

2018

A Land Use Regression Model for Explaining Spatial Variation in Air Pollution Levels using a Wind Sector Based Approach

Owen Naughton

Trinity College Dublin, Ireland

Aoife Donnelly

Technological University Dublin, aoife.donnelly@tudublin.ie

Paul Nolan

Irish centre for high end computing (ICHEC)

See next page for additional authors

Follow this and additional works at: <https://arrow.tudublin.ie/schfsehart>



Part of the [Applied Statistics Commons](#), [Environmental Health and Protection Commons](#), [Environmental Indicators and Impact Assessment Commons](#), [Environmental Monitoring Commons](#), and the [Other Environmental Sciences Commons](#)

Recommended Citation

Naughton, O. et al. (2018) A Land Use Regression Model for Explaining Spatial Variation in Air Pollution Levels using a Wind Sector Based Approach, *Science of The Total Environment*, Vol. 630, 15 July 2018, 1324–1334pp. doi.org/10.1016/j.scitotenv.2018.02.317

This Article is brought to you for free and open access by the School of Food Science and Environmental Health at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, vera.kilshaw@tudublin.ie.

Funder: Environmental Protection Agency

Authors

Owen Naughton, Aoife Donnelly, Paul Nolan, Francesco Pilla, Bruce Misstear, and Brian Broderick

A land use regression model for explaining spatial variation in air pollution levels using a wind sector based approach

Naughton, O.¹, Donnelly, A.², Nolan, P.^{3,4}, Pilla, F.⁵, Misstear, B.D.¹, Broderick, B.¹

¹Department of Civil, Structural and Environmental Engineering, University of Dublin Trinity College, Dublin 2, Ireland

²School of Food Science and Environmental Health, Dublin Institute of Technology, Cathal Brugha Street, Dublin.

^{3,2}Irish Centre for High-End Computing (ICHEC), National University of Ireland Galway, Ireland

⁴Met Éireann, Research and Applications Division, Dublin, Ireland.

⁵School of Architecture, Planning and Environmental Policy, University College Dublin

Corresponding Author: Owen Naughton

Email: naughto@tcd.ie, *Telephone:* +353 1 896 2781, *Fax:* +353 1 677 3072

Abstract

Estimating pollutant concentrations at a local and regional scale is essential for good ambient air quality information in environmental and health policy decision making. Here we present a land use regression (LUR) modelling methodology that exploits the high temporal resolution of fixed-site monitoring (FSM) to produce viable air quality maps. The methodology partitions concentration time series from a national FSM network into wind-dependent sectors or “wedges”. A LUR model is derived using predictor variables calculated within the directional wind sectors, and compared against the long-term average concentrations within each sector. This study demonstrates the value of incorporating the relative position of emission source and receptor into the empirical LUR model structure. In our specific application, a model based on 15 FSM training sites captured 78% of the spatial variability in NO₂ across the Republic of Ireland. This compares favourably to traditional LUR models based on purpose-designed monitoring campaigns despite using approximately half the number of monitoring points in model development. We applied the LUR equation at a high-resolution across the Republic of Ireland to enable applications such as the study of environmental exposure and human health, assessing representativeness of air quality monitoring networks and informing environmental management and policy makers.

1. Introduction

Public exposure to poor ambient air quality can have a damaging effect on people’s health and development, contributing to a significant global burden of respiratory and allergic diseases and increased mortality rates (EEA, 2012; Guerreiro et al., 2014; Hettelingh et al., 2013; Kunzli et al., 2000). Epidemiological studies have demonstrated the increased risk of negative health effects and natural-

cause mortality due to exposure to ambient pollutants such as nitrogen dioxide (NO₂), ozone (O₃) and fine particulate matter (PM) (Beelen et al., 2014; Laumbach and Kipen, 2012; Pascal et al., 2013; Raaschou-Nielsen et al., 2013). Pollutant emissions can also impact on sensitive ecosystems by contributing to acidification and eutrophication (Guerreiro et al., 2014), with nitrogen oxides (NO_x), sulphur dioxide (SO₂) and ammonia (NH₃) representing the main acidifying compounds causing excess nitrogen deposition. These pollutants also play a major role in the formation of ground-level O₃, which can negatively affect human health, ecological status and agricultural productivity (Galloway et al., 2004; Staelens et al., 2012). The critical need for good ambient air quality information in environmental and health policy decision making led to the introduction by the European Union (EU) of the Ambient Air Quality and Cleaner Air for Europe (CAFE) Directive (2008/50/EC). In addition to requiring compliance with limit values for priority pollutants set out in the CAFE Directive, EU Member States are also obliged to provide current air quality information to the public and encouraged to interpret pollution levels in terms of geographical variation and collective population exposure.

The field of deterministic modelling to predict local, regional and trans-boundary transport of air pollutants on a national and international scale is well established (Carmichael et al., 2008; Levitin et al., 2005; Savage et al., 2013; Solazzo et al., 2012). However, the technical and financial resources required to implement such models are often beyond the reach of many regulatory authorities, policy-makers and health researchers and their effectiveness is based on the availability of suitable input data such as an emissions inventory. Instead, less resource-intensive stochastic models, geo-statistical interpolation and land use regression have been employed successfully in recent years for the production of high-resolution air quality maps and forecasts (Beelen et al., 2013; Briggs et al., 1997; Denby et al., 2010; Janssen et al., 2008; Stedman et al., 1997). These maps provide a valuable source of information for policy making, evaluating spatial representativeness of monitoring stations and providing accurate exposure assessments for epidemiological studies (Denby et al., 2010; Janssen et al., 2012; Jerrett et al., 2004).

One modelling method, Land use regression (LUR), has grown in popularity in recent years (Beelen et al., 2013). LUR compares favourably with alternatives of geo-statistical interpolation and dispersion modelling, typically producing equivalent or better results in a relatively simple, cost-effective manner (Hoek et al., 2008; Vienneau et al., 2010). LUR uses spatial variables, such as traffic or land use, to predict pollutant concentrations at a given location (Hoek et al., 2008). Typically, variables used for calibrating LUR models are extracted using circular buffers drawn around measurement sites. However, this ignores the spatial orientation of source and receptor and the influence of meteorology; the pollutant concentration at a monitoring site can vary significantly depending on the prevailing wind direction and the relative position of local and regional emission sources (Donnelly et al., 2012; Donnelly et al., 2011).

Whilst previous LUR studies (Arain et al., 2007; Bertazzon et al., 2015; Chen et al., 2012; Kim and Guldmann, 2011; Li et al., 2015; Su et al., 2008) have demonstrated the value of incorporating the direction of emission sources into the empirical LUR model structure, this paper describes novel geospatial and statistical techniques used for the development of a national scale model for the key pollutant NO₂. Two meteorological variables, wind speed and wind direction, were incorporated into the LUR modelling process. The LUR model has been spatially refined by calculating predictor variables within wind-dependent sectors or “wedges”, and comparing against long-term average pollutant concentrations from fixed-site monitoring stations (FSMs) within each sector. Wind speed within each

sector was also included as a temporally-variable predictor in the regression equation. Following model fitting, the LUR equation was applied at a high-resolution across the Republic of Ireland (hereafter referred to as Ireland). The model itself developed as part of this research is unique to Ireland but the methodology could easily be applied elsewhere to develop unique site-specific air quality maps once a national air quality monitoring network exists together with records of wind speed and direction.

2. Materials and Methods

The basis of LUR mapping is a multiple linear regression that uses summaries of spatial variables surrounding the monitoring point to predict air pollutant concentrations. Typically, spatial indicators are calculated within circular buffers of varying radii around the monitoring point and the most significant used as predictor variables in the regression equation. However, circular buffers effectively apply equal weights to emission sources around a receptor, irrespective of the prevailing meteorological conditions and the relative positions of receptor and source. This limitation may be minor when LUR is applied at a coarse spatial and temporal resolution, but when developed at finer resolution and on a national scale as in this study, varying regional wind patterns may lead to poor model performance.

As regional prevailing wind conditions can vary substantially and thus impact the applicability of ordinary LUR techniques on a national scale, a novel LUR methodology was devised which incorporates wind effects using angular sectors or “wedges”. The 360° wind field is discretised into a set of eight 45° wind sectors; average pollutant concentrations and predictor variables are then calculated for each sector and used in the LUR process. The use of continuous data from fixed-site monitoring stations (FSMs), rather than short-term passive monitoring, allows the calculation of average concentrations within each wind sector. However, as prevailing wind directions vary geographically, seasonally and diurnally, a biased sectoral average may be obtained in some instances. For example, if a particular wind direction was more frequent during winter than summer months the raw sector average would be excessively high. Consequently, a non-parametric regression correction method has been applied to remove diurnal and seasonal bias from the data prior to sector averaging (Donnelly et al., 2016). There are four key steps involved in the wind sector land use regression (WS-LUR) model development and mapping process:

1. Calculate annual average pollutant concentrations within each wind sector at each monitoring site using a combination of hourly meteorological inputs and continuous monitoring data.
2. Generate predictor variables from geospatial datasets for each directional sector within a GIS environment.
3. Select predictor variables for LUR equation using a supervised stepwise approach.
4. Apply regression equation on a national scale and weight using interpolated local wind frequency.

2.1 Study Area

This study was undertaken in Ireland (Fig. 1), a country covering an area of just over 70,000km² and containing 4.6 million inhabitants. Pollutant levels of nitrogen dioxide (NO₂) are of concern in the Irish context (O'Dwyer, 2016). Significant reductions in national NO_x emissions have been achieved in recent decades due to improvements in the power generation sector as older peat- and coal-burning

plants were phased out or converted to cleaner fuel alternatives. However, an over-reliance on vehicle transport means that national NO_x emissions from transport have remained high; 45% of national NO_x emissions are derived from road transport (Duffy et al., 2015). This has resulted in a failure to achieve NEC emissions ceiling targets in 2010, as well as causing NO₂ levels close to the specified EU limit values for air quality in traffic-impacted areas (Duffy et al., 2015). The lack of a suitably detailed national scale emissions inventory coupled with appropriate spatial estimates of background pollution concentrations in Ireland has seriously impeded air quality modelling applications, limited health exposure assessments and restricted evaluation of the abatement potential of air quality management plans.

2.2 Measured NO₂ Concentration Data

NO₂ monitoring data were obtained from the national ambient air quality network operated by the Environmental Protection Agency (EPA), the designated Competent Authority for the monitoring of air quality in accordance with all Irish and EU ambient air quality legislation. The network comprises 12 active fixed-site monitoring stations (FSMs) measuring contaminant concentrations on an hourly basis. Additional FSMs decommissioned from the network post-2012 were included in the analysis; these provided a further three NO₂ sites bringing the total to 15 NO₂ sites over the analysis period of 2010 to 2012 inclusive. FSM locations are shown in Figure 1 (left). All sensors conformed to EU guidelines on background concentration monitoring and compliance with limit values (CEU, 2008). NO₂ was measured on an hourly basis using chemiluminescence samplers (API M200 NO_x analyser). Average concentrations at all monitoring sites were within the detection ranges of the respective instrumentation.

Wind direction and speed are key parameters in the WS-LUR model development; the former is used to calculate average pollutant concentrations while the latter is included as a predictor variable within the regression analysis. Hourly wind speed and wind direction data for each site were obtained from the national network of synoptic stations operated by the Irish Meteorological Service (Met Eireann). While the synoptic stations and FSMs are not always co-located, spatial variations in wind speed and direction in Ireland tend to be low and each FSM and synoptic site pair were investigated for major changes in geographic topography to ensure sites were paired appropriately. Furthermore, national FSMs were intentionally positioned in areas that would be representative of background conditions; sites affected by major local emission sources (e.g. directly adjacent to a highly-trafficked road), or that experienced complex local wind effects (e.g. street canyons), were avoided.

Continuous monitoring data were used to calculate wind sector means for each FSM. The division of a concentration time series at a point into eight sectors maximises the number of data points available for the LUR; however, it also reduces data points available for long-term mean value calculation. Diurnal and seasonal concentration variations may lead to a biased annual sector average estimate when calculated from sub-annual datasets. Concentrations tend to be higher in winter months than summer months (in Ireland) so, for example, if data within a sector were comprised 20% from winter and 80% from summer months, an unrealistically low value for the annual average would be obtained. Consequently, a non-parametric regression correction method was applied to remove bias from the concentration data prior to sector averaging. A brief overview of the correction method is provided here; a comprehensive explanation is provided in Donnelly et al. (2015).

The basis for the correction factors is a non-parametric kernel regression model which is applied to describe diurnal and seasonal variations in concentration levels. The outputs from the regression are a seasonal factor (S_f) and a diurnal factor (D_f). The normalised seasonal factors are defined as follows:

$$S_f = \left(\frac{\tilde{C}(\alpha, h)}{\bar{C}} \right)$$

where \bar{C} is the average concentration for the input data used in model development and $\tilde{C}(\alpha, h)$ is the average concentrations of a pollutant for a given day of the year (α) calculated as a weighted average of the data in a window (of width defined by smoothing parameter h) using weighted Gaussian kernel function $K_1(\alpha, h)$ around (α) and defined as follows:

$$\tilde{C}(\alpha, h) = \frac{\sum_{i=1}^N K_1\left(\frac{(\alpha - S_i)}{h}\right) C_i}{\sum_{i=1}^N K_1\left(\frac{(\alpha - S_i)}{h}\right)}$$

where C_i are de-trended concentrations, S_i is the day of the year for the i^{th} observation in a time period starting at time t_i . For circular data the Gaussian kernel (K) is the preferred method used to weight the observations (Henry et al., 2002) and is defined as follows:

$$K(x) = (2\pi)^{-1/2} \exp(-0.5x^2) \quad -\infty < x < \infty$$

The bandwidth is calculated based on the number of days in a year. As discussed in Silverman (1986) a bandwidth of $0.9\sigma n^{-1/5}$ was employed, where σ is the standard deviation of the predictor variable data (in this instance day of the year) and n is the number of data points.

In developing the D_f the data are first subdivided into four categories distinguishing between winter and summer, and between weekdays and weekends. The resulting factors are developed in exactly the same way as S_f but in this instance hours are used in replacement of days (i.e. S_i is replaced by H_i where H_i is the hour of the day).

To adjust the data, raw hourly or daily concentration values ($NO_{2(raw)}$) are firstly divided by the relevant seasonal factor (S_f) to obtain a seasonally adjusted value ($NO_{2(s)}$):

$$NO_{2(s)} = \frac{NO_{2(raw)}}{S_f} \quad (\text{Eqn. 1})$$

D_f values are determined for each season separately and do not therefore, account for the seasonal variation. To obtain the diurnally and seasonally adjusted concentration $NO_{2(s,d)}$, therefore, the seasonally adjusted concentration can be divided by the normalised D_f .

$$NO_{2(s,d)} = \frac{NO_{2(s)}}{D_f} \quad (\text{Eqn. 2})$$

Applying the factors in this way does not change the mean of the total data set. The correction factors do, however, have the potential to change the mean values within a given sector and remove any bias which has arisen due to uneven distribution of data across sectors. Following seasonal and diurnal

adjustment the hourly data were divided and averaged to provide an annual mean concentration for each wind sector.

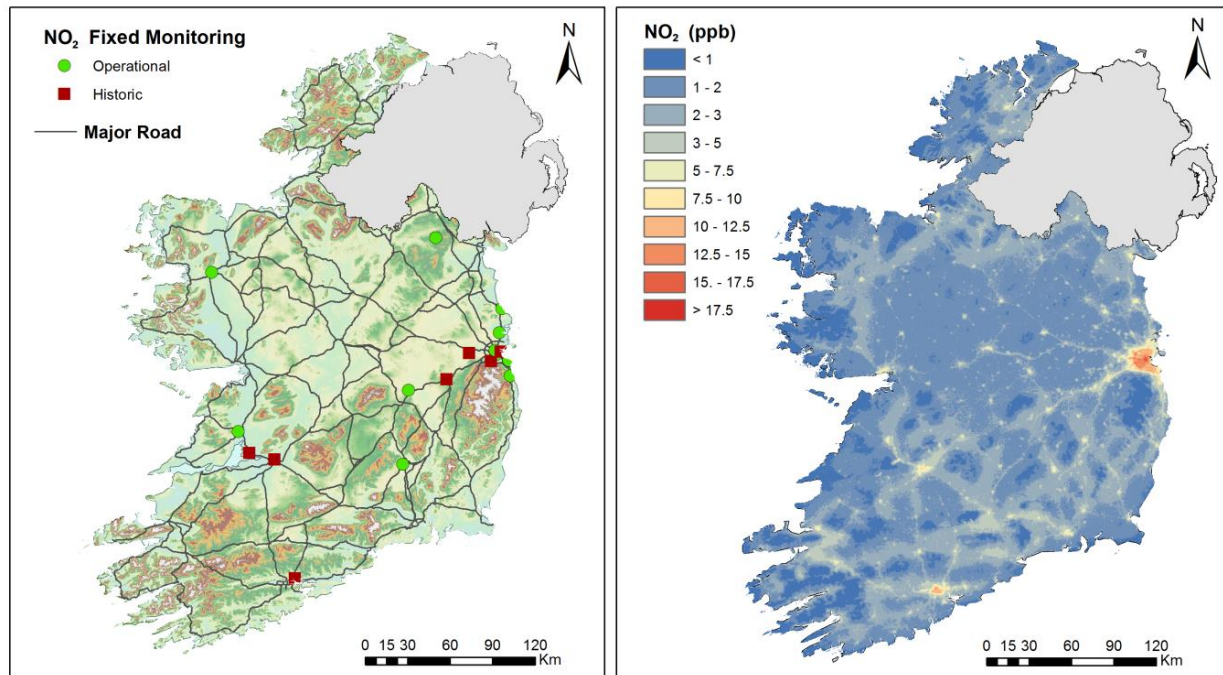


Figure 1: (left) Fixed site monitoring station locations and (right) national map of average annual NO₂ concentrations for 2010 – 2012 (resolution 50m)

2.3 Spatial Predictors

Geospatial predictor variables (Table 1) were calculated within each sector from nationally available, spatially homogeneous datasets for all sites using the ArcGIS 10.0 software package (ESRI, 2011). Eight circular buffers of variable radii were defined around each monitoring site, ranging from 25 metres to 5 kilometres, and further subdivided into eight 45° wind direction sectors (e.g. N, NE, E, etc.) (Figure 2 (a)).

Table 1: Predictor variables with variable names, units, and sector size

Category	Units	Sector Radius (m)	Subcategory	Variable Name	No. of Variables
Traffic Variables					
Road Length	Km	25, 50, 100, 250, 500, 1000, 2000, 5000	All roads National road Regional road Local road Major road	ALLRDLEN NRDLEN RPDLEN LRDLEN MRDLEN	56
Proximity to Road	Km ⁻¹ , Km ⁻²	N/A	Nearest road Nearest major road	DistARdInv DistARinv2 DistMRinv DistMRinv2	8
Traffic Flow	Vehicle Km	25, 50, 100, 250, 500, 1000, 2000, 5000	N/A	Vkm_(radius)	8
Weighted Traffic Flow	Vehicle Km	N/A	Inverse distance Gaussian	Vkmi _{invw} Vkm _{gw}	2
Background Variables					
Land Cover	Hectares	25, 50, 100, 250, 500, 1000, 2000, 5000	High density residential Low density residential Industry Port Urban green Semi-natural and forested Natural Sea/Ocean	HDRES LDRES INDUSTRY PORT URBGREEN SEMNNATURAL NATURAL SEA	64
Population Density	Persons/km ²	25, 50, 100, 250, 500, 1000, 2000, 5000	N/A	Pdens	8
Property Density	No. properties	25, 50, 100, 250, 500, 1000, 2000, 5000	Residential Commercial	Geo_Res Geo_Com	16
Residential Heating	Properties per heating type	25, 50, 100, 250, 500, 1000, 2000, 5000	Solid Gas Electricity Oil	CSO_SCH CSO_GCH CSO_ECH CSO_OCH	32
Household Cars	Cars	25, 50, 100, 250, 500, 1000, 2000, 5000	N/A	CSO_Cars	8
Proximity to Coast	Km	N/A	N/A	Coast	1
Point Source (PRTR)	Kg	25, 50, 100, 250, 500, 1000, 2000, 5000	N/A	PS_(radius)	8
Elevation	m	N/A	N/A	ELEV	1
Wind Speed	m/s	N/A	N/A	WS	1

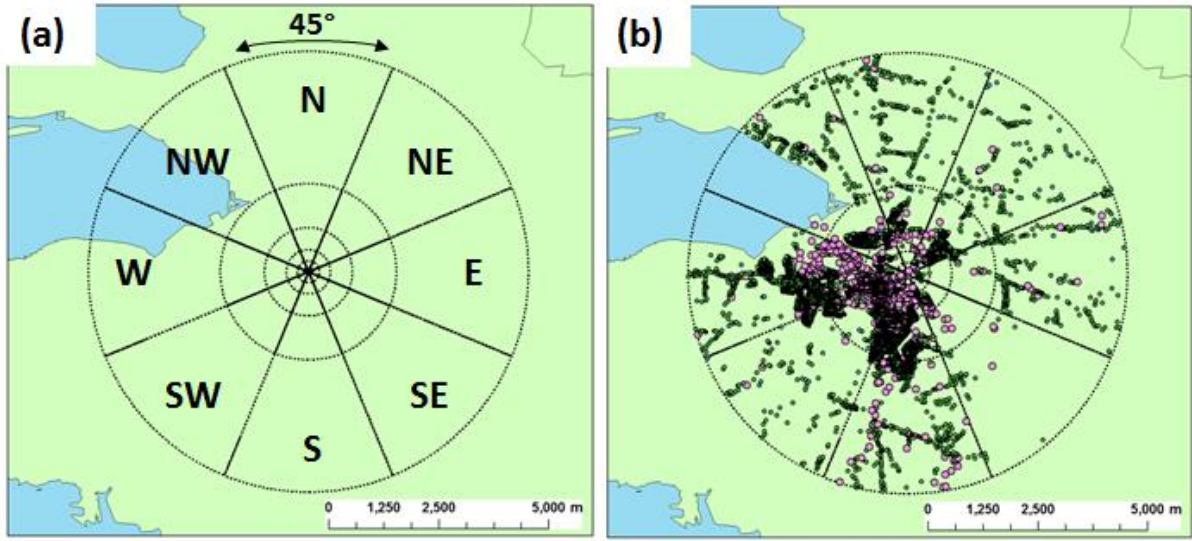


Figure 2: (a) Wind direction sectors, (b) residential and commercial properties

Residential and commercial property variables were derived using geographical coordinates from GeoDirectory, the national property register maintained by An Post and Ordnance Survey Ireland (OSI) (figure 2(b)). Traffic network and flow data were obtained from the National Traffic Model (NTM) developed by Transport Infrastructure Ireland (TII). The NTM provides statistics on traffic volumes and composition for all motorways, and national primary, secondary and regional roads in Ireland. Road length variables were calculated within each sector for each road category, and the length of each major road link passing through the sector was multiplied by the link Annual Average Daily Traffic (AADT) to give annual Vehicle km (Vkm). Due to the high correlation between NO_2 concentration and traffic parameters across the range of buffer radii, a weighted Vkm (Vkm_{invw}) parameter was developed. The weighting applied to each sector is related to the inverse of the distance of the sector from the monitoring point, with the closest sector (i.e. 25 m) carrying the highest weighting. The inverse-distance weighted Vkm factor was calculated as:

$$Vkm_{invw} = \frac{1}{r_0} \times Vkm_0 + \sum_{i=1}^N \frac{1}{r_i} \times (Vkm_i - Vkm_{i-1}) \quad (\text{Eqn. 3})$$

where $i = 0$ to $i = N$ represent each of the sectors considered, r is the distance from the monitoring point to the centre of a given sector, and d_i is the road density in each sector. In this study eight sector sizes are considered, 25 m, 50 m, 100 m, 250 m, 500 m, 1 km, 2 km and 5 km. A Gaussian weighted road density factor (G_i) was also developed for each monitoring site sector. The weighting applied to each sector is related to the distance of the sector from the monitoring point. The closest sector (i.e. 25 m) carries the highest weighting.

$$Vkm_{gw} = G_0 \times Vkm_0 + \sum_{i=1}^N G_i \times (Vkm_i - Vkm_{i-1}) \quad (\text{Eqn. 4})$$

where:

$$G_i = \frac{1}{\sqrt{2\pi}\sigma} \times e^{-(x-\mu)^2/2\sigma^2} \quad (\text{Eqn. 5})$$

Here, σ is the standard deviation or spread of the distribution (optimised to provide the highest correlation coefficient), and x is the distance from the monitoring site to the centre of the sector under consideration.

Population and residential combustion data were derived from Census data and spatially disaggregated based on residential property locations, whereby average household statistics were calculated within each Census Small Area (SA) using the total number of occupied residential properties within the SA. Sector variables were then calculated by summing the total of households contained within each sector. Large point source pollutant emissions were derived from the Pollutant Release and Transfer Register (PRTR), operated by the EPA. Point emission totals were assigned to each sector based on PRTR point locations and the annual emissions during the year (or years) for which monitoring data were available in the sector. Land cover variables were derived from CORINE (COoRdination of INformation on the Environment) land cover data for the year 2006 the European Environment Agency (EEA). Following the methods outlined in Vienneau et al. (2010) and Beelen et al. (2013) the 44 land cover classes in CORINE were regrouped into six (High density residential, low density residential, industry, port, urban green, and semi-natural and forested areas) as well as additional land use class representing areas of sea and open ocean.

2.4 Model Calibration and Validation

The LUR model is a multiple linear regression of the form:

$$C = C_0 + \alpha_1 X_1 + \alpha_2 X_2 \dots + \alpha_n X_n + \epsilon \quad (\text{Eqn. 6})$$

where C = Average annual NO_2 concentration; X_i = predictor variable i ; α_i = coefficient for predictor variable i and ϵ = error. Selection of the most appropriate explanatory variables within suitable sector sizes is important for defining final model performance. Variable selection was carried out using a supervised stepwise approach. Firstly, each predictor variable was assigned a plausible direction of effect and univariate regression analyses were carried out for all predictor variables. The model with the highest adjusted R^2 having an appropriate slope as predefined by the direction of effect was considered as the start model. Additional predictor variables were then added consecutively to the model and maintained if the following three conditions were met:

1. The R^2 value increased by at least 1%
2. The direction of effect of the new variable was as *a priori* defined
3. The direction of effect of previously included variables did not change

The large number of predictor variables examined meant that many of them were correlated. The variance inflation factor (VIF) was used to assess how much the variance of an estimated regression coefficient increases if predictors are correlated; it is equal to 1 if no factors are correlated. Variables with high VIF were removed from the model ensuring that each variable removed was redundant in the explanation of concentration. The set of predictor variables giving the highest adjusted R^2 value which conformed to *a priori* defined directions of effect were selected for inclusion in the final model. As a final step, variables with a p-value of greater than 0.05 were removed from the model.

Standard diagnostic tests for ordinary least squares regression were carried out. These included assessing residuals for heteroscedasticity and normality. Residuals were also analysed for influential or controlling observations or outlying data points. In a small number of instances this led to removal of certain data points after detailed investigations of the baseline data. The full regression procedure as described above was repeated until residual analysis was satisfactory.

The limited number of FSMs available made it unfeasible to hold aside sufficient sites for model validation. Instead, following the approach taken by Beelen et al. (2013) a leave one out cross validation method was used. The final model was fitted to N-1 sites and the predicted concentration compared to the actual concentration at the omitted site. This was repeated for all N sites and the overall level of fit between the predicted and measured concentration was assessed. The correlation coefficient for the cross validated data is known as the CV R².

2.5 Regression Mapping

The LUR equation was applied on a gridded basis for the entire land area of Ireland. All grid calculations were carried out within the ArcGIS 10.1 environment (ESRI, 2011). At each grid point eight raster layers representing the eight wind sectors were derived for each regression variable. A schematic diagram of the gridded wedge calculation for the north and north-west sectors is shown in Figure 3, where R is the critical sector radius as identified during the model fitting process. Each predictor variable raster was multiplied by the relevant regression coefficient and summed to give eight wind-dependent background concentrations for each grid point. The overall background concentration at each point was then calculated as the weighted sum of sector averages, weighted by the wind frequency in each sector. Thus, the ambient wind-dependent background concentration is given by:

$$C = \alpha_0 + \sum_{i=0}^8 \sum_{j=1}^M Wf_i \alpha_j P_j \quad (\text{Eqn. 7})$$

where $i = 0$ to $i = 8$ represent each of the eight wind-directional sectors, α_0 is the regression constant, $j = 1$ to $j = M$ represent each of the terms in the regression equation, Wf_i is the fraction of the reference period where the prevailing wind is within wind sector i , α_j is the regression coefficient and P is the predictor variable value in direction i . If the regression equation produced negative concentrations, cells were set to zero.

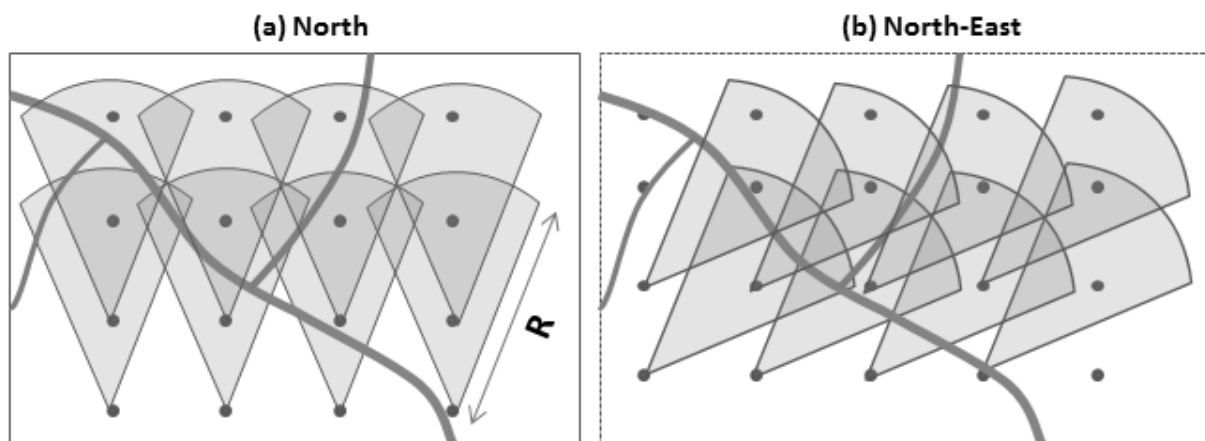


Figure 3: Schematic diagram of sector road length calculation for (a) north and (b) north-east sectors

The gridding of predictor variables within wind directional wedges was implemented at a 50 m spatial resolution using the Neighbourhood analysis tools in ArcGIS 10.1. The Focal Statistics and Point Statistics tools allow the calculation of spatial statistics within wedge-shaped neighbourhoods, wherein the start and end angles of each wedge are defined by the limits of each 45° wind direction bin and the search radius is determined by the predictor variable sector radius. Line and area predictor variables were gridded using the Focal Statistics tool while point-based features were gridded using the Point Statistics tool. Predictor variables derived from line vectors, such as road density, required the transformation of the vector layers into high resolution raster layers first, where after the Focal Statistics tool was used. Road length and Vkm were initially rasterised at a 5 m resolution to retain maximum detail from the road network, before being aggregated to 25 m resolution to reduce processing time and file storage requirements.

As wind speed and direction are monitored at a limited number of locations nationally, it was necessary to derive these variables from a high-resolution historical simulation of the Irish climate. The climate of Ireland was simulated by dynamically downscaling the ECMWF ERAInterim reanalysis dataset (Dee et al., 2011) for the period 1981-Present using the COSMO-CLM5 model with maximum grid-spacing of 1.5 km (Nolan and Baehler, 2016). COSMO-CLM5 is a non-hydrostatic and fully compressible regional climate model, based on the numerical weather prediction model COSMO (Baldauf et al., 2011). A more detailed description of the model and its parameterization schemes can be found in Rockel et al. (2008), Vautard et al. (2013) and Kotlarski et al. (2014). The data were archived at one-hour intervals. The benefit of such datasets is that they provide the best estimate of the four-dimensional atmospheric state and parameters, which are not routinely provided on a homogeneous grid (e.g., wind speed and direction).

Local estimates of mean wind speed and frequency of occurrence were downscaled using bilinear interpolation from a 1.5km grid of each of the eight 45° wind direction bins over the 3-year period from 2010 to 2012. Mean wind speed was used as a predictor variable in each of the eight wind direction LUR equations. Wind frequency was then used to weight these eight direction values to calculate overall mean NO₂ concentrations across the country.

3. Results

3.1 Impact of wind direction on NO₂

The impact of wind direction on NO₂ concentrations was first explored to validate the underlying approach. Overall mean (St. Dev.) concentrations were 11.7 (3.2) ppb, 6.4 (2.4) ppb and 3.0 (1.6) ppb for urban, suburban and rural sites respectively (Table 2). The range of NO₂ concentrations, as represented by the ratio of (maximum – minimum) / mean, was 187% which shows the relatively high variability between monitoring sites. Pollutant concentrations showed significant asymmetry depending on wind conditions across all location types (Figure 4), with differences between the highest and lowest sector NO₂ concentrations ranging from 1.4 to 9.5 ppb. An ANOVA test was used to test sector concentrations for consistency across the FSM network, to ensure sites were not overly influenced by winds of one direction or a large-scale systematic variation due to other external factors (e.g. transboundary pollution). No statistically significant trend was found, supporting the hypothesis that local land use characteristics were causing the spatial variations in NO₂.

Table 2: Summary statistics of observed sectoral NO₂ values

ID	Site Name	Coverage	Sector NO ₂ Concentration (ppb)				
			Max	Min	Range	StDev	Mean
U1	Coleraine Street	Urban	18.2	10.6	7.6	2.7	13.4
U2	Dun Laoghaire	Urban	12.9	7.6	5.3	2.0	8.5
U3	Knocklyon	Urban	13.4	6.8	6.6	2.3	12.9
U4	Ringsend	Urban	13.3	7.4	5.9	1.9	8.6
U5	Winetavern Street	Urban	19.8	13.6	6.1	2.0	15.9
SU6	Balbriggan	Suburban	9.1	4.4	4.7	1.4	6.0
SU7	Castlebar	Suburban	6.2	3.5	2.7	0.9	4.4
SU8	Limerick (Park Road)	Suburban	15.1	7.6	7.5	2.6	7.1
SU9	Rathmines	Suburban	16.7	7.5	9.3	3.4	10.4
SU10	Kilkenny (Seville Lodge)	Suburban	4.1	2.0	2.1	0.9	2.7
SU11	Shannon Town	Suburban	7.6	6.0	1.7	0.6	6.7
SU12	Swords	Suburban	11.4	5.2	6.2	2.3	7.6
R13	Emo Court	Rural	3.0	1.6	1.4	0.5	1.8
R14	Glashaboy	Rural	6.7	3.9	2.8	1.1	4.8
R15	Kilkitt	Rural	3.7	1.7	2.0	0.7	2.5

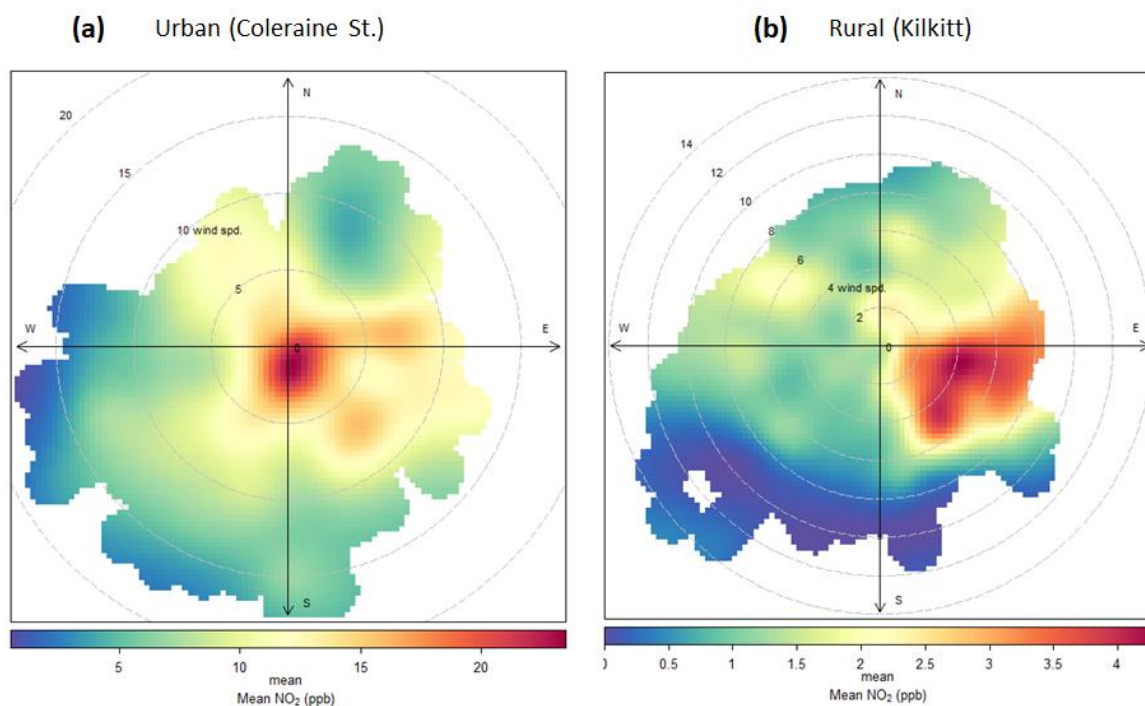


Figure 4: Polar plots of NO₂ concentration at an urban site (Coleraine St., left) and a rural background site (Kilkitt, right)

3.2 Regression Model

Parameters and relevant statistical metrics for the final LUR model are shown in Table 1. The final model explained over 78% of the spatial variability in NO₂ (Figure 5), while the cross validation R² was found to be slightly lower at 77.4%. NO₂ validation reveals a good explanation of variation by the model at all sites. The spread of data is well captured and the relationship between modelled and measured values is close to linear.

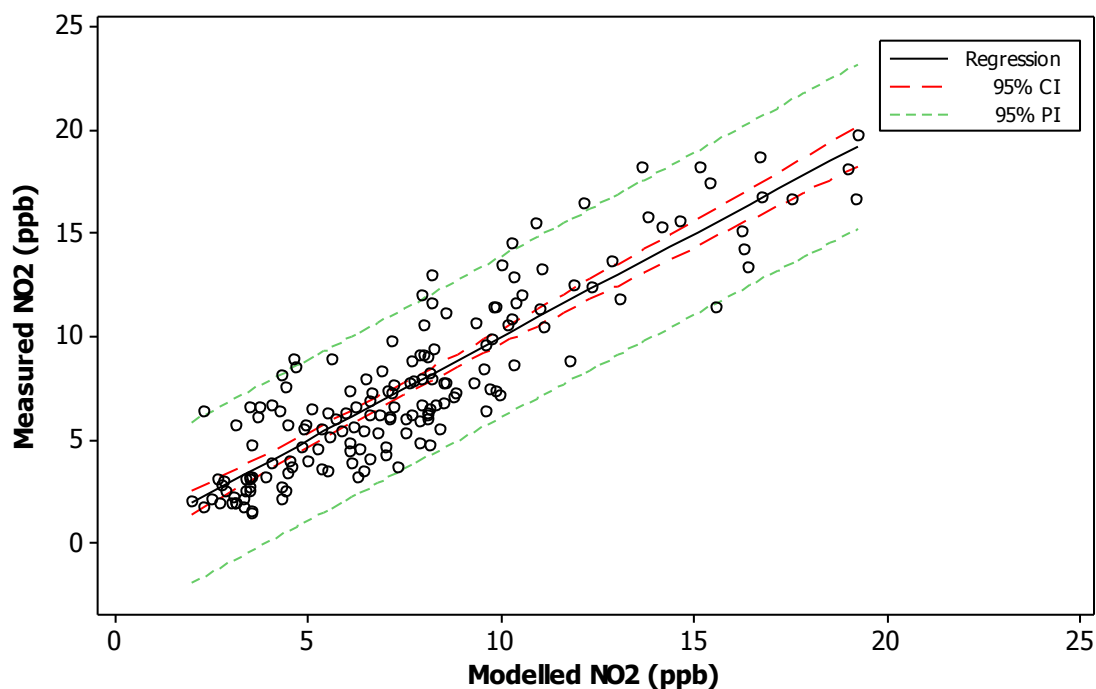


Figure 5: Comparison of modelled versus measured NO₂ for all wind sectors

The best set of predictor variables included terms for distance-weighted vehicle kilometres from 25m to 5km, commercial properties within 1km, percentage agricultural land within 1km, road density within 250 m and average wind speed. In a recent trans-European study, LUR models were developed in 36 study areas across Europe as part of the ESCAPE project (Beelen et al., 2013). The ESCAPE study model R² ranged from 55% to 92% (median 82%) for NO₂; model results using the wind sector LUR in the present Irish study are thus comparable, despite using only half the number of monitoring points in model development.

Table 1 NO₂ regression R²=78.7%

Predictor	Sector Radius	Coefficient	SE Coefficient	T	P	VIF
Constant		8.9535	0.74	12.1	0	
Inverse distance weighted Vehicle km travelled		2.88E-05	2.5E-06	11.51	0	1.80
Commercial buildings	1 km	0.002753	0.000833	3.3	0.001	1.56
Natural/agricultural land use	1 km	-9.1E-06	1.28E-06	-7.11	0	1.56
Average wind speed		-0.8304	0.1511	-5.49	0	1.13
Road density	250 m	0.002664	0.000886	3.01	0.003	1.38

4. Discussion and Conclusions

An important aspect of air quality modelling is determining which technique offers the best use of the resources and data that are available. The best model is not necessarily the most detailed or technically advanced; moreover, fundamental to the success of a given model, is the availability at sufficient resolution of the necessary data to drive it. In this context, the study described here used data available from routine FSMs rather than purpose-designed monitoring campaigns. In common with many countries, Ireland operates a FSM network primarily to assess compliance with limit values and to provide public information on the current state of air quality. This study expands the use of data from routine FSMs beyond compliance, or as a means of adjusting and limiting bias in passive sampling campaigns (Beelen et al., 2013; Henderson et al., 2007). The greater temporal coverage and short averaging period of FSM data (Cordioli et al., 2017) allowed us to divide the data wind sectors that, in addition to allowing the isolation of relevant source effects, also reduced site biases and allowed the introduction of a time variant aspect into the model. Furthermore, the reliability of the FSM measurements is typically higher than would be the case for many low-cost passive sampling techniques. By using FSM data, the method presented here thus minimises additional resources required to produce a viable background air quality map.

A necessity for quantifying personal exposure in epidemiological studies is the estimation of pollutant concentration at unmonitored locations. A significant benefit of the modelling process presented here is that a greater understanding of the effects of natural and anthropogenic effects on air quality is obtained. A model which maps air quality across the country provides a useful tool for knowing which areas to target or areas which require improvement. Maps are also a useful way to increase public awareness and alter behaviours. In this way, the air quality model can be considered to supplement the air quality monitoring network. Figure 1 (right) shows the spatial variability in NO₂ across Ireland. The NO₂ map illustrates the dominant influence of traffic emissions on national (and urban) NO₂ concentrations. Focussing in on the main urban centre in Ireland, Dublin, elevated NO₂ values are evident along the major traffic routes into and around the city. This reflects the role of road transport as the principal source of NO_x emissions in Ireland, contributing 45% of the national total in 2013 (Duffy et al., 2015).

A further use of the proposed approach is in assessing the spatial representativeness of the FSM network itself. Due to the high operational costs FSM networks often have limited spatial coverage and are heavily weighted to urban and suburban sites where pollutant concentrations and population exposure are higher. When using such networks to assess air quality at a regional or national level it is imperative to know how spatially representative the existing network is (Janssen et al., 2012). LUR models derive relationships between annual averaged concentrations and key land use characteristics; by mapping the range of land use variables across the region of interest and comparing to that surrounding the FSM, a quantitative measure of network representativeness can be provided. Furthermore, the inclusion of source-receptor orientation in the model using wind direction allows us to identify areas locally that may be strongly influenced by pollution derived from a single source/direction.

LUR is a statistical modelling methodology that relies on changes (correlations) in historical data and there are no implicit physical or chemical changes included in the model structure. While the LUR methodology described here includes elements that change over time (wind speed, direction, traffic volumes), the model still needs periodic updates to ensure validity. For example, changes over time in the composition national vehicle fleet, e.g. a gradual transition towards electric vehicles, would significantly change the relationship between traffic levels and consequent pollutant concentrations. This periodic recalibration, if carried out in conjunction with FSM network management, can form an integral part of an iterative design and improvement process to ensure future sustainability.

Acknowledgements

Research presented in this paper was carried out as part of the Science, Technology, Research and Innovation for the Environment (STRIVE) Programme 2014-2020 under grant number 2013_EH-FS-7. The programme is financed by the Irish Government under the National Development Plan 2007-2013 and administered on behalf of the Department of the Environment, Heritage and Local Government by the Environmental Protection Agency which has the statutory function of coordinating and promoting environmental research.

References

- Araín, M.A. et al., 2007. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. *Atmospheric Environment*, 41(16): 3453-3464. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2006.11.063>
- Baldauf, M. et al., 2011. Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. *Monthly Weather Review*, 139(12): 3887-3905.
- Beelen, R. et al., 2013. Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmospheric Environment*, 72(0): 10-23. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2013.02.037>
- Beelen, R. et al., 2014. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *The Lancet*, 383(9919): 785-795.
- Bertazzon, S., Johnson, M., Eccles, K., Kaplan, G.G., 2015. Accounting for spatial effects in land use regression for urban air pollution modeling. *Spatial and spatio-temporal epidemiology*, 14: 9-21.

- Briggs, D.J. et al., 1997. Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Science*, 11(7): 699-718. DOI:10.1080/136588197242158
- Carmichael, G.R. et al., 2008. Predicting air quality: Improvements through advanced methods to integrate models and measurements. *Journal of Computational Physics*, 227(7): 3540-3571. DOI:<http://dx.doi.org/10.1016/j.jcp.2007.02.024>
- CEU, 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on Ambient Air Quality and Cleaner Air for Europe. *Official Journal of the European Union*, L 152/1: 44 pages.
- Chen, L. et al., 2012. A land use regression model incorporating data on industrial point source pollution. *Journal of Environmental Sciences*, 24(7): 1251-1258. DOI:[http://dx.doi.org/10.1016/S1001-0742\(11\)60902-9](http://dx.doi.org/10.1016/S1001-0742(11)60902-9)
- Cordioli, M. et al., 2017. Combining land use regression models and fixed site monitoring to reconstruct spatiotemporal variability of NO₂ concentrations over a wide geographical area. *Science of The Total Environment*, 574(Supplement C): 1075-1084. DOI:<https://doi.org/10.1016/j.scitotenv.2016.09.089>
- Dee, D.P. et al., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656): 553-597.
- Denby, B. et al., 2010. Spatial Mapping of Ozone and SO₂ Trends in Europe. *Science of the Total Environment*, 408(20): 4795-4806. DOI:<http://dx.doi.org/10.1016/j.scitotenv.2010.06.021>
- Donnelly, A., Broderick, B., Misstear, B., 2012. Relating Background NO₂ Concentrations in Air to Air Mass History Using Non-Parametric Regression Methods: Application at Two Background Sites in Ireland. *Environmental Modelling and Assessment*, 17(4): 363-373.
- Donnelly, A., Misstear, B., Broderick, B., 2011. Application of nonparametric regression methods to study the relationship between NO₂ concentrations and local wind direction and speed at background sites. *Science of the Total Environment*, 409(6): 1134-1144.
- Donnelly, A., Naughton, O., Misstear, B., Broderick, B., 2016. Maximizing the spatial representativeness of NO₂ monitoring data using a combination of local wind-based sectoral division and seasonal and diurnal correction factors. *Journal of Environmental Science and Health, Part A*, 51(12): 1003-1011.
- Duffy, P., Hanley, E., Barry, S., Hyde, B., Alam, M.S., 2015. Informative Inventory Report 2015: Air pollutant emissions in Ireland 1990-2013 reported to the Secretariat of the UN/ECE Convention on long-range transboundary air pollution. Environmental Protection Agency, Johnstown Castle, Wexford, Ireland.
- EEA, 2012. Evaluation of Progress under the EU National Emissions Ceilings Directive. Progress Towards EU Air Quality Objectives. European Environment Agency, Copenhagen. EEA Technical Report No 14/2012.
- ESRI, 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.
- Galloway, J.N. et al., 2004. Nitrogen Cycles: Past, Present, and Future. *Biogeochemistry*, 70(2): 153-226. DOI:10.2307/4151466
- Guerreiro, C.B.B., Foltescu, V., de Leeuw, F., 2014. Air quality status and trends in Europe. *Atmospheric Environment*, 98: 376-384.
- 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(7): 2422-2428.

- Henry, R.C., Chang, Y.S., Spiegelman, C.H., 2002. Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction. *Atmospheric Environment*, 36(13): 2237-2244.
- Hettelingh, J.-P. et al., 2013. Assessing interim objectives for acidification, eutrophication and ground-level ozone of the EU National Emission Ceilings Directive with 2001 and 2012 knowledge. *Atmospheric Environment*, 75(0): 129-140. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2013.03.060>
- Hoek, G. et al., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, 42(33): 7561-7578. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2008.05.057>
- Janssen, S. et al., 2012. Land use to characterize spatial representativeness of air quality monitoring stations and its relevance for model validation. *Atmospheric Environment*, 59(0): 492-500. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2012.05.028>
- Janssen, S., Dumont, G., Fierens, F., Mensink, C., 2008. Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmospheric Environment*, 42(20): 4884-4903. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2008.02.043>
- Jerrett, M. et al., 2004. A review and evaluation of intraurban air pollution exposure models. *J Expo Anal Environ Epidemiol*, 15(2): 185-204.
- Kim, Y., Guldmann, J.-M., 2011. Impact of traffic flows and wind directions on air pollution concentrations in Seoul, Korea. *Atmospheric Environment*, 45(16): 2803-2810. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2011.02.050>
- Kotlarski, S. et al., 2014. Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, 7(4): 1297-1333.
- Kunzli, N. et al., 2000. Public-health impact of outdoor and traffic-related air pollution: a European assessment. *Lancet*, 356: 795-801.
- Laumbach, R.J., Kipen, H.M., 2012. Respiratory Health Effects of Air Pollution: Update on Biomass Smoke and Traffic Pollution. *Journal of Allergy and Clinical Immunology*, 129(1): 3-13.
- Levitin, J., Härkönen, J., Kukkonen, J., Nikmo, J., 2005. Evaluation of the CALINE4 and CAR-FMI models against measurements near a major road. *Atmospheric Environment*, 39(25): 4439-4452. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2005.03.046>
- Li, X. et al., 2015. The application of semicircular-buffer-based land use regression models incorporating wind direction in predicting quarterly NO₂ and PM₁₀ concentrations. *Atmospheric Environment*, 103(Supplement C): 18-24. DOI:<https://doi.org/10.1016/j.atmosenv.2014.12.004>
- Nolan, P., Baehler, T., 2016. High Resolution Solar and Wind Energy Datasets for Ireland.
- O'Dwyer, M., 2016. Air Quality in Ireland 2015: Key Indicators of Ambient Air Quality, Environmental Protection Agency, Johnstown Castle, Wexford, Ireland.
- Pascal, M. et al., 2013. Assessing the public health impacts of urban air pollution in 25 European cities: Results of the Aphekom project. *Science of the Total Environment*, 449(0): 390-400. DOI:<http://dx.doi.org/10.1016/j.scitotenv.2013.01.077>
- Raaschou-Nielsen, O. et al., 2013. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *The Lancet Oncology*, 14(9): 813-822.
- Rockel, B., Will, A., Hense, A., 2008. The regional climate model COSMO-CLM (CCLM). *Meteorologische Zeitschrift*, 17(4): 347-348.

- Savage, N.H. et al., 2013. Air quality modelling using the Met Office Unified Model (AQUM OS24-26): model description and initial evaluation. *Geosci. Model Dev.*, 6(2): 353-372. DOI:[10.5194/gmd-6-353-2013](https://doi.org/10.5194/gmd-6-353-2013)
- Silverman, B.W., 1986. *Density estimation for statistics and data analysis*. Chapman & Hall/CRC.
- Solazzo, E. et al., 2012. Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII. *Atmospheric Environment*, 53(0): 60-74. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2012.01.003>
- Staelens, J. et al., 2012. Trends in atmospheric nitrogen and sulphur deposition in northern Belgium. *Atmospheric Environment*, 49(0): 186-196. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2011.11.065>
- Stedman, J.R., Vincent, K.J., Campbell, G.W., Goodwin, J.W.L., Downing, C.E.H., 1997. New high resolution maps of estimated background ambient NO_x and NO₂ concentrations in the U.K. *Atmospheric Environment*, 31(21): 3591-3602. DOI:[http://dx.doi.org/10.1016/S1352-2310\(97\)00159-3](http://dx.doi.org/10.1016/S1352-2310(97)00159-3)
- Su, J.G. et al., 2008. An innovative land use regression model incorporating meteorology for exposure analysis. *Science of the Total Environment*, 390(2-3): 520-529. DOI:<http://dx.doi.org/10.1016/j.scitotenv.2007.10.032>
- Vautard, R. et al., 2013. The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Climate dynamics*, 41(9-10): 2555-2575.
- Vienneau, D. et al., 2010. Comparison of land-use regression models between Great Britain and the Netherlands. *Atmospheric Environment*, 44(5): 688-696. DOI:<http://dx.doi.org/10.1016/j.atmosenv.2009.11.016>