Representativeness of air pollution monitoring networks in an urban setting

2010

By

Daniela Fecht

Department of Epidemiology and Biostatistics

Imperial College London

Representativeness of air pollution monitoring networks in an urban setting

1. Introduction

Over the past decades, a large number of studies identified effects of air pollution on public health *(Brunekreef & Holgate, 2002; Pascal, 2009; World Health Organisation (WHO), 2004)*. Exposure to ambient air pollution has been linked to various health outcomes varying in severity from mild effects on the respiratory tract and pulmonary function to emergency room visits, hospital admissions and even mortality *(Pope & Dockery, 2006)*. Serious risks have been documented especially for the very young and older population as well as those with cardiorespiratory diseases such as asthma and heart disease. Epidemiological evidence also suggests a significant public health burden through the reduction of life expectancy of the average population by one or more years due to exposure to high levels of outdoor air pollution. A recent WHO review underlined that current levels of particulate matter concentrations measured across Europe pose a significant risk to human health *(World Health Organisation (WHO), 2003)*.

Monitoring air pollution levels, especially in urban areas, is therefore important in order to assess the impact on human health. Routine monitoring networks exist in most urban areas and are used to monitor the general pollution level within a city and, to determine if ambient air quality standards are exceeded. Routine monitoring networks consist of only a few stations, and in the case of PM_{10} of only one sampler in most British cities. These sample densities are not efficient enough to assess pollution levels in an effective manner with regard to human health impact. Air pollution levels can change dramatically within a short space. Some research suggests that the intra-urban variability in air pollution concentration may be larger than between cities *(Miller et al., 2007).* Very high levels are measured in short distances from emitting sources such as roads and industrial sites whilst, giving the right meteorological conditions, the pollution levels can drop very steeply with distance from a polluting source *(Seinfeld, 1986)*. Most routine monitoring networks are thus inadequate for representing the spatial variability in exposure that exists in urban areas *(Gilbert et al., 2005)*.

Many epidemiological studies, therefore, do not solely rely on the routine monitoring network. As mentioned above, routine monitoring networks are created in order to monitor air quality standards. The monitors are therefore mostly located in areas of high concentrations and hot spots such as heavily trafficked street locations and industrial areas. Epidemiologists, however, are mostly interested in the effect of air pollution on the health of the general population or individuals. Their interest, therefore, lies in the spatial variation of the concentration range as well as the pollutant concentration close to the population under study. Thus, one of the main shortcomings of epidemiological studies is often the correct exposure attribution. Relying on the routine monitoring network, exposures have been characterised by only one measured concentration across a city, assuming homogenous exposure within an urban area *(Dockery et al., 1993)*. But this crude proxy may result in a significant error in exposure, which may lead to substantial bias in exposure response relationships *(Monn, 2001).* Some epidemiological studies, therefore, use a targeted, study specific monitoring program. This might have the disadvantage of additional cost and limited temporal coverage but given the correct monitoring set-up should result in improved exposure estimates.

The type of pollutant monitored strongly influences the character and density of the monitoring networks. NO_2 concentrations are generally measured using NO_2 tubes. These are passive samplers, which can be set up very easily in the field and so are very cost effective. With little expenditure of time and experience, NO_2 tubes can be mounted at almost any location such as lampposts or rain pipes (see Figure 1a). PM_{10} monitors, on the other hand, are active samplers that use pumps and need electricity to run, which limits the locations where they can be set up. Typical locations might include shop fronts or residential gardens with easy access to electricity supply (see Figure 1b).

Figure 1: Example of monitor locations: a) NO_2 tube mounted at rain pipe, b) PM_{10} monitor (Harvard Impactor) in residential garden

PM_{10} monitors are more time intensive to set up and significantly more expensive than NO_2 tubes. These factors influence the density of monitoring stations to measure air pollution levels in a city. In the past, epidemiological studies have typically used between 40 and 100 NO_2 tubes across a city depending on its size and the number of inhabitants (see Table 1). The monitoring sites are typically categorised into different site types with approximately 60% of monitoring stations being traffic sites, 35% urban background sites and 5% would be used to measure regional background NO_2 concentrations. When collecting PM_{10}, budget and time constraints would usually only permit the collection at less than five sites for most epidemiological studies. Given the same monitoring density, the price for a PM_{10} active sampler network is on average three times higher than the cost for a NO_2 passive sampler network *(Hoek et al., 2008)*. Study specific PM_{10} monitoring networks are, therefore, not frequently used in epidemiological studies and PM_{10} measurements are more often taken from routine monitoring networks instead *(Briggs et al., 2010; Glorennec & Monroux, 2007; Pope et al., 2002)*.

Practical reasons are commonly the main motivation behind the monitoring networks used in epidemiological and indeed most environmental studies. Costs, ease of access, time effort and so forth play an overriding role in the set-up of these networks than the representativeness of the air pollution variation within an urban area. But what would be the 'right' monitoring strategy if none of these factors have to be considered? Which set-up strategy would provide the best results to most effectively attribute correct exposure estimates to the population? Is there a drop-off point in the number of monitoring stations that result in a dramatic decrease in accuracy? Researchers rely mostly on past experience and guidance from other researchers when setting up study specific monitoring networks. No rigorous methodology has been determined that would answer these questions and provide guiding principles for the set-up of an effective study specific monitoring network.

1.1 Aims and Objectives

The objective of this study is, therefore, to explore and assess the representativeness of air pollution monitoring networks in an urban setting. Particular focus is given to the misclassification of peoples' exposure to particulate matter (PM_{10}). Exposure misclassification is assumed to be due to the monitoring network. Other causes of exposure misclassification such as air pollution modelling error and uncertainty, temporal variation or people's different time-activity patterns, are not considered in this particular study but are discussed elsewhere in the literature *(Baxter et al., 2010; Zeger et al., 2000)*. Although environmental concentration levels do not necessarily equate to individual exposures they will be used in this context throughout this study, as is the case in most spatial epidemiological studies.

The specific aims of this study are a) to identify different environmental sampling strategies and evaluate their use for epidemiological analysis, and b) to experiment with different air pollution monitoring networks and evaluate them both in terms of predicting air pollution concentration in an urban environment and in terms of exposure misclassification. In particular, the study provides answers to the question of the best monitoring network strategy to predict both the exposure distribution within an urban area as well as individual exposure.

1.2 Monitoring network strategies

Various monitoring network strategies have been discussed in the literature. They range from haphazard or random sampling over monitoring in a grid pattern or purposive sampling in areas of interest *(Gilbert, 1987; Lioy, 1999)*. Haphazard sampling is a technique where any discretionary location can potentially become a sampling location. This encourages the selection of air pollution monitors at sites conveniently located in terms of both costs and effort. Haphazard sampling, however, is only appropriate if the monitored concentration surface is homogeneous in space and time because otherwise systematic bias can be introduced which hinders or even invalidates study results *(Piegorsch & Bailer, 2005)*. A homogeneous concentration surface is, however, not the case for most urban areas and therefore this technique is rarely applied in epidemiological studies.

A second sampling strategy referred to in the literature is probabilistic sampling (*Gilbert 1987)*. This is the general term for sampling strategies based on various degrees of randomness. Probabilistic sampling includes strategies as diverse as simple random sampling or systematic gridded sampling. The term simple random sampling describes the technique where each location within an urban area has an equal chance of being chosen as a monitoring site. This underlies the logic that if samples are taken randomly it balances out any uncontrolled systematic bias. Simple random sampling is only effective where there are no major trends or spatial variation pattern in air pollution concentrations. It is therefore only occasionally applied in urban areas because the random sampling error, which arises due to environmental variability not being picked up by the random selection process, is potentially very high. Although this approach has been used in an epidemiological context *(Rotko et al., 2000)*, generally it is confined to an environmental context when, for exampling, sampling soil or water *(Downes, 2010; Mattuck et al., 2005; Niemi & Niemi, 1990)*.

Gridded monitoring networks, also known as systematic monitoring networks, select the first monitoring station at random. Based on this first location, all further monitoring locations are systematically allocated based on a grid of a specified distance. Systematic sampling provides uniform coverage

Table 1: Characterisation of monitoring strategies of selected epidemiological studies

Reference	*Study area*	*No of monitoring sites*	*Monitoring strategy*	*Pollutant*
(Hoek *et al.*, 1997)	Umea, Sweden	2	Haphazard influenced by purposive decisions: one urban and one rural location	PM_{10}, Black smoke
	Malmoe, Sweden	2		
	Oslo, Norway	2		
	Kuopio, Finland	2		
	Amsterdam, Netherlands	2		
	Berlin, Germany	2		
	Hettstedt, Germany	2		
	Budapest, Hungary	2		
	Katowice, Poland	2		
	Cracow, Poland	2		
	Prague, Czech Republic	2		
	Teplice, Czech Republic	2		
	Pisa, Italy	2		
	Athens, Greece	2		
(Oanh *et al.*, 2006)	Bangkok, Thailand	4	Haphazard influenced by purposive decisions based on land use	$PM_{2.5}$, PM_{10}
	Bandung, Indonesia	5		
	Beijing, China	4		
	Chennai, Indonesia	3		
	Manila, Philippines	5		
	Hanoi, Vietnam	3		
(Rotko *et al.*, 2000)	Athens, Greece	50	Simple random	$PM_{2.5}$, CO, VOCs
	Basel, Switzerland	50		
	Grenoble, France	54		
	Helsinki, Finland	201		
	Milan, Italy	50		
	Prague, Czech Republic	50		
(Franco-Marina *et al.*, 2003)	Mexico City, Mexico	501	Simple random	Indoor radon
(Hirsch *et al.*, 2000)	Dresden, Germany	182	Systematic: 1 x 1 km grid	Benzene
(Martin *et al.*, 2006)	Valladolid, Spain	-	Systematic: 250 x 250 m grid	Noise
(Lebret *et al.*, 2000)	Amsterdam, Netherlands	80	Purposive based on distance to roads, includes background stations	NO_2
	Huddersfield, UK	80		
	Poznan, Poland	40		
	Prague, Czech Republic	80		
(Hoek *et al.*, 2002)	Netherlands	40	Purposive based on traffic density, includes background stations	$PM_{2.5}$
	Stockholm, Sweden	40		
	Munich, Germany	42		
(Kanaroglou *et al.*, 2005)	Toronto, Canada	100	Purposive based on population density using a location-allocation model	NO_2
(Jerrett *et al.*, 2007)	Toronto, Canada	100	Purposive based on population density using a location-allocation model	NO_2
(Madsen *et al.*, 2007)	Oslo, Norway	80	Purposive based on population density, includes background stations	NO_2
(Wheeler *et al.*, 2008)	Windsor, Ontario, Canada	54	Purposive based on population and traffic density, includes background stations	NO_2, SO_2, VOCs

of the study area and therefore is likely to result in more accurate exposure estimates. The frequency of the samples, i.e. the spacing of the grid, is a key factor. As with the simple random sampling method, this approach is rarely adopted in epidemiological studies *(Martin et al., 2006)*.

Purposive sampling, or judgemental sampling, is the subjective selection of monitoring sites by individuals based on knowledge and/or the objectives of the study. Purposive sampling is suitable when evaluating specific environmental concentrations such as high-end exposures or a specific risk associated with a target population. Monitoring sites that are representative for the investigated concentration range or the average population under study are selected. It is important, therefore, that the sample purpose and objectives are clearly defined before deciding on monitoring locations. This monitoring strategy is frequently used in environmental epidemiological studies amongst others to model air pollution using land use regression models *(Madsen et al., 2007; Wheeler et al., 2008)*.

Other sampling methods often applied in environmental sampling but rarely used in environmental epidemiological studies such as line-intercept sampling *(Khan, 2008; Mackey & Hodgkinson, 1995; Piegorsch & Bailer, 2005)* or geostatistical sampling methods *(Brus & Heuvelink, 2007)* are summarised elsewhere *(Wang et al., 2008)* and not considered here.

As these examples show, the main points to consider if setting-up a monitoring program are a) the objective of the sampling program, b) the cost effectiveness that ideally would allow achieving an acceptable level of representativeness at a specific cost and c) the spatial pattern of concentrations. The last point is an important concern when monitoring air pollution because of the complex nature particularly in urban areas where topography, surface roughness and meteorology combine to create a complex spatial and temporal pattern. It is, therefore, vital to evaluate the magnitude of sampling error for different designs and levels of effort.

2. *Methods*

2.1 The urban simulation

An urban simulation is used to investigate the hypothetical question of the representativeness of various monitoring networks. A simulation has an advantage over a real-world urban setting in that the exact air pollution concentration is known for every location in the city. Modelled PM_{10} concentrations based on a monitoring network can be compared to the known concentrations. Such a comparison is not possible in a real-world setting because the air pollution concentrations are obviously largely unknown apart from the concentrations measured at the monitoring sites, which invariably inform the PM_{10} model.

The urban simulation used in this study is SIENA, a GIS-based **SI**mulation to support **EN**vironmental health **A**nalysis. SIENA is a representation of an average medium sized city in Great Britain. Real-world sample cities around Great Britain were explored in terms of their spatial structure. Design rules were then extracted which informed the modelling process of SIENA within a GIS. SIENA consists of core data that includes information on land cover, topography, transportation networks and population densities, all at a 25 x 25 m grid level (see Figure 2). These data were modelled using a probabilistic modelling approach based on rules derived from the real-world sample cities. Additional, contextual data was added to SIENA, including information on daily traffic counts for the road network as well as on meteorological conditions such as average temperature, wind speed and wind direction. Hourly PM_{10} concentrations were then derived using both core and contextual data as input. The widely used dispersion model ADMS-Urban was employed to model the average hourly PM_{10} concentrations for a two-week period in August at the 25 x 25 m grid level.

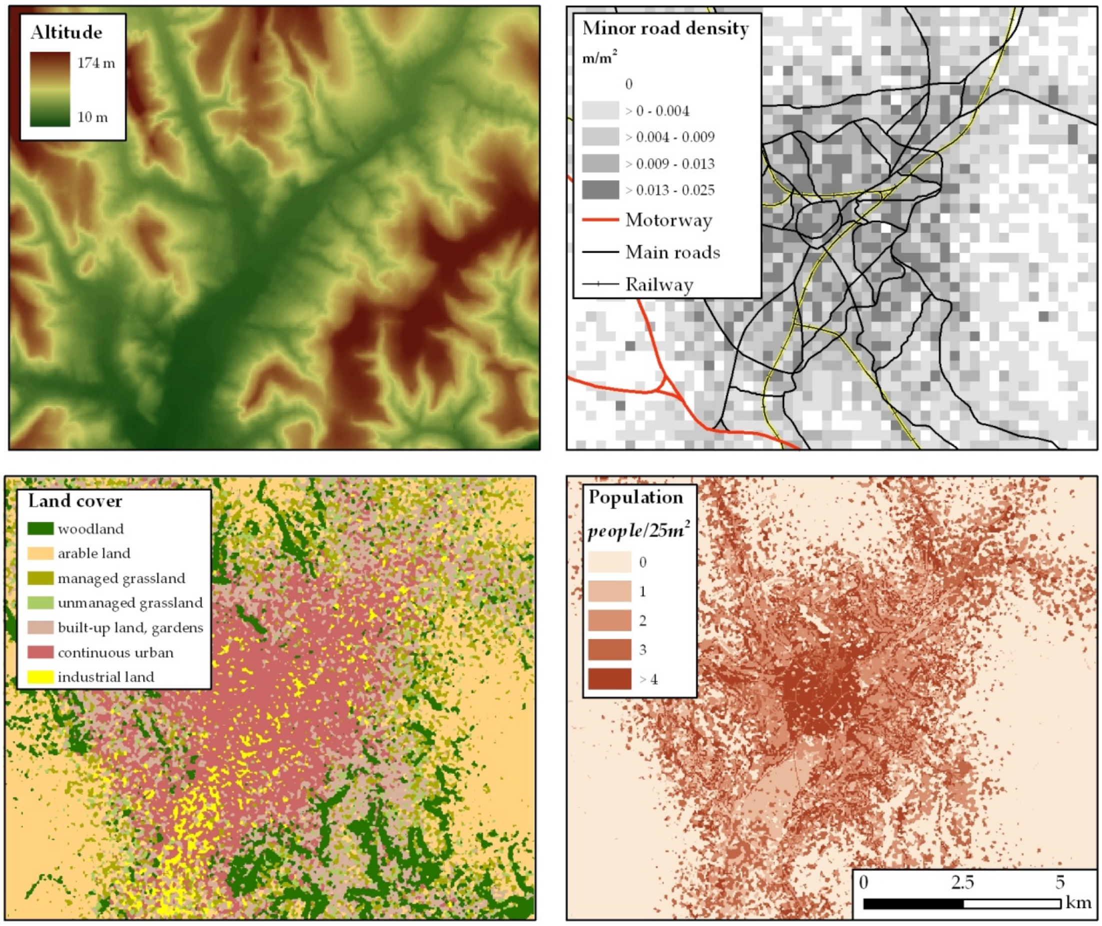


Figure 2: Data structure of the urban simulation: topography (upper left), transportation network and minor road densities (upper right), land cover (lower left) and population (lower right)

2.2 Observed exposure distribution

Based on the given data structure of SIENA, exposures for each individual in the city can be established. These exposures are the observed exposures to which all consequently modelled or estimated exposures can be compared. PM_{10} exposures are established for each individual at their home address by extracting the PM_{10} concentration for each address location from the PM_{10} concentration surface. Average PM_{10} concentrations measured over a two-week period at 10am on a weekday morning are used for this study.

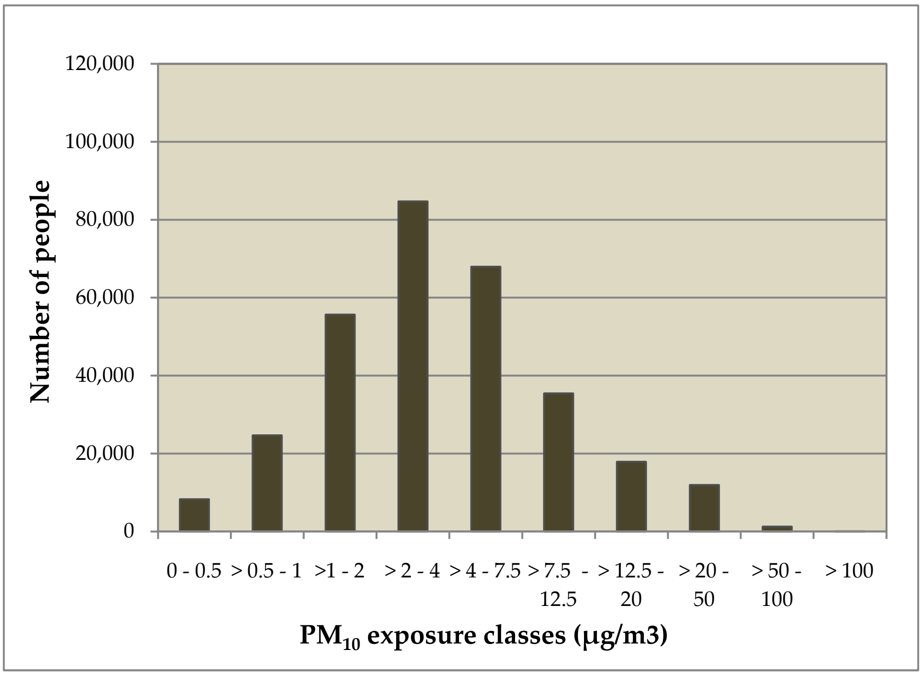


Figure 3: Observed exposure profile

Concentrations peak around this time of the day due to the concentration built-up from the morning rush hour and have, therefore, important health implications. Concentrations in the urban area range from 17 μg/m^3 in the urban hinterland to 157 μg/m^3 close to the motorway. Individuals are grouped into ten exposure groups as shown in Figure 3.

Descriptive statistics such as mean, standard deviation and skewness are calculated in order to describe the distribution. This observed exposure distribution provides the baseline distribution to which all other estimated exposure distributions will be compared.

2.3 The monitoring network set-ups

To explore the representativeness of monitoring networks the three main sampling strategies applied in epidemiological studies – purposive sampling, simple random sampling and systematic gridded sampling - are explored using various monitoring set-ups for each sampling strategy. Based on these monitoring set-ups, exposure profiles are calculated for the urban population of the study area based on PM_{10} concentrations, measured at the nearest monitoring stations. These exposure estimates are compared to the observed exposures in order to evaluate the performance of the different monitoring networks by assessing accuracy and sampling error.

Deriving accurate exposure estimates for an urban population is one of the most important and demanding issues in an environmental epidemiological study. Various methods have been reported in the literature of how best to attribute concentrations to individuals. They range from simple measures like proximity to roads or monitoring stations to more sophisticated interpolation techniques such as kriging and land use regression modelling to personal monitoring of individuals. All these methods have study-specific advantages and disadvantages and there is no golden standard, which can be applied to every study scenario. This study attributes individuals with the ambient PM_{10} concentration measured at their nearest monitoring station to avoid uncertainty introduced by the various interpolation and modelling techniques. Any errors in exposure assignment will be due to the chosen monitoring station network rather than the applied interpolation technique.

Four different purposive sampling approaches, using non-probabilistic sampling technique where monitoring locations are chosen based on prior knowledge, are applied here. Three approaches locate the monitoring sites in areas of high emissions from road traffic. Road emissions are estimated with increasing in accuracy as:

a) proximity to roads, measured as the distance of the 25x 25 m grid cell centroid to the nearest road,

b) high road density, measured as main and minor road density per 25x 25 m grid cell (m/m^2),

c) high traffic density, expressed as vehicle kilometres travelled per 25 x 25 m grid cell on minor and main roads.

The last approach places the monitoring stations in areas of high population density, measured as the number of people per 25m grid cell.

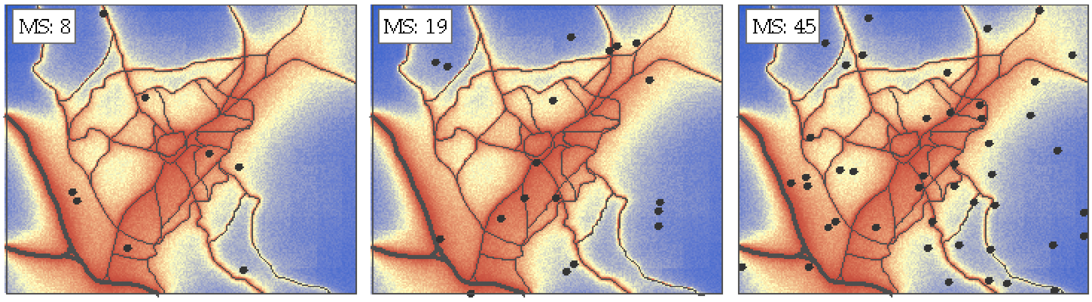
All four approaches locate the monitoring stations in randomly selected 25 x 25 m grid cells of SIENA that fall in the highest percentile, i.e. nearest to road or highest density. Using each method, a number of monitoring networks are designed and implemented with monitor numbers ranging from 5, a situation not unfamiliar in most studies, to 50 PM_{10} monitoring stations, a situation desirable but not applied in many studies (see Table 1).

In addition to locating monitoring stations based purely on the purposive sampling strategy, many epidemiological studies also use background stations in order to estimate pollution concentrations away from the main emitting sources *(Hoek et al., 2002; Madsen et al., 2007; Wheeler et al., 2008)*. To reflect this procedure, background stations are added to the 25 best performing purposive monitoring networks. The background stations are randomly located in areas of PM_{10} concentrations below the SIENA mean and added to the already established monitoring network. Between one and five background monitoring stations are added to each specified monitoring network. This results in a total of 309 different purposive monitoring networks, of which 125 networks have urban background stations.

For the systematic gridded sampling strategy, monitoring grids were developed that range from 1 x 1 km to 5 x 5 km cells increasing in 125 m increments. In total, 49 gridded monitoring networks are established. The 25 x 25 m grid cells of the urban simulation are used as building blocks where the grid cells are aggregated to form the monitoring grid. In all cases, the centroids of the monitoring grid are used as location for the PM_{10} monitors. The distances are chosen to obtain a realistic number of monitoring stations varying from 6 to 154 stations. A denser grid than 1 x 1 km results in a network of too many PM_{10} monitoring sites to be practically and financially feasible in most epidemiological studies, while a density lower than 5 x 5 km would result in fewer than five stations, the threshold for this study.

The simple random monitoring strategy selects a representative sample by using chance selection so that biases will not systematically alter the sample. Using this approach, stations are randomly selected from the 25 x 25 m grid cells without any outside input, varying in numbers from 5 to 50 monitoring stations. Again, PM_{10} monitoring sites are located at the 25 x 25 m grid cell centroid. In order to explore how representative the obtained exposure profiles are, the selection process is repeated five times, always extracting random stations. This results in 230 monitoring networks based on the simple random strategy. For each established monitoring network (Figure 4) exposure profiles are then estimated for the population in SIENA by assigning each individual the PM_{10} concentration measured at the nearest monitoring station to their residential address.

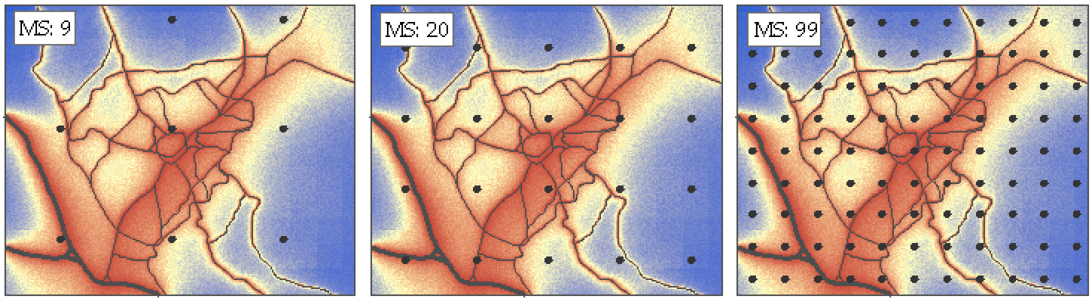
Random sampling strategy



Strategy specifications:

Randomly selected monitoring locations; number of monitoring stations: N = 5, ..., 50;
Total number of monitoring networks analysed: 230

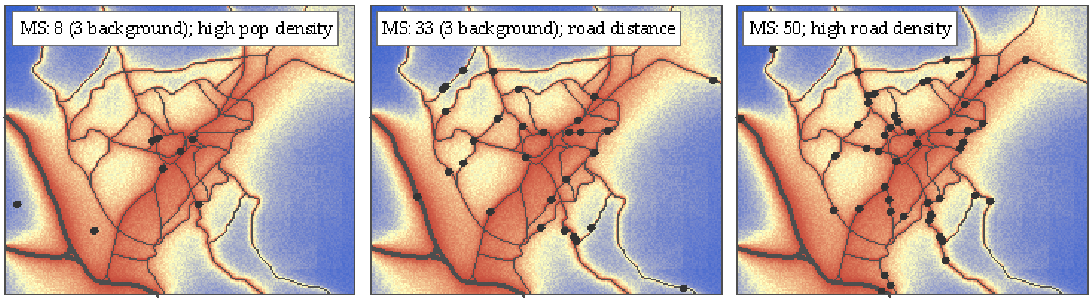
Gridded sampling strategy



Strategy specifications:

Gridded monitoring locations; lowest grid density 5 x 5 km (6 stations), highest grid density 1 x 1 km (154 stations); increments 125 m
Total number of monitoring networks analysed: 49

Purposive sampling strategy



Strategy specifications:

Four approaches based on closeness to road, high road density, high traffic density and high poplutaion density;
Number of stations for each approach: N = 5, ..., 50; Background stations (N = 1, .., 5) added for selected networks
Total number of monitoring networks analysed: 309

Figure 4: The monitoring network set-ups

2.4 Statistical analysis

The representativeness of the various monitoring networks is assessed from two different angles. The statistical analysis assesses a) which monitoring networks result in the best exposure distribution and b) which monitoring networks allow the best estimate of individual exposure. These two aspects do not necessarily have to be the same. The distribution of the estimated PM_{10} exposures can be very similar to the distribution of the observed PM_{10} exposures but highly exposed individuals, for example, are attributed with low exposure estimates and vice versa. This difference is especially important in an epidemiological context. Correct exposure distributions are important for estimating exposures for ecological or population based analysis, when for example comparing population's exposure in different urban areas (Dockery *et al.*, 1993; Kousa *et al.*, 2002; Kunzli *et al.*, 2000). Individual or small scale analysis, on the other hand, need exposure estimates to be as accurate as possible at the individual level, which is much more difficult to achieve (Iniguez *et al.*, 2009; Oglesby *et al.*, 2000). It is therefore important to investigate both aspects of potential exposure misclassification.

Several statistical measures are required to evaluate the performance of the monitoring networks in terms of both exposure distribution and individual exposure. A hierarchical model is developed for the statistical analysis to evaluate different aspects of network performance such as differences in shape and location of derived exposure distributions, correlation of individual exposure or spatial variation in performance. All of these aspects are important to assess the representativeness in terms of both exposure distribution and individual exposure, but the hierarchical model provides an effective method to evaluate the different monitoring approaches by excluding incrementally poorly performing monitoring networks and further in-depth analysis of the better performing networks (see Table 2).

Table 2. Statistical measures used in hierarchical analysis of exposure distribution and individual exposure

	Exposure distribution assessment	Individual exposure assessment
1. layer	• Q-Q plot	• Pearson's r correlation • Spearman's rho correlation
2. layer	• Kolmogorov-Smirnov test • Independent samples t-test • Descriptive statistics: minimum, maximum, median, mean, 5th/95th percentile ratio • Cumulative frequency plot	• R^2 • RMSE, NMSE • Mean bias, Normalized mean bias, Mean fractional bias, Fractional bias • Fa2
3. layer	• Descriptive statistics for each exposure tertile: minimum, maximum, median, mean, 5th/95th percentile ratio	• Error map • Moran's I: global and local • Kappa statistic

In the first hierarchical layer, monitoring networks are identified that show poor representativeness of PM_{10} exposure. Basic measures of distribution comparison and correlation are used to eliminate poorly performing monitoring networks from the further statistical analysis. In order to test if the observed and estimated exposures have similar distributions, Q-Q plots are used. Individual level assessment is carried out using Pearson's r and Spearman's rho correlation. These relatively basic measures of performance are used to give a first indication of the performance of the monitoring networks (*Chambers et al., 1983*).

Q-Q plots are produced to explore the shape of the exposure distributions derived from the various monitoring networks and compare those to the shape of the observed exposure distribution. The Q-Q plot tests if two distributions are similar in which case the Q-Q plot will approximately be a straight line (x = y). Here, the Q-Q plots map the various exposure distributions against a normal distribution. The observed PM_{10} exposure distribution is extremely skewed but approximately follows a log-transformed Gaussian distribution (Figure 5) - a pattern also described for urban PM_{10} exposure distributions around Europe *(Giavis et al., 2009)*. All exposure distributions are, therefore, log-transformed as well as centred and Q-Q plots generated for the observed exposure distribution as well as the 488 estimated exposure distributions based on the various monitoring networks.

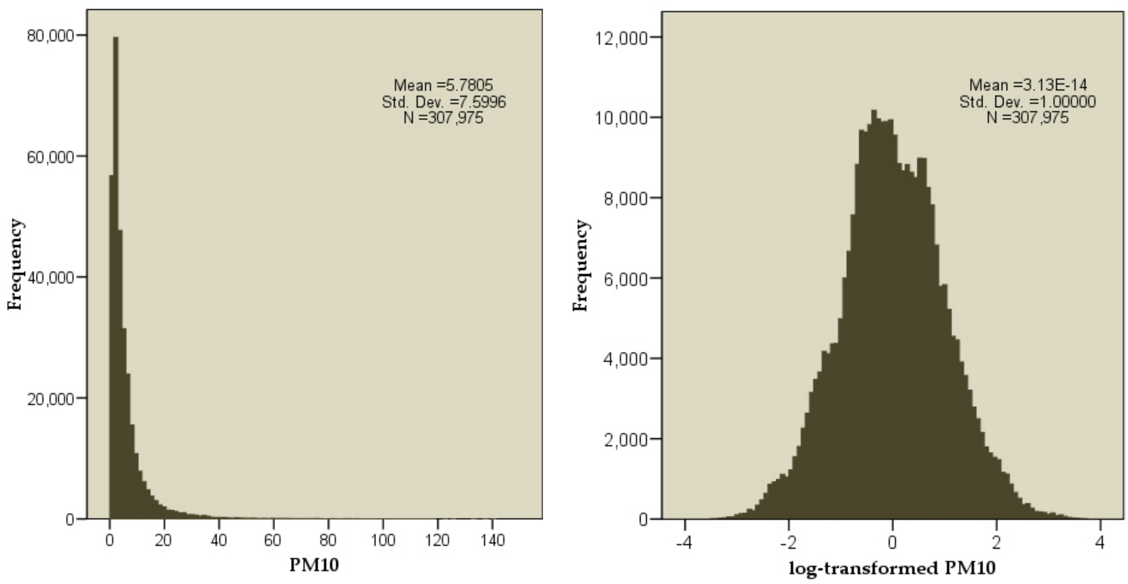


Figure 5: Observed PM_{10} exposure distribution and log-transformed and centred distribution

Both Pearson product-moment correlation coefficient (Pearson's r) and Spearman's rank correlation coefficient (Spearman's rho) are computed to assess the correlation between the observed and estimated PM_{10} exposures. Pearson's r is sensitive to a linear relationship between the observed and estimated variables and best works if the data is normal distributed. Spearman's rho on the other hand, is a non-parametric measure of correlation and does not make assumptions about the particular relationship between the variables.

Only those monitoring networks whose derived exposure distribution is similar to the shape of the observed exposure distribution (Q-Q plot approximately follow a straight line) and in addition have a medium to large correlation coefficient are considered for further statistical analysis. Correlations coefficients are considered medium to large if Pearson's $r > 0.3$ and Spearman's rho $r > 0.5$ *(Cohen, 1988)*.

In the second hierarchical layer, any remaining monitoring networks are assessed using tests chosen to evaluate statistically and visually the exposure distributions including both shape and location. Individual exposure is examined in more detail by quantifying the differences between observed and estimated individual exposure derived from the different monitoring networks and determining the direction of this difference (*Hanna, 1993*).

The performance of the monitoring networks in terms of predicting exposure distribution is assessed using the Kolmogorov-Smirnov test and Independent samples t-test. The Kolmogorov-Smirnov test establishes if two independent samples come from the same population. The two samples are represented by the observed exposure values and the estimated exposure values. PM_{10} concentrations are again log-transformed and centred because the test is sensitive to diversion from normal distribution. The Independent samples t-test tests the null hypotheses that the mean of the two normally distributed populations are equal. Because normal distribution is assumed, again, the log-transformed data is used. But the data is not centred because the mean for each sample would be zero. In addition, the distributions as statistically described in terms of minimum and maximum PM_{10} concentrations, median and mean. The $5^{th}/95^{th}$ percentile ratio is compared for exposure distributions derived from the monitoring networks and the original PM_{10} concentration surface. Cumulative frequency plots are produced to explore the distributions visually.

As a measure of individual exposure, correlations between the observed PM_{10} exposures and the estimated PM_{10} exposures are evaluated using the coefficient of multiple determination (R^2). The root mean square error (RMSE) and the normalised mean square error (NMSE) allow quantification of the difference between the observed and the estimated individual exposure values. Furthermore, the mean bias, normalised mean bias and the mean fractional bias are calculated. The fractional bias is a measure of performance often recommended in the model evaluation literature to determine the direction of the error. Values for factual bias are between -2.0 (extreme under-prediction) and 2.0 (extreme over-prediction). Values of the fractional bias that are equal to -0.67 are equivalent to under-prediction by a factor of two, while values that are equal to 0.67 are equivalent to over-prediction by a factor of two. Root mean square error scores are robust to variations in shape of the data distributions. R^2 is sensitive to the distribution of data. No one indicator can evaluate all aspects of the model and there is no consensus in the literature regarding which of the above is the best measure to evaluate model performance. All performance measures discussed are therefore used in conjunction because this increases the ability to evaluate all aspects of performance of the monitoring networks. The metrics used for this part of the assessment are summarised in Table 3.

Table 3: Calculated measures of model performance

Root mean square error (μg/m³) (RMSE)	$\boldsymbol{RMSE = \sqrt{\frac{\sum_{i=1}^{N}(E_o - E_e)^2}{N}}}$
Normalised mean square error (μg/m³) (NMSE)	$NMSE = \frac{1}{N}\sum_{i=1}^{N}\frac{(E_o E_e)^2}{E_o E_e}$
Mean bias (μg/m³) (MB)	$MB = \frac{1}{N}\sum_{i=1}^{N}(E_e\, E_o)$
Normalised mean bias (%) (NMB)	$NMB = \frac{\sum_{i=1}^{N}(E_e - E_o)}{\sum_{i=1}^{N} E_o}$
Mean fractional bias (%) (MFB)	$MFB = \frac{1}{N}\sum_{i=1}^{N}\frac{(E_e - E_o)}{(\frac{E_o + E_e}{2})}$
Fractional bias (FB)	$FB = \frac{1}{N}\frac{(E_o - E_e)}{0.5(E_o + E_e)}$
Fraction within a factor of two (fa2)	$0.5 \leq E_e/E_o \leq 2.0$

The third analysis level focuses on the best performing monitoring networks identified in the previous statistical analysis. This hierarchical layer focuses on the magnitude of discrepancy between observed and derived exposure distribution in different segments of the PM_{10} concentration range and on the spatial variation of individual exposure error.

In order to analyse different segments of the concentration range, descriptive statistics are calculated for each exposure tertile separately. Again, minimum and maximum PM_{10} exposures as well as the mean and median, and $5^{th}/95^{th}$ percentile ratio are computed. This provides an indication if different network set-up strategies perform better in different areas of low or high concentrations.

To look at the performance of the monitoring networks spatially, error maps are produced that show the absolute error, i.e. difference, in PM_{10}

concentrations between observed and estimated concentration at each residential address. The Moran's I index establishes if the errors are clustered, dispersed or randomly distributed over the city *(Moran, 1950)*. Both the global Moran's I as well as the Anselin Local Moran's I are calculated using the Toolbox in ArcGIS. The global Moran's I is a measure of spatial autocorrelation considering both error location and error value simultaneously. For each error map a Moran's I value is derived whereas a value near 1.0 indicates clustering and an index near -1.0 dispersion. A zero value indicates random spatial error patterns. Anselin Local Moran's I is a measure to identify clusters of estimation errors similar in magnitude. Even if no global clustering is detected, clusters at local level can still be found using local spatial autocorrelation. A further method to compare the estimated to the observed PM_{10} exposures spatially at the small scale is the Kappa analysis. The Kappa statistic is a measure of agreement between observed and estimated categorizations of two maps while correcting for the chance agreement between the two categories *(Congalton, 1991)*. For this purpose, PM_{10} concentrations are grouped into ten categories based on the exposure classes given in Figure 37. Observed concentration classes are mapped against the estimated concentrations classes for each 25m grid cell and the total accuracy is calculated by dividing the number of correctly classified cells. The accuracy assessment is performed within the ArcView extension Kappa Analysis 2.0 *(Jenness & Wynne, 2005)*.

The hierarchical approach to assess monitoring network performance and representativeness described here ensures a detailed analysis of network performance achieved by the different set-ups. Excluding incrementally poorly performing networks allows a more in-depth analysis of the remaining networks and permits conclusions about the monitoring networks that provide sound exposure estimates.

3. Results

The first statistical analysis conducted is the comparison of the Q-Q plots. The Q-Q plot of the observed PM_{10} exposure distribution almost follows a straight line (Figure 6a) and is therefore very similar to a normal distribution. Similar Q-Q plots can be achieved with different monitoring networks. Figure 6 (d-f) show as example three estimated exposure distributions which also follow a normal distribution. Other monitoring networks, however, result in exposure distributions that depart from the normal distribution (see Figure 6 b,c). These monitoring networks are the first to be eliminated from the further statistical analysis because of their poor representativeness of the observed exposure distribution. When looking at the number of eliminated monitoring networks clear differences can be detected between the three different monitoring strategies. Purposive monitoring networks show very different results based on the approach taken. Exposure distributions based on monitoring networks that focus on area of high road density and on areas that are close to main roads perform better than networks concentrated in areas of high traffic or population density.

22% of exposure distributions based on monitoring networks focusing on areas with high road density, for example, follow a normal distribution whereas only 4% result from the monitoring networks based on high traffic density show this pattern. The gridded sampling strategy provides the highest number of monitoring networks that result in normally distributed exposure estimates (23%), while with the systematic random sampling strategy 12% of exposure estimates follow a straight line in the Q-Q plot. The number of monitoring stations within the network does not seem to influence the performance, a trend present for all sampling strategies.

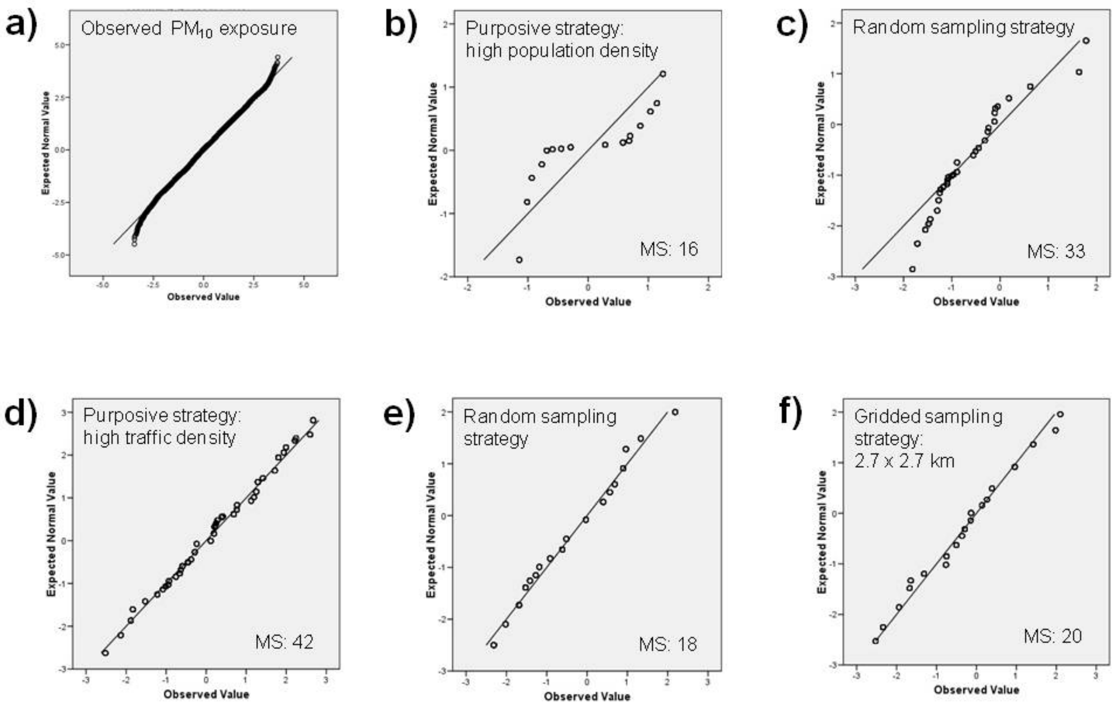


Figure 6: Q-Q plots of selected PM_{10} exposure distributions mapped against a normal distribution: a) observed PM_{10} exposure distribution; b-c) estimated exposure distribution based on different monitoring strategies not following a normal distribution; d-f) normally distributed exposure estimates (MS = number of monitoring stations)

Correlations obtained from Pearson's r and Spearman's rho support the general pattern observed in the Q-Q plots. Looking at the correlations between observed PM_{10} exposures and estimated exposures based on monitoring networks following the purposive sampling strategy, Pearson's and Spearman's correlation coefficient follow each other with Spearman's correlation being on average 10% better than Pearson's r. In general, there is a gradual tendency towards more accurate exposure estimates with increased station density but no clear threshold in the number of stations can be established below which the correlation coefficient drops steeply. Exposure estimates based on networks focusing on areas with high road density or closeness to roads tend to perform better than networks with stations concentrated in highly populated areas. The high traffic density strategy provides the poorest results. Adding background stations to the monitoring network does not imperatively improve the exposure estimates. In some cases, the distribution does depart from normal distribution after the introduction of one or more background stations but the correlation improves. Monitoring networks with one or more background stations seem to result in better exposure estimates than networks with four or five background stations.

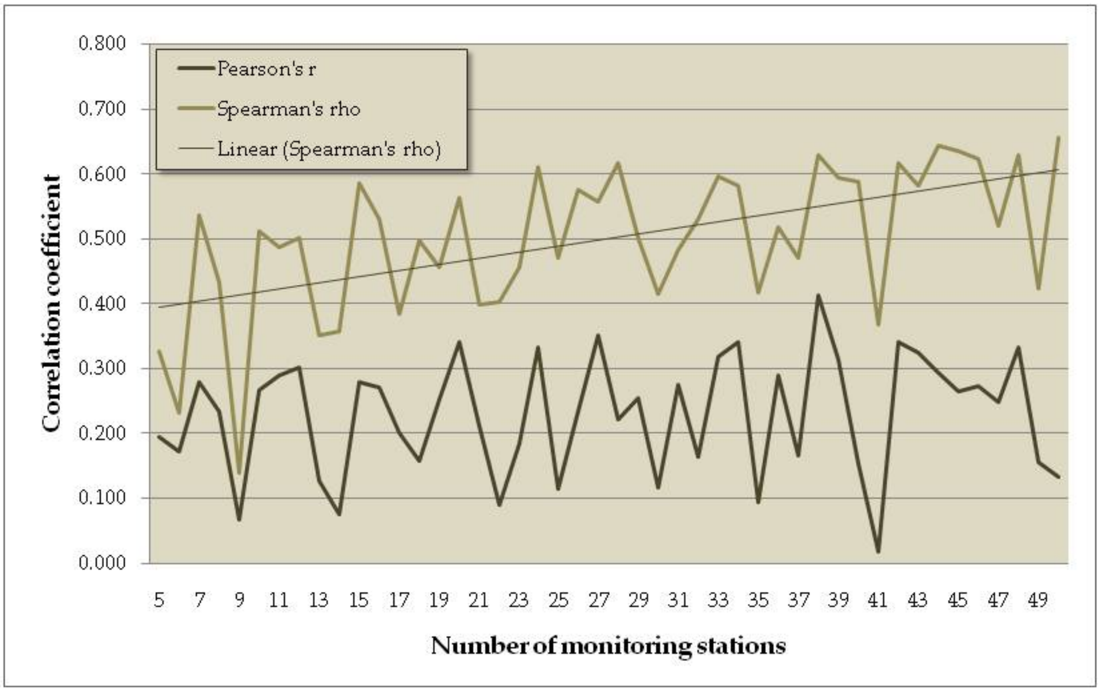


Figure 7: Correlation coefficients comparing observed with estimated exposure estimates based on random monitoring networks

A similar pattern arises for the probabilistic monitoring strategies. For exposure estimates, obtained using the systematic gridded as well as the simple random monitoring strategy, the Pearson's and Spearman's correlation coefficient follow each other with Spearman's rank correlation performing on average 30% better (Figure 7). In general, the Pearson correlation coefficient provides medium correlations ($r > 0.3$) and Spearman's rho large correlations ($r > 0.5$). Some monitoring networks result in a very poor representativeness of exposures, which could be caused by missing monitoring sites close to the city centre, i.e., highly polluted and populated areas are not represented.

Based on this first level of statistical analysis 431 monitoring networks are dropped because of poor representativeness of exposure values. 92% of all purposive networks, 80% of gridded and 90% of all random monitoring networks are excluded from the further statistical analysis because the exposure estimates do not correlate with the observed PM_{10} exposures and the distributions achieved by these monitoring networks do not follow a normal distribution. This is of significance, not only in terms of representativeness of these monitoring networks, but normally distributed data is a requirement for many of the following statistical test. Excluding almost 90% of considered monitoring networks is a large number but this will allow for a more detailed statistical exploration of the better performing monitoring networks. A visual exploration of some of the excluded monitoring networks in relation to the PM_{10} concentration surface as well as land cover and the road network in the urban simulation does not give any conclusive indications of why the monitoring networks would result in low representativeness.

The exposure estimates of the remaining 57 monitoring stations are investigated more closely in the second layer of the statistical analysis. The first statistical test performed is the Kolmogorov-Smirnov test, which checks the null hypothesis that two samples are drawn from the same distribution. This hypothesis has to be rejected for all estimated exposure distributions because of very low p-values (0.000). As the Kolmogorov-Smirnov test does not provide conclusive results, it cannot be reasoned that the estimated exposures and the observed exposures are drawn from the same distribution.

A reason for this could be that Kolmogorov-Smirnov test is a very conservative test, able to detect very small differences in the samples. The high number of exposure estimates also gives the test high statistical power to detect these small differences. The Independent samples t-test produces a similar result. Only one monitoring network, a random monitoring set-up with 32 stations, results in an exposure distribution, which, if compared with the observed exposures produces a p-value > 0.00. A better way, therefore, is the visual comparison of the distribution shapes. The cumulative frequency plots give an indication of the shape of the estimated exposure distributions compared to the observed PM_{10} distribution.

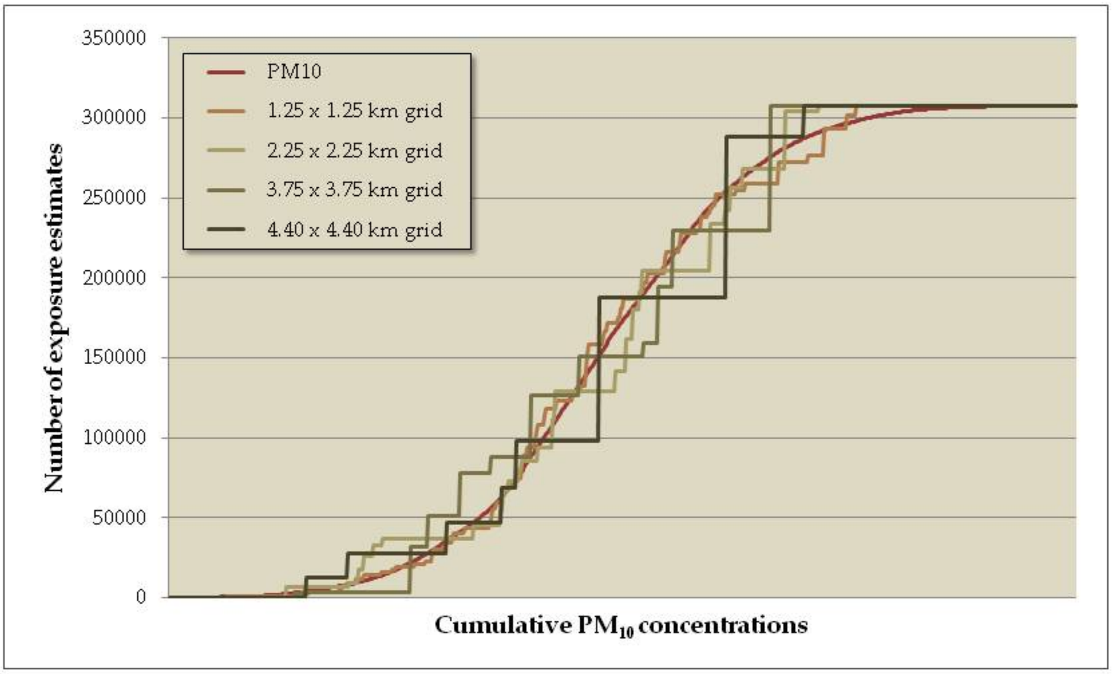


Figure 8: Cumulative frequency plots of selected exposure distributions based on the gridded monitoring strategy

Figure 8 shows as an example selected estimated exposure distributions based on the gridded monitoring strategy in comparison to the observed cumulative PM_{10} exposure distribution.

Gridded monitoring networks below a 2 x 2 km station density result in exposure estimates that follow the observed cumulative PM_{10} distribution very closely. Above that threshold, the increments increase and the curve departs more and more from the observed PM_{10} curve. This trend, higher stations density result in better fitting cumulative distribution, can also be seen for the other monitoring strategies. But no clear threshold in the number of stations can be detected which would result in a significantly decreased performance.

The descriptive statistics shows clear differences between the network strategies. Both minimum PM_{10} concentrations as well as the city average concentrations are very close to the observed vales for exposure estimates derived from gridded and random monitoring stations. But maximum values do not reach the levels observed in some high concentration locations close to roads. Exposure estimates based on purposive strategies on the other hand, highly over-predict the minimum concentrations for most networks as well as the city average. Maximum concentrations are higher but still do not reach the level of the observed PM_{10} concentrations. Again, no clear trend can be detected in terms of number of stations and network performance for any of the sampling strategies.

Results of the performance valuation with regard to introduced bias and errors are summarised in Table 4. Statistics are displayed for the monitoring networks that result in the best and the poorest performance for each of the three monitoring strategies. In general, the gridded and random strategies perform better than the purposive sampling strategy.

Table 4: Model evaluation results and descriptive statistics for best and worst performing monitoring set-ups for each strategy

Monitoring set-up strategy	No of stations	R^2	RMSE µg/m³	NMSE µg/m³	MB µg/m³	NMB (%)	MFB (%)	FB	Fa2 (%)	Min µg/m³	Max µg/m³	Mean µg/m³	5th/95th ratio
1.25 x 1.25 km grid	99	0.15	7.42	0.80	-0.620	-10.7	-5.8	-0.11	74.85	0.19	20.98	5.2	0.04
2.25 x 2.25 km grid	30	0.08	7.57	1.31	-1.329	-23.0	-15.3	-0.26	62.49	0.23	15.29	4.5	0.05
3.75 x 3.75 km grid	12	0.08	7.63	1.39	-0.704	-12.2	-3.8	-0.13	55.71	0.32	11.81	5.1	0.07
4.4 x 4.4 km grid *	9	0.08	7.92	1.75	-3.042	-52.6	-46.2	-0.71	49.10	0.26	7.58	2.7	0.05
random	8	0.12	7.53	0.94	-2.191	-37.9	-4.8	-0.47	64.89	1.18	6.39	3.6	0.18
random	32	0.10	7.81	1.06	-0.165	-2.9	1.0	-0.03	63.28	0.51	18.73	5.6	0.03
random	45	0.15	7.56	0.99	-2.633	-45.6	-26.1	-0.59	68.08	0.34	6.58	3.1	0.09
random *	31	0.08	7.84	1.71	-2.573	-44.5	-37.8	-0.57	51.76	0.29	19.43	3.2	0.09
purposive – population density	43 (incl. 5 background)	0.07	10.02	1.82	1.427	24.7	2.9	0.22	52.83	0.30	35.84	7.2	0.01
purposive – road density	50 (incl. 1 background)	0.11	22.28	7.09	18.484	319.8	130.8	1.23	10.04	1.73	55.42	24.3	0.17
purposive – road distance	22 (incl. 4 background)	0.11	27.72	5.97	17.813	308.2	100.4	1.21	22.51	0.48	106.34	23.6	0.04
Purposive – road distance *	16	0.09	28.27	9.47	23.822	412.1	139.6	1.35	6.65	9.25	81.87	29.6	0.17

* Worst performing monitoring set-up for each strategy

All exposure estimates based on random or gridded monitoring networks result in means square errors and mean biases of approximately the same magnitude (NMSE min: 0.84 μg/m³, max: 2.44 μg/m³, Fa2: 49.1% - 75.4%). Monitoring networks based on purposive strategy, however, seem to result in an extreme over-prediction of PM_{10} exposures, especially in terms of mean bias (NMSE min: 1.82 μg/m³, max: 9.68 μg/m³, Fa2: 6.65% - 52.83%), as has been seen in the descriptive statistics. Only two exposure estimates based on a monitoring network with 43 monitoring stations which concentrate on areas of high population density give similar low errors to the gridded and random networks. The maximum mean fractional bias, for example, is up to 3-fold higher in monitoring stations based on the purposive approach (MFB max: 139.6% over-prediction based on purposive network with 16 stations close to roads) than the exposure estimates derived from gridded and random networks (MFB max: 46.2% under-prediction based on gridded network with 9 stations). In general, gridded and random networks seem to slightly under-predict PM_{10} exposures while purposive networks even with five background stations greatly over-predict exposures. The twelve monitoring networks listed in Table 19 will be further investigated in terms of spatial error distribution in the third layer of statistical analysis. The three best performing monitoring networks for each of the three sampling strategies are explored, as is the worst-case monitoring network in order to find the reasons for the poor performance.

The pattern observed for the descriptive statistics also holds when looking at each concentration tertile separately. The first and second PM_{10} tertiles are well represented by all considered monitoring networks. Minimum and maximum PM_{10} concentrations as well as the mean and median of the estimates are very close to the observed exposures, except for purposive monitoring networks focusing on closeness to roads and high road density. These networks result in an up to 8-fold over-prediction of exposure estimates. Only maximum PM_{10} concentrations in the third tertile are better represented by the road based purposive networks then with any other monitoring network. Overall, gridded monitoring networks with a high station density (number of monitoring stations > 20) result in exposure estimates whose descriptive statistics are very similar to the observed values for all three tertiles.

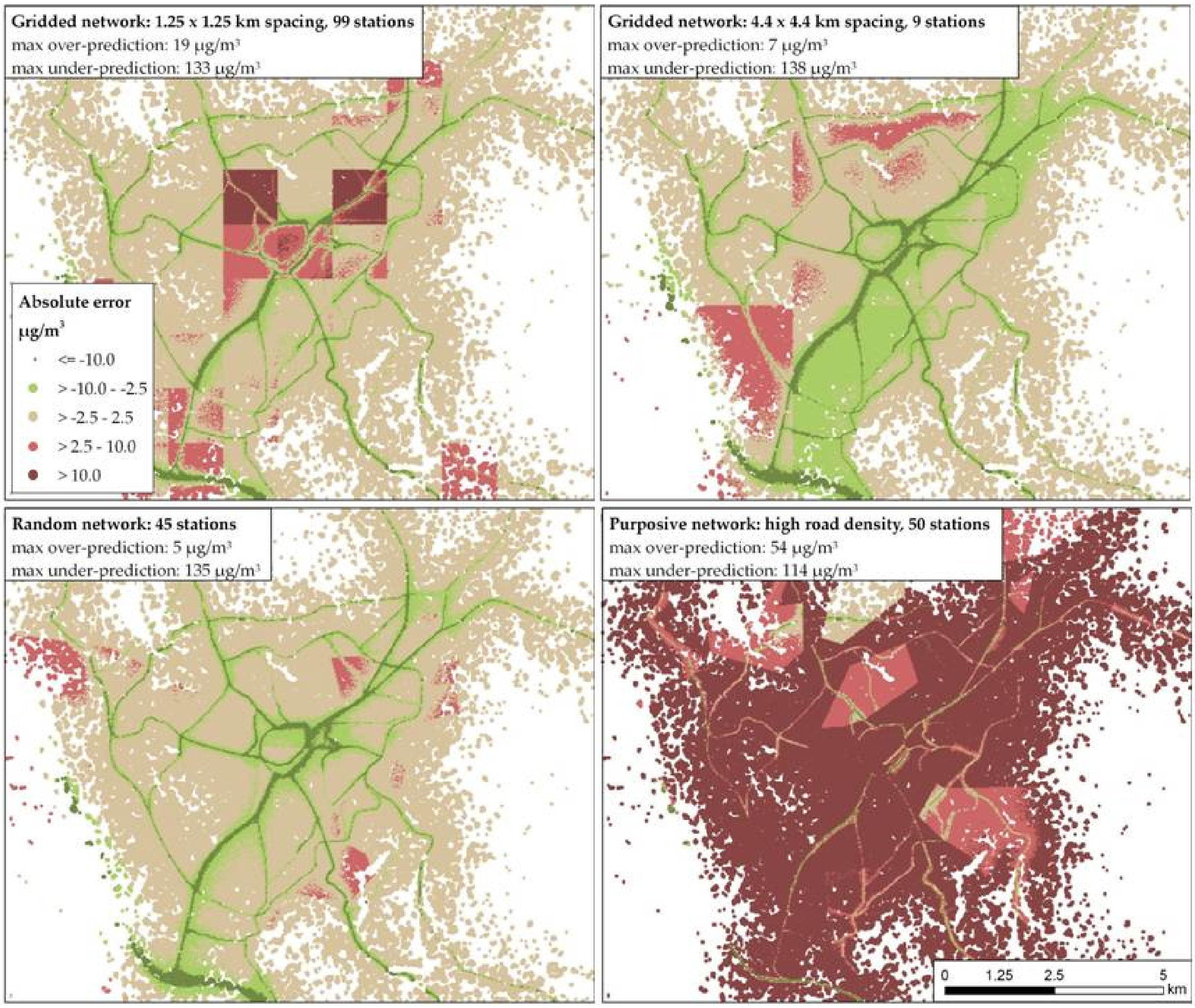


Figure 9: Spatial distribution of absolute error in PM_{10} exposures for selected monitoring networks

When looking at the performance of the different monitoring networks visually, a similar picture emerges. Figure 9 shows a selection of error maps resulting from different monitoring networks. Areas shaded in green under-predict

PM_{10} exposures and areas shaded in red over-predict PM_{10} exposures at the given residential location, beige indicates a good fit.

The maps confirm that the gridded and random monitoring networks generally provide good results although they under-predict the very high PM_{10} exposures very close to main roads and the motorway. The main errors here seem to be located close to the city centre (slight over-prediction in case of the 1.25 x 1.25 km gridded network) and along the main valley through the city from North East to South West. This is also the axis of the main transport route through the city and along which most of the industrial land can be found. Concentrations around this area are therefore comparable high, which causes in case of some monitoring networks, such as the 4.4 x 4.4 km gridded network, a slight exposure under-prediction. The purposive monitoring strategies, on the other hand, result in a very widespread over-prediction of PM_{10} exposures. Only the highest concentrations close to the main roads are represented well.

The Moran's I confirms with high significance that the absolute errors in estimated PM_{10} exposure resulting from different monitoring networks are spatially clustered. At a local level, the Anselin Local Moran's I further underlines that the errors cluster in areas of high PM_{10} concentrations, the city centre and close to the main transport routes.

The Kappa statistic allows for a more detailed investigation of the exposure estimates in the different exposure ranges. The ten exposure categories (see Figure 37) are evaluated in terms of class accuracy, the corresponding class membership for observed and estimated exposures, and in terms of overall, city wide, accuracy. The lowest five exposure classes, representing PM_{10} concentrations below 7.5 $\mu g/m^3$, are represented very well by gridded monitoring networks with a class accuracy of >50% in case of the 1.25 x 1.25 km gridded network. For higher exposure classes the accuracy declines gradually. The highest exposure classes >50 $\mu g/m^3$ are not represented. The overall accuracy for gridded monitoring networks increases with the number of monitoring stations and ranges from 22% to 45%. Similar pattern can be described for the random networks with >40% class accuracy for the lower exposure classes and no representation of the high exposure classes. Some random monitoring networks, however, also do not represent lower exposure classes very accurately (<30%) but the overall accuracy is slightly better than for the gridded networks with a range of 27% to 44%. No clear trend can be seen towards higher accuracy with increasing number of stations. For purposive monitoring networks, the class accuracy tends to decrease further, <30% for most exposure classes, but the represented exposure range increases. Higher exposure classes are represented by most purposive monitoring networks although with a low accuracy <20%. The overall accuracy is very low, ranging from 4% for networks, which focus on high road density, to 27% for networks with focus on highly populated areas.

4. *Discussion*

The study presented here provides an attempt to quantify the representativeness of PM_{10} monitoring networks within an urban setting. Three different monitoring strategies described in the environmental sampling literature are considered for the analysis. Purposive sampling, systematic gridded sampling and monitoring networks with randomly defined station locations are investigated with regard to their representation of urban exposure estimates. Some important conclusions can be drawn from this analysis. One of the main findings is that there is no clear trend towards a better performance of monitoring networks with increasing number of stations. Some tests indicate a slight trend in this direction, such as an increase in correlation between observed and estimated exposures, but overall the evidence is not strong enough to support this theory. This suggests that the location of the stations within a monitoring network is more important than the number of stations.

Monitoring networks performing best in the analysis do not concentrate on one area but have a widespread monitoring coverage in areas of higher PM_{10} concentration as well as effectual background coverage. This is not achieved with the purposive modelling strategy. The purposive monitoring strategy is sufficient for targeting certain populations or exposure categories but not to represent the overall city population, even if used in conjunction with background stations. This is true for all four approaches tested in this analysis. Overall, the approach, which targets areas with high road density, provides the best results. Networks with focus on high trafficked areas, on the other hand, represent exposures inadequately for all analysed networks. Responsible for this pattern is the overrepresentation of highly polluted areas close to the motorway and along the main transport axes, which are heavily monitored because of their high traffic volume. In contract, the high road density approach also includes areas of low traffic counts such as minor roads and therefore represents a broader PM_{10} concentration range, as is the case for the proximity to road approach. The purposive approach, which concentrates the monitors in areas of high population density, provides results similar to the high traffic density approach. The estimation of the exposure distribution in the city is generally inaccurate and individual exposure is highly overestimated, except for two of the analysed monitoring networks. Population density is highest close to the city centre, which also hosts some of the highest PM_{10} concentrations in the city. These areas are therefore overrepresented in the networks, which consequently results in biased exposure estimates. The introduction of background stations does not counterbalance this trend.

Most gridded as well as several random monitoring networks outperform the purposive networks in terms of representativeness of exposure. Gridded monitoring networks generally provide good results in terms of both estimating exposure distribution and individual exposure. The analysis further suggests that a station density of 2 x 2 km or higher provides the best exposure estimates. Randomly selected monitoring networks have the potential to result in exposure estimates close to the observed because they counter systematic bias. This is confirmed in the analysis. The random strategy results in some very representative monitoring networks. This result is however not reproducible because chance plays a very important role in distributing the monitoring stations within the urban area *(Fernandez et al., 2005)* which is reflected in some very poorly performing monitoring networks. This might be overcome by a significant increase in the number of monitoring stations.

In order to achieve a high level of representativeness in a monitoring network, the right balance between the representation of the spectrum of PM_{10} concentrations and the population density has to be found. The results of this analysis suggest a gridded monitoring strategy of an adequate station density (>20 monitoring sites) in conjunction with some stations in areas of high population density and a few stations in highly polluted areas. This would allow to measure concentration peaks and give a solid coverage of the overall concentration variation in the urban area. Thereby, it is very important to measure background concentrations effectively because otherwise the exposure estimates for a large part of the population will be over-predicted, as shown in the analysis.

Some of the considered monitoring networks result in exposure estimates that are very far from the actual exposure scenario in SIENA. This is true for all three monitoring strategies. The analysis could not find any conclusive reasons for this pattern apart from the ones discussed above. Further investigation is needed to determine the underlying effects that result in some very poor performances by monitoring networks that are apparently very similar in composition than networks that perform very well.

One of the factors influencing the results of this analysis is surely the lack of temporal variability in PM_{10} concentrations. The analysed concentration patterns present a snapshot in time and both daily concentration fluctuations and long-term temporal changes are largely ignored. Possible effects of temporal variability are not included in this study to avoid the concealment of any design specific effects. Monitoring programs implemented for epidemiological studies typically include sampling periods of one or two weeks *(Aguilera et al., 2008; Henderson et al., 2007; Wheeler et al., 2008)*, or in some cases even longer *(Ryan et al., 2007)* and inevitabely contain some temporal variability. This fact has to be reflected in the study specific network design.

Another factor, which influences the results, is the method used to assign people with exposure values. The method chosen here, to assign people with concentrations measured at their nearest monitoring station, is used with the incentive to reduce the errors potentially introduced by more sophisticated modelling techniques which would make interpretation of the results very difficult. The use of different exposure matrixes and proxies is explored elsewhere in the literature *(Huang & Batterman, 2000; Marshall et al., 2008)* and within INTARESE.

Reference List

AGUILERA I., SUNYER J., FERNANDEZ-PATIER R., HOEK G., ALFARO A.A., MELIEFSTE K., BOMBOI-MINGARRO M.T., NIEUWENHUIJSEN M.J., HERCE-GARRALETA D. & BRUNEKREEF B. (2008) Estimation of outdoor NOx, NO2, and BTEX exposure in a cohort of pregnant women using land use regression modeling. *Environmental Science & Technology* **42**, 815-821.

BAXTER L.K., WRIGHT R.J., PACIOREK C.J., LADEN F., SUH H.H. & LEVY J.I. (2010) Effects of exposure measurement error in the analysis of health effects from traffic-related air pollution. *Journal of Exposure Science and Environmental Epidemiology* **20**, 101-111.

BRIGGS D., DE HOOGH C. & GULLIVER J. (2010) Comparative assessment of GIS-based methods and metrics for modeling exposure to air pollution. *Journal of Toxicology and Environmental Health* **in press**,

BRUNEKREEF B. & HOLGATE S.T. (2002) Air pollution and health. *Lancet* **360**, 1233-1242.

BRUS D.J. & HEUVELINK G.B.M. (2007) Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* **138**, 86-95.

CHAMBERS J.M., CLEVELAND W.S., KLEINER B. & TUKEY P.A. (1983) *Graphical methods for data analysis*. Wadsworth Brooks/Cole.

COHEN J. (1988) *Statistical power analysis for the behavioral sciences*. 2nd Ed. New Jersey: Lawrence Erlbaum.

CONGALTON R.G. (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* **37**, 35-46.

DOCKERY D.W., POPE C.A., III, XU X., SPENGLER J.D., WARE J.H., FAY M.E., FERRIS B.G., Jr. & SPEIZER F.E. (1993) An association between air pollution and mortality in six US cities. *New England Journal of Medicine* **329**, 1753-1759.

DOWNES B.J. (2010) Back to the future: little-used tools and principles of scientific inference can help disentangle effects of multiple stressors on freshwater ecosystems. *Freshwater Biology* **55**, 60-79.

FERNANDEZ J.A., REAL C., COUTO J.A., ABOAL J.R. & CARBALLEIRA A. (2005) The effect of sampling design on extensive bryomonitoring surveys of air pollution. *Science of the Total Environment* **337**, 11-21.

FRANCO-MARINA F., VILLALBA-CALOCA J., SEGOVIA N. & TAVERA L. (2003) Spatial indoor radon distribution in Mexico City. *Science of the Total Environment* **317**, 91-103.

GIAVIS G.M., KAMBEZIDIS H.D. & LYKOUDIS S.P. (2009) Frequency distribution of particulate matter (PM10) in urban environments. *International Journal of Environment and Pollution* **36**, 99-109.

GILBERT N.L., GOLDBERG M.S., BECKERMAN B., BROOK J.R. & JERRETT M. (2005) Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model. *Journal of the Air & Waste Management Association* **55**, 1059-1063.

GILBERT R.O. (1987) *Statistical methods for environmental pollution monitoring*. John Wiley and Sons. pp. 1-336.

GLORENNEC P. & MONROUX F. (2007) Health impact assessment of PM10 exposure in the city of Caen, France. *Journal of Toxicology and Environmental Health-Part A-Current Issues* **70**, 359-364.

HANNA S.R. (1993) Uncertainties in Air-Quality Model Predictions. *Boundary-Layer Meteorology* **62**, 3-20.

HENDERSON S.B., BECKERMAN B., JERRETT M. & BRAUER M. (2007) Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environmental Science & Technology* **41**, 2422-2428.

HIRSCH T., NEUMEISTER V., WEILAND S.K., VON MUTIUS E., HIRSCH D., GRAFE H., DUHME H. & LEUPOLD W. (2000) Traffic exposure and allergic sensitization against latex in children. *Journal of Allergy and Clinical Immunology* **106**, 573-578.

HOEK G., BEELEN R., DE HOOGH K., VIENNEAU D., GULLIVER J., FISCHER P. & BRIGGS D. (2008) A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment* **42**, 7561-7578.

HOEK G., FORSBERG B., BOROWSKA M., HLAWICZKA S., VASKOVI E., WELINDER H., BRANIS M., BENES I., KOTESOVEC F., HAGEN L.O., CYRYS J., JANTUNEN M., ROEMER W. & BRUNEKREEF B. (1997) Wintertime PM10 and black smoke concentrations across Europe: Results from the peace study. *Atmospheric Environment* **31**, 3609-3622.

HOEK G., MELIEFSTE K., CYRYS J., LEWNE M., BELLANDER T., BRAUER M., FISCHER P., GEHRING U., HEINRICH J., VAN VLIET P. & BRUNEKREEF B.

(2002) Spatial variability of fine particle concentrations in three European areas. *Atmospheric Environment* **36**, 4077-4088.

HUANG Y.L. & BATTERMAN S. (2000) Selection and evaluation of air pollution exposure indicators based on geographic areas. *Science of the Total Environment* **253**, 127-144.

INIGUEZ C., BALLESTER F., ESTARLICH M., LLOP S., FERNANDEZ-PATIER R., AGUIRRE-ALFARO A. & ESPLUGUES A. (2009) Estimation of personal NO2 exposure in a cohort of pregnant women. *Science of the Total Environment* **407**, 6093-6099.

JENNESS J. & WYNNE J.J. (2005) Kappa analysis (kappa_stats-avx) extension for ArcView 3.x. *Jenness Enterprises*

JERRETT M., ARAIN M.A., KANAROGLOU P., BECKERMAN B., CROUSE D., GILBERT N.L., BROOK J.R., FINKELSTEIN N. & FINKELSTEIN M.M. (2007) Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *Journal of Toxicology and Environmental Health-Part A-Current Issues* **70**, 200-212.

KANAROGLOU P.S., JERRETT M., MORRISON J., BECKERMAN B., ARAIN M.A., GILBERT N.L. & BROOK J.R. (2005) Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. *Atmospheric Environment* **39**, 2399-2409.

KHAN A. (2008) A combination of adaptive and line intercept sampling applicable in agricultural and environmental studies. *Journal of Statistics* **15**, 44-53.

KOUSA A., KUKKONEN J., KARPPINEN A., AARNIO P. & KOSKENTALO T. (2002) A model for evaluating the population exposure to ambient air pollution in an urban area. *Atmospheric Environment* **36**, 2109-2119.

KUNZLI N., KAISER R., MEDINA S., STUDNICKA M., CHANEL O., FILLIGER P., HERRY M., HORAK F., PUYBONNIEUX-TEXIER V., QUENEL P., SCHNEIDER J., SEETHALER R., VERGNAUD J.C. & SOMMER H. (2000) Public-health impact of outdoor and traffic-related air pollution: a European assessment. *Lancet* **356**, 795-801.

LEBRET E., BRIGGS D., VAN REEUWIJK H., FISCHER P., SMALLBONE K., HARSSEMA H., KRIZ B., GORYNSKI P. & ELLIOTT P. (2000) Small area variations in ambient NO2 concentrations in four European areas. *Atmospheric Environment* **34**, 177-185.

LIOY P.J. (1999) Exposure analysis and assessment in the 21st century. *Inhalation Toxicology* **11**, 623-636.

MACKEY A.P. & HODGKINSON M.C. (1995) Concentrations and Spatial-Distribution of Trace-Metals in Mangrove Sediments from the Brisbane River, Australia. *Environmental Pollution* **90**, 181-186.

MADSEN C., CARLSEN K.C.L., HOEK G., OFTEDAL B., NAFSTAD P., MELIEFSTE K., JACOBSEN R., NYSTAD W., CARLSEN K.H. & BRUNEKREEF B. (2007) Modeling the intra-urban variability of outdoor traffic pollution in Oslo, Norway - A GA(2)LEN project. *Atmospheric Environment* **41**, 7500-7511.

MARSHALL J.D., NETHERY E. & BRAUER M. (2008) Within-urban variability in ambient air pollution: Comparison of estimation methods. *Atmospheric Environment* **42**, 1359-1369.

MARTIN M.A., TARRERO A., GONZALEZ J. & MACHIMBARRENA M. (2006) Exposure-effect relationships between road traffic noise annoyance and noise cost valuations in Valladolid, Spain. *Applied Acoustics* **67**, 945-958.

MATTUCK R., BLANCHET R. & WAIT A.D. (2005) Data representativeness for risk assessment. *Environmental Forensics* **6**, 65-70.

MILLER K.A., SISCOVICK D.S., SHEPPARD L., SHEPHERD K., SULLIVAN J.H., ANDERSON G.L. & KAUFMAN J.D. (2007) Long-term exposure to air pollution and incidence of cardiovascular events in women. *New England Journal of Medicine* **356**, 447-458.

MONN C. (2001) Exposure assessment of air pollutants: a review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone. *Atmospheric Environment* **35**, 1-32.

MORAN P.A.P. (1950) Notes on continuous stochastic phenomena. *Biometrika* **37**, 17-23.

NIEMI R.M. & NIEMI J.S. (1990) Monitoring of Fecal Indicators in Rivers on the Basis of Random Sampling and Percentiles. *Water Air and Soil Pollution* **50**, 331-342.

OANH N.T.K., UPADHYAYA N., ZHUANG Y.H., HAO Z.P., MURTHY D.V.S., LESTARI P., VILLARIN J.T., CHENGCHUA K., CO H.X., DUNG N.T. & LINDGREN E.S. (2006) Particulate air pollution in six Asian cities: Spatial and temporal distributions, and associated sources. *Atmospheric Environment* **40**, 3367-3380.

OGLESBY L., KUNZLI N., ROOSLI M., BRAUN-FAHRLANDER C., MATHYS P., STERN W., JANTUNEN M. & KOUSA A. (2000) Validity of ambient levels of fine particles as surrogate for personal exposure to outdoor air pollution - Results of the European EXPOLIS-EAS study (Swiss Center Basel). *Journal of the Air & Waste Management Association* **50**, 1251-1261.

PASCAL L. (2009) Short-term health effects of air pollution on mortality. *Revue Francaise D Allergologie* **49**, 466-476.

PIEGORSCH W. & BAILER J. (2005) *Analyzing environmental data*. Chichester: John Wiley & Sons.

POPE C.A., BURNETT R.T., THUN M.J., CALLE E.E., KREWSKI D., ITO K. & THURSTON G.D. (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama-Journal of the American Medical Association* **287**, 1132-1141.

POPE C.A. & DOCKERY D.W. (2006) Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association* **56**, 709-742.

ROTKO T., OGLESBY L., KUNZLI N. & JANTUNEN M.J. (2000) Population sampling in European air pollution exposure study, EXPOLIS: comparisons between the cities and representativeness of the samples. *Journal of Exposure Analysis and Environmental Epidemiology* **10**, 355-364.

RYAN P.H., LEMASTERS G.K., BISWAS P., LEVIN L., HU S.H., LINDSEY M., BERNSTEIN D.L., LOCKEY J., VILLAREAL M., HERSHEY G.K.K. & GRINSHPUN S.A. (2007) A comparison of proximity and land use regression traffic exposure models and wheezing in infants. *Environmental Health Perspectives* **115**, 278-284.

SEINFELD J. (1986) *Atmospheric chemistry and physics of air pollution*. New York: Wiley.

WANG G., GERTNER G., ANDERSON A.B. & HOWARD H. (2008) Repeated measurements on permanent plots using local variability sampling for monitoring soil cover. *Catena* **73**, 75-88.

WHEELER A.J., SMITH-DOIRON M., XU X., GILBERT N.L. & BROOK J.R. (2008) Intra-urban variability of air pollution in Windsor, Ontario - Measurement and modeling for human exposure assessment. *Environmental Research* **106**, 7-16.

WORLD HEALTH ORGANISATION (WHO) (2003) Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide. In *http://www.euro.who.int/document/e79097.pdf, accessed 8 November 2007*. AnonymousCopenhagen:

WORLD HEALTH ORGANISATION (WHO) (2004) Health aspects of air pollution - results from the WHO project "Systematic review of the health aspects of air pollution in Europe". In *http://www.euro.who.int/document/e83080.pdf, accessed 01 November 2007*. Anonymous

ZEGER S.L., THOMAS D., DOMINICI F., SAMET J.M., SCHWARTZ J., DOCKERY D. & COHEN A. (2000) Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environmental Health Perspectives* **108**, 419-426.