential address can be linked to that person. This kind of census data is often desirable when conducting exposure studies with high spatial resolution. However, the disadvantage is that this kind of study poses a threat to personal privacy due to the possibility of geographical identification in addition to sensitive information, such as health and socio-economic factors. Therefore, access to these data is often restricted.

3.1.2 Spatially aggregated census data

To avoid privacy problems, or when a lack of geographical coordinates makes it impossible to conduct studies with high spatial resolution, census data are often aggregated into statistical areas or into grid cells of varying sizes. It is important to be aware that these two methods of aggregating census data are quite different. The statistical areas are often created based on infrastructure or the population pattern within an area, for example postcodes or household density. This method of aggregating census data is advantageous for studies where there is a need to describe the population structure within an area. However, a serious disadvantage of statistical areas is that the structure and demarcations often change as the areas are frequently altered due to changes in the population structure. Another problem regarding statistical areas in the same data set is that the sizes can vary considerably, i.e. statistical areas in city centres tend to be relatively small, due to
the high density of households, in comparison to statistical areas in rural areas where the household density is lower.

The problem of varying area sizes does not involve grids since their cell size is constant to each other. In certain studies this might be a disadvantage since blocks and neighbourhoods are divided into grid cells without taking into account population density or the physical characteristics of the area. This makes it difficult to geographically identify an area or present results statistically for different sub-areas within a region.

Varying numbers of individuals in different areas or grids may also cause problems due to the fact that in areas with few individuals, a difference in one or two cases can make a substantial difference, compared with more densely populated areas. This is often denoted the small number problem (Cromley and McLafferty, 2002).

3.1.3 Census data used in these studies
Two categories of population data were used in these studies (referred to as Population data set 1 and Population data set 2) for the GIS-based exposure studies. Both originate from the Swedish National Registry, and were obtained through the Regional Office of Scania, Sweden. Both data sets are on individual level for the registered population in Scania during 2001 (1,128,211 inhabitants), but they differ in spatial resolution and the amount of attribute data linked to them.

Population data set 1
In this data set the individuals location are represented by points at the coordinates of their real estate as listed in the National Registry and located by the Swedish Land Survey (Lantmäteriet), and the attributes sex and age are linked to them (Figure 3.2). Consequently, a family of three listed at the same residence will appear as one point on the map, since their real estate is the same, but in fact the data consist of three individual points, one on top of the other, linked to individual attributes, as illustrated in Figure 3.2.
Chapter 3

Figure 3.2: Example of the positioning of Population data set 1 and its attributes. Lines represent the boundaries of individuals’ real estate and shaded polygons represent their houses. The black dots show the location of the individual’s real estate according to the Swedish Land Survey.

Unfortunately, for individuals living in high-rise blocks, this generalised method of positioning them may cause substantial misplacement since their real estate may consist of several buildings (Figure 3.3).

Figure 3.3: Example of positioning errors that may occur in Population data set 1. Lines represent the boundaries of the real estate and shaded polygons represent their houses. Black dots indicate the locations of the residents.

Population data set 2
In this data set the location of each individual is given as the centre point in a 1 km x 1 km grid. Each individual is located at the centre point of the 1 km x 1 km grid cell in which their real estate is situated (Figure 3.4). The data are linked with socio-economic attributes from Statistics Sweden, and the data set contains

<table>
<thead>
<tr>
<th>No</th>
<th>Sex</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>112320</td>
<td>F</td>
<td>28</td>
</tr>
<tr>
<td>000565</td>
<td>M</td>
<td>29</td>
</tr>
<tr>
<td>025831</td>
<td>M</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No</th>
<th>Sex</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>895732</td>
<td>M</td>
<td>29</td>
</tr>
<tr>
<td>234127</td>
<td>F</td>
<td>28</td>
</tr>
<tr>
<td>006778</td>
<td>M</td>
<td>92</td>
</tr>
<tr>
<td>0746265</td>
<td>M</td>
<td>43</td>
</tr>
<tr>
<td>6943001</td>
<td>F</td>
<td>41</td>
</tr>
</tbody>
</table>
the following set of attributes: sex, year of birth, country of birth, marital status, income for the previous year (2000) and educational level.

Figure 3.4: Example of positioning Population data set 2 and its attributes. Shaded polygons represent the individuals’ house. The small black dots show the location of the residents in population data set 1, while the large black dots represent the corresponding position of the individuals in Population data set 2.

3.2 Methods of exposure estimation

The way in which individuals and population groups are exposed to various health hazards is highly dependent on where they live and work, and how they spend their spare time. During recent years, a great number of studies have been conducted to increase the accuracy in describing spatial variations in air pollution in relation to population location, in order to better estimate their level of exposure (Larsson et al., 2005). The various ways of tackling these problems have been summarised by Jerrett et al. (2005). They divide existing methods into six classes of intra-urban (i.e. on city scale) exposure methods depending on their level of complexity.

1. Proximity-based methods

These methods are the most basic exposure assessment methods. The assessment is based on the distance between a pollution source and the individual, assuming that proximity to emission sources is proxy for exposure in human populations. The advantage of these methods is that they are fairly straightforward and easy to execute. The disadvantages, however, are many since the methods often neglect a number of covariates which can have substantial impact on the results.
Chapter 3

2. Statistical interpolation
These methods use deterministic and stochastic geostatistical techniques to interpolate the levels of air pollutants over an area based on measurements obtained at a set of monitoring sites. Their advantages compared with the proximity-based methods are that they use real measurements of air pollution in the computations, and that they can quantify the difference in level of exposure between individuals. A major disadvantage, however, is the fact that geostatistical interpolations are highly dependent on the quality and number of monitoring sites available for the interpolation. This is especially a problem regarding pollutants that vary significantly over small areas, such as NO\textsubscript{2}.

3. Land use regression models
This method employs regression models to calculate levels of air pollution based on emission sources and land use. The advantages of this method are that it is associated with relatively low costs and that it can be used in areas without additional monitoring and data collection. However, studies have shown that the method may produce poor correlations when it is used to study areas with diverse land use or topography.

4. Dispersion models
A variety of dispersion models are in use (cf. Chapter 2). They are based on assumptions about deterministic processes, and data on, for example emission sources and meteorological and topographic conditions to estimate levels of air pollution. The strength of dispersion models is that spatial and temporal variations in air pollutants can be combined without using a network of monitoring stations. Among the disadvantages, however, is the fact that data and hardware are costly; furthermore, the implementation of these models requires a high level of programming and GIS expertise. There is also a need for cross-validation with monitoring stations, and temporal mismatches can cause substantial errors in the estimates.

5. Integrated emission-meteorological models
These models are a combination of meteorological and chemical modules that are related to each other to enable simulations of dynamic atmospheric pollutants. The meteorological data in these models are used as input for the chemistry modules during simulation. These models have a considerable potential to model complex dynamic atmospheric processes. The possibility of incorporating chemical transport and pathways enables the models to simulate the development of secondary pollutants and to more precisely estimate the likely pollutant mixture. However, the disadvantage is that these models are costly due to the need for complex computational facilities, software and highly trained personnel.
6. Hybrid models
There are two classes of hybrid models.

a) Hybrid models combining personal or household exposure monitoring (i.e. personal monitoring equipment attached to the study subject’s clothes or fixed a monitoring station placed near or within the person’s home) with one of the preceding methods.

b) Hybrid models combining two or more of the preceding methods with regional monitoring. The benefit of using hybrid models is that they provide measurement validation. The weaknesses of the models depend on the combination of models used.

3.3 Spatial resolution in exposure modelling
For exposure studies, methods of estimating the levels of air pollution, in time and space, must be combined with models for the estimation of a population’s position in time and space. Regardless of the data sets used, the spatial resolution of the data is crucial for GIS applications (Paper II; Cromley and McLafferty, 2002). This is a matter not only of spatial, but also of temporal aggregation. As shown in Paper II, the accuracy of a particular spatial resolution may depend as much on the level of temporal aggregation as on the spatial resolution itself. As stressed in Paper II, the level of agreement for coarser resolutions increases with an increase in temporal aggregation. The temporal influence on the required level of resolution is therefore an important factor, and should be borne in mind. However, for exposure analyses with a long duration temporal aggregation of daily, weekly or monthly means may be sufficient, and thus also lower levels of spatial resolution. It is also important to consider spatial differences. Concentrations modelled with too low a spatial resolution in urban areas, i.e. areas with high contrasts, may generate high error estimates since they fail to reflect small-scale variations. However, in rural areas a larger grid size can generate reasonable results.

Many health data sets used in exposure studies deal with individuals on a detailed geographic scale. These details often include addresses, which enables the analyst to use the GIS process of address match geocoding in order to convert each address into a point on a map. However, these geographical identifiers are often difficult to obtain due to privacy and confidentiality considerations (Cromley and McLafferty, 2002). Therefore, many exposure methods and analyses are based on health and census data that are aggregated into predefined geographical units, as described in Section 3.1.2.

One way to conduct exposure estimates for a geographically aggregated population is through areal interpolation (Cromley and McLafferty, 2002). This is a set
Chapter 3

of techniques used to estimate the distribution of a phenomenon (for example asthmatics or children younger than 10 in an area) across one set of spatial units called source units (the area unit into which the population is aggregated) in terms of a second set of spatial units called target units (for example, area units consisting of different levels of NO$_x$ or buffer areas around roads). It is then possible to use an area-weighting method to calculate the proportion of, in this example, asthmatics/level of NO$_x$. This is achieved by using the proportion of the source unit areas that lies within the target area and thereafter weighting the number of asthmatics within the source units according to this proportion and then add the number to the target units.

3.4 A comparative study of exposure estimates

Within the project BHM financed by The Swedish Energy Agency, I have conducted a comparative study of exposure estimates using census data sets with different levels of aggregation. The aim of this study was to obtain an indication of the range of errors that could be expected, depending on the level of aggregation used for the population data set. The following section provides an extension and brief description of the main findings within this study.

3.4.1 Data

The study was conducted on the population of the city of Malmö in Scania, southern Sweden (Figure 1.1, Section 1.1).

Firstly, a pollutant grid was generated in order to assign exposure estimates to the population. This grid had a resolution of 250 metres and was modelled using the dispersion model ENVIMAN (Section 2.5.3). The pollutant grid consisted of annual mean NO$_x$ values during the year 2001. The modelled concentrations included local emissions and regional contributions from sources in the whole county of Scania, as well as contributions from long-distance sources, e.g. Zealand and the rest of Denmark.

The census data used in this study can be divided into three groups (based on their geometric shapes).

1. Points: The population living inside Malmö’s city boundary: 238,451 individuals (see Figure 3.5a). These were extracted from population data set 1 (Section 3.1.3) and is hereafter referred to as RefPop.

2. Polygons: Statistical area units within Malmö city obtained from Malmö Municipality (see Figure 3.5). These consisted of three levels:
   2a) Statistical unit level 4 (341 areas). Average area size 0.3 km$^2$ with a span of 0.02 km$^2$ - 6.3 km$^2$ (Figure 3.5b).
2b) Statistical unit level 3 (119 areas). Average area size 0.9 km\(^2\) with a span of 0.09 - 11.8 km\(^2\) (Figure 3.5c).

2c) Districts (10 areas). Average area size approximately 16 km\(^2\), but the variation within this group is extensive and the area sizes ranged from 3 to 52 km\(^2\) (Figure 3.5d).

3. Grids: Two raster data sets with different resolutions were created: 1 kilometre resolution and 500 metre resolution. Both sets covered the area of Malmö city.

![Figure 3.5: Four data sets of census data (Malmö city shown in light grey):](image)

- A) The population located as points according to their residential address
- B) The boundary of statistical level 4
- C) The boundary of statistical level 3
- D) The boundary of the districts

3.4.2 Methodology
To be able to compare and evaluate the exposure estimates it is important that the census data used originate from the same data set. Therefore, the census data in the data set RefPop were spatially joined with the polygon and grid data. The population within each area unit was summed and added to the different data sets. The individuals in the RefPop data set were also assigned the statistical area’s area codes and the grid cells’ ID numbers through spatial join. This linkage
made it possible to refer the estimated exposure levels back to the individuals living in the different areas. The RefPop data set thus contained the concentrations of NO\textsubscript{X} from all the different exposure estimates studied.

The following methods were used to add levels of exposure to each data set.

**Points:** The census data set consisting of points, RefPop, was spatially joined with the modelled concentrations of NO\textsubscript{X} by assigning the concentration of the grid cell in which the individual was located.

**Polygons:** The concentrations of NO\textsubscript{X} for the polygon areas were calculated as the mean of modelled concentrations in the cells that fell within, or intersected, the area’s borders.

**Grids:** Two pollutant values were assigned to the population grids. The first pollutant value was calculated as the mean level of NO\textsubscript{X} in those cells in the pollutant grid that fell within the cells in the population grid, i.e. 2x2 for the 500 metre grids and 4x4 for the 1 kilometre grid (Figure 3.6).

The second pollutant value was based on the observation that the method of applying the mean concentration of NO\textsubscript{X} for the square kilometre (or 500 x 500 m) overlapping an individual’s residence might not always be the most suitable approximation of the actual level of exposure of an individual. The numbers of residences within a grid cell are, for example, not evenly distributed in space. Therefore, the centroid of the population density within each grid cell was calculated (Figure 3.7). Thereafter, the level of NO\textsubscript{X} concentration was calculated by estimating the influence of the four nearest grid cells to the centroid using bilinear interpolation. This level was then assigned to the individuals within the cell.

Finally, the relative difference in the exposure estimates was calculated for each individual and area unit by applying Eq. 3.1:

\[
\frac{\text{(Area unit level of NO}_X\text{ - Reference level of NO}_X)}{\text{Area unit level of NO}_X} = \text{(3.1)}
\]

where:

*Area unit level of NO\textsubscript{X} –* is the pollutant concentration calculated for the area units in the various data sets (statistical area 4, statistical area 3, districts, 500 metre grid cells, etc.) and,

*Reference level of NO\textsubscript{X} –* is the pollutant concentration obtained from RefPop.
Exposure Modelling

Figure 3.6: Modelling the first pollutant value in the grid cells. The heavy black lines correspond to the 1 km population grid, while the dashed lines correspond to the 500 metre population grid. The light lines and the shades of grey correspond to the underlying pollutant grid of NO\textsubscript{X} (µg/m\textsuperscript{3}).

Figure 3.7: Calculation of the population density centroid for the 500 metres and 1 kilometre grid. The small black dots correspond to the location of each individual’s real estate (obtained from the RefPop data set). The large black dot corresponds to the population density point within the 1 km grid cell (black lines). White dots correspond to the population density points within the 500m grid cells (white lines). The shaded cells (4x4) correspond to the different levels of NO\textsubscript{X} (µg/m\textsuperscript{3}).
3.4.3 Results and discussion

Comparing the exposure estimates for the statistical area units showed that concentrations estimated for individuals aggregated within statistical level 4 were most consistent with the reference levels (RefPop), while the largest discrepancy was found when aggregating individuals at the district level. The relative differences between these aggregations showed that the district level had approximately half the percentage of individuals with the same levels of NO\textsubscript{X} as the reference level, compared with the other statistical areas (Table 3.1). The aggregation into districts also causes underestimation of the levels for 19% of the individuals (10% + 9%) by more than 10% and overestimation of the levels for 11% of the individuals. One possible explanation of this effect may be that the size of the area units within the statistical areas affects the results. The district areas are much larger and vary much more in size than the other two statistical areas. Due to this, the average concentration for any of the larger district areas might be unrepresentative of the actual exposure to NO\textsubscript{X} for the majority of individuals living there, while smaller areas tend to have more accurate exposure estimates, due to the smaller size of the unit. The variation in the area of the districts causes a variation in the error which, in turn, leads to a data set of non-uniform quality. Decreasing and unifying the area sizes should also decrease the error in the estimate.

Table 3.1: The relative difference for the proportion of individuals/area unit (All rows do not add up to 100% as a consequence of rounding off)

<table>
<thead>
<tr>
<th>Area</th>
<th>&lt; -20%</th>
<th>-20 - -11%</th>
<th>-10 - -1%</th>
<th>0%</th>
<th>1-10%</th>
<th>11-20%</th>
<th>&gt; 20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical area 4</td>
<td>0%</td>
<td>3%</td>
<td>20%</td>
<td>58%</td>
<td>18%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>Statistical area 3</td>
<td>0%</td>
<td>5%</td>
<td>24%</td>
<td>45%</td>
<td>22%</td>
<td>4%</td>
<td>0%</td>
</tr>
<tr>
<td>Districts</td>
<td>9%</td>
<td>10%</td>
<td>19%</td>
<td>25%</td>
<td>26%</td>
<td>11%</td>
<td>0%</td>
</tr>
<tr>
<td>500 m grid *</td>
<td>0%</td>
<td>0%</td>
<td>22%</td>
<td>57%</td>
<td>19%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>500 m grid **</td>
<td>0%</td>
<td>0%</td>
<td>19%</td>
<td>63%</td>
<td>16%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>1 km grid *</td>
<td>0%</td>
<td>7%</td>
<td>31%</td>
<td>35%</td>
<td>24%</td>
<td>3%</td>
<td>0%</td>
</tr>
<tr>
<td>1 km grid **</td>
<td>0%</td>
<td>6%</td>
<td>23%</td>
<td>41%</td>
<td>25%</td>
<td>4%</td>
<td>0%</td>
</tr>
</tbody>
</table>

* Mean level of NO\textsubscript{X} (µg/m3)
** Exposure estimate for population centroid/grid cell calculated through bilinear interpolation
Exposure Modelling

The pollution estimates for the 500 meter grid cells followed statistical level 4 with only minor variations. The pollution estimates obtained using the population density centroid/grid cell showed the best results and gave the same exposure levels as the RefPop for 63% of the individuals (Table 3.1). The exposure estimates for the individuals aggregated into kilometre grid cells were less accurate than the 500 metre grid cells, but were distributed the same way.

The best agreement with the reference level for both the 500 metre and the 1 kilometre grid was obtained with exposure estimation using the population density centroid/grid cell (Table 3.1). In general, the grid pollutant concentrations show similar trends. The difference lies in the lower proportion of “correct” exposure estimates for the 1 km grid, probably due to the fact that the grid cells cover a larger geographical area and are therefore assigned a more averaged concentration. This may also be due to the fact that the approximation of pollutant levels is applied to a larger number of individuals, which increases the size of the error.

The results of this study suggest that area division into high-resolution grids, preferably in combination with the bilinear interpolation method, is recommended for this study design. The main reason for using high-resolution grids instead of smaller statistical areas is that the geographical extent of the statistical areas varies substantially, making it difficult to compare the exposure estimates for individuals within the different sub-areas and the resulting inability to validate the exposure accuracy.
CHAPTER 4

HEALTH EFFECTS

Many steps are required to conduct a reliable exposure-response study, as described in previous chapters. When the levels to which a population or a group of individuals is exposed have been modelled, the task of analysing these levels in relation to their health status begins (Figure 4.1). This procedure entails a number of uncertainties.

This chapter starts with a description of sources of health data and the factors and uncertainties that must be dealt with (Section 4.1). The following two sections describe the health effects caused by air pollutants (Section 4.2) and noise (Section 4.3).

4.1 Health data

Health data used in exposure-response studies can be gathered from a number of different sources, including hospital records or data from self-reported health surveys or questionnaires. Data may also consist of area-based statistical figures gathered by the medical authorities or estimates obtained from statistics regarding the number of smokers, individuals on sick leave, etc. When the required health data for a population, group, or individual have been collected they are statistically analysed together with the exposure level assigned to the unit of analysis. If, for example, the population within an area seems to suffer from a higher number of respiratory diseases and, at the same time, is exposed to higher levels of air pollutants than the population in surrounding areas, the elevated levels of pollutants may be the cause of this increase in respiratory cases.

Figure 4.1: Illustration of the method used to estimate the health effects on the exposed population.
Chapter 4

4.1.1 Epidemiological studies

The type of epidemiological study to be carried out determines the exposure data required. Exposure-response studies are usually divided into two main groups (Silva, 1999):

1. Ecological studies
   These studies are concerned with investigating the frequency of a health outcome in relation to the level of exposure of an aggregated group of individuals. Thus the group rather than the individual is the unit of observation and analysis. Since ecological studies are based on aggregated areas, or population groups, the level of exposure, or health outcome, cannot be linked to a particular individual. Grouping can be based on place of residence, place of birth, level of income, occupation, etc. The main advantage of ecological studies is that they are relatively easy and cost efficient, and therefore often used to generate hypotheses. The limitation is that it is difficult to predict whether there is a true relationship or not, since the results refer to aggregated groups and not individuals.

2. Individual studies
   These studies are concerned with individuals, and link personal attributes, such as health outcome or socio-economic variables, to exposure. Individual studies have the advantage of being able to directly link the exposure level to health outcome.

Regardless of which sort of epidemiological study chosen problems may arise when analysing diseases and health hazards with long latency periods, such as cancer. Samples and measurements made when a disease manifests itself do not usually represent historical exposure levels (Cromley, 2003). Therefore, it is important to store data in order to be able to model past exposure levels.

4.1.2 Sources of health data for exposure studies

There are many possible sources of health data, but two main kinds of data source are used for most epidemiological studies.

1. Register data
   This source includes hospital records, information from other treatment facilities (cancer centres, private clinics, hospices, old people’s homes, etc.), information from diagnostic services (such as pathology departments, biochemical and immunological laboratories, X-ray and ultrasound departments, etc.) and death certificates from the death registers.
   
   Many of these registers collect or updated their systems and databases routinely, including personal attributes such as address and socio-economic attributes.
Therefore, these databases and systems enable epidemiological studies on individuals where both the exposure and health outcomes of interest are obtained for each subject (Silva, 1999). Since the data are gathered routinely, studies based on them can be conducted relatively inexpensively. A major limitation, however, is that the data available are not always suited to the design of the study and therefore limit the study in various ways.

The use of personal identity numbers, required in the Scandinavian countries, enables unique possibilities to link various register data, and thereby to perform high-quality epidemiological studies.

2. Self-reported data

Many countries lack the possibility to link health and socio-economic data with the residential address of an individual, and thereby to conduct “address-specific” exposure studies. Directed or undirected surveys are therefore another means of obtaining information on individuals, while providing the possibility to link the information to a specific address. However, in this approach the individuals are often free to choose whether to participate or not, which may result in a selection bias of individuals included in the study. Participants in a survey are often required to take an active part in the gathering of information, such as answering questionnaires, keeping diaries, biological sampling, personal monitoring or being interviewed, which might cause the answers to be of a more subjective nature than objective measures, and therefore more difficult to evaluate and compare.

4.1.3 Confounding, effect modification and vulnerability

An increase in the number of diseases/cases in an area with high levels of pollutants can not be directly attributed to the effect of the pollutants. The increase in disease could also be caused by other factors, which are correlated to exposure as well as outcome, i.e. confounding factors. It is important to control for confounding factors in an exposure-response study. Proposing a non-existent connection due to various effects of exposure together with the effect of another variable may lead to erroneous conclusions. According to Briggs (2005) the associations between air pollution and health could be confounded by many social, demographic and environmental factors. The demographic and socio-economic attributes of the exposed population, such as age, sex and income, are often helpful in quantifying the magnitude of such bias. Also, socio-economic data are often used as surrogate variables in an attempt to control for differences in smoking, dietary and exercise habits. Even if such factors are not correlated with exposure levels they can be associated with the outcome, and thereby they may modify the health effects, i.e. effect modification.

Specific groups within the population may have a higher vulnerability to the effect of exposure to air pollutants. This includes those naturally more susceptible
to the effects of exposure than others (i.e. individuals with respiratory disorders, children or elderly people) and those who become more susceptible, for example, as a result of occupational, environmental or social factors (WHO, 2004). The socio-economic and demographic factors of the individuals or population groups that are studied should therefore be evaluated with care (Paper I). Such factors may increase the impact of air pollutants on health, but they can also act as confounders when studying health effects, making it difficult to quantify the health effects of air pollutants alone.

Several studies have been conducted regarding the levels of exposure to pollutants and the socio-economic status of individuals. This field of research is sometimes called environmental injustice. Although the empirical foundations of many of these articles are poorly established (Bowen, 2002) the implication of most articles is that the activities of society usually impose a greater health risk upon populations living in neighbourhoods that are characterised by large proportions of minorities, low-income earners and the otherwise disadvantaged (Bowen, 2002).

The results of the current (Paper I and Paper III) show that there is a covariation between socio-economic status and the levels of NO$_2$ and noise in residential areas in Scania. The results thus confirm that socio-economic factors may confound the results, and create a bias when investigating the health effects of pollution. However, the study in Paper I also show that the exposure to air pollutants of different socio-economic groups varies considerably between areas, and that the relationships can differ, not only in magnitude but also in direction. Even the relations seen were the opposite in cities of similar sizes and populations. The results indicate that there is no consistency in how individuals in different socio-economic classes are exposed to air pollutants over a large area, and that both the size of the area and the choice of socio-economic indices affect the associations observed. Therefore, it is important to thoroughly analyse the study area with regard to socio-economic factors before trying to control for them.

4.2 Health effects of air pollution

There are many environmental health hazards around us that could be avoided simply by avoiding exposure to them. However, it is often difficult to avoid being exposed to air pollutants and in many cases we are not even aware that we are being exposed.

The various health effects of air pollution have been the subject of extensive studies during recent years. Many new findings suggest that health effects can be seen at low concentrations of air pollution (Brunekreef and Holgate, 2002; Rosenlund, 2005). Not only short-term and long term respiratory effects, but also cardiovascular diseases, such as myocardial infarction and stroke, have
been linked to exposure to air pollutants. According to Brunekreef and Holgate (2002) the air pollution in Austria, France and Switzerland (combined) is believed to be responsible for 40,000 deaths per year out of a population of 74.5 million. In Sweden, it has been estimated that the annual number of premature deaths due to anthropogenic particle exposure is around 5000, corresponding to a decrease in life expectancy of between 3 and 10 months in the different counties of Sweden (Forsberg et al., 2005).

Some studies also claim that air pollution is (in part) responsible for the rapid increase in asthma in the developing countries during the past 20 years (Maynard, 2004; Ferguson et al., 2004). Associations between urban air pollution, especially those emitted from vehicles, and an increased risk of lung cancer have also been reported (Nyberg et al., 2000). There are also possible associations between exposure to air pollution during pregnancy and reproductive outcome (Gouveia et al., 2005).

Air pollution is a mixture of substances that may vary in composition in space and time. Inevitably, it is difficult to separate the effects of each component in an epidemiological study, in contrast to experimental studies. I have chosen to focus on the health effects resulting from two groups of substances: NO\textsubscript{X} (mainly NO\textsubscript{2}) and particles. The main reasons for this are firstly that NO\textsubscript{X} are good indicators of other combustion substances, so when the levels of NO\textsubscript{X} are high the levels of other air pollutants are also high (Welinder et al., 2003). This is true not only for gaseous pollutants, but also for ultrafine particles. Secondly, the emission sources that were available for this study are mainly concerned with the emission of these substances.

4.2.1 Nitrogen dioxides

Nitrogen dioxide reaches the deeper regions of the respiratory tract, such as the bronchotracheal and alveolar regions, through the respiratory system (Kraft et al., 2005) where it causes inflammation of the lungs leading to decreased lung function. Animal studies have revealed that acute high exposure (≥10 ppm) is followed by death through lung oedema, while lower concentrations caused morphological damage to the lungs (Kraft et al., 2005).

Nitrogen oxides are not only toxic in themselves, but they also have the potential to become oxidants, either by direct effects on lipids and proteins or indirectly through activation of intracellular oxidant pathways (Brunekreef and Holgate, 2002).

In the report “Health Aspects of Air Pollution” (WHO, 2004) it is stated that health effects resulting from short-term exposure to nitrogen dioxide include effects on pulmonary function (particularly in asthmatics), an increase in airway allergic inflammatory reactions, an increase in hospital admissions and increase in mortality. The effects of long-term exposure on health include
reduction in lung functions and increased probability of respiratory symptoms (WHO, 2004).

A study conducted by Hwang and Chan (2002) in Taiwan revealed that nitrogen dioxide (and PM$_{10}$) had a significant effect on daily clinic visits due to lower respiratory tract illness. Their analysis also showed that elderly people (aged 65 or over) were more susceptible than other age groups to the effects of nitrogen dioxide and PM$_{10}$. Also, individuals with already decreased lung function, such as asthmatics, are especially vulnerable to high levels of nitrogen oxides (Spix et al., 1998).

4.2.2 Particles

Exposure to particles has been linked to a number of health effects. The negative effect of these particles seems to be linked to their characteristics, such as size, surface area and chemistry (Kobach et al., 2006). It has not been possible to establish a causal relationship between particle-related health effects and one single particle component; although a number of epidemiological and toxicological studies suggest that some types of emissions are more strongly associated with health effects, especially those from motor vehicle and other combustion products. However, the particle fraction $\leq 2.5 \, \mu m$ (PM$_{2.5}$) and ultrafine particles are probably more strongly related to health effects than PM$_{10}$ (WHO, 2000). Ultrafine particles are characterised by a high number concentration, low mass concentration and a large surface area. They also have a higher deposition rate in the peripheral lung, compared to larger particles, and can cross the pulmonary epithelium to reach the interstitium (Alessandrini et al., 2006). Another danger associated with ultrafine particles is that they have specific toxicity, which may result in the induction of oxygen radicals and the catalysis of chemical reactions (Kreyling et al., 2006).

There is evidence that the non-respiratory effects of air pollutants, especially particles, are mediated through the induction of inflammation. According to Donaldson et al. (2005) there are three potential pathways that could lead to adverse cardiovascular effects.

1. Inflammation in the lung leading to systemic inflammation, which might cause sudden death through cardiovascular causes.
2. The inflammation might create an imbalance in coagulation factors, which might favour creation of thrombi if it has already been initiated.
3. Inflammation might affect the autonomic nervous system, which in turn could alter the heart rhythm.
Brunekreef and Holgate (2004) reviewed a number of major studies on the effects of air pollution on health. They concluded that effects resulting from exposure to particles can be seen at very low levels of exposure. Short-term health effects of exposure to particles may include pulmonary inflammatory reactions, respiratory symptoms, adverse effects on the cardiovascular system, increase in medication usage, increase in hospital admissions, and increase in mortality (WHO, 2004). It has been estimated that the short-term effects of particulate exposure correspond to an increase in the number of hospital admissions of 1 - 1.5 % and an increase in mortality of 0.5 - 0.6% with an increase in particle levels with 10 µg/m$^3$ (Pope et al., 1999). Other studies have reported associations between long-term exposure to air pollutants (particles) and increased risk of mortality (Filleul et al., 2002) and lung cancer (Kobach et al., 2006). According to the WHO report (2004) the most severe effect, in terms of the overall health burden, is a significant reduction in life expectancy of the average population by a year or more, owing mainly to cardiopulmonary mortality and probably also to lung cancer. This effect is linked to long-term exposure to particular matter, and also includes the effects of an increase in lower respiratory tract symptoms, reduction in lung function in children, increase in chronic obstructive pulmonary disease and a reduction in lung function in adults (WHO, 2004).

4.3 Health effects of noise

According to King and Davis (p.123, 2003) a sound is identified as a noise in the ear of the beholder and cannot simply be identified by amplitude and frequency. Health effects are thus not only a question of amplitude and frequency, but also the quality and timing of the noise, causing the health effects to be immediate, delayed or barely perceptible (King and Davis, 2003).

The most obvious health effect of noise is hearing impairment. According to the WHO (Berglund et al., 1999) noise-induced hearing impairment is the most prevalent irreversible occupational hazard. They also state that in developing countries environmental noise poses an increasing risk. Together with hearing impairment, associated findings may be loudness recruitment and tinnitus.

An indirect health effect of noise is sleep disturbance. A long-term study by Öhrström (2003) concluded that sleep quality was drastically reduced for residents exposed to road traffic at noise levels below 60 dB and that sleep quality could be significantly improved by an extensive reduction in noise. According to King and Davis (2003) several studies suggest that sleep disturbance can be avoided by restricting the nigh-time value of LA$_{eq}$ to 30 dB for continuous noise such as traffic. The most frequently attributed effects of noise on sleep are increased sleep latency, night-time awakening and changes in sleep stages; all of which are indicators of chronically disturbed biological sleep rhythms (King
Other primary physiological effects of sleep disturbance are increased blood pressure, increased heart rate, changes in respiration, cardiac arrhythmia and an increase in body movements during sleep (Berglund et al., 1999). According to Berglund et al. (1999) there are also secondary effects such as reduced sleep quality, increased fatigue, depressed mood and poorer performance, which might lead to long-term effects on psychosocial well-being. According to Rosenlund (2005) the relation between community noise exposure and subjective complaints, annoyance and sleep disturbance is well documented, and there is growing concern that persistent stress caused by living in areas subjected to noise might cause undesirable health effects in residents, such as hypertension and cardiovascular disease.

Other physical functions are also affected by noise. These are mostly stress-related disorders, such as migraine, hypertension and coronary heart disease (King and Davis, 2003). Acute noise exposure activates the autonomic and hormonal systems, leading to temporary changes such as increased blood pressure, increased heart rate and vasoconstriction (Blindlund, et al., 1999). According to the WHO’s guidelines for community noise, prolonged exposure may cause permanent effects, such as hypertension and ischaemic heart disorder in susceptible individuals in the population (Berglund et al., 1999). Groups that tend to be exceptionally vulnerable to noise disturbance are those with reduced personal functions (old, ill or depressed people) but also people with particular medical problems, people dealing with complex cognitive tasks, blind people or babies and young children, since these individuals may be less able to deal with the effects of noise exposure and they are thus at greater risk of harmful effects (Berglund et al., 1999). According to the study described in Paper III, 29% of the population in Scania was exposed to an average level of 55 dB or more from road traffic. A high prevalence of disturbance of daily activities and a high prevalence of frequent annoyance due to noise from road traffic were recorded. The observed trends between annoyance, disturbance of daily activities and exposure level were striking, and consistent for average noise levels, as well as maximum noise levels. This is a concerning finding since approximately 1/3 of the study population was estimated to be in the highest exposure category. However, no apparent associations were found between the maximum level of road traffic noise and self-reported general health problems. For the average noise level, there was a weak overall association with extensive sleeping disturbance during the past week, whereas an association with insufficient sleep was not generally apparent. An association was also found between average road noise level and treatment for hypertension in females, and also in males who reported “fairly much” or “much” annoyance due to road traffic, as well as some negative health effects resulting from road traffic noise in subgroups that could be more vulnerable due to exposure to social stressors, and this association should thus be considered with care.
5.1 Paper I
To prevent the results of exposure-response studies from being biased, the effects of socio-economic variables must be considered. This study analysed the exposure levels of NO$_2$ for different socio-economic groups to determine whether there was any bias. The results of the study demonstrated that there was no consistency in how individuals in different socio-economic classes were exposed to air pollutants. The study was conducted for the whole county of Scania (Figure 1.1) and also for the five major cities in the county (Malmö, Helsingborg, Lund, Kristianstad and Trelleborg).

The exposure to air pollutants for different socio-economic groups varied considerably between the areas studied, and the relationships seen differed, not only in magnitude but also in direction. It was found that the extent and character of the area had significant effects on the results. For the whole county of Scania the results indicated that the area was heterogeneous regarding the association between air pollution and socio-economic status. However, the relationship differed when the analysis was performed on city level. It also differed depending on whether the cities were analysed separately or together. The main conclusions drawn from this study are therefore that it is inadvisable to determine and analyse associations between socio-economic factors and exposure to air pollutants on county level, and that the size and choice of study area is of great importance. Also, the choice of socio-economic indices (in this study: country of birth and educational level) is of importance and affects the associations observed.

5.2 Paper II
This study investigated the optimal spatial resolution with respect to temporal resolution for the establishment of a pollutant database in Scania. A program was developed to calculate the error estimates for pollutant grids of various resolutions. Since the time duration of future studies, using the database pollutant,
Chapter 5

will be at least one day, the spatial resolution must be set in relation to this time resolution. The results showed that the coarser resolutions reflected temporal variations well, and that the level of spatial agreement between coarser and finer grids increased as the temporal resolution increased. This shows that a coarser grid can be used for studies of longer duration, while a finer grid is more suitable for studies with a shorter duration. The character of the study area was also found to be of importance. If the concentration levels varied considerably, as is the case in urban areas with a high influence from traffic, a coarser grid failed to detect these variations. However, for areas with less spatial variation the coarser grids accurately reflected the concentrations. This implies that a pollutant database that allows different spatial resolution for urban and rural areas should be developed.

5.3 Paper III

This study was performed to evaluate the effects of road traffic noise using self-reported health effects, annoyance and disturbance of daily activities. The data were gathered from a large public health survey conducted in southern Sweden, 1999-2000, which supplied data on demography, annoyance and disturbance of daily activities, and on general health problems regarding concentration, sleep, stress and treatment for hypertension. GIS was also used to estimate the average exposure to road traffic for the study population in Scania.

The results showed that around 29% of the population was affected by noise levels of at least 55 dB(A), which is close to the overall estimates for the European Union. This is of concern since the results show that the observed trends between “annoyance”, “disturbance of daily activities” and exposure level in our self-reported health study were consistent for average noise levels as well as maximum noise levels, being generally at least doubled from the lowest to the highest exposure categories. The results also showed that “disturbed sleep or relaxation (sometimes or frequently)” was reported by as many as 19 - 32% of the individuals in the highest exposure categories. High exposure to road noise was also more common among those not born in Sweden, living alone and not employed. Within the subgroup that reported annoyance from road traffic noise, associations were found with concentration problems past week as well as with treatment for hypertension. However, positive findings in these subgroups should be interpreted with care. We observed no associations between the maximum level of road traffic noise and self-reported general health problems. For average noise levels, we found a weak overall association with extensive sleeping disturbance during the past week. We also found an association between average road noise level and treatment for hypertension in females, and in males who reported “fairly much” or “much” annoyance from road traffic.
References


DMU (Denmark’s National Environmental Research Institute), 2006. (http://www2.dmu.dk/1_viden/2_Miljoe-tilstand/3_luft/4_spredningsmodeller/5_OSPM/default_en.asp)


IVL (Swedish Environmental Research Institute) , 2006. (http://www.ivl.se/)


Krzyzanowski, M., Kunna-Dibbert, B., Schneider, J., 2005. Health effects of transport-related air pollution. WHO.


Opsis, 2006 (http://www.opsis.se/)


SNAP., 2006. Swedish National Air Pollution and Health Effects Program. www.snap.se


The most exciting phrase to hear in science, the one that heralds new discoveries, is not ‘Eureka!’ but ‘That’s funny.....’

Isaac Asimov (1920-1992)
Are associations between socio-economic characteristics and exposure to air pollution a question of study area size? An example from Scania, Sweden

Emilie Stroh*, Anna Oudin, Susanna Gustafsson, Petter Pilesjö, Lars Harrie, Ulf Strömberg and Kristina Jakobsson

Address: 1 GIS Centre & The Department of Physical Geography and Ecosystem Analysis, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden and 2 Department of Occupational and Environmental Medicine, Lund University Hospital, SE-221 85 Lund, Sweden

Email: Emilie Stroh* - emilie.stroh@giscentrum.lu.se; Anna Oudin - anna.oudin@skane.se; Susanna Gustafsson - susanna.gustafsson@giscentrum.lu.se; Petter Pilesjö - petter.pilesjo@giscentrum.lu.se; Lars Harrie - lars.harrie@lantm.lth.se; Ulf Strömberg - ulf.stromberg@ymed.lu.se; Kristina Jakobsson - kristina.jakobsson@ymed.lu.se

* Corresponding author

Abstract

Background: Numerous studies have shown that exposure to air pollutants in the area of residence and the socio-economic status of an individual may be related. Therefore, when conducting an epidemiological study on the health effect of air pollution, socio-economy may act as a confounding factor. In this paper we examine to what extent socio-economic status and concentrations of NO₂ in the county/region of Scania, southern Sweden, are associated and if such associations between these factors differ when studying them at county or city level. To perform this study we used high-resolution census data and modelled the annual exposure to NO₂ using an emission database, a dispersion modelling program and a geographical information system (GIS).

Results: The results from this study confirm that socio-economic status and the levels of NO₂ in the area of residence are associated in some cities. The associations vary considerably between cities within the same county (Scania). Even for cities of similar sizes and population bases the associations observed are different. Studying the cities together or separately yields contradictory results, especially when education is used as a socio-economic indicator.

Conclusion: Four conclusions have been drawn from the results of this study. 1) Adjusting for socio-economy is important when investigating the health effects of air pollution. 2) The county of Scania seems to be heterogeneous regarding the association between air pollution and socio-economy. 3) The relationship between air pollution and socio-economy differs in the five cities included in our study, depending on whether they are analysed separately or together. It is therefore inadvisable to determine and analyse associations between socio-economy and exposure to air pollutants on county level. This study indicates that the size and choice of study area is of great importance. 4) The selection of socio-economic indices (in this study: country of birth and education level) is important.

Published: 16 November 2005


Received: 25 August 2005

Accepted: 16 November 2005

This article is available from: http://www.ij-healthgeographics.com/content/4/1/30

© 2005 Stroh et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Are associations between socio-economic characteristics and exposure to air pollution a question of study area size? An example from Scania, Sweden

Emilie Stroh*, Anna Oudin, Susanna Gustafsson, Petter Pilesjö, Lars Harrie, Ulf Strömberg and Kristina Jakobsson

Address: 1GIS Centre & The Department of Physical Geography and Ecosystem Analysis, Lund University, Solvegatan 12, SE-223 62 Lund, Sweden and 2Department of Occupational and Environmental Medicine, Lund University Hospital, SE-221 85 Lund, Sweden

Email: Emilie Stroh* - emilie.stroh@giscentrum.lu.se; Anna Oudin - anna.oudin@skane.se; Susanna Gustafsson - susanna.gustafsson@giscentrum.lu.se; Petter Pilesjö - petter.pilesjo@giscentrum.lu.se; Lars Harrie - lars.harrie@lantm.lth.se; Ulf Strömberg - ulf.stromberg@ymed.lu.se; Kristina Jakobsson - kristina.jakobsson@ymed.lu.se

* Corresponding author

Abstract

Background: Numerous studies have shown that exposure to air pollutants in the area of residence and the socio-economic status of an individual may be related. Therefore, when conducting an epidemiological study on the health effect of air pollution, socio-economy may act as a confounding factor. In this paper we examine to what extent socio-economic status and concentrations of NO$_2$ in the county/region of Scania, southern Sweden, are associated and if such associations between these factors differ when studying them at county or city level. To perform this study we used high-resolution census data and modelled the annual exposure to NO$_2$ using an emission database, a dispersion modelling program and a geographical information system (GIS).

Results: The results from this study confirm that socio-economic status and the levels of NO$_2$ in the area of residence are associated in some cities. The associations vary considerably between cities within the same county (Scania). Even for cities of similar sizes and population bases the associations observed are different. Studying the cities together or separately yields contradictory results, especially when education is used as a socio-economic indicator.

Conclusion: Four conclusions have been drawn from the results of this study. 1) Adjusting for socio-economy is important when investigating the health effects of air pollution. 2) The county of Scania seems to be heterogeneous regarding the association between air pollution and socio-economy. 3) The relationship between air pollution and socio-economy differs in the five cities included in our study, depending on whether they are analysed separately or together. It is therefore inadvisable to determine and analyse associations between socio-economy and exposure to air pollutants on county level. This study indicates that the size and choice of study area is of great importance. 4) The selection of socio-economic indices (in this study: country of birth and education level) is important.
Background
Humans are inevitably, although we constantly seem to forget this, part of the environment we live in. The way we interact with and affect the environment will have consequences not only for our surroundings but also ultimately on ourselves. Air pollution is an example of an anthropogenic effect that has become one of the major health hazards of our time. Airborne pollutants generated by traffic, industry, energy consumption, combustion, etc. are believed to cause not only respiratory and cardiovascular diseases in those exposed to them, but they may also cause premature death in certain groups of vulnerable people. Exposure to different air pollutants and associations to various respiratory and cardiovascular diseases have been the subject of intense study during recent years [1-5]. It has been shown that short-term variations in air pollution levels are related to lung function and asthma symptoms [6], as well as mortality and hospital admissions due to cerebrovascular disease and heart disease [7]. Particular matter ≤ 10 µm (PM10) and ozone are usually considered to be the air pollutants having the greatest effects on health but NO₂ is a good indicator of air pollutants in general [8], and we have thus chosen this component as the indicator of exposure element in this study.

There is an element of discrimination in the way we are exposed to air pollutants. The area affected and the people exposed may be miles away from the source of the pollutants, which makes it difficult not only to control and regulate for pollution, but also to analyse its effects on health. The effect of the pollution may also differ depending on weather conditions or, for example, the socio-economic characteristics of the people exposed. Another difficulty is that the health effects of air pollution are similar to those of many other diseases caused by social factors such as poor dietary habits, lack of exercise and stress. Since both air pollutants and socio-economic status strongly influence health [9-11] socio-economic factors can act as confounders when studying the health effects of air pollutants [6], making it difficult to quantify the health effects of air pollutants alone.

Numerous studies have shown that groups with low socio-economic status tend to live in areas that are exposed to air pollutants to a greater extent than groups with high socio-economic status [12-16]. Several studies on the relation between geographical location and health have been conducted during the past 30 years; many of them confirming that socio-economic status can act as a confounder when investigating the exposure to different health hazards in the area of residence. According to Bowen (2002), who has reviewed 42 empirical research studies on the subject, many of these studies are based on such poor empirical foundations that the results should be viewed as unreliable [17]. Bowen also states that many of these studies fail to detect the underlying spatial process involved due to badly chosen geographical units. Willis, Jerrett, Burnett and Krewski (2003) studied the impact of the size of the study area (metropolitan areas versus county areas) on the results of a study on the relationship between long-term exposure to sulphate air pollution and mortality [18]. Their results clearly shows that the size of the study area can have an impact on the results of an epidemiological study, but that there are both advantages and disadvantages of large and small study areas.

It is important to be able to describe socio-economy on a group level when evaluating possible contextual health effects. A contextual effect is an effect where the group influences the individual, for example, when the socio-economic characteristics of a neighbourhood influence the effect on the health of an individual caused by air pollution. In a multi-level model, where data are analysed according to an individual's group affiliation, contextual effects are of particular interest. Here, group affiliation is determined by area of residence. Individuals from the same group are assumed to both influence, and be influenced by, group membership. This assumption might not be valid if a group is very heterogeneous. In the future we intend to analyse health effects of air pollution in Scania with multi-level models, as the level of socio-economy within an area of residence might have a contextual influence. Clearly, the definition of the area can affect the level, as well as the heterogeneity of socio-economy within that area. In this work, we investigated the relationship between air pollution and socio-economy in Scania, in...
order to define relevant contextual areas for further studies.

Improved technologies and more detailed data sources have made it possible to model personal exposure to air pollutants and socio-economic status. This study analyses the association between socio-economic status and mean annual concentrations of NO\textsubscript{2} in Scania, as well as the possible influence of the choice of geographical level and study area on the results.

The present study was carried out on the population of the whole county of Scania in southern Sweden (described below) in 2001, separately for the five major cities in the region: Malmö, Helsingborg, Lund, Kristianstad and Trelleborg (Figure 1), and for the five cities grouped together in one data set. These cities differ in geographical location, infrastructure and population size, as well as in socio-economic structure. In order not to confound the analysis with a rural-urban gradient, the analyses of the five cities were performed strictly on residents within the city limits.

**Scania**

Scania is the southernmost county in Sweden. It covers around 11,350 km\textsuperscript{2}, which is approximately 2% of the total Swedish land area. The region is densely populated, with more than 1.1 million people living in the area, which is approximately 11% of the total Swedish population. In Scania, approximately 67% of the population lives along the west coast. Most road traffic (passenger cars as well as trucks) from the European continent to Sweden and Norway passes through this area, and five motorways run through the region. There are also several harbours in the region and a considerable amount of cargo shipping and ferry transport along the coast. These factors, and the closeness to Copenhagen in Denmark, and the European continent, contribute to high concentrations of air pollutants in the region, compared with most other regions in Sweden.

**Malmö**

Malmö is the largest city in Scania, with a population of about 260,000 (the third largest city in Sweden) and an area of 66 km\textsuperscript{2}. The city is the residential focus of the rich agricultural area of Scania, and used to be one of Sweden’s most important industrial and trading cities. During the past 20 years industry and commerce in the area have decreased quite considerably, resulting in a sudden rise in the proportion of unemployed, to almost twice the national average [19]. During recent years, this number has begun to slowly decrease as the city has begun to adapt to new areas of the labour market. The completion of the Øresund bridge in 2000, connecting the mainland of Sweden with Denmark, has expanded the labour market for the residents in both Malmö and Copenhagen. The commuting between Sweden and Denmark now dominates commuter traffic, and this almost doubled between 2001 and 1997 [20,21]. Malmö is known for its high proportion of immigrants, with more than 20% of its residents born outside Sweden [22]. The city is considered to be one of Sweden’s most segregated [23].

**Helsingborg**

Helsingborg is the second largest city in Scania and covers approximately 35 km\textsuperscript{2}. It has a population of 89,000. The city has one of the most important harbours in Sweden and is of great importance as a transportation link for trains. Both the ferry traffic and commercial traffic are of great importance to the city. The distance to the Danish mainland across the strait is only 6 km, and the ferry traffic between the two countries is intense. The city is considered to be segregated, with a north/south socio-economic gradient [24].

**Lund**

Lund has a population of 76,000 residents and covers an area of 23 km\textsuperscript{2}. The city is in many ways characterised by the university, and around one third of the residents are students. Thus the city's residents are younger than the national average, and especially the age group 20–29 is highly overrepresented. Due to the high number of students and young people, relative to the population size, migration is much higher than in the other cities. The university, the science park and the university teaching hospital are three of the major employers. The city is not as segregated as Malmö and Helsingborg which, in many ways, can be explained by the high migration to and from the city [25].

**Kristianstad**

Kristianstad is situated on the east coast of Scania. It has a population of 32,000 and an area of 19 km\textsuperscript{2}. The city is the centre of Scania’s eastern region and is situated in one of Sweden’s largest fruit-growing districts. Kristianstad is an active centre of commerce in the region and most of its residents are involved in the trade or service sectors. Although the city is not as segregated as Malmö and Helsingborg, there are still distinct differences in socio-economy between the city’s neighbourhoods [26].

**Trelleborg**

Trelleborg is the most southerly, as well as the smallest, of the cities included in the study. It has a population of only 24,000 inhabitants and an area of 10 km\textsuperscript{2}. The city is highly affected by the busy trade and ferry traffic with the German harbours in Travemünde, Sassnitz and Rostock. Although the city is small, the fact that it is situated in a lively metropolitan area and the hectar trade traffic result in the same social problems as in the larger cities in the...
region. The proportion of residents with a high educational level is lower than the national average [27].

Aims

The main aim of this study was to describe associations between the levels of mean annual concentrations of air pollution (NO$_2$) and two socio-economic indices (“country of birth” and “level of education”) in the region of Scania in Sweden. These associations are of special interest for future studies of health effects resulting from exposure to air pollutants, since socio-economy might act as a confounding factor, both individually and contextually in such a study.

A secondary aim of this study was to investigate the possible influence of differences in size (or level) of the study area on any associations observed. Do the size and choice of the study area affect the associations seen, and if so, to what extent? The answer to this question is of utmost importance, not only for this study, but for health effect studies in general. It is important to establish the effects of socio-economy to ensure that the results of epidemiological studies in general are not biased. If the associations observed are dependent on the size and choice of the study area, generalising the contextual effect of socio-economy for too large an area may lead to erroneous results.

Results

The results of this study confirm the hypothesis that associations exist between socio-economic status and NO$_2$ concentration in our study area (Table 1, 2 and Figure 2, 3, 4, 5, 6, 7). According to the Spearman correlation analysis there is a statistically significant correlation between both country of birth and the level of education, not only in Scania as a whole, but also in all five cities included in the study, regardless of whether they are studied separately or together. The correlation coefficients are often low, however.

The associations seen are not consistent between cities. Also, the sign of the associations differs between the two socio-economic indexes, i.e. “country of birth” and “level of education” show opposite correlations to the level of pollution.

The stronger correlation coefficients in the study implied that associations exist between country of birth and NO$_2$ in Scania (negative), in all cities together (negative), in Malmö (negative), and in Lund (positive). The association between country of birth and level of NO$_2$ implies that being an immigrant in Scania, Malmö or the five study cities (analysed as a group) is associated with elevated levels of NO$_2$. In Lund, however, the association between NO$_2$ concentrations and country of birth is the opposite, implying that being born in Sweden is associated with higher levels of NO$_2$.

A correlation was observed between education level and concentration of NO$_2$ in the cities together (negative), in Lund (positive) and Trelleborg (positive). This implies that less educated groups of people, living in the five cities studied and in Scania in general, are exposed to higher levels of NO$_2$. This is, however, not the case in Lund and Trelleborg when studied separately, where highly educated residents seemed to be exposed to higher concentrations.

---

**Table 1: Descriptive statistics for the population in Scania, Sweden. Statistics for the population in Scania, Sweden (2001) and the five cities in the region (Malmö, Helsingborg, Lund, Kristianstad and Trelleborg).** Statistics for the socio-economic subgroup relating to educational level were calculated for the age group 25–44 years of age, while the subgroup “country of birth” includes all the inhabitants of the region/city.

<table>
<thead>
<tr>
<th>Inhabitants</th>
<th>Scania</th>
<th>Malmö</th>
<th>Helsingborg</th>
<th>Lund</th>
<th>Kristianstad</th>
<th>Trelleborg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of inhabitants</td>
<td>1,165,411</td>
<td>258,140</td>
<td>88,992</td>
<td>76,334</td>
<td>32,295</td>
<td>24,254</td>
</tr>
<tr>
<td>Born in Sweden (%)</td>
<td>87</td>
<td>75</td>
<td>83</td>
<td>83</td>
<td>86</td>
<td>84</td>
</tr>
<tr>
<td>Born in OC (%)</td>
<td>10</td>
<td>20</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Low education (%)</td>
<td>82</td>
<td>81</td>
<td>84</td>
<td>49</td>
<td>82</td>
<td>92</td>
</tr>
<tr>
<td>High education (%)</td>
<td>18</td>
<td>19</td>
<td>16</td>
<td>51</td>
<td>18</td>
<td>8</td>
</tr>
</tbody>
</table>

**Concentrations of NO$_2$ (µg/m$^3$)**

| Median level | 14 | 17 | 13 | 13 | 9 | 12 |
| Quartiles | 9/17 | 15/19 | 12/20 | 12/14 | 8/10 | 11/18 |
| Min | 3 | 10 | 10 | 9 | 6 | 7 |
| Max | 23 | 23 | 21 | 17 | 11 | 22 |
of NO\textsubscript{2}. We found that the sign of the correlation coefficient for education was positive in all cities when studied separately, but negative when analysed as one data set.

**Discussion**

This study confirms that associations exist between socio-economy and level of air pollution (Table 2). It is important to distinguish between a low p-value and a strong correlation. A low p-value does not imply that the association is strong, only that the association observed is not likely to occur due to chance, regardless of how weak the association actually is. The p-values in this study are all < 0.001 as a result of the many observations, while many of the correlation coefficients are below 0.5, which can be considered as indicating weak associations.

The results are not consistent; neither between cities nor between the socio-economic indexes, nor when analysing all cities together or separately. The consistency of socio-economic indexes will not be discussed further here, but we conclude that it appears important to gather as much information as possible about socio-economy to describe socio-economic status.

The results obtained when analysing the cities together in one data set were not at all consistent with the results when considering them separately, especially regarding level of education, where even the sign of the association differed between the combined data set and the separate cities (Table 2). Therefore, generalizing associations between socio-economy and air pollution from a regional level to a city level can give erroneous results. Although there seems to be a strong correlation between NO\textsubscript{2} concentration and country of birth (Figure 2) this correspondence is probably strongly influenced by a rural-urban gradient, resulting in a biased relation between this socio-economic variable and the exposure level. Most immigrants in Sweden tend to move to cities, especially larger cities, rather than settle in the countryside [28]. Since the levels of NO\textsubscript{2} are significantly lower outside cities, this gradient probably makes a major contribution to the strong correlation observed between the concentration of NO\textsubscript{2} and “country of birth” in the whole county of Scania. On the regional level there may also be a risk of relations cancelling each other out, since they are not consistent between cities.

In Malmö, the negative correlation for the group born in Sweden implies that this group tends to live in areas with lower concentrations of NO\textsubscript{2} than the group born in other countries (Figure 4). The proportion of immigrants in the different areas of Malmö is uneven which makes Malmö a segregated city. The neighbourhoods with high proportions of immigrants are located in the outskirts of the city. Since Malmö is surrounded by a ring road at which three of the five main motorways in the area converge, the emis-

---

**Table 2: Correlation between socio-economic subgroups and mean annual exposure to NO\textsubscript{2} in Scania, Sweden (2001).** The correlation coefficients for those born in OC (Other Country) and those with a low level of education are not presented since they are the exact inverse of the correlation coefficients for those born in Sweden, and those who are highly educated, respectively. The proportion with: “high level of education and concentration of NO\textsubscript{2}” is calculated for the age-group 25–64 years.

<table>
<thead>
<tr>
<th>Area</th>
<th>Proportion born in Sweden</th>
<th>Proportion with high level of education and concentration of NO\textsubscript{2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scania</td>
<td>Correlation coefficient</td>
<td>-0.9</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>All 5 cities together</td>
<td>Correlation coefficient</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Malmö</td>
<td>Correlation coefficient</td>
<td>-0.8</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Helsingborg</td>
<td>Correlation coefficient</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lund</td>
<td>Correlation coefficient</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Kristianstad</td>
<td>Correlation coefficient</td>
<td>-0.3</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Trelleborg</td>
<td>Correlation coefficient</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
sions in these neighbourhoods tend to be higher than in other parts of Malmö.

In Helsingborg the correlation coefficients show no correlation at all between the socio-economic variables analysed and levels of NO$_2$. This is probably due to differences in directions between the socio-economic gradients and the concentration gradient for NO$_2$. Helsingborg stretches along the coast of the strait between Sweden and Denmark, making the city quite long and narrow in a north/south direction, with a distinct socio-economic gradient in the same direction. The major source of air pollutants in the area is the lively ferry traffic along and across the strait, as well as from the three motorways that converge and run alongside the city. These factors cause the city to have the same variation in NO$_2$ levels in the north south direction but vary in the west easter direction. This results in evening out of the exposure to air pollutants between the different socio-economic groups.

In Lund the correlation is positive for both the socio-economic variables, implying that people born in Sweden and individuals with a high level of education tend to live in areas with higher levels of NO$_2$ than people born outside Sweden or with lower levels of education (Figure 6 and 7). This is probably the result of a phenomenon observed in many metropolitan areas, i.e. it is very desirable to live in the city centre where the cost of accommodation is high, and can usually only be afforded by people with high economic status. Since pollution levels are higher in the city centre these people tend to be more highly exposed. It is worth noting that this phenomenon is probably present in Malmö as well, although it does not compensate for the other socio-economic and NO$_2$ gradients mentioned above.

In Kristianstad, only weak associations were seen between socio-economy and air pollution. Compared with the other cities in the study, the levels of NO$_2$ in Kristianstad are, much lower, and less spread (Table 1). The small range in exposure decreases the possibilities of detecting an association.

The positive correlation between the level of education and the exposure to NO$_2$ seen in Trelleborg is weaker but implies a similar explanation to that in Lund, where highly educated people tend to live in the more central areas of the city. The fact that Trelleborg is located along the coast with a large, busy industrial harbour in the middle of the city also increases the risk of people living in the city centre being highly exposed to NO$_2$ and other air pollutants derived from the ferry traffic.

As already mentioned, the socio-economic indexes country of birth and educational level do not show the same associations to concentration of NO$_2$ in our study. Strong associations are seen between country of birth and concentration of NO$_2$ and regarding education and concentration of NO$_2$, but not in the same cities (with Lund as an
The analysis regarding education was performed on the population between 25 and 64 years old only, while the analysis regarding country of birth was performed on the whole population. In order to investigate whether the difference in results between the two socio-economic indexes is a result of different age distributions, the analysis of the association between country of birth and concentration of NO$_2$ was also performed on the age group 25–64 years. This did not alter the results. Thus the different age distributions do not explain why "country of birth" and "level of education" do not show the same associations with concentration of NO$_2$.

We chose to analyse the data with simple correlation analysis. Geographical Weighted Regression (GWR) models provide a more sophisticated analysis, but such models are not suited to our purpose. GWR analyses spatially varying relationships by weighting them to form one surface for each relationship studied. Since one of the aims of our study was to investigate socio-economic associations with air pollution in separate cities, GWR is not suitable.

Valid exposure assignments are naturally of particular importance in reducing bias. Despite the fact that this study does not focus on this problem, we can not stress enough the importance of using a functional dispersion models and valid exposure data for exposure/concentration response relationship studies.

The inconsistency in the relationship between socio-economy and NO$_2$ between cities, as well as over larger areas, may conceal existing associations [29]. Performing an analysis for a large area and then generalising the results to the whole country or region is not recommended since the socio-economic factors in larger cities in the area or the rural-urban gradient might dominate the results.

Valid exposure assignments are naturally of particular importance in reducing bias. Despite the fact that this study does not focus on this problem, we can not stress enough the importance of using a functional dispersion models and valid exposure data for exposure/concentration response relationship studies.

The inconsistency in the relationship between socio-economy and NO$_2$ between cities, as well as over larger areas, may conceal existing associations [29]. Performing an analysis for a large area and then generalising the results to the whole country or region is not recommended since the socio-economic factors in larger cities in the area or the rural-urban gradient might dominate the results.

The results of this study clearly show that the size and type of the study area have an impact on the findings when investigating possible associations between socio-economy and exposure to air pollution. Relations observed on one area level may be almost the opposite on a larger scale. This is the case for the associations seen between country of birth and exposure to NO$_2$ for the residents of Lund, where the correlation coefficient implies that being born in Sweden is associated with high levels of NO$_2$. When Lund is included in the group of five cities studied the opposite association would, however, be implied namely that being an immigrant and living in any of these five cities would increase the risk of being highly exposed to NO$_2$. This would be even more misleading if these cities were to be included in a regional study of the whole area of Scania. The smallest area studied was a city. It would have been interesting to divide the data further into districts, to see how such a division influenced the relationships.

Conclusion

The results of this study show a covariation between socio-economic status and the levels of NO$_2$ in the area of residence in Scania, and thereby confirm that socio-econ-
Omy could act as a confounding factor and create bias when investigating health effects of air pollution. In order to quantify the magnitude of bias, several other assumptions have to be made, such as risk levels for various groups of air pollution, smoking habits and other important variables (age, sex, etc).

In this study the exposure to air pollutants of different socio-economic groups varies considerably between areas. It has been shown that the relationships can differ, not only in magnitude but also in direction, whether splitting the study area into small areas or not, as seen for level of education. Relations can be the opposite in cities with similar sizes and populations. Larger areas, such as the county of Scania, are influenced by a large number of socio-economic gradients. The urban-rural gradient makes it especially difficult to determine and analyse associations between socio-economic variables and exposure to air pollutants on a regional level.

In epidemiological studies on the health effects of air pollution, incorporation of individual-level and area-level data on various socio-economic variables (i.e. country of birth, educational level, etc.) is important. Our results demonstrate that there is no consistency in how individuals in different socio-economic classes are exposed to air pollutants over a larger area, and that both the size of the area, and the choice of socio-economic indexes affect the associations observed.

**Methods**

The procedure used to analyse the covariation between socio-economic factors and exposure to air pollutants is described below:

1. **Modelling the concentration of NO\textsubscript{2}**

An emission database and a GIS program were used to model the levels of NO\textsubscript{2} in Scania. This was done using a dispersion modelling program and emission data for the year 2001 together with meteorological data, resulting in a grid giving mean annual concentrations of NO\textsubscript{2} for Scania in 2001.

2. **Modelling the exposure to NO\textsubscript{2}**

Mean annual NO\textsubscript{2} exposure to each individual was modelled in a GIS program using the mean annual concentrations of NO\textsubscript{2} and population data sets.

3. **Division of the population into socio-economic groups**

The population data were divided into socio-economic groups using the indexes **country of birth** and **level of education**.

4. **Statistical analysis**
The data were statistically analysed depending on each individual’s socio-economic status and exposure to NO$_2$ using a weighted Spearman rank correlation.

**Modelling the concentration of NO$_2$**

To model the concentrations of NO$_2$ for Scania an NO$_x$ emission database with approximately 24,000 sources for the area was used [30]. The different sources of emissions consist of roughly 23,500 line sources corresponding to the major roads and ship routes, 500 point sources corresponding to most of the larger industrial plants and approximately 50 area sources corresponding to emissions from heating, wood burning, farming machinery, construction machinery, etc. The emission database includes time variations for different pollutants depending on hour, day and month of the year. The emission database does not include any continental contribution, i.e. sources outside Sweden.

The software Enviman was used to model the concentrations of NO$_2$ (later converted into NO$_x$ levels). Enviman uses a Gaussian model: AERMOD, a USEPA model [31]. The meteorological parameters in the calculation were based on climatological series for the period 1995–2003 gathered from a meteorological weather station in central Malmö [32].

In this study, the mean annual value of NO$_2$ was modelled with a spatial resolution of 250 × 250 m for the county of Scania. Both the levels and the spatial variation of the modelled concentrations of air pollutants are of interest. Therefore, the geographic resolution of the modelling grid is of vital importance. If the resolution is too low the model might generate a grid with lower variations in concentration between the cells. This might increase the uncertainty in the following exposure estimates. Modelling with higher resolution increases the opportunity to detect variations in space as well as extreme values. The capacity of the modelling tool and the computer, as well as the resolution and quality of the data in the emission database, set the limit for the highest resolution that is reasonable to model. To be able to model with this relatively high resolution the county was divided into three parts by splitting the region into three equally sized areas stretching from north to south. Each of these three areas had a 10-kilometre overlap with their neighbouring areas to prevent modelling errors that might otherwise appear along the borders (Figure 8). For each of the three areas two modelling operations were carried out: one to model the concentrations from local sources within that specific area, and a second to model the regional contribution of the NO$_x$s concentrations in each area contributed by the two neighbouring areas in the region. The local and regional concentrations for each cell were summed, and
removing the 10-kilometre overlap between the areas and merging them together created a continuous raster.

To convert the concentration of NO\textsubscript{x} to NO\textsubscript{2} the following equation were used [32]:

\[
\text{NO}_2 \leq 13.85 \text{µg/m}^3 \Rightarrow \text{NO}_2 = \text{NO}_x \\
\text{NO}_x > 13.85 \text{µg/m}^3 \Rightarrow \text{NO}_2 = 8.8 \cdot (\text{NO}_x^{0.44} - \text{NO}_x^{0.18})
\]

To take the continental contribution into account, 2.5 µg/m\textsuperscript{3} NO\textsubscript{2} was added to the modelled concentration. The continental contribution was calculated as a regional mean from measurements from ten different reference stations in the area during the period 2000–2002 [30].

Finally, the concentration modelling was validated with measurements from 23 measuring stations in Scania. The yearly means for these were compared with the modelled annual concentrations of NO\textsubscript{2} for the year 2001. The validation of the concentration modelling (Figure 9) gave a correlation coefficient of 0.69. For concentrations up to 10 µg/m\textsuperscript{3} the modelled concentrations are quite close to the measured values, while at concentrations above 10 µg/m\textsuperscript{3} the model seems to underestimation concentrations ranking from 10 to 15 µg/m\textsuperscript{3} and overestimation concentrations ranking from 15 to 25 µg/m\textsuperscript{3}.

The concentration modelling (Figure 10) shows that there are considerably higher levels of NO\textsubscript{2} in the western regions than in the less populated eastern part of the county. The emissions of NO\textsubscript{2} from the three major cities in the region (Malmö, Helsingborg and Lund) can be seen clearly, as well as the major roads in the county. Also, the ferry traffic along the coast and some of the contributing emissions from Denmark in the west are visible. Areas with NO\textsubscript{2} levels lower than 5 µg/m\textsuperscript{3} have a very low population density and consist mainly of agricultural land or forests.

**Modelling the exposure to NO\textsubscript{2}**

In this study, two different sets of population data were used (henceforth referred to as PD1 and PD2). Both of the data sets were obtained from the Regional Office of Scania, Sweden, and are based on the Swedish National Registry. The data sets are on individual level for the registered population in Scania during 2001, but they differ in spatial resolution and the amount of attribute data linked to them.
1. PD1

In PD1 individuals’ location are represented by points at the centre coordinates of their real estate (listed in the National Registry). The only attributes linked to this data set were sex and age.

2. PD2

In PD2 the location of each individual’s real estate is given as a centroid in a 1 km grid. PD2 was linked with socio-economic attributes from Statistics Sweden and contains the following attributes: sex, year of birth, country of birth, marital status, income for the previous year (2000) and highest educational level.

The task is to add exposure values to and all individuals in PD2. Since we cannot distinguish between the persons living in a 1 x 1 km cell all of them will get the same exposure value. This could be performed by only using only the information in PD2. However, the exposure value can be improved by using the detailed positions of the individuals in PD1 using the methodology described below.

The annual mean of NO$_2$ in Scania for 2001 was modelled with a 250 x 250 m grid resolution (small coloured grid in Figure 11). Since the residences of the individuals within each square kilometre are not evenly distributed in space, it was decided to relocate the centre coordinate to a position within each square kilometre which better corresponded to the total population density within that specific square kilometre.

To find the geographical centre of the individuals’ residences in PD2 the data set was combined with PD1. PD1 contained the exact location of each individual’s real estate (small black dots in Figure 11). By creating a grid (large bold grid cell in Figure 11) out of the kilometre points of census data obtained in PD2 and applying a ID number to each of these cells (consisting of each specific cells X and Y centre coordinates) this ID number could then be transferred to the points in PD1, depending on within which square kilometre grid cell the points in PD1 were positioned. For each set of points with the same ID number (i.e. those that fell within the same square kilometre grid cell) in PD1, the geographical centre was then estimated by calculating the average X and Y coordinate for all the points. The coordinates for this new centre point were then transferred back to PD2 and the individuals within this data set were subsequently repositioned from their centroids to these new coordinates (white crossed circle in Figure 11).

An approximation of the individual’s average annual exposure to NO$_2$ in $\mu g/m^3$ for these new locations was assigned to each individual in the repositioned population data set, PD2, by using bilinear interpolation (white edged grid cells in Figure 11) [33]. This relocation also increases the accuracy in the estimation of the NO$_2$ concentration to which most of the individuals within each square kilometre were exposed [34].

**Division of the population into socio-economic groups**

From the repositioned population set, PD2, which was linked with socio-economic attributes and NO$_2$ concentrations, different subsets were created based on the individual’s country of birth and their educational level.

**Country of birth**

In the field of environmental justice it is common to study ethnicity and exposure to various pollutants. Sweden is not as segregated as some countries regarding race, and therefore race as a measure of socio-economic status is not applicable here. Nevertheless, segregation between immigrants, as a mixed ethnic group, and native Swedes exists to various extents. In 2000 around 11% of the Swedish population had been born outside the country, and most of these immigrants were concentrated in metropolitan areas or larger cities [35].

Since immigrants from other countries tend to have difficulties getting into the labour market they often belong to the lower social and income groups. Therefore, depending on country of birth, the individuals were classified into two different groups: Sweden – individuals born in Sweden (1,007,958 individuals) OC: Other Countries (110,214 individuals) The OC group includes individuals from countries outside the Scandinavia, countries and the European Union in 2001 and those not from: Australia, Canada, Japan, New Zealand or the USA. Individuals who were born in Scandinavian countries other than Sweden (2% of the total population in Scania), countries belonging to the European Union in 2001 (1% of the total population in Scania) and the major economies: Australia, Canada, Japan, New Zealand and the USA (0.2% of the total population in Scania) were excluded from the analysis. This was done as individuals from these countries might immigrate to Sweden under different conditions and for different reasons than immigrants from the other countries.

In a preliminary study it was examined whether or not this exclusion would affect the results of the study. It was found that leaving out this group did not significantly affect the results of the study, due to the small number of individuals in this group, compared to the other two categories,
Level of education

The Swedish educational system has changed over the years, and the definition of an individual with a high educational level has varied considerably. To reduce birth-cohort bias and to focus on the working population the analysis of this index was only carried out on individuals in the age group 25–64 years. Information on the highest level of education was missing for some individuals. In total around 70% of the total population of Scania was included in this analysis.

The attributes regarding education were divided into seven different categories:

1. Pre-secondary education for less than nine years (9% of the population)
2. Pre-secondary education for nine years (11% of the population)
3. Secondary education for maximum two years (19% of the population)
4. Secondary education for three years or longer (13% of the population)
5. Post-secondary education for less than three years (9% of the population)
6. Post-secondary education for three years or longer (9% of the population)
7. Postgraduate studies (0.6% of the population)

These seven different classes were grouped into two generalised classes representing individuals with a high level education (groups 6 and 7) and those with a low level of education (groups 1 to 5).

Statistical analysis

The data were analysed by calculating the Spearman rank correlation coefficients for the exposure to NO₂ and the proportion of the population being highly educated and born in Sweden. A correlation coefficient is a number between -1 and 1 which measures the degree to which two variables are related. Spearman’s correlation is based on ranks, and can be used to describe non-linear relationships. The levels of NO₂ were rounded off to single-decimal numbers. The correlation coefficients were determined calculating the proportion with high education and the proportion born in Sweden for each NO₂ level. The correlation coefficients for the association between level of NO₂ and proportion being highly educated or born in Sweden were then calculated. Consequently, a positive correlation coefficient here shows that the level of NO₂ increases with an increasing proportion of the population having a high socioeconomic status. Correlation coefficients in the range of -0.4 to 0.4 were regarded as weak correlations of minor importance.

The data were plotted in bubble diagrams, with the level of NO₂ on the x-axis and the proportion belonging to each socio-economic group on the y-axis. The size of the bubbles represents the inverse variance of the proportion estimates, and thus represents the certainty in the estimate. A large bubble represents a proportion estimate that was calculated based on a larger group of people than a small bubble. The weights, \( w_i \), for NO₂ concentration strata \( i \) was calculated according to:

\[
\text{var}(\hat{p}_i) = \frac{n_i (1 - \hat{p}_i)}{\hat{p}_i n_i}
\]

\[
w_i = \frac{1}{\text{var}(\hat{p}_i)}
\]

where \( \hat{p}_i \) is the estimated population proportion and \( n_i \) is the number of people in stratum \( i \).

Table 1 presents the size of the population, the proportion belonging to the socio-economic subgroups, the median, range, and quartiles of the concentration of NO₂ in the population for the whole of Scania and for each city in the study.

Authors’ contributions

PP, KJ and US conceived the study and participated in its design and coordination.

US assisted with the statistical analysis.

LH helped to draft the manuscript.

SG carried out the NO₂ concentration modelling.

AO performed the statistical analysis, the alignment and wrote part of the paper.

ES carried out the NO₂ exposure modelling and wrote the final version of the paper.

Acknowledgements

We would like to acknowledge:

Our reviewers for helpful comments and suggestions.

The Swedish National Air Pollution and Health Effects Program (SNAP) for financial support.
Åke Boalt, Region Skåne, for providing us with population data for the region of Scania.

Niclas Olofsson, Department of Social and Economic Geography at Lund University, for providing us with useful help and information.

References


15. Resor och pendling över Oresund [http://www.skane.se/temp/]


17. Statistik om Lund [http://www.lund.se/templates/]

18. Kristianstad municipality information [http://www.kristianstad.se/]


22. Exempel på segregationsanalys.


24. Kristianstad municipality information [http://www.kristianstad.se/]


Paper II

A study of spatial resolution in pollution exposure modelling

Authors:
Emilie Stroh, Lars Harrie and Susanna Gustafsson,

Abstract
This study is part of several ongoing projects concerning epidemiological research into the effects on health of exposure to air pollutants in the region of Scania (Skåne), southern Sweden. The aim is to investigate the optimal spatial resolution, with respect to temporal resolution, for a pollutant database. The results of this study suggest that the spatial resolution is dependent on a number of variables such as study design and quality, temporal resolution and the characteristics of the areas studied.

Key words
Spatial resolution, air pollutants, dispersion models, exposure modelling, GIS.

1 Introduction

1.1 Background
The air we breathe contains a vast number of chemical pollutants. Some originate from natural sources such as soil or salt particles generated from nearby fields or the sea, but most of them originate from man-made sources such as vehicles, industrial plants, combustion and heating. These chemicals not only interact with each other but also with the cells and compounds within our respiratory and cardiovascular systems. Studies have shown that exposure to air pollutants such as fine particles and nitrogen dioxide contributes to excess diseases and mortality among people suffering from cancer, respiratory and cardiovascular diseases (WHO 2003). It has also been shown that air pollution leads to poorer health among vulnerable groups such as asthmatics, old people and children. The health effects on children are of special concern since these health problems might have repercussions later in life (Chaix 2005).
Geographic analyses have previously been used in epidemiology to study the correlation between exposure to air pollutants and health (Jerrett et al. 2005; Briggs 2005). These analyses were conducted in three steps: (1) modelling of the pollutant field using a dispersion model and an emission source database, (2) using the pollutant field to estimate the exposure to pollutants of the study group, and (3) correlation of the exposure with health status of the study group.

This study is part of several ongoing projects involving epidemiological research in the region of Scania in southern Sweden. These projects employ population databases obtained from the Regional Office of Scania and are based on the Swedish National Population Register. These population databases are unusually detailed regarding both the high degree of detail in the socio-economic and health data they contain, but also in their resolution. All the databases contain information on individuals, including the exact location of their residence. This location is either represented as a point at the centre of their real estate or as a centroid in a 500 m - 1 km grid. This detailed information regarding the population in Scania enables us to perform exposure studies with high resolution.

An emission source database including traffic, industry, and combustion was established (Gustafsson et al. 2006) and dispersion modelling software implemented. However, it was found to be impractical to conduct dispersion modelling for each epidemiological study. Therefore, we plan to establish a pollutant database; this database will contain pollutant values in a fixed grid covering Scania for the time period 2001 - 2003. The advantage of a pollutant database, rather than an emission source database, is that it will make epidemiological studies much easier to perform for users who are not accustomed to using emission source databases and dispersion modelling. The disadvantage is that a pollutant database is not as flexible as an emission source database, mainly in that the pollutant database has a fixed resolution in time and space.

It is important to take the time factor into consideration when conducting geographical analysis in exposure studies with the aim of analysing diseases and health hazards with long latency periods, such as cancers. Samples and measurements made when a disease manifests itself do not usually represent historical exposure levels (Cromley 2003). It is therefore important to collect data in order to be able to model accumulated exposure levels from previous years. This need to store data raises the issues of storing capacity and modelling resources. Since digital data for these studies are often required as high-resolution data with a low level of spatial aggregation, the storage capacity needed and the modelling time required increase rapidly. This problem is apparent in exposure studies that take the movement of a dynamic population into account; not only do these studies demand digital data with high spatial resolution, but there is also a need for high temporal resolution in time.

The aim of this study was to evaluate the loss of information due to the use of
a coarse space resolution in the pollutant database. This loss of information must be seen in relation to the accuracy of the estimated pollutant field.

1.2 Related studies

The impact of scale and aggregation factors in epidemiological analysis can not be ignored. As Krieger (2003) states: “in epidemiological studies completeness in geocoding does not equal success. Accuracy – and choice of geographic level – matter as much, if not more” (Krieger 2003). Cromley (2003) also concludes that, in many health studies, the aggregated level of available health data is a major limitation since data are often aggregated inappropriately for the purpose of the study. A study by João (2002) shows that scale, in terms of both detail and extent, is an important issue, which to date has been largely neglected in environmental impact assessments. This is further emphasized by Setton et al. (2005), who stated that many exposure assessments and epidemiological analyses of the impacts of air pollution on health have been undertaken on regional scales, and that only recently have researchers begun to investigate neighbourhood-level variations in pollutant levels.

Finding the optimal scale for studies concerned with factors varying in space is a complex matter of great concern. If large areas are defined they reveal more global structures, while local variations are obscured (Ali et al. 2005). This means that only superficial assessments may be possible, and the uncertainty will increase (CEAA 1996). However, although small neighbourhood sizes yield detailed patterns (Ali et al. 2005) and a more detailed examination may be feasible, the understanding of the broad context may be lost (CEAA 1996). As a result, the choice of optimal size is important when investigating ecological associations with diseases, since these patterns might otherwise be concealed or misjudged. As Gregorio et al. (2006) concludes: “spatial analysis results may be rightfully considered conditional upon the particular geography selected for the study”.

Apart from the risk of altering the results of a study by choosing the “wrong” scale or aggregation level Ali et al. (2005) also concluded that the optimal neighbourhood size is data dependent. Their study suggest that, depending on the factors and correlations included, different neighbourhood sizes may be optimal for different variables. This is supported by findings from our previous study (Stroh et al. 2005). Therefore, it can not always be assumed that the choice of study area for one set of parameters is optimal for another, even though they may be similar. According to Gregorio et al. (2006) nationwide efforts to promote regional health information networks/organisations that cross traditional geo-political boundaries demand a greater understanding of how aggregating health and population data may affect the analysis and interpretation of disease patterns.

Due to the four-dimensional nature of the distribution of atmospheric pol-
lutants, the importance of scale and aggregation level applies not only to spatial but also temporal variations. According to Pénard-Morand et al. (2006), who conducted a study on schoolchildren’s exposure to traffic-related air pollution in France, there is extreme variability in the level of exposure that exists from one place to another. Briggs (2005) also concludes that pollution surfaces in urban areas are often extremely complex, with steep gradients away from ground-level sources, and often highly localised hot-spots and peaks. Also, Cyrys et al. (2005) state that several studies have documented important variations in pollutant concentrations within cities, especially related to motorised traffic and location within the city – for example city centre versus suburbs. These variations are not only variations in space, but also in time mostly due to human behaviour such as rush-hour traffic and industrial activities. Results from a study by Wentz et al. (2002) suggest that there is a distinct relationship between levels of CO₂ and spatial patterns of human activities in urban areas. The results of their study also show that the temporal and spatial patterns of CO₂ levels correspond closely to the density of traffic, population and workplaces. Consequently, the optimal choice of resolution for a pollutant database may be mainly dependent on the spatial distribution of the concentration of air pollution in the area, but also on the temporal distribution.

João shows in her study that the chosen scale should be considered as part of the observation bias since the choice of scale, both in the meaning of the details analysed and the spatial extent of study area, can have important repercussions on the results, such as the determination of impact significance and the measurement of environmental parameters (João, 2002). It is therefore of fundamental importance to define the scale of observation. In other words, scale choice should be explained, justified and explicitly stated in all environmental impact assessment studies (João, 2002).

1.3 Objectives
The overall objective of this study was to determine the optimal resolution for a pollutant database for the county of Scania, southern Sweden. The database will contain concentration values of nitrogen oxides (NOₓ), and possible also other pollutants, over a period of three years (2001-2003).

The optimal spatial resolution is dependent on which time resolution is considered in the epidemiological study. Since the pollutant database will be used mainly for epidemiological studies with durations of days, weeks or even longer periods, we have evaluated the results using different time resolutions.

The difference between modelled and measured values, i.e. the error estimate, of air pollutants has also been evaluated since this together with the error estimate generated from the choice of resolution will combine to give the total error estimate for the pollutant database.
2 Material & Methods

To investigate suitable spatial resolution for our pollutant database we modelled pollutant fields (NO\textsubscript{X} values) for several spatial and time resolutions. This enabled us to compare the loss of information when using coarser resolutions. This loss of information should be seen in relation to the accuracy of the modelled values. To estimate this quality we compared modelled values with values measured at a meteorological station. Our approach consisted of the following steps (Figure 1).

1) **Comparison of modelled values and observed values** (section 2.2). Hourly NO\textsubscript{X} values were generated using the dispersion model. These modelled values were compared with observed values recorded at the meteorological station.

2) **Generation of pollutant fields with a time resolution of one hour** (section 2.3). A dispersion model was used to estimate pollutant fields. The pollutant field with the highest spatial resolution is below denoted *the fine grid*. The other grids are denoted *coarse grids*. Hourly NO\textsubscript{X} values for four weeks were estimated for all grids.

3) **Generation of pollutant fields with a spatial resolution of one day and one week** (section 2.4). Mean values of the original hourly data were used to create new pollutant fields with time resolutions of one day and one week.

4) **Interpolation of the coarse pollutant fields** (section 2.5). The coarse grids were interpolated to obtain the same spatial resolution as the fine grid.

5) **Comparison of the fine grid and the coarse grids** (section 2.6). The values of the fine grid and the interpolated coarse grids were plotted in a time series. The spatial distribution of the difference between the grid values was also studied.

6) **Evaluation.** The results obtained from Step 5 were evaluated concerning the loss of information due to using a coarser spatial resolution. This loss of information is placed in relation to the time resolution (hour, day and week) and the accuracy of the model values (Step 1).
Figure 1: Main flowsheet of the study.
2.1 Study area

The area studied was Scania, the southernmost county of Sweden. The county covers around 11,350 km$^2$, which is approximately 2% of the total Swedish land area. The region is relatively densely populated, with more than 1.1 million people living in the area, which constitutes approximately 11% of the total Swedish population. In Scania, approximately 67% of the population lives close to the west coast. Most road traffic (passenger cars as well as trucks) from the European continent to Sweden and Norway passes through this area, and five motorways run through the region. There are also several harbours in the region and a considerable amount of cargo shipping and ferry transport along the coast. These factors, and the closeness to Copenhagen in Denmark, and the European continent, contribute to high concentrations of air pollutants in the region, compared with most other regions in Sweden.

We studied two sites in Scania, one urban and one rural site (Figure 2). Both of the study sites were 12.8 x 12.8 km. The urban study site covers the city of Lund and its surroundings, while the rural study site covers the area of a small village called Genarp, located approximately 15 km south-east of Lund.

Lund is the third largest city in Scania with approximately 76,000 inhabitants, and the city covers approximately 23 km$^2$. Although the city centre consists mainly of narrow streets and pedestrian precincts, the traffic can be quite intense in some areas, and there are three motorways along the south, east and north boundaries of the city. This creates rather steep gradients in the concentrations of traffic-generated pollutants between the city, enclosed by the motorways, and the surroundings, which consist mainly of agricultural land.

The area around Genarp consists mainly of rural and agricultural land with small clusters of buildings. A major road with a large traffic volume passes through the northern part of the study area.

The validation analysis between modelled concentrations and measured concentrations was performed using data from a meteorological station in Malmö (Figure 2). Malmö is the largest city in Scania with a population of about 260,000. The city is located on the west coast facing Denmark and covers an area of

![Figure 2. Map showing the two study sites: Lund (urban area) and Genarp (rural area) and the location of Malmö.](image-url)
66 km². Malmö is surrounded by a ring road at which three of the five main motorways in the area converge.

2.2 Comparison of modelled values and observed values

NO\textsubscript{X} values were modelled using a dispersion software and an emission source database for air pollutants. This database was constructed by the GIS Centre, Lund University, supported by the Swedish National Air Pollution and Health Effects Programme (SNAP) (Gustafsson et al. 2006). The database includes emission data for all major sources for the years 2001-2003; the main pollutants included are nitrogen oxides (NO\textsubscript{X}) and particular matter. The emission source database covers the whole Öresund region, where Scania, Zealand (Denmark) and the sea around Scania. The purpose of extending the area is to obtain more precise description of air quality in Scania, since emissions from Zealand and shipping affect the air quality in the western part of Scania. The most important sources in the database are road traffic, shipping, aviation, rail transport, industry, power plants, small-scale heating and machinery. Information, statistics and emissions were collected from official sources. The database includes data on about 23,000 road sources, 500 point sources and nearly 100 area sources.

For modelling we used the dispersion software Enviman (Opsis 2006). Enviman estimates the pollutant values over a grid in which all emission sources are located in the centre of the cell. A box model is used to compute the contribution to the cell in which the emission sources are located. The contribution to other cells is calculated using a Gaussian dispersion method (Aermod; USEPA 1998). Enviman has a default time resolution of one hour. To perform dispersion calculations a meteorological data set is needed. In this study we used observed meteorological data, such as wind speed, wind direction, temperature and solar radiation. The meteorological measurements were performed on a 24 m high mast, in an open field a few kilometres from central Malmö. In the calculations, we allowed the plume from each source inside the calculation area travel 6 hours. After six hours, we assumed that the concentration was so low that it had no significant effect on the air pollution levels.

Enviman dispersion software and the emission source database were used to estimate the hourly NO\textsubscript{X} values at a meteorological station in Malmö during four weeks in January 2003. Hourly NO\textsubscript{X} values for the same time period were collected from the meteorological station. The measurements were performed by the local environment protection agency in Malmö and the measuring technique was chemiluminescence.
2.3 Generation of pollutant fields with a time resolution of one hour

Enviman was used to produce five pollutant fields of NO\textsubscript{X} for each of the two study sites, with spatial resolutions of 100 m, 200 m, 400 m, 800 m and 1.6 km. Below the pollutant field with 100 m spatial resolution is denoted the *fine grid*, and the other grids are denoted *coarse grids*. The time resolution for all pollutant fields was one hour and the total time modelled was four weeks in January 2003. In the modelling we used meteorological data from Malmö. It should be noted that the modelling of these pollutant fields does not include background emissions; the reason being that the aim was only to estimate the effect of using different spatial resolutions. Also background emissions produce a long-wavelength component in the pollutant field that is of no interest when studying the effect of the spatial resolution in modelling.

Figure 3 illustrates the 100 m pollutant fields computed for the two study sites. For the urban study site the mean level of NO\textsubscript{X} (from local sources during the study period January 2003) varies between 0.74 and 30.07 µg/m\textsuperscript{3}. It can be seen that the highest values are along the motorways and other major roads, and in the city centre. For the rural study site the level, as well as the gradient, of air pollution is low in comparison with the urban study site; the mean concentrations of NO\textsubscript{X} in the study area during January 2003 were between 0.16 and 4.85 µg/m\textsuperscript{3}.

![Figure 3. Illustration of the modelled NO\textsubscript{X} values (µg/m\textsuperscript{3}) at the urban study site (Lund) and the rural study site (Genarp).](image-url)
2.4 Generation of pollutant fields with time resolutions of one day and one week

From the one-hour data we computed pollutant fields of NO\textsubscript{X} values with temporal resolutions of one day and one week. This step was simply performed by taking the mean values of the one-hour data.

2.5 Interpolation of the coarse pollutant fields

To estimate the loss of information when using a coarser grid it is necessary to know the NO\textsubscript{X} values for the same locations for all pollutant fields. The basic approach used was to estimate the NO\textsubscript{X} values in the coarse grid for all the points in the fine grid; this was performed by interpolation. The dispersion software (Enviman) estimates the NO\textsubscript{X} value at the centre of each cell in the grid. From Figure 4, we can see that there are no common points modelled in both the fine grid and in the coarse grids (if the cell sizes differ by a factor of 2 and the grids have the same extent).

The ideal interpolation method for our application should generate a continuous surface that passes through all the modelled values (i.e., the input values for the interpolation), this and the interpolated surface should be as close as possible to the modelled surface. In this study, we tested two interpolation methods: bilinear interpolation and polynomial interpolation (see, e.g., Zhizhou 1990; Burrough and McDonnel 1998; Press et al. 2002).

Bilinear interpolation requires that the input points be given in a regular grid, and is defined in Eq. 1 (using the notation in Figure 5):

\[ z(x_p, y_p) = z_1 + (z_2 - z_1) \cdot w + (z_3 - z_1) \cdot u + (z_1 - z_2 - z_3 + z_4) \cdot uw \]  

where

\( z(x_p, y_p) \) = interpolated value at the point \( x_p, y_p \) (i.e. the point at which the value is to be interpolated),
$z_1, z_2, z_3, z_4 = \text{values of the four closest points (to } x_p, y_p) \text{ in the coarse grid,}$

$w = (x_p - x_i)/\Delta x,$

$u = (yp - y_i)/\Delta y \text{ and}$

$\Delta x, \Delta y = \text{the spatial resolution of the coarse grid.}$

Bilinear interpolation has the advantage that it is simple and creates a continuous surface that passes through all the input points (points 1-4 in Eq.1). But it is not capable of modelling extreme values between the input points.

![Figure 5. Bilinear interpolation. The value at point P (with coordinates $x_p, y_p$) is interpolated from the values at points 1-4.](image)

Polynomial interpolation can be implemented by computing a two-dimensional polynomial surface. However, this is too computationally intensive for our application (we have to interpolate around 100 million points) and therefore we decided to use a succession of one-dimensional interpolations. We used polynomials of degree three and 4x4 input points in a regular grid (equal to the 16 closest points in the coarse grid) for the interpolation. The approach adopted was to start computing a one-dimensional polynomial for a fixed $y$-value ($y_i$) using Eq. 2 (with the notation in Figure 6):

$$z(x_p|y = y_i) = a_0 + a_1 \cdot x_p + a_2 \cdot x_p^2 + a_3 \cdot x_p^3$$  \hspace{1cm} (2)

Since only four values (along the line $y=y_i$) are used to estimate the polynomial parameters ($a_i$) there is a unique solution (in reality the polynomial is written somewhat differently, see Press et al. 2002, but here we prefer this simple form
for clarity). Equation 2 is then applied for the other three \( y \)-values; this implies that we have four estimations of \( z \) for \( x = x_p \). Finally, these values are used to interpolate the point \((x_p, y_p)\) using a polynomial in the \( y \)-direction (Figure 6).

![Figure 6. Interpolation of the value at point P \((x_p, y_p)\) (grey dot) by a succession of polynomial interpolations. First, the values at points \((y_i, x_p)\) \((i=1,2,3,4)\) (white dots) are interpolated (in the \( x \)-direction) using Eq. 2. These new values are then used as input to interpolate in the \( y \)-direction to determine the value at the point P \((x_p, y_p)\).](image)

Polynomial interpolation provides a continuous surface as long as the same input points are used. However along the boundary lines where the input points are changed the surface is discontinuous. Polynomial interpolation (of degree three) is capable of modelling extreme values between the input values. This can be both an advantage and a disadvantage. For some types of surface characteristics the extreme values will decrease the difference between a modelled surface and an interpolated surface; for other types the difference will increase.

2.6 Comparison of the fine grid and the coarse grids

Interpolation provides pollutant fields with values for the same points. For the fine grid the values are modelled and for the coarse grid the values are interpolated. Furthermore, we have pollutant fields with time resolutions of \textit{one hour, one day and one week}.

A number of methods are available to compare the values from the grids. Of special interest for our objective is the discrepancy between the values from the modelled fine grid and the interpolated values from the coarse grids. The dis-
crepancy is estimated for each cell by its mean value and the standard deviation. The mean value of the discrepancy for each cell \((\bar{\Delta}^{i,j})\) is estimated by:

\[
\bar{\Delta}^{i,j} = \frac{1}{T_{\text{hour}}} \sum_{t=1}^{T_{\text{екс}}} \left( \text{FineNOX}^{i,j}_{\text{hour},t} - \text{CoarseNOX}^{i,j}_{\text{hour},t} \right)
\]

(3)

where:

\(i, j\) denotes that the cell in row \(i\) and column \(j\) in the grid, \(T_{\text{hour}}\) is the number of time intervals for a spatial resolution of one hour, \(t\) denotes the time interval, \(\text{FineNOX}^{i,j}_{\text{hour},t}\) are the NO\(_X\) values for time interval \(t\) and cell \(i,j\) in the fine grid, and \(\text{CoarseNOX}^{i,j}_{\text{hour},t}\) are the NO\(_X\) values for time interval \(t\) and cell \(i,j\) in the interpolated coarse grid.

The mean value will be the same regardless of whether we study hourly, daily or weekly NO\(_X\) values (since the daily and weekly data were computed as mean values from the hourly data). Behind this lies an approximation that the mean values of the discrepancies are constant throughout the day and between different days. This approximation is fairly good.

The standard deviation will vary depending on whether we study hourly, daily or weekly data. For each cell and time resolution, the standard deviation \((s_x^{i,j})\) is given by:

\[
s_x^{i,j} = \sqrt{\frac{1}{T_x - 1} \sum_{t=1}^{T_x} (\Delta_x^{i,j} - \bar{\Delta}^{i,j})^2}
\]

(4)

where:

\(x\) denotes the time resolution (\(\text{hour, day or week}\)), and

\(\Delta_x^{i,j} = \text{FineNOX}^{i,j}_{x,t} - \text{CoarseNOX}^{i,j}_{x,t}\).

Finally, we also use the mean value for all cells, i.e.:

\[
\bar{\Delta} = \frac{1}{NM} \sum_{i}^{N} \sum_{j}^{M} \bar{\Delta}^{i,j}
\]

\[
\bar{s}_x = \frac{1}{NM} \sum_{i}^{N} \sum_{j}^{M} s_x^{i,j}
\]

(5)

where \(N, M\) are the number of rows/columns.
2.7 Software environment

A program in C++ was developed to compute the daily/weekly values, perform the interpolation and compute the statistics. Input data for the program were binary files containing modelled NO\textsubscript{X} values from the dispersion software Enviman. For the implementation of the polynomial interpolation the program utilizes functions from *Numerical Recipes in C* (Press et al. 2002). The time series were generated by a standard spreadsheet program (MS Excel) and the maps by a standard GIS program (ArcGIS).

3 Comparison with measured values

3.1 Results

To estimate the quality of the modelled concentrations of NO\textsubscript{X}, measured values from a meteorological station in Malmö were compared with the concentration values modelled with the same emission database and dispersion model used to generate values for the study sites. The results were plotted as a time series of daily mean values (Figure 7). As can be seen in the figure, the divergence between the daily time series is large during the first week, and thereafter follows the pattern of the measured values, with some exceptions, reasonably well.

The mean difference between the modelled and the measured value of NO\textsubscript{X} is 0.6 µg/m\textsuperscript{3} and the standard deviation of the differences is 12.3 µg/m\textsuperscript{3} for the hourly difference, 5.9 µg/m\textsuperscript{3} for the daily difference and 4.4 µg/m\textsuperscript{3} for the weekly difference.

![Modelled & Measured values of NO\textsubscript{X} (January 2003)](image)

*Figure 7. Modelled and measured NO\textsubscript{X} values for the meteorological station in Malmö.*
3.2 Discussion

Despite the temporal variations seen in Figure 7 the mean difference is low, with a value of approximately 0.6 µg/m³ NOₓ. However, the standard deviation of the differences reveals that the divergence for the hourly values is high. This number decreases considerably when comparing the concentrations as daily or weekly means. This indicates that the general trend predicted by the dispersion model follows the actual pattern of concentrations of air pollutants in Malmö, but that the dispersion model is too coarse to reflect finer temporal variations than daily means. If the highest and lowest quartiles are removed from the data set the standard deviation of the means is lowered considerably, and in the case of daily means the standard deviation of the difference is reduced from 5.9 µg/m³ to 2.4 µg/m³ NOₓ.

The main contribution to both the modelled and measured values is traffic generated. The dispersion model uses predefined patterns of traffic flow in the area to estimate the concentration and levels of emitted air pollutants. This may be the reason for the large divergence between modelled and measured values at the beginning of the month (days 1-6). During the first week in January many people are still on holiday due to the Christmas and New Year holidays, causing the traffic flow to diverge from its normal pattern. This will probably lead to lower levels of measured traffic-generated emissions than the modelled results. However, this does not explain the large divergence between modelled and measured values during days 9-11 and 23-26, which probably originates from errors related to emission sources or the calculation of the dispersion model.

4 Comparison of interpolation methods

4.1 Results

The coarse grids (with spatial resolutions of 200, 400, 800 and 1600 m) of hourly data were interpolated by bilinear interpolation (Eq. 1) and polynomial interpolation (Eq. 2) for the urban study site. The mean and standard deviation of the difference between interpolated grids and the fine grid (100 m resolution) were computed. In Table 1 the mean values for all cells given by Eq. 5 are listed. Figure 8 illustrates the daily mean, during January, of NOₓ concentrations calculated for the urban study site with the resolutions of 400 and 800 metres using the two interpolation methods.
Table 1. Mean ($\overline{\Delta}$) and standard deviation ($\overline{s_{hour}}$) of the difference between interpolated and modelled hourly NO$_X$ values for the whole urban study site, January 2003.

<table>
<thead>
<tr>
<th>Resolution (m)</th>
<th>Bilinear $\overline{\Delta}$ [ug/m$^3$]</th>
<th>Bilinear $\overline{s_{hour}}$ [ug/m$^3$]</th>
<th>Polynomial $\overline{\Delta}$ [ug/m$^3$]</th>
<th>Polynomial $\overline{s_{hour}}$ [ug/m$^3$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.141</td>
<td>0.143</td>
<td>0.140</td>
<td>0.142</td>
</tr>
<tr>
<td>400</td>
<td>0.309</td>
<td>0.346</td>
<td>0.307</td>
<td>0.345</td>
</tr>
<tr>
<td>800</td>
<td>0.473</td>
<td>0.682</td>
<td>0.465</td>
<td>0.678</td>
</tr>
<tr>
<td>1600</td>
<td>0.394</td>
<td>1.119</td>
<td>0.342</td>
<td>1.125</td>
</tr>
</tbody>
</table>

Figure 8. NO$_X$ values [µg/m$^3$] obtained by bilinear and polynomial interpolation for the urban study site with spatial resolution 400 and 800 metres and a time resolution of one hour during the whole of January 2003. The polynomial interpolation method requires 16 (4x4) input values, unlike the bilinear interpolation method, which requires only 4 input values, and therefore a large proportion of the grid cells close to the boundary of the modelled area are not computed.
4.2 Discussion

The interpolated surface should preferably be as close as possible to the modelled surface of the fine grid. As can be seen in Table 1 the differences and standard deviations between the two interpolation methods are quite similar. Both the mean and the standard deviation are slightly better for the polynomial interpolation method than the bilinear.

Studying the different methods visually reveals, however, that the interpolation methods differ spatially. As can be seen from the polynomial interpolation shown in Figure 8, this method gives low values less than zero. This is caused by the fact that polynomial interpolation of degree three can model local maxima/minima between the input values where the contrast for the modelled values is high (this implies that the parameter \( a_3 \) in Eq. 2 is large if the input values vary too much). In Figure 8, the local minima for the polynomial interpolation are seen as black dots close to areas with high concentrations (i.e. bright areas).

Although the polynomial surface seems to follow the fine surface slightly better on average than the bilinear interpolation, we chose to use the bilinear interpolation method for this study (and recommend bilinear interpolation be used in future studies). The main reason is that polynomial interpolation can produce poor-quality output if there is high contrast in the input data.

5 Comparison of interpolated values and modelled values

5.1 Results

To estimate the loss of information using a coarse grid instead of a fine grid we compared the fine grid and the interpolated coarse grids by plotting their daily mean NO\(_X\) concentrations against each other. In Figure 9 the modelled values for the 100 metre grid (fine grid) and for the interpolated coarse grids (200, 400, 800 and 1600 metres) for the urban study site are plotted in a time series of daily values. As can be seen in Table 2 and Figure 9 almost all of the coarser grids, except for the 1600 m interpolated grid, gave lower or equal concentrations when they were interpolated to the same resolution as the fine grid. The grid with the coarsest resolution, 1600 metres, overestimates the concentrations during periods with low concentration. The accuracy and fit to the fine grid seem to be dependent on the resolution, the higher resolution, the better the accuracy compared to the fine grid.
Table 2. Mean ($\overline{\Delta}$) and standard deviation ($\overline{S}_{\text{hour}}$, $\overline{S}_{\text{day}}$, $\overline{S}_{\text{week}}$) of concentrations of NO$_X$ ($\mu$g/m$^3$) between the interpolated coarse resolutions and the fine grid for the whole month of January.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>$\overline{\Delta}$</th>
<th>$\overline{S}_{\text{hour}}$</th>
<th>$\overline{S}_{\text{day}}$</th>
<th>$\overline{S}_{\text{week}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban study site: Lund</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 m</td>
<td>0.14</td>
<td>0.87</td>
<td>0.45</td>
<td>0.22</td>
</tr>
<tr>
<td>400 m</td>
<td>0.38</td>
<td>1.63</td>
<td>0.80</td>
<td>0.40</td>
</tr>
<tr>
<td>800 m</td>
<td>0.70</td>
<td>3.00</td>
<td>1.36</td>
<td>0.69</td>
</tr>
<tr>
<td>1600 m</td>
<td>0.96</td>
<td>5.52</td>
<td>2.52</td>
<td>1.33</td>
</tr>
<tr>
<td>Rural study site: Genarp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 m</td>
<td>0.03</td>
<td>0.16</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>400 m</td>
<td>0.09</td>
<td>0.27</td>
<td>0.14</td>
<td>0.07</td>
</tr>
<tr>
<td>800 m</td>
<td>0.17</td>
<td>0.46</td>
<td>0.23</td>
<td>0.10</td>
</tr>
<tr>
<td>1600 m</td>
<td>0.27</td>
<td>0.71</td>
<td>0.36</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Figure 9. Time series of NO$_X$ concentrations ($\mu$g/m$^3$) for different resolutions interpolated to 100 metres by bilinear interpolation. The graph shows the mean value for all cells.
Studying the standard deviations in Table 2 and Figure 10 reveals that when concentration values are aggregated in time (i.e. hours, days and weeks) the deviation and disagreement between the different resolutions is considerably reduced. It is also quite obvious that the standard deviation is much lower for the rural study site than for the urban one.

![Areal standard deviation of concentrations of NO\textsubscript{X} between the interpolations and the fine grid](image)

*Figure 10. Areal standard deviation of NO\textsubscript{X} concentrations (µg/m\textsuperscript{3}) for different resolutions interpolated to 100 metres by bilinear interpolation. The graphs show the values for the urban and rural study sites.*

5.1.1 Urban study site - Lund
The spatial distribution of standard deviations of differences in daily concentrations ($\bar{\sigma}_{\text{day}}^{i,j}$) between different resolutions of modelled daily concentrations for the urban study area can be seen in Figure 11. The dissimilarity at lower resolutions is quite apparent, and varies between 0 and 30 µg/m\textsuperscript{3}. The largest divergence is seen in areas of high contrast, such as along roads, near combustion sources and in the city centre. For the resolution of 200 metres the divergence is largest near the motorways, with values up to approximately 5 µg/m\textsuperscript{3}, and near combustion sources, 2.60 - 30 µg/m\textsuperscript{3}. For the 400 metre grid the values begin to diverge near major roads ($\bar{\sigma}_{\text{day}}^{i,j}$ approximately 5-10 µg/m\textsuperscript{3} ). For the lower resolutions of 800 and 1600 metres calculations fail to reflect the correct concentrations in the city centre, and for the coarsest resolution (1600 metres) the divergence for the entire city of Lund is 5.84 µg/m\textsuperscript{3}.
5.1.2 Rural study site - Genarp

The spatial distribution of the standard deviation of differences between different resolutions of modelled daily concentrations for the rural study area can be seen in Figure 12. The dissimilarity at lower resolutions is not as apparent as for the urban study area. The divergence varies between 0 and 5 µg/m³, but for all the resolutions most of the area has a standard deviation of less than 1 µg/m³. As in the case of the urban study area, the highest contrasts occur along roads and near the village centre. For the 200 m resolution the standard deviation of the difference is largest near the motorway 0 – 3µg/m³ and the standard deviation near the village of Genarp is well above 1 µg/m³. For the 400 and 800 metre
grids the changes in estimation along the motorway are somewhat more apparent and the standard deviation rises to a maximum of approximately 5 µg/m³. Using a resolution of 1600 metres the daily standard deviation for the village centre of Genarp increases to levels between 1.1 and 2 µg/m³.

5.2 Discussion
As can be seen from Table 2 and Figure 9 there is a systematic difference between the values from the fine grid and the interpolated values from the coarse grids. The larger the cell size, the lower the NO\textsubscript{X} value. It should be noted that this is not due to the interpolation but the dispersion modelling. This can be illustrated by studying the original modelled values. The mean values for all modelled
points (for all cells and time intervals); were computed and denoted Mean100, etc. The difference between the grids was then calculated to obtain: Mean100-Mean200 = 0.14 µg/m³, Mean100-Mean400 = 0.32 µg/m³, Mean100-Mean800 = 0.52 µg/m³, and Mean100-Mean1600 = 0.52 µg/m³. These values show a clear resemblance to the values in Table 2.

As can be seen in Table 2 and Figure 9 all the coarser resolutions seem to reflect the temporal variations in concentrations well when they are interpolated to a finer grid size, except for the coarsest grid (1600 m). The level of agreement increases with an increase in temporal aggregation. For hourly values a resolution above 400 metres may be too coarse since this gives values with a standard deviation of 3 µg/m³ or higher. However, if the temporal aggregation level is increased to daily or weekly means the standard deviation is immediately reduced to values of less than 3 µg/m³ for all the analysed resolutions. Data with a temporal resolution of hours might be necessary for dynamic exposure studies when exposure is studied for different time intervals during a day. For these kinds of studies a resolution above 200 metres for the modelled concentrations of air pollutants in an area might be too coarse. For exposure analyses with a longer duration of exposure a temporal aggregation of daily or weekly means is sufficient.

It is important to take into consideration the spatial variations in standard deviation. Concentrations modelled in urban areas with coarser resolution than 400 metres seem to generate larger error estimates in areas with high contrasts. However, for rural areas, where the variations in air pollutants are much lower, even a grid size of 1600 metres seems to generate reasonable results. Most individuals in Scania live in urban areas where both levels and gradients of air pollutants are high. When conducting exposure studies in these areas too low a resolution might fail to reflect these variations, thus giving incorrect exposure estimates for the inhabitants. Therefore, the results from the urban study site should have a greater impact on the choice of resolution. If data are to be stored as hourly values a resolution of, at least, 400 metres is preferable as lower resolutions will yield too high deviations in the cities and in areas with high gradients. However, this may lead to a huge increase in the need for generation and storage of data. One option could therefore be to generate a grid for the entire region of Scania with a resolution of 800-1600 metres, and in addition to this generate concentration grids with higher resolution, at least 400 metres, for the larger urban areas. This could lead to calculated concentrations in areas along major roads in the rural areas having a larger standard deviation than the rest of the region. The advantages, on the other hand, would be that the data storage and computer capacity needed would decrease considerably, while exposure estimates for the vast majority of the population in the region would be accurate.
6 Conclusions

When establishing a pollutant database certain criteria must be taken into consideration. The fact that a pollutant database has a fixed resolution in time and space makes the choice of resolution for these two parameters critical for the future use of the database. The aim of this study was to investigate the optimal spatial resolution with respect to temporal resolution for future epidemiological studies in Scania, Sweden. The pollutant database will be used mainly for epidemiological studies with durations of days, weeks or even longer periods. Some studies might model a dynamic population (e.g. modelling being at work and at home), and therefore we need an hourly time resolution in the pollutant database. However, since the duration of the studies is at least one day the spatial resolution should be set in relation to this time resolution.

The standard deviation for the accuracy of daily mean values of air pollutants is approximately 6 µg/m³; a value that could be decreased by about 50% if the extreme values are removed. The choice of spatial resolution should not considerably decrease this accuracy; therefore, we believe that an error due to coarse resolution greater than 1 µg/m³ is inadvisable. Based on Table 2 we conclude that for urban areas a spatial resolution of 200-400 m is suitable; and for urban areas the spatial resolution could be coarser (about 1600 m). This implies that we should develop a pollutant database that allows different spatial resolution for urban and rural areas.

Acknowledgements

We would like to acknowledge EMFO (Emission research programme conducted by the Swedish Road Administration) for financial support of this project, Anna Axmon for statistical advice, Kristina Jakobsson Petter Pilesjö for general support and Håkan Törnevik for consultation regarding the dispersion model Enviman.
References


This study investigated residential noise from road traffic and its relation to annoyance, disturbance of daily activities, and general health. 

Methods: A large public health survey in southern Sweden 1999-2000 supplied data (N = 13 557; 54% participation rate) on demography, annoyance and disturbance of daily activities, and on general health problems regarding concentration, sleep, stress, and treatment for hypertension. Residential road noise exposure was assessed using a geographic information system. Associations with 24 hr equivalent (average) and maximum road noise level were investigated for all participants and for selected subgroups using Cochran-Armitage trend test and Cox regression analysis.

Results: Annoyance from road traffic noise and disturbance of daily activities increased markedly with road noise exposure. More than 25% reported at least occasional disturbance from traffic noise during relaxation and sleep in the highest exposure category for each noise measure. No overall pattern between road noise exposure and general health problems emerged. Among subjects that reported annoyance from road traffic noise (N = 623), average road noise level was associated with concentration problems (p for trend = 0.03) and with hypertensive treatment (p for trend = 0.02). Positive associations between average road noise exposure and health problems were found among females (hypertension), subjects born outside Sweden (sleep), unemployed (stress), and among subjects that reported financial problems (concentration problems).
Conclusions: Exposure to road traffic noise at high levels was common and produced frequent disturbances of daily activities. Negative health effects from road traffic noise were observed in important subgroups.

Key terms
Epidemiology; geographic information systems; noise, transportation.

Background
With the exception of direct damage to the hearing organ caused by sound exposure, the individual perception of sound as negative, i.e. as noise, is likely to be an important determinant of its health effects (1, 2). Annoyance from noise exposure may stem from, or may generate, disturbance of daily life such as sleep, relaxation, concentration, conversation, listening to radio and watching TV. The prevalence of annoyance increases with increasing equivalent sound level (3), but is determined also by fear associated with the noise source and individual noise sensitivity (4). General demographic variables are less strong determinants of annoyance (5). It has been suggested that increased vulnerability for stress inducing factors could trigger annoyance from noise (3), and studies of critical subgroups and the role of annoyance as a mediator of health effects have been encouraged (6). Non-auditory physical health effects that are biologically plausible in relation to noise exposure and annoyance from noise exposure include changes in blood pressure, heart rate, and levels of stress hormones (7), which may increase the risk for hypertension and ischemic heart disease (1, 2, 8, 9).

Residential noise is experienced in closed rooms of dwellings as a result of noise sources unrelated to the dwelling and it is measured outside at the facade (10). Approximately 30% of the population in the European Union is exposed to a day-night equivalent level (L_{dn}; see appendix 1) of traffic noise at the residence exceeding 55 dB(A), which makes transports (on roads, by rail, and in the air) the most important source of community noise in Europe (11). Very high 24 hr equivalent levels (L_{Aeq,24}; appendix 1) greater than 65 dB(A), seem to have stabilised in some countries, while exposure in the range 55-65 dB(A) has significantly increased as a result of the fast growing volume of road traffic (12). Noise exposure has been forecast to worsen along ring roads and motorways, and at nearby regional airports, because of the growth in transportation, especially freight and air traffic (13). In Sweden 16% of the population was estimated to be exposed to equivalent noise levels from road traffic exceeding 55 dB(A) in the year 2000, with no certain over-all time-trend since 1995 (14). Frequent annoyance (at least once a week) due to noise from road traffic was reported by 9% in a Swedish survey in 1999 (15), and by 18% in our preliminary calculations based
on survey data from the region Skåne (the Scania region) in southern Sweden in 2004 (16).

The aim of the present study, which uses data from a large public health survey conducted in southern Sweden 1999-2000, was to investigate residential noise originating from road traffic and its relation to annoyance, disturbance of daily activities, and general health. It was of special interest to investigate health effects among subjects annoyed from road traffic noise and in subgroups that are possibly more vulnerable.

Materials and methods

Public health survey
In order to investigate the relation between road traffic noise and possible health effects, we used data from a large public health survey distributed as a mailed questionnaire (17). The study population for the survey was defined as all persons at least 18 and at most 80 years old and living in any of 33 municipalities in the Scania region in southern Sweden in November 1999. After stratifying the study population in 60 different geographical areas, samples of approximately equal size were randomly selected for each stratum from the population registry. The total sample comprised 24,945 persons and the questionnaires were mailed between November 1999 and April 2000. Three mailed reminders and one reminder by telephone were used in order to increase the participation rate. In the end, answers were obtained from 13,604 (54.5%) of the 24,945 persons selected for the health survey. The reasons for non-participation were: no reply (63.0%), refusal (25.2%), unable to answer due to illness, traveling, or other reason (6.9%), invalid address in registry (3.9%), and answer from wrong person (1.0%). The participation rates were generally higher among females than among males and generally increased with age. A detailed analysis of the non-participants showed that persons with low education and persons born outside the Nordic countries were underrepresented among the participants (18). Among the 13,604 participants, residential geocodes were possible to obtain for 13,557.

In the mailed questionnaire, detailed questions were asked regarding self-reported health, long-term diseases and sick-leaves, treatment with drugs, healthcare usage, annoyance from environmental factors (electrical equipment and smells (19)), social network, occupation and work environment, smoking habits, alcohol consumption, physical exercise, financial situation, education, civil status, country of origin, and residential environment. The health-sections of the questionnaire included questions about ability to concentrate during the last weeks (General Health Questionnaire – 12 (20)), sleeping disturbance during the last two weeks, insufficient sleep generally, stress, and treatment for hyper-
tension during the past 12 months. These general health questions were asked without referring to traffic noise or any other exposure. In the section about residential environment, specific noise-related questions were asked about traffic noise disturbances of daily activities, together with three general questions about annoyance from roads, trains, and aircrafts (see appendix 2).

Assessment of road traffic noise

No measurements of sound levels were conducted. Instead, we used a geographic information system (GIS) in order to assess the outdoor noise exposure from traffic noise. Geocoded residential addresses at the end of year 1999 for the participants in the public health survey and road traffic data were used. No data on train traffic or aircrafts were available. Road traffic data included 21 397 road segments, 17 339 administrated by Swedish Road Administration and 4 058 by local municipalities. The number of vehicles was available for 82% of the road segments. Speed limits were available for >95% of the segments. Some of the traffic data are not fully updated but 93% percent are from 1985 or newer and 71% are from 1995 or newer. For road segments without traffic data, mean values were assigned to each segment based on existing data for included road types (21).

Based on the road traffic data, we used a simplified version of the Nordic prediction method for road traffic noise [see refs. (22, 23) for a complete description] in order to estimate noise exposure for the residential locations of the study participants. In short, the Nordic prediction method first calculates the unattenuated noise level at 10 meters distance from the road center using the number of light and heavy vehicles and the speed limit of each road segment. Corrections are then calculated for:

- The distance between source (the road) and receptor, where the noise level decrease with 3 dB with a doubling of the distance.
- Attenuation due to ground surface type and noise barriers. The attenuation of the noise is depending of surface type with less attenuation for hard surfaces (asphalt, water, concrete) and more attenuation for soft surfaces (vegetation, grass etc).
- Additional corrections for special cases (including very steep topography, reflections from buildings etc).

In this study, we had to simplify the Nordic prediction method by using corrections for distance and surface type only. We were not able to correct for noise barriers and the additional special cases mentioned above, as no such data were available.

We assumed flat ground in all cases and soft surfaces between the residence and the road for the participants living on the countryside (N = 2 199), while hard surface was assumed for the participants living in more densely populat-
ed areas (N = 11 358). We had no data on which floor the residences in the apartment blocks were located and we therefore estimated the noise level at the ground floor for all residences. We estimated the A-weighted equivalent noise level during full day (24 hr; LA_{eq,24}), and the A-weighted maximum noise level with fast time weighting (LA_{fmax}; see appendix 1). Estimated noise levels at day- and night-time were too strongly correlated with noise level during full day to be used for separate analyses. Based on the number of vehicles (light and heavy) and the speed limit for each road segment, we calculated LA_{eq,24} and LA_{fmax} for each 25 meter zone up to 300 meter from the road center. Figure 1 illustrates the outdoor noise exposure for the residences in each 25 meter zone next to an arbitrary road segment.

![Figure 1. An example of estimated A-weighted equivalent (average) noise level during full day (LA_{eq,24}) outside the residences in each 25 meter zone next to an arbitrary road segment.](image)

As individuals may appear in noise zones for more than one road segment, the maximum values for LA_{eq,24} and LA_{fmax} across all road segments near the residence were extracted for each person and used for further analysis. Road noise exposure was categorized according to equivalent (average) noise level during full day (LA_{eq,24}; Low if < 50 dB(A), Medium if 50 – 54 dB(A), and High if ≥...
55 dB(A)) and according to maximum noise level (LA_{fmax}: Low if < 60 dB(A), Medium if 60 – 69 dB(A), and High if ≥ 70 dB(A)). Fifty percent of all geocoded participants were assigned the same exposure category (low/medium/high) for both average and maximum noise level.

Statistical analysis
For noise from each traffic source, subjects reporting fairly much or much annoyance were classified as “annoyed”, whereas subjects reporting no or not much annoyance were classified as “not annoyed” (see appendix 2). Associations between general background characteristics of the participants and exposure category (low/medium/high) and the prevalence of annoyance (fairly much or much) from road traffic noise were investigated using chi-square test. If data suggested monotonically increasing or decreasing trends in the prevalence of disturbance (occasional or frequent), annoyance, and self-reported health problems in relation to exposure category, we tested these trends with Cochran-Armitage trend test (24), using StatXact-6 (Cytel Software Corporation). Trends in health problems were investigated overall as well as separately among subjects who were annoyed by road traffic noise. In addition, we also analyzed self-reported health problems in relation to exposure category in the following subgroups that we hypothesized as being possibly more vulnerable due to exposure to social stressors and for which we observed higher prevalences of annoyance: females, subjects born outside Sweden, unemployed, and subjects that reported financial problems. Observed trends for health problems were investigated further by Cox regression with constant risk period (equal to one) (25), using SPSS 12.0.1 for Windows (SPSS Inc.). In the Cox regression analyses, we adjusted for gender, age in three broad groups (< 45 years, 45 – 64 years, and ≥ 65 years), and, when investigating treatment for hypertension, body mass index (BMI) in 5 groups (missing, < 20, 20 – 24, 25 – 29, and ≥ 30 kg / m2). We considered p-values below 0.05 and 95% confidence intervals (CIs) for prevalence ratios (PRs) that excluded unity as significant.

Population prevalences were estimated in weighted statistical analyses, which accounted for the stratified sampling scheme with respect to geographical area and the selective participation with respect to gender and age.

Results

Associations with exposure and annoyance
We estimated that 29% (95% CI 28 - 30%) of the study population, the Scania region in southern Sweden, had high average exposure (ie, LA_{eq,24} ≥ 55 dB(A)) to road traffic noise and that 37% (95% CI 36 – 37%) had high maximum exposure (ie, LA_{fmax} ≥ 70 dB(A)). Country of origin, civil status, smoking status, em-
ployment status, type of residence, and financial problems were all markedly associated with both average and maximum road noise exposure (p ≤ 0.002 for all of these associations). In particular, living in apartment blocks was much more common among subjects with high average or maximum exposure to road traffic noise than among others (Table 1). The proportions of subjects born in Sweden, married/cohabiting, and employed were somewhat lower, whereas the proportion of current smokers and the proportion of subjects that reported financial problems were somewhat higher among subjects with high road noise exposure. These differences were generally more marked when exposure was grouped according to average noise level rather than according to maximum noise level. Exposure did not differ noticeably with respect to age, gender, BMI, physical exercise, or educational level.

Fairly much or much annoyance from road traffic noise was reported by 4.7% (95% CI 4.4 – 5.1%; weighted analysis yielded an identical population estimate). Gender, country of origin, civil status, physical exercise, employment status, type of residence, and financial problems were all associated with the prevalence of annoyance (fairly much or much) from road traffic noise (p ≤ 0.005 for all these associations except for gender where p = 0.036). In particular, the proportion of employed were much lower among the annoyed than among others (table 1).

Annoyance from road traffic noise increased markedly with road noise exposure (table 2; p < 0.001 for all associations). No marked associations between exposure to road traffic noise and fairly much or much annoyance from aircraft or train noise were discerned.

**Disturbance of daily activities and self-reported health problems**

Disturbance of daily activities (sometimes or frequent) increased with both average and maximum road noise exposure (table 3; p < 0.001 for all associations). More than 25% reported at least occasional disturbance from traffic noise during relaxation and sleep in the highest exposure category for each noise measure.

Among all participants, no consistent pattern between road noise exposure and health problems emerged (table 4). However, extensive sleeping disturbance last two weeks was more common among subjects with medium or high average road noise exposure (p for trend = 0.01), the same trend was observed in the multivariable analysis using Cox regression. The PR for extensive sleeping disturbance last two weeks contrasting average exposure ≥ 55 dB(A) with exposure below 50 dB(A), was 1.2 both unadjusted and after adjustment for gender and age (95% CI 1.1 – 1.4).
Table 1. Characteristics of 13,557 participants of the public health survey in the Scania region in southern Sweden 1999-2000 in relation to exposure to and annoyance from road traffic. The exposure was categorized according to equivalent (average) noise level during full day (LA\textsubscript{eq,24}) of at least 55 dB(A) and according to maximum noise level (LA\textsubscript{fmax}) of at least 70 dB(A). Subjects who reported fairly much or much annoyance from road traffic noise were categorized as “annoyed”. All characteristics are given in percent if not otherwise stated.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N\textsuperscript{a}</th>
<th>All</th>
<th>N = 3,489</th>
<th>N = 4,509</th>
<th>N = 623</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>13,557</td>
<td>49 (19–78)\textsuperscript{b}</td>
<td>48 (19–78)\textsuperscript{b}</td>
<td>49 (20–78)\textsuperscript{b}</td>
<td>48 (20–77)\textsuperscript{b}</td>
</tr>
<tr>
<td>Females</td>
<td>13,557</td>
<td>54.4</td>
<td>54.9</td>
<td>54.9</td>
<td>58.4</td>
</tr>
<tr>
<td>Born in Sweden</td>
<td>13,377</td>
<td>89.3</td>
<td>85.1</td>
<td>86.7</td>
<td>83.5</td>
</tr>
<tr>
<td>Married/cohabiting</td>
<td>13,289</td>
<td>70.4</td>
<td>62.2</td>
<td>64.9</td>
<td>65.3</td>
</tr>
<tr>
<td>Body mass index (^{\text{kg/m}^{2}})</td>
<td>13,080</td>
<td>25 (19–34)\textsuperscript{b}</td>
<td>24 (19–34)\textsuperscript{b}</td>
<td>24.5 (19–34)\textsuperscript{b}</td>
<td>25 (18–34)\textsuperscript{b}</td>
</tr>
<tr>
<td>Physical exercise</td>
<td>12,971</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mainly sedentary</td>
<td>13,314</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td>12,971</td>
<td>15.1</td>
<td>15.7</td>
<td>16.1</td>
<td>19.7</td>
</tr>
<tr>
<td>Regular or hard</td>
<td>12,971</td>
<td>63.2</td>
<td>61.5</td>
<td>62.2</td>
<td>59.9</td>
</tr>
<tr>
<td>Smoking status</td>
<td>12,882</td>
<td>21.8</td>
<td>22.8</td>
<td>21.7</td>
<td>20.4</td>
</tr>
<tr>
<td>Never</td>
<td>12,882</td>
<td>33.0</td>
<td>31.7</td>
<td>33.5</td>
<td>33.2</td>
</tr>
</tbody>
</table>

\textsuperscript{a}N indicates the number of participants.

\textsuperscript{b}Values in parentheses indicate the range.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Na</th>
<th>All</th>
<th>Average (\geq 55 \text{ dB(A)}) ((N = 3,489))</th>
<th>Maximum (\geq 70 \text{ dB(A)}) ((N = 4,509))</th>
<th>Annoyed from road traffic ((N = 623))</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 - 12 years at school</td>
<td>30.7</td>
<td>30.4</td>
<td>29.8</td>
<td>29.9</td>
<td></td>
</tr>
<tr>
<td>Vocational training</td>
<td>10.2</td>
<td>10.2</td>
<td>10.6</td>
<td>9.9</td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>26.1</td>
<td>27.7</td>
<td>26.2</td>
<td>27.1</td>
<td></td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>375</td>
<td>51.8</td>
<td>54.5</td>
<td>46.5</td>
<td></td>
</tr>
<tr>
<td>Retired or on sick-leave</td>
<td>25.0</td>
<td>27.2</td>
<td>27.3</td>
<td>30.8</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>8.8</td>
<td>11.4</td>
<td>10.0</td>
<td>10.9</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>5.1</td>
<td>5.9</td>
<td>5.0</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>Works at home</td>
<td>3.6</td>
<td>3.6</td>
<td>3.2</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>Type of residence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private house</td>
<td>51.8</td>
<td>32.1</td>
<td>41.3</td>
<td>44.5</td>
<td></td>
</tr>
<tr>
<td>Apartment block</td>
<td>29.1</td>
<td>52.2</td>
<td>43.3</td>
<td>37.1</td>
<td></td>
</tr>
<tr>
<td>Terrace house</td>
<td>13.6</td>
<td>10.3</td>
<td>10.2</td>
<td>9.3</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>5.5</td>
<td>5.4</td>
<td>5.3</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td>Problems with paying bills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>71.2</td>
<td>67.9</td>
<td>69.5</td>
<td>61.1</td>
<td></td>
</tr>
<tr>
<td>Only very occasionally</td>
<td>19.0</td>
<td>20.4</td>
<td>19.6</td>
<td>22.3</td>
<td></td>
</tr>
<tr>
<td>At least every second month</td>
<td>9.8</td>
<td>11.6</td>
<td>10.9</td>
<td>16.7</td>
<td></td>
</tr>
</tbody>
</table>

\(a\) No. of answers  
\(b\) Median (2.5th – 95th percentiles)
Self-reported health problems among annoyed subjects

Subjects who were fairly much or much annoyed from road traffic noise (N = 623; table 4) experienced more health problems than others. Among these annoyed subjects, an association between average road noise exposure and concentration problems last weeks was observed (p for trend = 0.03). Concentration problems were also more common among annoyed subjects with moderate or high maximum road noise exposure. The unadjusted PR for concentration problems last weeks among the annoyed, contrasting average exposure ≥ 55 dB(A) with exposure below 50 dB(A), was 1.5 (95% CI 0.95 – 2.5) and the PR adjusted for gender and age was 1.7 (95% CI 1.0 – 2.7).

Among the subjects who were fairly much or much annoyed from road traffic noise, an association between average noise level from road traffic and hypertensive treatment was also observed (p for trend = 0.02; table 4). The unadjusted PR for hypertensive treatment among the annoyed, contrasting exposed ≥ 55 dB(A) with exposure below 55 dB(A), was 1.8 (95% CI 1.1 – 3.0) and the PR adjusted for gender, age, and BMI was 1.7 (95% CI 1.0 – 2.7). The association between average noise level from road traffic (< 50 dB(A), 50-54 dB(A), ≥ 55 dB(A)) and hypertensive treatment was present among annoyed males (3.8, 9.4, 13.8%; N = 255) but was not entirely consistent among annoyed females (11.0, 7.6, 15.5%; N = 359).

Self-reported health problems in possibly more vulnerable subgroups

When restricting the analyses to females, subjects born outside Sweden, unemployed, and to subjects that reported financial problems, no apparent associations with maximum road noise exposure were observed. However, the following positive associations between average road noise exposure (< 50 dB(A), 50-54 dB(A), ≥ 55 dB(A)) and the prevalence of health problems were found: hypertensive treatment among females (9.3, 9.8, and 11.1% treated, p for trend = 0.04), insufficient sleep among subjects born outside Sweden (10.4, 9.9, and 14.6%, p for trend = 0.04), frequent stress among unemployed (16.1, 20.6, 23.4%, p for trend = 0.04), and concentration problems last weeks among subjects that have had problems paying bills at least every second month (25.0, 27.9, and 32.1%, p for trend = 0.02). Multivariable analyses did not alter these trends (results not shown).
Table 2. Self-Reported annoyance (fairly much and much) from traffic noise at the residence in relation to road noise exposure among 13 557 participants of the public health survey in the Scania region in southern Sweden 1999-2000. Road noise exposure was categorized according to equivalent (average) noise level during full day \( \text{LA}_{eq, 24h} \): Low if < 50 \( \text{db(A)} \); \( N = 6 564 \), Medium if 50 – 54 \( \text{db(A)} \); \( N = 3 504 \), and High if \( \geq 55 \text{db(A)} \); \( N = 3 489 \) and according to maximum noise level \( \text{LA}_{\text{max}} \): Low if < 60 \( \text{db(A)} \); \( N = 3 528 \), Medium if 60 – 69 \( \text{db(A)} \); \( N = 5 520 \), and High if \( \geq 70 \text{db(A)} \); \( N = 4 509 \).

<table>
<thead>
<tr>
<th>Source of annoyance</th>
<th>Nb</th>
<th>Fairly much</th>
<th>Much</th>
<th>Annoyance (%) among all</th>
<th>Noise measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road traffic noise</td>
<td>13 168</td>
<td>3.7</td>
<td>1.0</td>
<td>Average</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Maximum</td>
<td>1.4</td>
</tr>
<tr>
<td>Aircraft noise</td>
<td>12 989</td>
<td>1.3</td>
<td>0.3</td>
<td>Average</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Maximum</td>
<td>1.5</td>
</tr>
<tr>
<td>Train noise</td>
<td>12 994</td>
<td>1.5</td>
<td>0.5</td>
<td>Average</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Maximum</td>
<td>1.3</td>
</tr>
</tbody>
</table>

The p-values for trend were below 0.001 for the prevalence of fairly much/much annoyance from road traffic in relation to both average and maximum road noise level. No positive associations were found between exposure to road traffic noise and annoyance from aircraft or train noise.

Total number of answers
Table 3. Self-Reported disturbance of daily activities (sometimes and frequent) from traffic noise at the residence in relation to road noise exposure among 13,557 participants of the public health survey in the Scania region in southern Sweden 1999-2000. Road noise exposure was categorized according to equivalent (average) noise level during full day ($L_{A eq,24}$): Low if < 50 db (A); N = 6,564, Medium if 50 – 54 db (A); N = 3,504, and High if ≥ 55 db (A); N = 3,489 and according to maximum noise level ($L_{A max}$); Low if < 60 db (A); N = 3,528, Medium if 60 – 69 db (A); N = 5,520, and High if ≥ 70 db (A); N = 4,509).

<table>
<thead>
<tr>
<th>Daily activity</th>
<th>N(^b)</th>
<th>Disturbance among all</th>
<th>Disturbance (%) in relation to exposure category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low exposure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sometimes</td>
<td>Frequent(^b)</td>
</tr>
<tr>
<td>Relaxation</td>
<td>12,947</td>
<td>18.4</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep</td>
<td>12,948</td>
<td>18.9</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Falling asleep</td>
<td>13,007</td>
<td>16.7</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hearing radio or TV</td>
<td>13,114</td>
<td>12.1</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conversation</td>
<td>12,974</td>
<td>5.2</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telephone conversation</td>
<td>13,016</td>
<td>4.4</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) The p-values for trend were all below 0.01 for the prevalences of disturbance of daily activities (sometimes or frequent) in relation to both average and maximum road noise exposure. \(^b\) Total number of answers. \(^c\) “Frequent” here refers to “every week”
<table>
<thead>
<tr>
<th>Health problem</th>
<th>Exposure category (average noise level)</th>
<th>Exposure category (maximum noise level)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>All participants (N)</td>
<td>6 564</td>
<td>3 504</td>
</tr>
<tr>
<td>Concentration problems last weeks (%)</td>
<td>13.3</td>
<td>13.3</td>
</tr>
<tr>
<td>Extensive sleeping disturbance last two weeks (%)</td>
<td>7.2</td>
<td>8.5</td>
</tr>
<tr>
<td>Insufficient sleep generally (%)</td>
<td>8.8</td>
<td>9.4</td>
</tr>
<tr>
<td>Frequently under stress (%)</td>
<td>18.9</td>
<td>18.8</td>
</tr>
<tr>
<td>Treatment for hypertension last 12 months (%)</td>
<td>10.0</td>
<td>10.1</td>
</tr>
<tr>
<td>Fairly much or much annoyance from road traffic noise (N)</td>
<td>129</td>
<td>171</td>
</tr>
<tr>
<td>Concentration problems last weeks (%)</td>
<td>16.3</td>
<td>19.9</td>
</tr>
<tr>
<td>Extensive sleeping disturbance last two weeks (%)</td>
<td>18.6</td>
<td>15.2</td>
</tr>
<tr>
<td>Insufficient sleep generally (%)</td>
<td>18.8</td>
<td>14.7</td>
</tr>
<tr>
<td>Frequently under stress (%)</td>
<td>27.3</td>
<td>27.1</td>
</tr>
<tr>
<td>Treatment for hypertension last 12 months (%)</td>
<td>7.9</td>
<td>8.3</td>
</tr>
</tbody>
</table>

* Monotonically increasing or decreasing trends in the prevalence of health problems in relation to exposure category were tested by Cochran-Armitage
Discussion

Our estimated prevalence (29%) of average exposure to road traffic noise of at least 55 dB(A) ($L_{eq, 24}$) is close to the overall estimates for the European Union, and much higher than previous national estimates for Sweden (16%). An advantage with our model as compared to national estimates is that we could include data on vehicles for road segments belonging to the local municipalities, which is important especially for those who live in an urban environment. Parts of our data on traffic intensity are not up-to-date, however, which may have produced some misclassification of exposure. We were able to differentiate in the model between urban (hard surface) and rural (soft surface) areas. This is of course an oversimplification since not all grounds in cities are hard, implying that we may to some extent have overestimated urban exposure. Lack of data on noise barriers and floor for residences in apartment blocks may also have produced overestimations of noise levels. Since a large proportion of the study subjects (1142 out of 13 557 subjects; 8.4%) were assessed by our model as being exposed to 55-56 dB(A), a systematic misclassification of, say, +2 dB(A) would inflate the estimated proportion exposed to >55 dB(A) substantially. However, we observed clear associations between modelled exposure and reported annoyance from road traffic noise, indicating that we managed to differentiate fairly well between different exposure levels. Furthermore, the high prevalence of disturbance of daily activities that we observed, and the high prevalence of frequent annoyance due to noise from road traffic in preliminary results from a more recent survey in 2004 (16), suggest that the levels of road noise exposure indeed is of concern in the Scania region in southern Sweden.

High exposure to road noise was more common among those not born in Sweden, living single and not employed. Lack of equity with regard to exposure to adverse environmental factors, including traffic noise, has previously been reported from the US (26), Germany (27), the Netherlands (28) and Birmingham in the UK (29). To the best of our knowledge, this is the first such report regarding road traffic noise from Scandinavia.

Fairly much, or much, annoyance due to road traffic noise was reported by 4.7% of all subjects, by 9.5 % of those exposed to >55 dB(A), and by 8.1% of those exposed to maximal levels of >70 dB(A). It is likely that the threshold for positive reporting was rather high in the general survey question about road traffic annoyance, since disturbed sleep or relaxation (sometimes or frequently) was reported by much higher proportions (19 – 32%) in the highest exposure categories. The exact phrasing of the question can be expected to be critical since higher prevalences of annoyance have been observed in studies where more specific questions about the frequency of the annoyance were asked. The preliminary estimate based on survey data from Scania in 2004 show that about 18% in the population are annoyed at least once a week from road traffic noise and
a recent Swedish report found that more than 15% of the subjects exposed to >55 dB(A) were annoyed frequently (every week) (3, 16). Traffic in general has been reported as the most important source of annoyance from noise in Sweden followed by noise from neighbours (15). A spill-over from being annoyed from noise from one means of transportation to another, may therefore be a concern. However, we found that subjects with high exposure to road noise did not report more annoyance from noise from trains or aircrafts than subjects with low exposure to road noise. The observed trends between annoyance, disturbance of daily activities and exposure level in our study were striking, and consistent for average noise levels as well as maximum noise levels, being generally at least doubled from the lowest to the highest exposure categories. Given that 29% of the population was estimated to be in the highest exposure category, this is a finding of concern. Disturbances of daily activities were somewhat more frequent for high average than for high maximum noise exposure.

We observed no associations between maximum level of road traffic noise and self-reported general health problems. For average noise level, we found a weak overall association with extensive sleeping disturbance during the last week, whereas an association with insufficient sleep generally was not apparent. Consistent data from other studies indicate that exposure to occupational noise and noise from aircrafts increases the risk for hypertension, but the findings are less consistent with regard to road traffic (2). We found an association between average road noise level and hypertensive treatment in females and also in males who reported fairly much or much annoyance from road traffic.

It is well known that sensitivity to noise varies widely between individuals. Within the subgroup that reported annoyance from road traffic noise, we found associations with concentration problems last weeks as well as with hypertensive treatment. Noise sensitive subjects have in several studies been shown to have poorer performance than non-sensitive subjects under noisy conditions (30), while no such difference was evident under silent conditions (31). The support for a higher risk for hypertension among subjects who are noise sensitive is weaker. However, an experimental study has shown a disruption on cortisol pattern from exposure to low-frequency noise that was evident only in subjects who rated themselves as noise-sensitive (32).

Positive findings in subgroups should be interpreted with care unless supported by other data. Some of the associations we observed for the self-reported general health problems in the 26 subgroup analyses (5 groupings of the participants × 5 different outcomes + annoyed males vs. hypertensive treatment) we conducted, generally with lower socio-economic status, may be due to chance. They may however also represent findings in especially vulnerable groups exposed to multiple social stressors. Socio-economic status may be inversely associated not only with exposure, but also with vulnerability. Among children
in rural areas in the US, adverse environmental exposure was not only associated with low income, but cumulative exposure was associated with effects, as monitored by increased levels of stress hormones, only among the low-income children and not in the middle-income sample (26). Among the unemployed in the present study, feeling frequently under stress was more common in those with high average exposure to road noise, as was concentration problems among subjects that reported financial problems. In addition, an association between insufficient sleep generally and average road noise level was suggested among subjects born outside Sweden.

The participation rate was low (54%) and differed with age, gender, education, and country of origin (18). Among these factors, country of origin varied with exposure among the participants. Thus, some association between exposure and participation was probably present, which may have produced bias if the participation was associated with any of the health parameters under investigation as well. The validity of the questionnaire should also be considered. A large study has shown high agreement between self-reported hypertension and medical records (kappa = 0.80, sensitivity = 82%, specificity = 92%; n = 2037) (33). Similar but smaller studies of self-reported hypertension have reported moderate to high agreement with medical records (34, 35). Misclassification of hypertensive treatment, if independent of noise exposure level, would tend to yield bias towards the null (36). The question about concentration ability during last weeks, in relation to their usual ability, was taken from a well-established instrument, the General Health Questionnaire - 12 (20). The questions about sleep and stress were specific for the present and similar surveys in Scania. The question about problems with paying bills, used to identify persons with financial problems in the subgroup analyses, has been linked to poor self-rated health (37).

In conclusion, exposure to road traffic noise at high levels was common in the study population and produced frequent annoyance and disturbances of daily activities. Associations between road traffic noise and negative health effects were observed among annoyed subjects and in other important subgroups.

Acknowledgements

We are grateful to Susanna Gustafsson, GIS Centre, Lund University for providing road traffic data and to Åke Boalt at the county council of the Scania region for assistance with the geocoding. The project was supported by grants from the Swedish environmental protection agency (Naturvårdsverket).
Appendix 1 – Measures of sound (noise) level

All measures of sound pressure level referred to in this article are A-weighted, which rates sound levels at different frequencies in a way that mimics the sensitivity of the human hearing organ (1). Sound level is expressed in dB(A). In order to average time-fluctuating sound, equivalent sound level is used, which is the corresponding steady noise level in a predefined time period that contains the same noise pressure as the fluctuating noise during the same time period.

**LA_{eq,24}**  
Equivalent sound level over 24 hr

**L_{dn}**  
Equivalent sound level over 24 hr where sound levels during the night (11 pm – 7 am) is increased by 10 dB(A), since noise during the night is usually perceived as more annoying.

**LA_{fmax}**  
Defined as the sound level exceeded by the loudest 5% of the vehicles passing a specific road segment. Fast time weighting is applied, which means that the sound level is obtained by integrating the instant sound level over a narrow time period (125 ms).
Appendix 2 - Questions about annoyance from traffic noise in the public health survey

Are you annoyed by noise from road traffic, trains or aircrafts?

1) Road traffic
☐ Not at all
☐ Not much
☐ Fairly much
☐ Much

2) Train noise
The same alternatives as for road traffic were used.

3) Aircraft noise
The same alternatives as for road traffic were used.
References


2. Christina Lindkvist-Scholten: Kvinnors försörjning på landsbygd: exempel från sydöstra mål 5(b) området (1999)


The use of GIS in Exposure-Response Studies
A regional study of Air pollution and Noise in southern Sweden