# Spatial air pollution modelling for a West-African town

Sirak Zenebe Gebreab,[1,2] Danielle Vienneau,[1,2] Christian Feigenwinter,[3] Hampâté Bâ,[4] Guéladio Cissé,[1,2] Ming-Yi Tsai[1,2,5]

[1]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel; [2]University of Basel, Basel; [3]MCR-Meteorology, Climatology and Remote Sensing, University of Basel, Switzerland; [4]Institut National de Recherches en Santé Publique, Nouakchott, Mauritania; [5]Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA, USA

Correspondence: Ming-Yi Tsai, Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland.
Tel: +41.61.2848310 - Fax: +41.61.2848105.
E-mail: m.tsai@unibas.ch

## Abstract

Land use regression (LUR) modelling is a common approach used in European and Northern American epidemiological studies to assess urban and traffic related air pollution exposures. Studies applying LUR in Africa are lacking. A need exists to understand if this approach holds for an African setting, where urban features, pollutant exposures and data availability differ considerably from other continents. We developed a parsimonious regression model based on 48-hour nitrogen dioxide ($NO_2$) concentrations measured at 40 sites in Kaédi, a medium sized West-African town, and variables generated in a geographic information system (GIS). Road variables and settlement land use characteristics were found to be important predictors of 48-hour $NO_2$ concentration in the model. About 68% of concentration variability in the town was explained by the model. The model was internally validated by leave-one-out cross-validation and it was found to perform moderately well. Furthermore, its parameters were robust to sampling variation. We applied the model at 100 m pixels to create a map describing the broad spatial pattern of $NO_2$ across Kaédi. In this research, we demonstrated the potential for LUR as a valid, cost-effective approach for air pollution modelling and mapping in an African town. If the methodology were to be adopted by environmental and public health authorities in these regions, it could provide a quick assessment of the local air pollution burden and potentially support air pollution policies and guidelines.

## Introduction

Air pollution is a major environmental health problem affecting large populations around the world. Exposure to air pollutants causes a number of adverse health outcomes including respiratory infections, cardiovascular diseases and lung cancer (Brunekreef and Holgate, 2002; Brook et al., 2004; Chen and Kan, 2008, Raaschou-Nielsen et al., 2013). Certain populations, especially those in the developing countries, are highly impacted by the health effects related to poor air quality, which has mostly been linked to indoor air pollution from biomass burning (Bruce et al., 2000; Lim et al., 2012). However, in these regions, the worsening outdoor air pollution, as a result of increasing urbanisation, population growth, motor vehicle use, unregulated industrial emissions and unsustainable policies, is also causing considerable problems (Mehta, 2003; HEI International Scientific Oversight Committee, 2010). According to the World Health Organization (WHO) report, an estimated 3.7 million premature

deaths are attributed to outdoor air pollution globally in 2012; more than 85% of these deaths occur in low and middle-income countries (LMICs) (WHO, 2014).

The health effects of outdoor air pollution have been investigated for a number of years in many epidemiological studies. Exposure assessment plays an essential role in these studies in understanding and defining the relationship between air pollutants and associated health effects (Zou *et al*., 2009; Brauer *et al*., 2012). Due to financial and technical constraints, however, air pollution exposure assessment studies conducted in the developing world are relatively scarce.

Land use regression (LUR) modelling has recently become a common approach used in air pollution epidemiological studies to assess exposures related to urban and traffic related air pollution (Briggs *et al*., 2007; Chen *et al*., 2010; Liu *et al*., 2012, Beelen *et al*., 2013). Models are derived by regressing air pollution measurements taken at a set of locations against geographical covariates that are potential proxies for emission sources or air pollutant dispersion processes. The purpose of these models is to estimate concentrations of pollutants in unmeasured locations, *e.g*., where people live and are exposed, thereby, enabling studies on the association between health outcomes and exposure (Cesaroni *et al*., 2012; MacIntyre *et al*., 2013). Furthermore, LUR models based on measurements that fully capture the heterogeneity and pollutant range of an area can provide maps of pollution distribution over an entire area, which would be infeasible with measurements alone. LUR models have also been shown to be especially important tools for characterising the spatial variability of pollutants that rapidly decay from the source (Jerrett *et al*., 2007; Brauer *et al*., 2008).

LUR modelling has thus far been employed in a variety of European and North American cities (Hoek *et al*., 2008; Ross *et al*., 2005; Beelen *et al*., 2013). To our knowledge, with the exception of a few studies in some Asian LMICs (Chen *et al*., 2010; Allen *et al*., 2013; Saraswat *et al*., 2013), this approach has not been applied in an African region. A need therefore exists to evaluate whether such an approach also holds for this part of the world, where urban features differ from those of other continents. We aimed to apply the LUR approach to model and map nitrogen dioxide ($NO_2$) for a medium sized West-African town. Our objective was to demonstrate that robust LUR models could be created using available and satellite derived geographical data, and to identify the inherent limitations and benefits of this approach in an African context.

## Materials and Methods

### Study site

The study was conducted in Kaédi (16.15°N, 13.50°S), a medium sized town located in the semi-arid Sahelian region of Mauritania (Figure 1). The Senegal River valley separates the town from the neighbouring country of Senegal. Daily temperature variation ranges from 16 to 41°C. The population was estimated to be 49,152 inhabitants in 2010 (ONS, 2013). Kaédi has a high population density where many live in congested housing with substandard basic services. There are two main traffic axes in the town with paved concrete roads, crossing the town roughly from north to south and east to west, and many secondary dirt roads. Dust storms, emission from cottage industries, road traffic and traditional cooking using biomass fuels are the main sources of outdoor air pollution in the town.

### Air pollutant measurements

In April 2011, in conjunction with the Kaédi-wide dry season vulnerability assessment multidisciplinary survey, in the framework of the *Ecohealth and adaptation to climate change in water and health sectors in West Africa* project funded by the International Development Research Centre (IDRC) (Cissé *et al*., 2011; Touray *et al*., 2012), outdoor $NO_2$ samples were collected in Kaédi. Monitoring locations were selected to achieve a broad, even coverage across the town and to capture the full variability of air pollution. Measurements were collected for 48-hours using passive samplers deployed at a height of 3 m at 40 systematically selected locations across the different neighborhoods of the town. The collected samples were stored in an icebox and transported to the Swiss Tropical and Public Health Institute, Switzerland for analysis (Passam AG, Männedorf, Switzerland). The sampling sites were selected for three site types representing street, urban and regional background sites. Street sites were near major roads and expected to have high road traffic flow. Urban background sites were mostly in built-up areas where pollution from different sources was readily apparent. Regional sites were located on the outskirts of the town, further away from pollution sources (traffic and high population settlements).

### Geographic information system analysis

Following previous LUR models, we initially aimed to obtain or derive GIS variables in six main categories: roads, land use, traffic intensity, population density, household density and physical geography (Ryan and Le Masters, 2007; Hoek *et al*., 2008). Traffic intensity and household data were not available for Kaédi from local sources and, hence, not considered in this study. Most data were obtained from online and satellite derived resources as described below. Spatial analyses were conducted using ArcGIS10 (ESRI, Redlands, CA, USA).

We used Open Street Map to digitize a spatial layer for roads, supplemented with information from Google Earth to include additional road detail and guarantee the completeness of the network across the town. Based on the Mauritanian road network classification system, two main road types were determined: *local paved roads* and *local feeder roads*. Road variables by type were calculated as total road length within 50, 100, 150 and 200 m buffers; the Euclidian distance between measurement sites and roads by type were also computed (Briggs *et al*., 2000; Beelen *et al*., 2013; Abernethy *et al*., 2013).

We created a study area specific land use/land cover (LULC) data set (Figure 2) using a standard image classification algorithm. We obtained and evaluated two different resolution satellite images as close to the sampling day as possible: the freely available multiband *Landsat-5 TM* (30 m) satellite image with spectral bands: blue 450-520 nm, green 520-600 nm, red 630-690 nm, NIR 760-900 nm, SWIR1 1550-1750 nm, SWIR2 2080-2350 nm (http://www.usgs.gov); and the multispectral resolution *RapidEye* image (5 m-pixel size, 5 spectral bands: blue 440-510 nm, green 520-590 nm, red 630-685 nm, red edge 690-730 nm, NIR 760-850 nm) (http://www.blackbridge.com). A multi-step supervised classification was applied in ArcGIS10: creating a priori training areas (samples) of land covers and applying the maximum likelihood classifier (Richards and Xiuping, 1999). The created training areas were evaluated by histograms and scatter plots for the degree of overlap between classes before classifying the image. This process of image classification was repeated a number of times, until the best possible classification was obtained. The quality of the LULC classification was inspected visually and verified using Google Earth as ground truth. The final classified image was again examined for the presence of flagrant misclassifications and corrected with a post classification

process. In order to provide a measure for the accuracy of the LULC classification, well recognizable regions of interests for each class were defined in the town of Kaédi and its closer surroundings and a confusion matrix was calculated (Manandhar *et al*., 2009). Based on the final LULC data set, geographical variables for each land class were calculated for buffer sizes of 100, 300, 500 and 1000 m.

Buffers (zones of influence) for all of the predictor variables were made to take into account the environmental processes of the variables and the geographic extent of our study area. For example, we restricted our maximum buffer size to 200 m for road variables as larger buffers could misrepresent the observed pollutant effect. Similarly, for land use variables, we set the maximum buffer size to 1km given the influence of the pollutant under study.

Population data for Kaédi were only available as census area totals for year 2000. To obtain a finer population representation, we spatially disaggregated the census totals using a dasymetric mapping technique. Dasymetric mapping is a thematic cartographic method used to redistribute spatially aggregated population data to a finer scale that better depicts the underlying distribution. The technique uses ancillary data to weight and spatially relocate populations (Mennis, 2003; Briggs *et al*., 2007). We used our LULC map as ancillary data to provide the contextual information on how the population is likely distributed across the town. Weights were applied to each LULC class in the ancillary data, based on the proportion of people expected to be living there. Greater population, for example, is expected to live in settlements compared to farmland. Census population (*i.e.* within census boundaries) was then proportioned across the LULC grid and aggregated to a 200 m grid. We applied *Focalsum* functions to the resulting 200 m grid to compute population within *buffers* of 200, 400, 600, 800 and 1000 meters around the measurement sites.

As elevation is often considered an important geographical factor affecting pollutant concentration (Shmool *et al*., 2014), we extracted elevation at measurement sites from SRTM 90 m digital elevation

model (http://www.cgiar-csi.org) and stored as a point predictor variable.

The database containing potential GIS predictor variables was exported to the statistical software R (http://cran.r-project.org/) for statistical analysis and model development.

## Land use regression modelling

A multiple linear regression approach was used to develop the LUR on the basis of untransformed $NO_2$ concentration (dependent variable) and the GIS predictor variables (independent variables). We followed a supervised forward stepwise regression procedure (Hoek *et al*., 2008; Eeftens *et al*., 2012), which allows for a logical selection of variables that maximize the model's percentage of explained variability ($R^2$). Briefly, each predictor variable was regressed against the $NO_2$ concentrations and the variable that gave the highest $R^2$ was selected as the first variable to be included in the model. Selection of the subsequent variables, tested in turn, was made based on the magnitude of their additional contribution to the model (in partial adjusted $R^2$, with at least a 1% increase). This process of evaluating and adding variables continued until inclusion of variables no longer improved the model. A variable was only kept in the model if it entered with the expected direction of effect and did not change the pre-specified direction of effect for variables already included. All included variables were checked for significant associations with $NO_2$; those with P values greater than 0.1 were sequentially removed from the model.

We checked for the presence of multicollinearity between the potential predictor variables using the variance inflation factor (VIF), where values greater than 3 indicate collinearity (O'Brien, 2007). Spatial autocorrelation in the residuals was tested using Moran's I statistic (Moran's I P>0.05 indicates no spatial autocorrelation). Cook's distance was also computed to detect the presence of influential sites (being both an outlier and leverage point). Sites having Cook's dis-
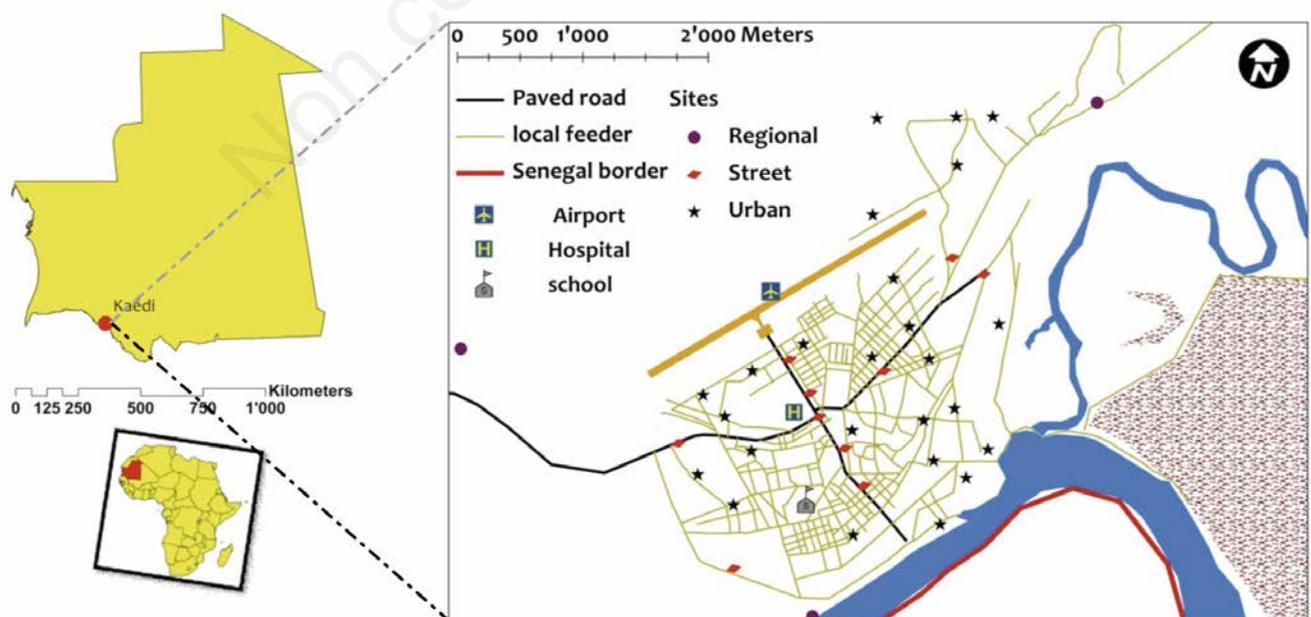


**Figure 1. Map of Kaédi, Southern Mauritania.**

tance values greater than 1, indicating an influential point, were further analysed.

The model was internally validated by leave-one-out cross-validation (LOOCV) (Hoek *et al*., 2008; Eeftens *et al*., 2012; Beelen *et al*., 2013). In LOOCV, one observation from the data is excluded and the model refitted and applied to predict for the site left out. This is repeated for all sites so that each site has a prediction based on a model that did not use its measurement. The magnitude of error, root mean square error (RMSE) and the explanatory power ($R^2$) of the cross-validated model were then assessed by regressing observations against predicted concentrations. The sensitivity of model parameters to the selection of samples was also evaluated with a non-parametric bootstrap analysis (Ross *et al*., 2005; Ryan *et al*., 2007). In the analysis, random n-5 samples were drawn from the observations with replacement; each time a sample is drawn, a model was run and coefficients were recorded. This procedure was repeated 10,000 times, and mean, standard error and confidence intervals of these coefficients were computed.

Finally, the validated model was rendered at a 100 m resolution (*i.e.* predicted at grid cell centroids) and spatially smoothed via bilinear interpolation to create a 48-hour $NO_2$ concentration map. The map was visually evaluated for plausibility.

## Results

### Measurements

During the measurement campaign, one sample went missing, thus samples from only 39 sites were retrieved. The 48-hr $NO_2$ concentrations at these sites ranged from 0.3 to 16 µg/m$^3$. The mean concentration value was 5.26±3.19 µg/m$^3$ with a median (interquartile range) concentration of 5.3 (3.3-6.5) µg/m$^3$. $NO_2$ concentrations were different between the three site types, with that of the street background showing the highest mean concentrations (7.0±4.4 µg/m$^3$), the urban background the next highest (5.1±2.3 µg/m$^3$), and the regional background lowest (2.1±1.7 µg/m$^3$). In Kaédi, relatively elevated concentrations were measured at the sites along the main paved north-south street with the highest level measured near the intersection of the two main traffic axes. Concentrations were also elevated around the town center; whereas, low values (0.3-4.0 µg/m$^3$) were observed in the northern part of the town.

### Geographic information system derived predictors

A total of 65 potential predictor variables (Supplementary Table 1)

were produced in the GIS analysis and tested in the land use regression modelling.

The digitised road network for Kaédi had a length of 14,519 m of paved and 90,149 m of feeder roads. Sites in the main town had the highest road densities ranging from 1100 to 3200 m within a 200 m buffer, while regional background sites had relatively lower density (0-547 m) and were further from paved roads (302-1052 m).

In deriving the LULC data set, we tested two satellite images of differing spatial resolution; our final data set was based on the higher resolution *RapidEye* imagery. The image classification resulted in a 5 m LULC data set (Figure 2) with 11 different classes (Supplementary Table 2). The overall accuracy of the classification was 89% with a kappa coefficient of 0.87. Main misclassifications were observed between classes of low spectral separability, *i.e.* between the high and low settlement class, between dense vegetation and pasture/fluvial, and between the two settlement classes and thoroughfare/track class. Some LULC classes such as water, farmland and sparse vegetation were substantially absent or dispersed around the monitoring sites while others like settlement and thoroughfare were clustered.

The overall altitude variability in Kaédi was very low; altitude at the measurement sites ranged between 12 and 22 m above sea level with the exception of one site measured at 32 m.

We estimated the small-scale population distribution in Kaédi to range from 0 to 1550 individuals per 200 m grid cell. At half of the sampling sites, the population within each 200 m grid cell was approximated to be more than 100 people.

### Final land use regression model

A parsimonious model with four predictors was developed, giving a final model with an adjusted $R^2$ of 0.68 (Table 1). The variable *paved road within 100 m buffer* (PAVBUF100) was the first variable to be selected with the highest adjusted $R^2$ of 0.35. Three additional variables entered the model: *area of dense settlement within 1000 m buffer* (SETH1000), *area of moderate settlement within 300 m buffer* (SETL300), and *all types of road within 200 m buffer* (ROADBUFF200).

The maximum VIF of the estimated regression coefficients was 1.23, and Moran's I statistic on model residuals was -0.03, P=0.89. Although site 15 was identified as a potential influential site (Cook's D of 1.44), we retained it as sensitivity analysis, in which it was removed, did not impact the model. This site is also in a location where emission from different sources occur and removing it would create a loss of information and statistical power. The LOOCV of the model yielded an adjusted $R^2$ of 0.54 and RMSE of 2.18 µg/m$^3$. Results from the bootstrap sensitivity analysis showed that the coefficient values of the intercept and

**Table 1. Final nitrogen dioxide land use regression model based on measurements collected in April 2011.**

| Model variable | Beta | SE | P value | VIF | Partial $R^2$ (%) | Beta (90-10)° |
|---|---|---|---|---|---|---|
| Intercept | 0.96 | 0.674 | 0.16 | - | - | - |
| PAVBUF100 | $1.91 \times 10^{-2}$ | $3.54 \times 10^{-3}$ | <0.01 | 1.23 | 35 | 3.45 |
| SETH1000 | $1.76 \times 10^{-5}$ | $4.48 \times 10^{-6}$ | <0.01 | 1.13 | 24 | 3.37 |
| SETL300 | $3.56 \times 10^{-5}$ | $1.34 \times 10^{-5}$ | 0.01 | 1.14 | 5 | 2.19 |
| ROADBUF200 | $8.7 \times 10^{-4}$ | $3.67 \times 10^{-4}$ | 0.02 | 1.13 | 4 | 1.98 |

Model $R^2$=0.71; model adjusted $R^2$=0.68; cross validated $R^2$=0.54; cross validated root mean square error=2.18; Moran's I=-0.03 (P=0.89).

SE, standard error; VIF, variance inflation factor; PAVBUF100, paved road within 100 m buffer; SETH1000, dense settlement within 1000 m buffer; SETL300, moderate settlement within 300 m buffer; ROADBUF200, all types of road within 200 m buffer. °Coefficients multiplied by the difference of the 10$^{th}$ and 90$^{th}$ percentile of the predictors.

three of the four model predictor variables were normally distributed. However, the coefficient of *paved road within a 100 m buffer* (PAVBUF100) appeared to be skewed and to have a some what bimodal distribution (Figure 3).

The predicted values in the smoothed $NO_2$ concentration map (Figure 4) ranged between 0.96 and 12 $\mu g/m^3$. Higher values are seen around paved roads and at their intersections; whereas, suburban areas outside the main city showed relatively low concentration values.

## Discussion

### Measurements

$NO_2$ concentrations measured in Kaédi are quite low compared to other cities in low and middle-income countries (LMICs) (Han and Naeher, 2006; El-Batrawy, 2011; Allen *et al*., 2013). Although there is no national air quality guideline to which we can directly refer these 48-hour $NO_2$ values, the concentrations were still much lower than WHO's
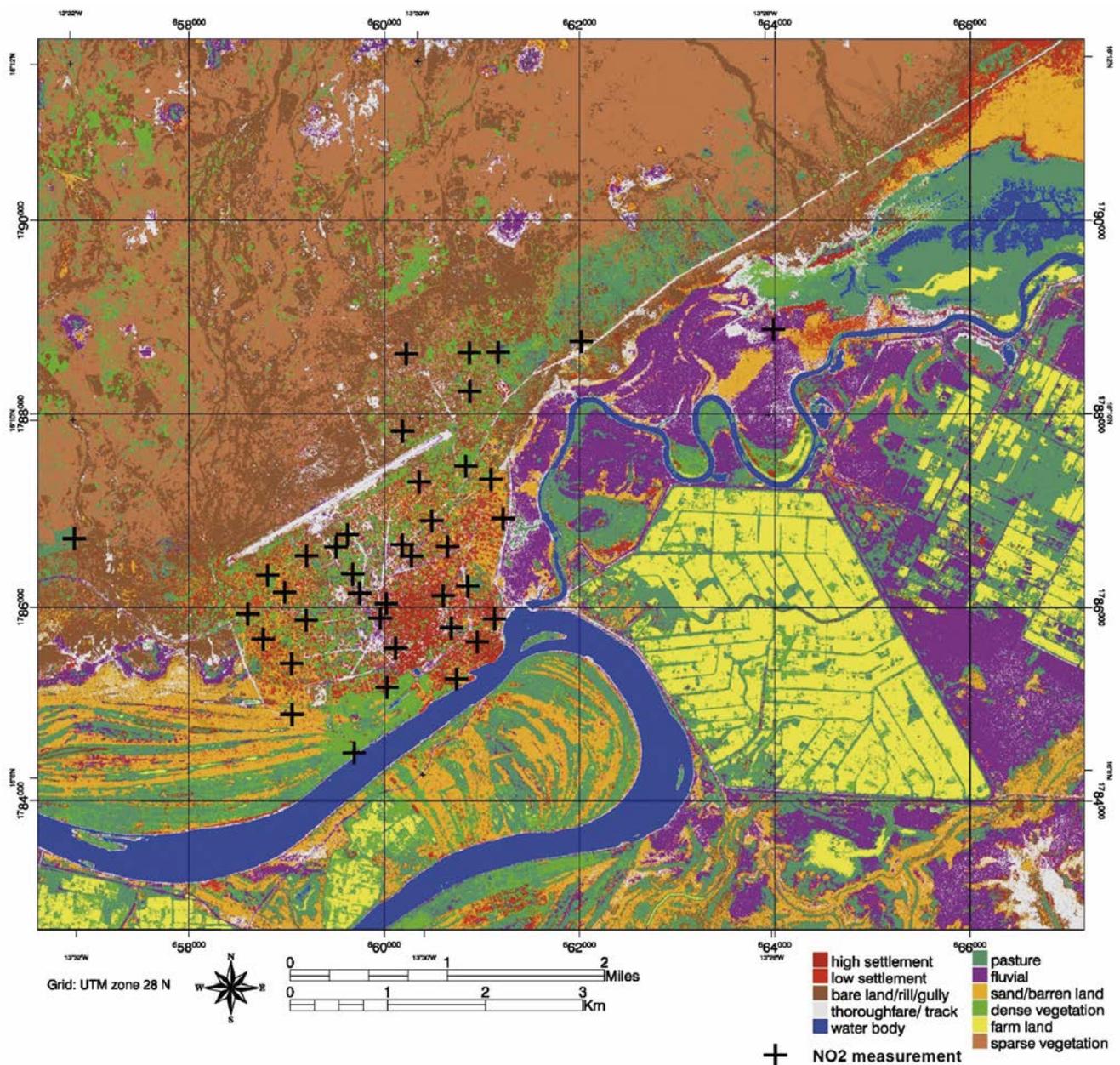


**Figure 2. Land use/land cover map of Kaédi developed with image classification technique, using *RapidEye* (5 m) satellite image. $NO_2$ measurement sites are also indicated.**

recommended values (annual mean=40 μg/m³, 1-hour mean=200 μg/m³) (Krzyzanowski and Cohen, 2008). However, variability of concentrations between locations is present: concentrations near roads and urban settlements are comparatively elevated relative to concentrations on the periphery of the town.

The low NO₂ concentrations observed in Kaédi are most likely due to a combination of low traffic density in the area and that the measurements were taken during the hot dry season when the mixing height is particularly high due to strong convective currents from surface heating. The case would likely have been different for other pollutants such as particulate matter, where the sources are more numerous, *e.g.*, road dust resuspension (only two paved roads), open waste burning, various cottage industries, and biomass solid fuel cooking (a common practice in the area).

## Generating geographic information system data

Availability of GIS data for our LUR model was one of the anticipated challenges in this research project. We overcame this challenge by integrating satellite derived information and publicly available data into a comprehensive GIS for Kaédi. Google Earth, Open Street Maps and other GIS web portals were the main data sources that we used. Taken together, these sources provided important GIS information that could be used as pollution surrogates.

Road density is an important surrogate for traffic-related NO₂ concentration. The majority of the papers reviewed by Hoek *et al*. (2008) have included road related variables in their models. As in our case, the lack of consolidated road network data is likely to be a challenge in other low-income countries. To manually digitize the road network was feasible for our small study; however, this approach may be time con-
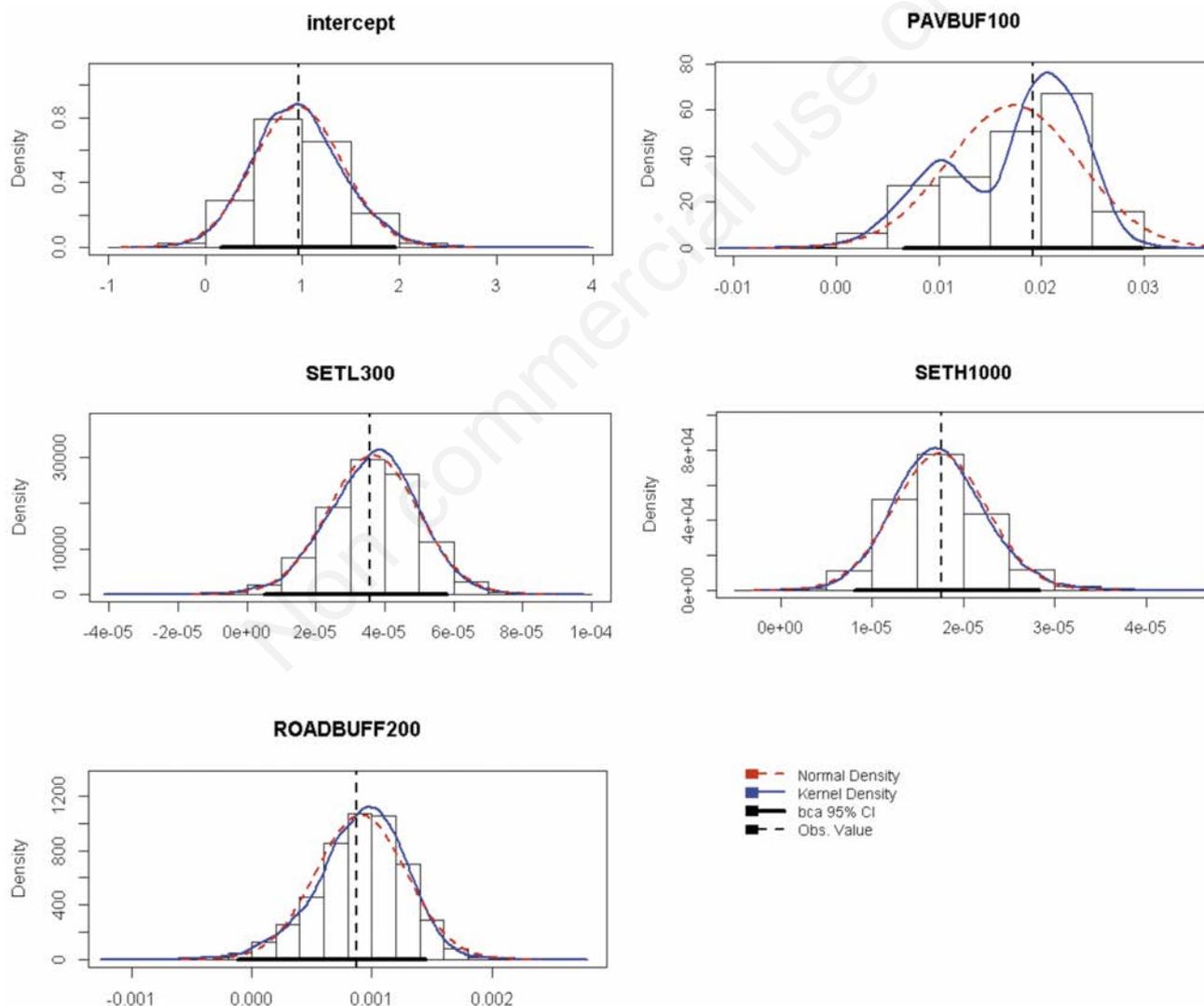


**Figure 3. Histograms of land use regression model parameters after bootstrap analysis with 10,000 iterations. PAVBUF100, paved road within 100 m buffer; SETL300, moderate settlement within 300 m buffer; SETH1000, dense settlement within 1000 m buffer; ROADBUF200, all types of road within 200 m buffer.**

suming and costly for areas with bigger and more complex road networks.

The other key surrogate that is considered important for air pollution modelling is land use. Land use data are typically derived from satellite images, and for many study areas are readily available GIS data sets. Previous LUR modelling studies have used land use and land cover variables that are, in one way or another, derived from satellite images (Allen *et al.*, 2013). In this study, we had to derive our own LULC data set using image classification of satellite imagery. In creating our LULC data set, we found the use of the 30 m resolution Landsat-5 TM data to be too coarse to identify small area differences in LULC in Kaédi. The finer resolution (5 m) RapidEye data, however, provided a better and more accurate classification. Hence, selection of suitable satellite data is important for extracting reliable and study-specific information from the raw data. While not appropriate for our study area, Landsat-5 TM may be suitable for larger cities than for small, congested areas like Kaédi. The other limitation that we recognized, which could potentially lead to a high level of misclassification, is the lack of ground truthing available to select training areas and confirm the classification. However, this issue was easily resolved through the use of Google Earth images and prior (on the ground) knowledge of the study area (acquired by the monitoring team). In general, the availability of high-resolution satellite images offer a good potential to derive appropriate land use predictors when readily available GIS data are lacking. The

ease of use of GIS and remote sensing tools, as has been shown by other researchers (Maxwell, 2010; Mao *et al*., 2012; Allen *et al.*, 2013), further facilitates the exploitation of satellite imagery in air pollution modelling and exposure assessments in low income countries.

Many applications, including air pollution modelling and exposure assessment, often require population data at finer spatial resolutions; however, population data are usually collected only at coarse administrative area levels. Different techniques have been used to redistribute census population to smaller scales, *e.g.*, area weighting being the simplest approach. We used a more involved method called dasymetric mapping, which enabled us to represent Kaédi's population at a finer scale and provided a more plausible population representation in the town. Where block level population data are not available or too costly to acquire, this method provides a viable alternative. Nevertheless, we were limited in our application of this approach by the need to assume weights rather than derive them deterministically. We further lacked the necessary data to undertake a validation of our final population distribution and thus aggregated the final population data set to 200 m to reduce the uncertainty.

Ultimately, the fact that we could build a LUR model using these data sources generally supports their use and this modelling approach for air pollution modelling and exposure assessment in developing countries. In the absence of locally available GIS data, they are also cost effective substitutes that are globally available.
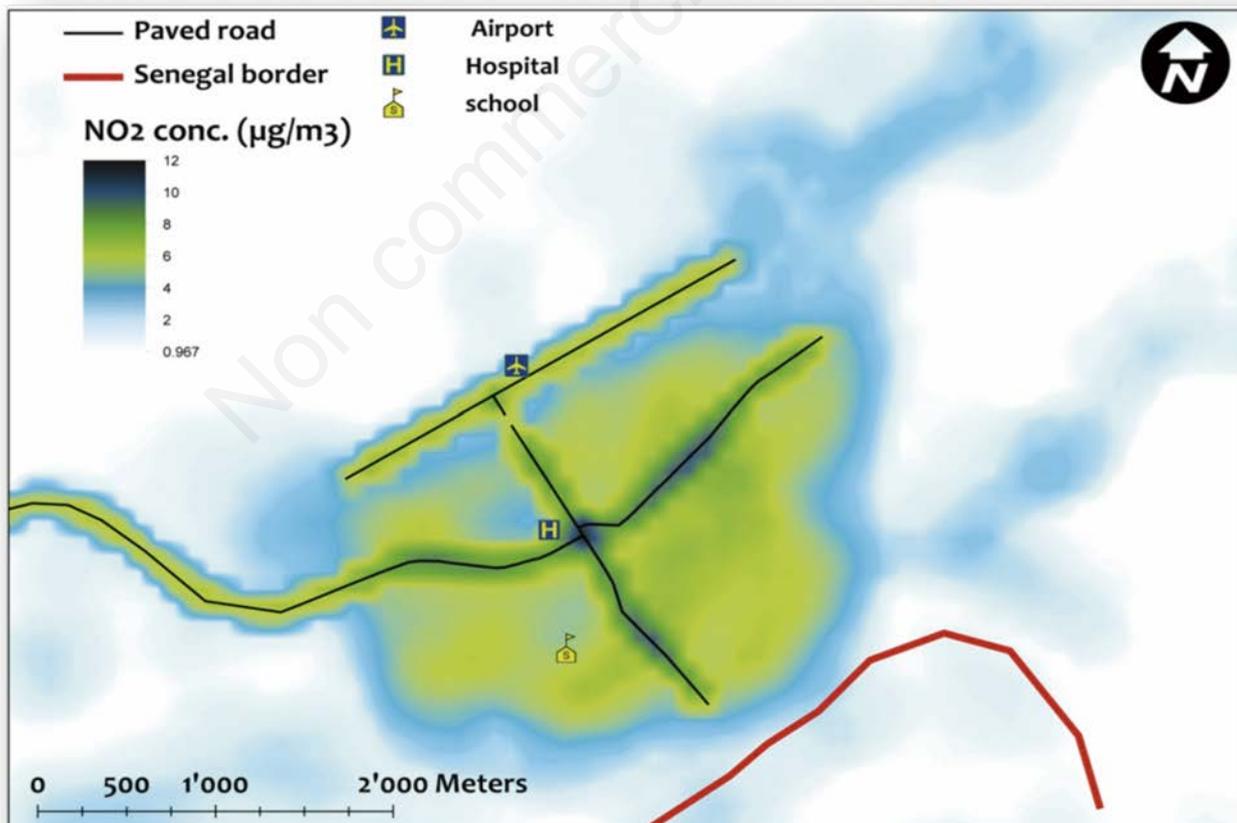


**Figure 4. Spatially smoothed 48-hour NO₂ concentration map for Kaédi (based on predictions at a 100 m pixel).**

## Land use regression modelling and validation

Our LUR model with an adjusted $R^2$ of 0.68 (LOOCV $R^2$ of 0.54) was based on a one-off monitoring campaign of 48hr $NO_2$ measurements during the hot dry season at 40 sites. We were able to select variables that together explained a substantial fraction of the variability of $NO_2$ concentrations in the area. These variables were determined to be plausible and stable in the model (low VIFs). Since the aim was to develop a predictive model, variables were allowed to enter into the model based on their predictive capacity and assessed by the model's explained variability ($R^2$). Even though some important predictor variables like traffic density were not available for Kaédi, 68% of $NO_2$ concentration variability in the town is explained by this model. This is moderately comparable to results found in previous European and North American studies. Variables included in the final model (*length of paved road, length of all types of roads, area of high and low settlement*) are also plausible with respect to the known $NO_2$ pollution pattern (Ryan *et al*., 2007; Hoek *et al*., 2008; Beelen *et al*., 2013).

Due to a limited number of sites, we did not reserve any sites for independent validation. Instead, we internally evaluated the performance and stability of our model using LOOCV (Hoek *et al*., 2008; Eeftens *et al*., 2012). Our cross-validated model has an explanatory power of 54%, which is 14% lower than the original prediction model. In most LUR studies, a model is considered to be stable if the coefficient of variation of the cross-validated model is similar with that of the original prediction model. For example, in the ESCAPE study (Beelen *et al*., 2013), the cross validated $R^2$ for most of the 36 areas was within 10% of the model $R^2$, documenting robustness; however, for a few areas, there was up to 15% difference. The 14% difference we observed in this study could be taken to indicate a moderately stable model. Further support for the performance of our LUR model was also evaluated by performing sensitivity analysis. Using the bootstrap method, we showed that our model parameters were robust to sampling variation. All of the variables except the variable paved road around 100 m buffer (PAVBUF100) had mean bootstrap values close to the model's coefficient values. Further analysis revealed that the skewed/bimodal distribution in the coefficient PAVBUF100 was caused by the influential site. An insignificant Moran's I of -0.03 demonstrated no spatial auto-correlation of the residuals indicating the suitability of the LUR approach. In general, the model's performance was reasonably good and was therefore applied to predict values in unmeasured areas in Kaédi.

The concentration map for Kaédi (Figure 4) has offered a good way to visualise the spatial structure of $NO_2$ pollution over the modelled 48-hour period. The effect of road traffic on the level of $NO_2$ concentration can be seen from the map as concentrations near major paved roads are relatively elevated. Mapping the spatial variability of the pollutant in general is useful to understand population exposures as well as plan relevant air quality measures.

The major challenge in developing this Kaédi LUR model was the lack of GIS data. Although we have tried to overcome this problem by using available online data sources and avoiding expensive and time consuming fieldwork, large data gaps still exist for other predictors, such as, point pollution sources, traffic density and household data. Future studies might consider such limitations and collect relevant data beforehand, for example, during the measurement campaign. We have also only considered 48-hour $NO_2$ measurements in the hot dry season, which leaves open the question of model stability/comparability when it comes to weekly $NO_2$ levels or weekend/weekday differences. Furthermore, additional measurements, *e.g.*, in the rainy season, would be needed to model seasonal patterns and long-term average concentrations.

## Conclusions

To date, LUR models have typically been applied in European and North American cities. African cities often have different urban features and air pollution sources than cities in high-income countries. Given the complexities of air pollution in urban areas of LMICs, and the roles of indoor/outdoor sources, much remains to be done. $NO_2$, a specific marker of traffic air pollution in the outdoor environment, in combination with other pollutants such as $PM_{10}$, $PM_{2.5}$, and ultrafine particles can greatly elucidate priority sources of air pollution in such settings. Information on exposures in these settings is a critical start to understanding and alleviating the considerable burden of air pollution in the developing world.

We have demonstrated that a single short-term saturation campaign of 40 measurements in combination with GIS data derived primarily from on-line and satellite data can provide a pollution map for a mid-size African town where no air pollution data have been previously measured. Moreover, we highlight several approaches to data acquisition and processing, which can be used to support LUR model development in similar low-income areas. If the described LUR approach were adopted (including more pollutants and seasonal differences) by environmental and public health authorities, it would be a cost-effective approach to model the spatial distribution of air pollution, which would inform/support local air pollution policies and guidelines in LMICs.

## References

Abernethy R, Allen R, Ian M, Brauer M, 2013. A land use regression model for ultrafine particles in Vancouver, Canada. Environ Sci Technol 47:17-25.

Allen RA, Gombojav E, Barkhasragchaa B, Byambaa T, Lkhasuren O, Amram O, Takaro KT, Janes C, 2013. An assessment of air pollution and its attributable mortality in Ulaanbaatar, Mongolia. Air Qual Atmos Health 6:137-50.

Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X, Tsai M-Y, Kuenzli N, Schikowski T, Marcon A, Eriksen KT, Raaschou-Nielsen O, Stephanou E, Patelarou E, Lanki T, Yli-Toumi T, Declercq C, Falq G, Stempfelet M, Birk M, Cyrys J, von Klot S, Nador G, Varro MJ, Dedele A, Grazuleviciene R, Moelter A, Lindley S, Madsen C, Cesaroni G, Ranzi A, Badaloni C, Hoffmann B, Nonnemacher M, Kraemer U, Kuhlbusch T, Cirach M, de Nazelle A, Nieuwenhuijsen M, Bellander T, Korek M, Olsson D, Stromgren M, Dons E, Jerrett M, Fischer P, Wang M, Brunekreef B, de Hoogh K, 2013. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe. The ESCAPE Project. Atmos Environ 72:10-23.

Brauer M, Ainslie B, Buzzelli M, Henderson S, Larson T, Marshall J, Nethery E, Steyn D, Su J, 2008. Air pollution modelling and its application. Springer, Amsterdam, The Netherlands.

Brauer M, Markus A, Burnett R, Cohen A, Dentener A, Ezzati M, Henderson S, Krzyzanowski M, Randall M, Dingenen R, Donkelaar A, Thurston R, 2012. Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. Environ Sci Technol 46:652-60.

Briggs D, de Hoogh K, Gulliver J, Wills J, Elliott P, Kingham S, Smallbone K, 2000. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. Sci Total Environ 253:151-67.

Briggs D, Gulliver J, Fecht D, Vienneau D, 2007. Dasymetric modelling of small-area population distribution using land lover and light emissions data. Remote Sens Environ 108:451-66.

Brook R, Franklin B, Cascio W, Hong Y, Howard G, Lipsett M, Luepker R, Mittleman M, Samet J, Smith S, Tager I, 2004. Air pollution and cardiovascular disease a statement for healthcare professionals from the expert panel on population and prevention science of the American Heart Association. Circulation 109:55-71.

Bruce N, Perez-Padilla R, Albalak R, 2000. Indoor air pollution in developing countries: a major environmental and public health challenge. B World Health Organ 78:1078-92.

Brunekreef B, Holgate S, 2002. Air pollution and health. Lancet 360:1233-42.

Cesaroni G, Porta D, Badaloni C, Stafoggia M, Eeftens M, Meliefste K, Forastiere F, 2012. Nitrogen dioxide levels estimated from land use regression models several years apart and association with mortality in a large cohort study. Environ Health 11:48.

Chen B, Kan H, 2008. Air pollution and population health: a global challenge. Environ Health Prev Med 13:94-101.

Chen L, Baili Z, Kong S, Han B, You Y, Ding X, Du S, Liu A, 2010. A land use regression for Predicting NO2 and PM10 concentrations in different seasons in Tianjin region, China. J Environ Sci 22:64-73.

Cissé, G, Koné B, Bâ H, Mbaye I, Koba K, Utzinger J, Tanner M, 2011. Ecohealth and climate change: adaptation to flooding events in riverside secondary cities, West Africa. In: Zimmermann KO, ed. Resilient cities. Springer, Amsterdam, The Netherlands, pp 55-67.

Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, Cirach M, Declercq C, Dėdelė A, Dons E, de Nazelle A, Dimakopoulou K, EriksenK, Falq G, Fischer P, Galassi C, Gražulevičienė R, Heinrich J, Hoffmann B, Jerrett M, Keidel D, Korek M, Lanki T, Lindley S, Madsen C, Mölter A, Nádor G, Nieuwenhuijsen M, Nonnemacher M, Pedeli X, Raaschou-Nielsen O, Patelarou E, Quass U, Ranzi A, Schindler C, Stempfelet M, Stephanou E, Sugiri D, Tsai M, Yli-Tuomi T, Varró M, Vienneau D, von Klot S, Wolf K, Brunekreef B, Hoek G, 2012. Development of land use regression models for PM(2.5), PM(2.5) absorbance, PM(10) and PM(coarse) in 20 European study areas. Results of the ESCAPE Project. Environ Sci Technol 46:195-205.

El-Batrawy O, 2011. Traffic related air pollution in residential environment, Damietta, Egypt. Available from: www.idosi.org/aejaes/jaes11(6)11/23.pdf

Han X, Naeher LP, 2006. A review of traffic-related air pollution exposure assessment studies in the developing world. Environ Int 32:106-20.

HEI International Scientific Oversight Committee, 2010. Outdoor air pollution and health in the developing countries of Asia: a comprehensive review. Health Effects Institute, Boston, MA, USA.

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. Atmos Environ 42:61-78.

Jerrett M, Arain MA, Kanaroglou P, Beckerman B, Crouse D, Gilbert N, Brook J, Finkelstein N, Finkelstein M, 2007. Modelling the intraurban variability of ambient traffic pollution in Toronto, Canada. J Toxicol Env Heal A 70:200-12.

Krzyzanowski M, Cohen A, 2008. Update of WHO air quality guidelines. Air Qual Atmos Health 1:7-13.

Lim S, Vos T, Flaxman A, Danaei G, Shibuya K, Adair-Rohani H, Amann M, Anderson HR, Andrews KG, Aryee M, Atkinson C, Bacchus LJ, Bahalim AN, Balakrishnan K, Balmes J, Barker-Collo S, Baxter A, Bell ML, Blore JD, Blyth F, Bonner C, Borges G, Bourne R, Boussinesq M, Brauer M, Brooks P, Bruce NG, Brunekreef B, Bryan-Hancock C, Bucello C, Buchbinder R, Bull F, Burnett RT, Byers TE, Calabria B, Carapetis J, Carnahan E, Chafe Z, Charlson F, Chen H, Chen JS, Cheng AT, Child JC, Cohen A, Colson KE, Cowie BC, Darby S, Darling S, Davis A, Degenhardt L, Dentener F, Des Jarlais DC, Devries K, Dherani M, Ding EL, Dorsey ER, Driscoll T, Edmond K, Ali SE, Engell RE, Erwin PJ, Fahimi S, Falder G, Farzadfar F, Ferrari A, Finucane MM, Flaxman S, Fowkes FG, Freedman G, Freeman MK, Gakidou E, Ghosh S, Giovannucci E, Gmel G, Graham K, Grainger R, Grant B, Gunnell D, Gutierrez HR, Hall W, Hoek HW, Hogan A, Hosgood HD 3rd, Hoy D, Hu H, Hubbell BJ, Hutchings SJ, Ibeanusi SE, Jacklyn GL, Jasrasaria R, Jonas JB, Kan H, Kanis JA, Kassebaum N, Kawakami N, Khang YH, Khatibzadeh S, Khoo JP, Kok C, Laden F, Lalloo R, Lan Q, Lathlean T, Leasher JL, Leigh J, Li Y, Lin JK, Lipshultz SE, London S, Lozano R, Lu Y, Mak J, Malekzadeh R, Mallinger L, Marcenes W, March L, Marks R, Martin R, McGale P, McGrath J, Mehta S, Mensah GA, Merriman TR, Micha R, Michaud C, Mishra V, Mohd Hanafiah K, Mokdad AA, Morawska L, Mozaffarian D, Murphy T, Naghavi M, Neal B, Nelson PK, Nolla JM, Norman R, Olives C, Omer SB, Orchard J, Osborne R, Ostro B, Page A, Pandey KD, Parry CD, Passmore E, Patra J, Pearce N, Pelizzari PM, Petzold M, Phillips MR, Pope D, Pope CA 3rd, Powles J, Rao M, Razavi H, Rehfuess EA, Rehm JT, Ritz B, Rivara FP, Roberts T, Robinson C, Rodriguez-Portales JA, Romieu I, Room R, Rosenfeld LC, Roy A, Rushton L, Salomon JA, Sampson U, Sanchez-Riera L, Sanman E, Sapkota A, Seedat S, Shi P, Shield K, Shivakoti R, Singh GM, Sleet DA, Smith E, Smith KR, Stapelberg NJ, Steenland K, Stöckl H, Stovner LJ, Straif K, Straney L, Thurston GD, Tran JH, Van Dingenen R, van Donkelaar A, Veerman JL, Vijayakumar L, Weintraub R, Weissman MM, White RA, Whiteford H, Wiersma ST, Wilkinson JD, Williams HC, Williams W, Wilson N, Woolf AD, Yip P, Zielinski JM, Lopez AD, Murray CJ, Ezzati M, AlMazroa MA, Memish ZA, 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the global burden of disease study, 2010. Lancet 380:2224-60.

Liu L-JS, Tsai M-Y, Keidel D, Gemperli A, Ineichen A, Hazenkamp M, Bayer L, Rochat T, Künzli N, Ackermann-Liebrich U, Straehl P, Schwartz J, Schindler C, 2012. Long-term exposure models for traffic related NO2 across geographically diverse areas over separate years. Atmos Environ 46:460-71.

MacIntyre E, Gehring U, Mölter A, Fuertes E, Klümper C, Krämer U, Quass U, Hoffmann B, Gascon M, Brunekreef B, Koppelman G, Beelen R, Hoek G, Birk M, de Jongste J, Smit H, Cyrys J, Gruzieva O, Korek M, Bergström A, Raymond M. Agius, de Vocht F, Simpson A, Porta D, Forastiere F, Badaloni C, Cesaroni G, Esplugues A, Fernández-Somoano A, Lerxundi A, Sunyer J, Cirach M, Nieuwenhuijsen M, Pershagen G, Heinrich J, 2013. Air pollution and respiratory infections during early childhood: an analysis of 10 European birth cohorts within the ESCAPE project. Environ Health Persp 122:107-22

Manandhar R, Inakwu O, Ancev T, 2009. Improving the accuracy of land use and land cover classification of landsat data using post-classification enhancement. Remote Sensing 1:330-44.

Mao L, Qiu Y, Kusano C, Xu X, 2012. Predicting regional space-time variation of PM2.5 with land-use regression Model and MODIS data. Environ Sci Pollut R 19:128-38.

Maxwell SK, 2010. Generating land cover boundaries from remotely sensed data using object-based image analysis: overview and epidemiological application. Spatial and Spatiotemporal Epidemiol 1:131-7.

Mehta S, 2003. Air pollution and health in rapidly developing countries. B World Health Organ 81:771.

Mennis J, 2003. Generating surface models of population using dasymetric mapping. Prof Geogr 55:31-42.

O'Brien RM, 2007. A caution regarding rules of thumb for variance inflation factors. Qual Quant 41:673-90.

ONS, 2013. Recensement général de la population et de l'habitat (RGPH2013). Available from: http://www.ons.mr/images/doc/publication/2014/brochure%20rgph_final%204%20aot%202014.pdf

Raaschou-Nielsen O, Andersen ZJ, Beelen R, Samoli E, Stafoggia M, Weinmayr G, Hoffmann B, Fischer P, Nieuwenhuijsen MJ, Brunekreef B, Xun W, Katsouyanni K, Dimakopoulou K, Sommar J, Forsberg B, Modig L, Oudin A, Oftedal B, Schwarze PE, Nafstad P, Faire UD, Pedersen NL, Östenson C-G, Fratiglioni L, Penell J, Korek M, Pershagen G, Eriksen KT, Sørensen M, Tjønneland A, Ellermann T, Eeftens M, Peeters PH, Meliefste K, Wang M, de Mesquita BB, Key TJ, de Hoogh K, Concin H, Nagel G, Vilier A, Grioni S, Krogh V, Tsai M-Y, Ricceri F, Sacerdote C, Galassi C, Migliore E, Ranzi A, Cesaroni G, Badaloni C, Forastiere F, Tamayo I, Amiano P, Dorronsoro M, Trichopoulou A, Bamia C, Vineis P, Hoek G, 2013. Air pollution and lung cancer incidence in 17 european cohorts: Prospective analyses from the European study of cohorts for air pollution effects (ESCAPE). Lancet Oncol 14:813-22.

Richards JA, Xiuping J, 1999. Remote sensing digital image analysis: an introduction. Springer-Verlag, Berlin, Germany.

Ross Z, English B, Scalf R, Gunier R, Smorodinsky S, Wall S, Jerrett M, 2005. Nitrogen dioxide prediction in southern California using land use regression modelling: potential for environmental health analyses. J Expo Sci Env Epid 16:106-14.

Ryan P, Le Masters G, 2007. A review of land-use regression models for characterizing intraurban air pollution exposure. Inhal Toxicol 19:127-33.

Ryan P, Le Masters G, Biswas P, Levin L, Hu S, Lindsey M, Bernstein D, Lockey J, Villareal M, Hershey G, Grinshpun S, 2007. A comparison of proximity and land use regression traffic exposure models and wheezing in infants. Environ Health Persp 115:278-84.

Saraswat A, Apte J, Kandlikar M, Brauer M, Henderson S, Marshall J, 2013. Spatiotemporal land use regression models of fine, ultrafine, and black carbon particulate matter in New Delhi, India. Environ Sci Technol 47:12903-11.

Shmool JLC, Michanowicz D, Cambal L, Tunno B, Howell J, Gillooly S, Roper C, Tripathy S, Chubb L, Eisl H, Gorczynski J, Holguin F, Naumoff Shields K, Clougherty J, 2014. Saturation sampling for spatial variation in multiple air pollutants across an inversion-prone metropolitan area of complex terrain. Environ Health 13:28.

Touray S, Bâ H, Bâ O, Koïta M, Ould Ahmed Salem BC, Keïta M, Traoré D, Sy I, Winkler M, Utzinger J, Cissé G, 2012. Absence of dry season Plasmodium parasitaemia, but high rates of reported acute respiratory infection and diarrhoea in preschool-aged children in Kaédi, southern Mauritania. Parasite Vector 5:193.

WHO, 2014. Burden of disease from ambient and household air pollution. Available from: http://www.who.int/phe/health_topics/outdoorair/databases/en/

Zou B, Wilson JG, Zhan FB, Zeng Y, 2009. Air pollution exposure assessment methods utilized in epidemiological studies. J Environ Monitor 11:475-90.