2016

Maximizing the spatial representativeness of NO2 monitoring data using a combination of local wind-based sectoral division and seasonal and diurnal correction factors

Aoife Donnelly
Technological University Dublin, aoife.donnelly@tudublin.ie

Owen Naughton
National University of Ireland, Galway

Bruce Misstear
Trinity College Dublin, Ireland

See next page for additional authors

Recommended Citation

Donnelly, Aoife & Naughton, Owen & Misstear, Bruce & Broderick, Brian. (2016). Maximizing the spatial representativeness of NO 2 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. 1-9. 10.1080/10934529.2016.1198174.

Funder: Government of Ireland

Authors
Aoife Donnelly, Owen Naughton, Bruce Misstear, and Brian Broderick

Maximising the Spatial Representativeness of NO_2 Monitoring Data Using a Combination of Local Wind-based Sectoral Division and Seasonal and Diurnal Correction Factors

Aoife Donnelly[1], Owen Naughton[2], Bruce Misstear[3], Brian Broderick[3],

[1]*School of Food Science and Environmental Health, Dublin Institute of Technology, Cathal Brugha St., Dublin 1, Ireland*

[2]*Civil Engineering, School of Engineering and Informatics, University of Ireland, Galway, Ireland*

[3]*Department of Civil, Structural and Environmental Engineering, Museum Building, Trinity College Dublin, College Green, Dublin 2, Ireland*

[*] Address correspondence to Aoife Donnelly, School of Food Science and Environmental Health, Dublin Institute of Technology, Cathal Brugha Street, Dublin 1, Ireland; E-mail: aadonnel@tcd.ie

Abstract

This paper describes a new methodology for increasing the spatial representativeness of individual monitoring sites. Air pollution levels at a given point are influenced by emissions sources in the immediate vicinity. Since emissions sources are rarely uniformly distributed around a site, concentration levels will inevitably be most affected by sources in the prevailing upwind direction. The methodology provides a means of capturing this effect and providing additional information regarding source/pollution relationships. The methodology allows for the division of the air quality data from a given monitoring site into a number of sectors or wedges based on wind direction and estimation of annual mean values for each sector, thus optimising the information that can be obtained from a single monitoring station. The method corrects for short-term data, diurnal and seasonal variations in concentrations (which can produce uneven weighting of data within each sector) and uneven frequency of wind directions. Significant improvements in correlations between the air quality data and spatial air quality indicators were obtained post application of the correction factors. This suggests the application of these techniques would be of significant benefit in land use regression modelling studies. Furthermore the method was found to be very useful for estimating long-term mean values and wind direction sector values using only short-term monitoring data. The methods presented in this paper can result in cost savings through minimising the number of monitoring sites required for air quality studies while also capturing a greater degree of variability in spatial characteristics. In this way more reliable, but also more expensive monitoring techniques can be used in preference to a higher number of low cost but less reliable techniques. The methods described in the paper have application in local air quality management, source receptor analysis, land use regression mapping and modelling and population exposure studies.

Keywords: Air pollution, seasonality, non-parametric regression, monitor representativeness, land use regression.

Introduction

Concentrations of air pollutants such as nitrogen dioxide (NO_2) are not constant but vary temporally and spatially arising from changes in meteorology, topography and also natural and anthropogenic emission sources. Air quality monitoring has the ability to capture these variations. Diversified monitoring objectives exist which include: demonstration of compliance with regulatory values, evaluation of general air quality in a region, epidemiological studies, estimation of long-term trends and source apportionment. Many stochastic air quality modelling methods used to produce air quality maps and forecasts, such as geo-statistical interpolation and land use regression, rely on relationships between measurements at monitoring stations and traffic, land use or population density parameters. [1-5] However, obtaining concentration data of sufficient spatial and temporal representativeness to achieve modelling objectives can pose significant cost and logistical obstacles. Methods that optimise information that can be obtained from an individual air quality monitor or monitoring network can provide a better understanding of the cause and effect in a given region and thus allow for improved air quality modelling and management.

Continuous temporally-resolved data from routine fixed-site monitoring (FSM) networks allow air quality modelling and mapping applications at a resolution generally unattainable through passive monitoring. A sufficient number of monitoring sites is required to capture the true relationships between the forcing variables and the concentration levels and Basagaña *et al.* [6] showed that in small samples, model fit tended to be highly inflated when compared to validation data sets. The high cost in maintaining routine networks, however, means that they are generally of restricted spatial scale, frequently covering a minimum of site types stipulated under EU regulations and not dense enough for the purpose of model development. [7] While a limited number of studies have used continuous measurements over the course of a full calendar year (or years) from routine networks [8-11], many studies focus on the spatial density requirement of monitoring campaigns at the expense of temporal resolution. [12-15]

Co-location of short-term sampling with background reference sites is generally used to extrapolate long-term averages [2, 13, 16]. Adequate correction therefore depends on dual assumptions holding true: that the continuous reference site(s) is representative of the temporal variation in concentrations and that the air pollutant pattern across the study area is stable over time. [7] The principal use of circular buffers in many land use regression (LUR) models ignores the spatial orientation of source and receptor and the influence of meteorological variables such as wind. In practice, source types and densities will inevitably vary in different directions at a local and regional scale and there will be a prevailing wind, meaning that sources in this direction relative to the monitoring site will have a disproportionally large influence on overall concentrations. As noted by Wheeler *et al*. [17] who modelled NO_2 in Windsor Ontario using a LUR, the impact of being downwind of any localised industrial or traffic source would be expected to have an impact on concentrations and they therefore suggest that including information on wind direction may help resolve some of the remaining variance in their model.

This paper presents a robust method which maximises the information from a single monitoring site and corrects data for both temporal (seasonal and diurnal) effects and meteorological (wind direction) effects. The methodology outlined in this paper is applicable for the following main objectives:

- Improving the long-term applicability of short-term monitoring
- Identifying local sources
- Estimating annual mean values for a given site typology and region
- Partitioning long-term monitoring data into wind dependant sectors to improve spatial representativeness

The application of these techniques to monitoring data to achieve these objectives can, in turn, be used to improve and refine land use regression models air quality mapping studies and population exposure studies.

The method is based on dividing the monitoring data into wind dependent sectors or wedges, and subsequently averaging the concentrations. The objective of dividing monitoring data in this manner is to relate average concentrations within a given sector to land use types or sources in the upwind direction and to maximise the (effective) number of monitoring locations. Prevailing wind directions vary seasonally and hence a biased sectoral average may be obtained in some instances. Furthermore, diurnal variations illustrate significant peaks and troughs in NO_2 concentrations and if there is a higher weight of observed values for rush-hour peaks in a given sector, an unrealistically high overall average will be obtained. In order to adequately model spatial variation in concentration levels, these temporal biases must be removed from the data prior to averaging.

The methods presented in this paper can provide cost savings by minimising the number of monitoring sites required for air quality studies. Dividing the data into a number of wind direction dependent sectors not only increases the number of monitoring points available but also captures a greater degree of variability in spatial characteristics. The correction factors developed allow these data to be extrapolated to represent annual mean concentrations. Studies can then utilise a smaller number of reliable (but often expensive) monitoring techniques in preference to a greater number of less reliable low cost techniques.

Methodology and Data

Overview

Fig. 1 outlines the methodology for increasing the spatial representativeness of a monitoring site and how it is applied to a dataset from a particular air quality monitoring site. A long-term dataset of hourly NO_2 values can be directly averaged to give a reasonable estimate of the mean concentration at that site (assuming there is no trend in the data). This provides us with one data point with which to compare spatial characteristics in the area. The new methodology involves splitting this dataset to provide additional data points. Firstly, the raw data are corrected for short-term fluctuations due to known seasonal and diurnal variations (s_t factors). Secondly, the data are split into subsets based on the wind direction for a given hour. For example, all NO_2 measurements for hours when the wind was blowing from a northerly direction will be assigned to the northerly subset. The concentrations within each subset are averaged to give a mean concentration for that wind sector. These average values are then corrected for uneven weighting that may occur due to unequal distribution of data between sectors (L_t factors). This then provides multiple annual mean values, one for each wind sector.

A smoothing kernel method was previously used by the authors to describe the variation in concentrations at background sites with local wind speed and direction and air mass history. [18, 19] Perez *et al.* [20] also used a smoothing kernel to describe the daily and seasonal cycle in CO_2 concentrations at a rural site. Based on these methods and the techniques described in Henry *et al.* [21] and Yu *et al.* [22] a short temporal correction factor (s_t) has been developed for raw data to remove these variations from the data prior to wind sector averaging. Long temporal (L_t) correction factors have also been developed, to apply to the data after the sector division in order to account for uneven weighting of data from each season that may arise within different sectors. These have been defined through a regression using monthly mean concentrations as predictor variables and annual mean concentrations as response variables from a range of urban, suburban and rural sites.

Monitoring Data

The Environmental Protection Agency (EPA) is the responsible authority in Ireland for the implementation of EU ambient air quality legislation. The EPA operates and maintains the national ambient air quality monitoring network [23] which measures statutory pollutants using reference or equivalent methods as outlined in the CAFÉ Directive. [24] These data are used for demonstrating compliance with air quality limit values and to give public information on ambient air quality. These sites were selected because their classification ranges from urban centre to rural background, but their proximity to urban conurbations and major roads is variable. In each case monitor placement conforms to the guidelines on monitoring laid out in EU directive 2008/50/EC. [24] NO_x is measured on an hourly basis at a 19 sites throughout Ireland using chemiluminescence samplers (Figure 1S).

s_t factors were derived based on historical variations in concentrations at each of the sites. Where possible, the most recent 5 years of validated data were used (typically 2007-2012). L_t factors were developed based on two years of data from each site so as to avoid an unfair weighting of any given site that has a greater length of monitoring record. 2011 and 2012 were used where available (or where monitoring has ceased the most recent years were chosen). In the development of the L_t factors, monthly data were only used if they had > 80% data capture and a number of years were omitted where data were missing for more than 25% of any individual month.

Short-term Correction (S_t)

The basis for the s_t correction factors is a non-parametric kernel regression model. Nonparametric regression relaxes the functional form assumed in parametric regression, the object being to estimate the regression function directly, rather than to estimate parameters. The authors previously used the techniques to define diurnal and seasonal variations in NO_2 concentrations for use as inputs to a statistical point wise air quality forecast model the outputs of which have been validated in Donnelly *et al*. [25]

In this paper the regression 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). Accurate quantification of the long-term variations can be made, and their effects removed from the raw data, thus supporting the improved definition of concentrations for use in spatial air quality modelling.

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 [21] and is defined as follows:

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

The bandwidth is calculated based on the number of days in a year. As discussed in Silverman [26] 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 $(NO_{2(s)})$:

$$NO_{2(s)} = \frac{NO_{2(raw)}}{S_f}$$

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

$$NO_{2(s,d)} = \frac{NO_{2(s)}}{D_f}$$

Applying the factors in this way does not change the mean of the total data set (i.e. the mean of all the raw hourly values is approximately the same as the mean of all the seasonally and diurnally adjusted hourly values). 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.

Long-term Correction (L_t)

When the raw data have been seasonally and diurnally adjusted they are divided into groups for each identified wind direction sector. While this maximises the number of data points available, it also reduces the number of data points that are used to predict long-term mean values at a given location. Annual average values are generally calculated using a full year of monitoring data. If less than a full year of data are used for the calculation of the annual mean, certain sites and sectors may have an unrepresentative weighting of data from a particular season. Concentrations tend to be higher in winter months than summer months so, for example, if data were available for 20% of winter and 80% of summer months, an unrealistically low value for the annual average would be obtained. In the present study L_t correction factors have been used to adjust the sectoral averages at each site and ensure they are representative of the annual mean.

Monthly and annual average concentrations were calculated for each monitoring site. For each individual month a linear regression analysis was carried out using linear least squares on each month (e.g. Figure 2) of data to produce an equation for each month, i:

$$\widehat{C_i} = a_i + b_i \times M_i$$

where $\widehat{C}_i$ is the estimated annual mean concentration at a given site, a_i and b_i are the intercept and slope of the regression, respectively and M_i is the monthly average concentration. Table 1 shows the derived correction factors for each month based on this regression.

Data tend to be available for a range of durations from different months. Therefore, a weighted average technique is employed to predict the overall mean using the relevant adjustment factors from each month. Unlike the seasonal and diurnal correction factors above, this correction can change the long-term mean of the total data set, particularly if monitoring was only carried out for a short period of time.

Results and Discussion

In the following sections the results of the method are going to be discussed according to a number of potential applications which have been identified.

Applicability of methods

The methods presented in this paper have a number of applications which makes them useful for many air quality studies. The methods:

- Allow predictions of annual mean based on short-term data sets. The short-term correction factors from similar stations with long runs of data can be used directly on the raw data. The long-term correction factors are then applied to the averaged data, be it in total or in sub divided wind sectors.
- Provide a greater amount of information about air quality at a given station.
- Provide the ability to analyse air pollution levels in relation to the land use types, roads and emissions in the surrounding area and account for wind direction.

- Can increase the spatial density of data points used in LUR. Data can be divided into wind dependent sectors and the factors developed in this paper can be applied to estimate annual mean values for each sector, thus increasing the number of response variable/predictor variable pairs. The methods are particularly important when considering a LUR over a large spatial area such as in the creation of national scale background maps

In the following sections, some applications of the method are demonstrated followed by a validation of the S_t and L_t factors.

Wind sector averaging

In order to extrapolate additional data from a given monitoring site the data can be grouped into equal sized wind dependent sectors. The aim of this is to capture specific spatial characteristics that are influencing the air quality in the region. Depending on the resolution of the wind direction data the number of sectors can directly reflect the number of peaks but in this study eight sectors have been used. Figure 3 shows these eight sectors marked around a monitoring site in Kilkenny town. Land use typology and the road network are also indicated on the map and it is clear that each of the different sectors represents a separate land use composition. This is mirrored in the polar plot in Figure 4 which shows increases in concentrations for northerly, north easterly and easterly wind directions. If a circular buffer was applied here to infer relationships between land use and concentrations, the results would have been misleading. Using the sector-based approach it can be seen that areas with increased coverage of artificial surfaces and roads tend to lead to higher concentrations. In order to quantify these effects it is necessary to estimate the long term mean concentration for each sector. Direct averaging of data within each wind sector would not necessarily result in reasonable predictions of the long run mean due to seasonal differences in wind direction frequencies and other external forcing factors such as variation in sunshine hours and stability conditions. To account for this, the s_t and L_t correction factors can be applied to the data to account for uneven weighting of data from each season that may arise within different sectors.

Correcting data

To demonstrate the application of the S_t and L_t factors where only short term monitoring data are available, a subset of data was taken from a full 5-year dataset (Rathmines site, Dublin). Two random weeks were chosen from each season in 2011. These were weeks beginning on:

- 3rd January
- 2nd April
- 27th June
- 25th October

A different site was used to develop the S_t factors since in practice, long-term data would not generally be available for the site at which the methods were being applied (a nearby urban site where long term data were available, Winetavern Street, was used). These factors were applied to the data subset to smooth out seasonal and diurnal variations. Using the L_t factors an adjusted annual mean value was calculated based on the subset of data (10.43ppb). This is compared to the actual annual mean from the full year (10.24bbp) and the average of the raw data (11.06ppb). Applying the correction factors to the subset of data results in a closer estimation of the true annual mean.

The data subset was then divided into eight equally sized wind sectors. Thereafter, the L_t factors were used to estimate the long run mean for each of the sectors. This was compared to the results using the raw subset data and the full year of data. Results are illustrated in Figure 5. The raw subset of data underestimates the concentrations for the higher sectors and overestimates concentrations for the lower sectors. Applying the S_t and L_t factors results in an improved estimation of the variation among sectors.

Validation of L_t factors

The L_t factors have been validated by applying them to predictions of annual means using individual monthly values from data not used in the formation of the equations. Table 2 shows the correlation between predictions of annual mean using the defined equations and the measured annual mean values. The sites used in the validation encompassed urban centre to rural background sites and were Dun Laoghaire, Rathmines, Kilkitt, Winetavern Street, Coleraine Street and Park Road. Data from 2007 and 2008 were used, indicating that the equations are stable over time.

NO_2 model calibration and validation yielded strong results for February, March, June, July, August, September, October and November. In Ireland, December and January tend to experience well established low pressure systems which can bring strong winds and rainfall. On occasion, cold anticyclonic air can extend its influence westwards to Ireland from continental Europe, which can lead to long cold periods. This variability in meteorological conditions contributes to the monthly variability in observed concentrations and slightly poorer explanation of the annual mean as shown by R_{cal}^2. Validation for both January and December 2011 showed good R^2_{val} values. Although winds were predominantly westerly during the month of December there was some ingress of easterly air masses which affected the east coast of Ireland. Towards the end of December ozone concentration increased significantly and during this time period air mass back trajectory analysis using the Hysplit model revealed that re-circulated air masses from over northern Europe were affecting the east coast of Ireland. [27] Global solar radiation totals for the December period were slightly above average for the Dublin region. [28] This suggests that the equation is robust but also that care must be taken when applying it to data collected under atypical meteorological conditions.

April to May is also a period of change in Irish weather conditions as temperatures rise slowly and various air masses have the potential to facilitate long range transport of emissions which could explain why May was found to be the least indicative month for annual mean NO_2 prediction. This suggests that short-term monitoring campaigns should not be carried out during May in isolation where possible. Towards late June the rise in pressure over the Ocean and fall in pressure over Europe results in a predominantly westerly surface air flow with a long ocean track which tends to be associated with lower contributions of anthropogenic emissions. [27]

To account for the variability in meteorological conditions from month to month and the inevitable departure from mean conditions, monitoring should be carried out for a number of months (or part thereof) and a weighted average calculated using the modelled equation for each month. Where this is not feasible the best result from the earlier regression analysis as presented in Table 1 should be taken and if a conservative value is required, the upper confidence limit can be calculated using the variance.

Validation of S_t factors

Correlations between NO_2 concentrations and typical spatial descriptors used in land use regression studies were calculated using both raw NO_2 concentrations and corrected NO_2 concentrations at each of the 23 NO_2 sites in Ireland. Hourly monitoring data at each site were corrected using the S_t correction factors. Data were then divided into eight equal sized wind direction sectors corresponding to N, NE, E, SE, S, SW, W and NW. Average values (NO_2Cor) were calculated for each sector and corrected using the L_t correction factors. Raw hourly NO_2 data were then averaged directly for each of the wind sectors to produce NO_2Raw values.

Parameters considered in the study were length of all roads (R_l), total vehicle kilometres (Vkm), the area of continuous urban land use (urb), the number of residential properties (res), the number of commercial properties (com), the population and area of agricultural land use type ($agri$). Table 3 shows the resulting correlations for buffer sectors of various radii. For all buffer sizes, the correlations are significant in both the raw and corrected data, with highest correlations being observed with R_l and Vkm in the 2km sector size. Improvements are observed in correlation p-values for all variables after data correction (except for urb in the 250m sector). The removal of concentration fluctuations due to meteorological and seasonal factors allows the isolation of external forcing factors and thus improved quantification of spatial variability in concentration levels using spatial descriptors. These results illustrate that improvements in land use regression studies can be observed by applying the correction factors developed as part of this paper.

Conclusions

This paper presents a simple but effective means of extrapolating numerous representative annual average concentrations from a single fixed air quality monitoring site. The methods presented in the paper were developed so as to make more use out of the high quality data collected as part of Ireland's air quality monitoring regulatory requirements under EU legislation. This paper describes the application of the methods to NO_2 but the techniques can be applied to any pollutant measured at an hourly or daily temporal resolution. Seasonal and diurnal correction short-term correction factors were developed using non parametric regression techniques. Validated measured data are initially corrected for short-term fluctuations using these factors. These data are then divided into subsets for 8 (or more) wind direction sectors. Concentrations within each sector are then averaged. These averages are corrected using long-term correction factors which were derived and validated using historical air quality data sets from the Irish air quality monitoring network. These factors account for seasonal biases that may arise and facilitate the accurate estimation of annual mean values from the sectoral averages. Calibration R^2 ranged from 74.8 % to 97.3 % for NO_2 while validation R^2 values ranged between 79 % and 98.9 % for individual monthly predictions. Correlations between spatial predictor variables (such as road density) and NO_2 concentration were compared using raw data and data post correction. Significant improvements were observed in all correlations through the use of the correction factors. This suggests the application of these techniques would be of significant benefit in land use regression modelling studies. Furthermore the method was found to be very useful for estimating long-term mean values and wind direction sector values using only short-term monitoring data. The methods presented in this paper can result in cost savings through minimising the number of monitoring sites required for air quality studies while also capturing a greater degree of variability in spatial characteristics. In this way more reliable, but often cost prohibitive monitoring techniques can be used in preference to a higher number of low cost but less reliable techniques.

Acknowledgements

This paper has been written as part of the Science, Technology, Research and Innovation for the Environment (STRIVE) Programme 2007–2013. The programme is financed by the Irish Government under the National Development Plan 2007–2013. It is administered on behalf of the Department of the Environment, Heritage and Local Government by the Environmental Protection Agency which has the statutory function of co-ordinating and promoting environmental research.

References

[1] Dons, E.; Van Poppel, M.; Kochan, B.; Wets, G.; Int Panis, L. Modeling temporal and spatial variability of traffic-related air pollution: Hourly land use regression models for black carbon. Atmospheric Environment, **2013**, *74*, 237-246.

[2] Briggs, D.J.; Collins, S.; Elliott, P.; Fischer, P.; Kingham, S.; Lebret, E.; Pryl, K.; Van Reeuwijk, H.; Smallbone, K. and Van Der Veen, A. Mapping urban air pollution using GIS: a regression-based approach. International Journal of Geographical Information Science, **1997**, *11*, 699-718.

[3] Janssen, S.; Dumont, G.; Fierens, F.; Mensink, C. Spatial interpolation of air pollution measurements using CORINE land cover data. Atmospheric Environment, **2008**, *42*, 4884-4903.

[4] Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M.-Y.; Künzli, N.; Schikowski, T.; Marcon, A.; Eriksen, K.T.; Raaschou-Nielsen, O.; Stephanou, E.; Patelarou, E.; Lanki, T.; Yli-Tuomi, T.; Declercq, C.; Falq, G.; Stempfelet, M.; Birk, M.; Cyrys, J.; von Klot, S.; Nádor, G.; Varró, M.J.; Dėdelė, A.; Gražulevičienė, R.; Mölter, A.; Lindley, S.; Madsen, C.; Cesaroni, G.; Ranzi, A.; Badaloni, C.; Hoffmann, B.; Nonnemacher, M.; Krämer, U.; Kuhlbusch, T.; Cirach, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Bellander, T.; Korek, M.; Olsson, D.; Strömgren, M.; Dons, E.; Jerrett, M.; Fischer, P.; Wang, M.; Brunekreef, B.; de Hoogh, K. Development of NO_2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. Atmospheric Environment, **2013**, *72*, 10-23.

[5] Wang, R.; Henderson, S.B.; Sbihi, H.; Allen, R.W.; Brauer, M. Temporal stability of land use regression models for traffic-related air pollution. Atmospheric Environment, **2013**, *64*, 312-319.

[6] Basagaña, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bouso, L.; Elosua, R.; Foraster, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Vila, J. Effect of the number of measurement sites

on land use regression models in estimating local air pollution. Atmospheric Environment, **2012**, *54*, 634-642.
[7] Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmospheric Environment, **2008**, *42*, 7561-7578.
[8] Hooyberghs, J.; Clemens, M.; Gerwin, D.; Frans, F. Spatial interpolation of ambient ozone concentrations from sparse monitoring points in Belgium. Journal of Environmental Monitoring, **2006**, *8*, 1129-1135.
[9] Hoek, G.; Fischer, P.; Van Den Brandt, P.; Goldbohm, S.; Brunekreef, B. Estimation of long-term exposure to outdoor air pollution for a cohort study on mortality. Journal of Exposure Analysis and Environmental Epidemiology, **2001**, *11*, 459-469.
[10] Beelen, R.; Hoek, G.; Fischer, P.; van Den Brandt, P.; Brunekreef, B. Estimated long-term outdoor air pollution concentrations in a cohort study. Atmospheric Environment, **2007**, *41*, 1343-1358.
[11] Meng, X.; Chen, L.; Cai, J.; Zou, B.; Wu, C.-F.; Fu, Q.; Zhang, Y.; Liu, Y.; Kan, H. A land use regression model for estimating the NO 2 concentration in shanghai, China. Environmental research, **2015**, *137*, 308-315.
[12] Lee, J.-H.; Wu, C.-F.; Hoek, G.; de Hoogh, K.; Beelen, R.; Brunekreef, B.; Chan, C.-C. Land use regression models for estimating individual NO x and NO 2 exposures in a metropolis with a high density of traffic roads and population. Science of The Total Environment, **2014**, *472*, 1163-1171.
[13] Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; de Nazelle, A. Development of land use regression models for PM2. 5, PM2. 5 absorbance, PM10 and PMcoarse in 20 European study areas; results of the ESCAPE project. Environmental science & technology, **2012**, *46*, 11195-11205.
[14] Hoek, G.; Meliefste, K.; Cyrys, J.; Lewné, M.; Bellander, T.; Brauer, M.; Fischer, P.; Gehring, U.; Heinrich, J.; van Vliet, P. Spatial variability of fine particle concentrations in three European areas. Atmospheric Environment, **2002**, *36*, 4077-4088.
[15] Jerrett, M.; Burnett, R.T.; Ma, R.; Pope III, C.A.; Krewski, D.; Newbold, K.B.; Thurston, G.; Shi, Y.; Finkelstein, N.; Calle, E.E. Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology, **2005**, *16*, 727-736.
[16] Brauer, M.; Hoek, G.; van Vliet, P.; Meliefste, K.; Fischer, P.; Gehring, U.; Heinrich, J.; Cyrys, J.; Bellander, T.; Lewne, M. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. Epidemiology, **2003**, *14*, 228-239.
[17] Wheeler, A.J.; Smith-Doiron, M.; Xu, X.; Gilbert, N.L.; Brook, J.R. Intra-urban variability of air pollution in Windsor, Ontario—measurement and modeling for human exposure assessment. Environmental Research, **2008**, *106*, 7-16.
[18] Donnelly, A.; Misstear, B.; Broderick, B. Application of nonparametric regression methods to study the relationship between NO_2 concentrations and local wind direction and speed at background sites. Science of the Total Environment, **2011**, *409*, 1134-1144.
[19] Donnelly, A.; Broderick, B.; Misstear, B. Relating background NO_2 concentrations in air to air mass history using non-parametric regression methods: application at two background sites in Ireland. Environmental Modeling & Assessment, **2012**, *17*, 363-373.
[20] Pérez, I.A.; Sánchez, M.L.; García, M.Á.; Pardo, N. Carbon dioxide at an unpolluted site analysed with the smoothing kernel method and skewed distributions. Science of the Total Environment, **2013**, *456*, 239-245.

[21] Henry, R.C.; Chang, Y.S.; Spiegelman, C.H. Locating nearby sources of air pollution by nonparametric regression of atmospheric concentrations on wind direction. Atmospheric Environment, **2002**, *36*, 2237-2244.
[22] Yu, K.N.; Cheung, Y.P.; Cheung, T.; Henry, R.C. Identifying the impact of large urban airports on local air quality by nonparametric regression. Atmospheric Environment, **2004**, *38*, 4501-4507.
[23] O'Dwyer, M.; Air Quality in Ireland 2013. Environmental Protection Agency; Johnstown Castle Estate, Wexford; **2014**.
[24] CEU, 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. Commission, E. Ed.; Official Journal of the European Union; **2008**.
[25] Donnelly, A.; Misstear, B.; Broderick, B. Real time air quality forecasting using integrated parametric and non-parametric regression techniques. Atmospheric Environment, **2015**, *103*, 53-65.
[26] Silverman, B.W. Density estimation for statistics and data analysis; Chapman & Hall/CRC, 1986.
[27] Donnelly, A.; Broderick, B.; Misstear, B. The effect of long range air mass transport pathways on PM_{10} and NO_2 concentrations at urban and rural background sites in Ireland: Quantification using clustering techniques. Journal of Environmental Science and Health Part A, **2015**, *50*, 647-658.
[28] Met Eireann, Monthly Weather Bulletin December 2011. 2012.

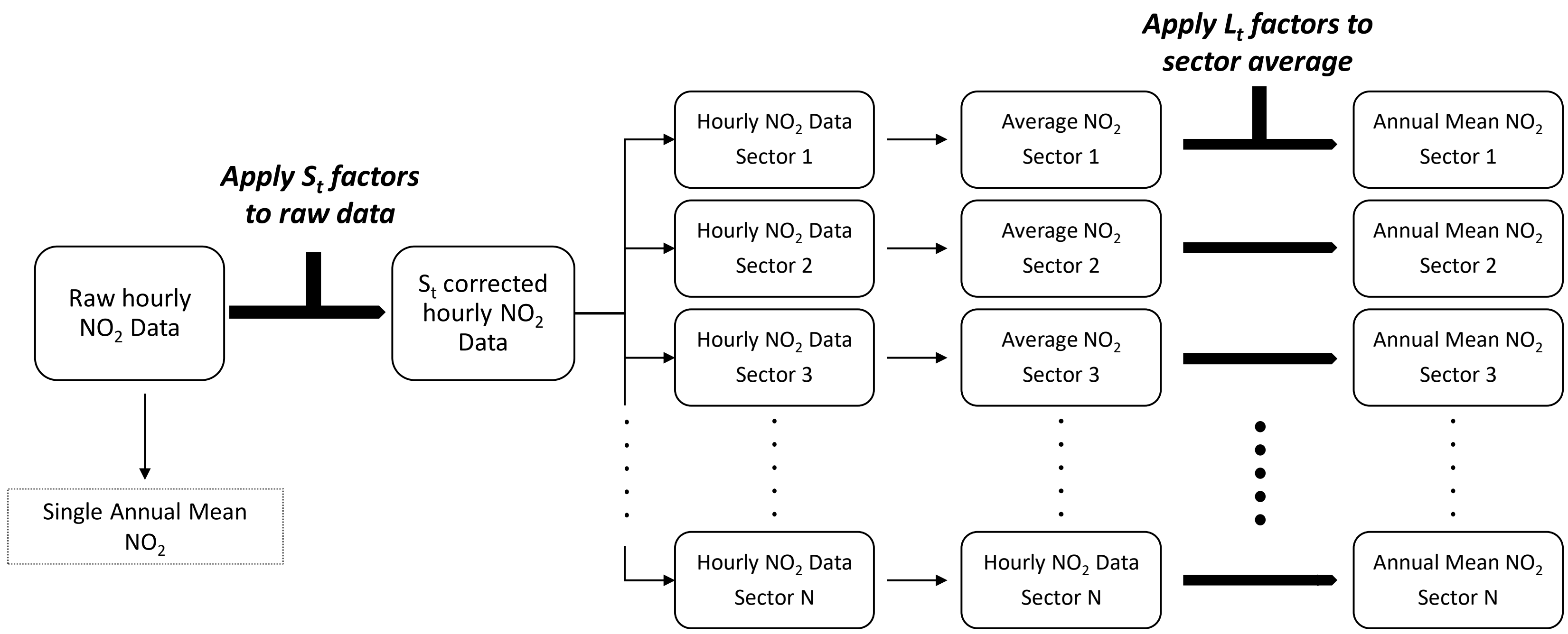
Apply L_t factors to sector average
Apply S_t factors to raw data
Raw hourly NO_2 Data
S_t corrected hourly NO_2 Data
Single Annual Mean NO_2
Hourly NO_2 Data Sector 1
Hourly NO_2 Data Sector 2
Hourly NO_2 Data Sector 3
Hourly NO_2 Data Sector N
Average NO_2 Sector 1
Average NO_2 Sector 2
Average NO_2 Sector 3
Hourly NO_2 Data Sector N
Annual Mean NO_2 Sector 1
Annual Mean NO_2 Sector 2
Annual Mean NO_2 Sector 3
Annual Mean NO_2 Sector N

Fig. 1

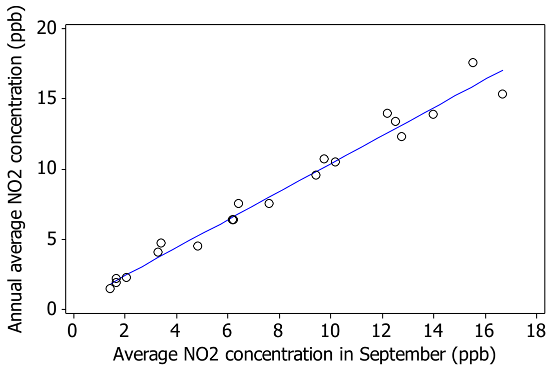
Annual average NO2 concentration (ppb)
Average NO2 concentration in September (ppb)
0
5
10
15
20
0
2
4
6
8
10
12
14
16
18

Fig. 2

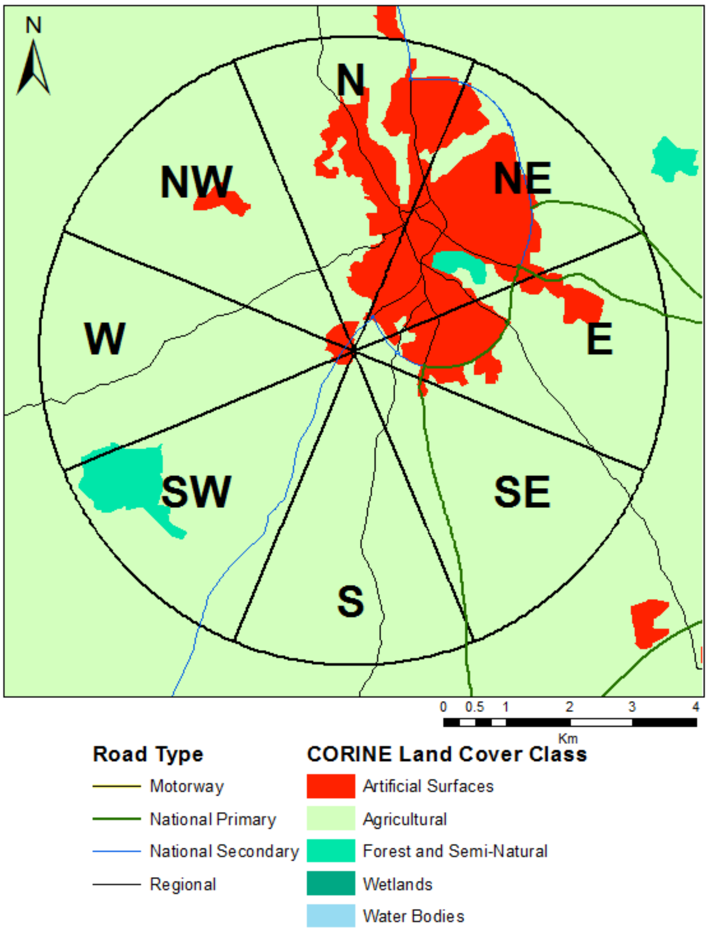
N
NW
NE
W
E
SW
SE
S
0 0.5 1 2 3 4
Km
Road Type
Motorway
National Primary
National Secondary
Regional
CORINE Land Cover Class
Artificial Surfaces
Agricultural
Forest and Semi-Natural
Wetlands
Water Bodies

Fig. 3

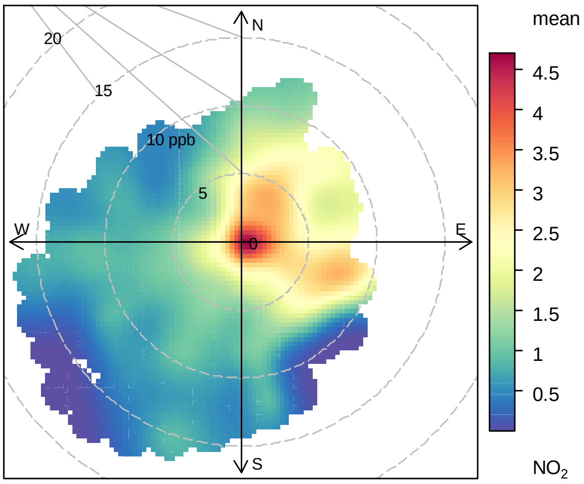
mean
4.5
4
3.5
3
2.5
2
1.5
1
0.5
NO_2
N
S
W
E
20
15
10 ppb
5
0

Fig. 4

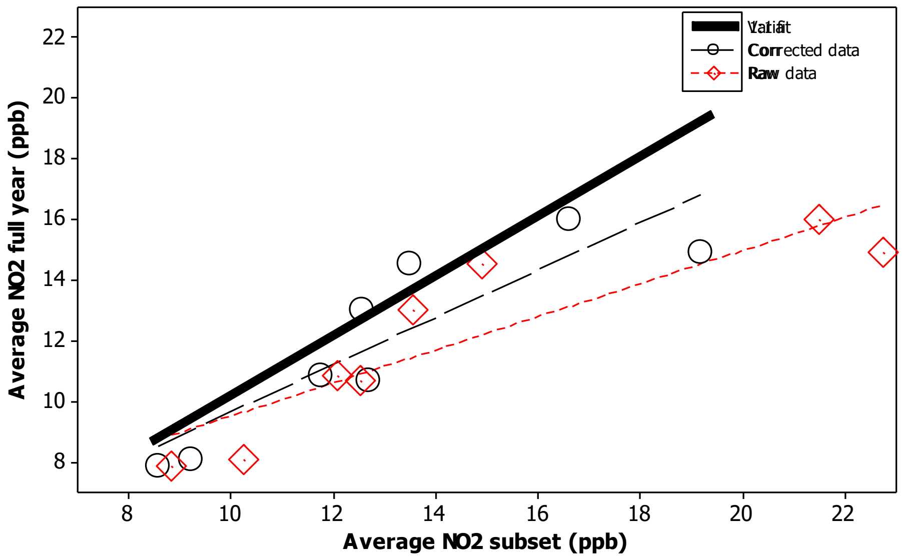
Corrected data
Raw data
Average NO2 full year (ppb)
Average NO2 subset (ppb)
8
10
12
14
16
18
20
22

Fig. 5

Table 1. Correction factors for NO_2

Month	Correction constant	Correction multiple	R^2_{cal}	s
Jan	1.74	0.75	81.3	2.15
Feb	0.25	1.04	95.0	1.11
Mar	0.13	0.89	97.0	0.87
Apr	0.95	0.92	89.6	1.60
May	2.14	0.80	74.8	2.43
Jun	0.45	0.96	95.4	1.07
Jul	0.41	0.92	95.3	1.09
Aug	0.57	0.91	93.8	1.17
Sep	0.38	1.00	97.3	0.81
Oct	0.34	0.94	91.8	1.42
Nov	0.60	0.87	95.5	1.04
Dec	0.42	1.04	89.7	1.68

Table 2. Validation of equations

Month	R^2_{val}
Jan	95.7
Feb	91.3
Mar	95.9
Apr	95.9
May	79
Jun	98.9
Jul	97.8
Aug	85.1
Sep	90.2
Oct	97
Nov	98.1
Dec	95.5

Table 3. Correlation of NO_2 concentrations with land use variables using raw and corrected NO_2 data

Buffer size	**250m**		**500m**		**1km**		**2km**		**5km**	
	NO_2Raw	NO_2Cor	NO_2Raw	NO_2Cor	NO_2Raw	NO_2Cor	NO_2Raw	NO_2Cor	NO_2Raw	NO_2Cor
$\boldsymbol{R_l}$	0.36	**0.379**	0.456	**0.521**	0.517	**0.591**	0.525	**0.618**	0.456	**0.539**
$\boldsymbol{Vkm}$	0.41	**0.466**	0.443	**0.53**	0.407	**0.509**	0.477	**0.598**	0.469	**0.563**
$\boldsymbol{urb}$	**0.514**	0.472	0.511	**0.517**	0.476	**0.53**	0.45	**0.539**	0.344	**0.396**
$\boldsymbol{res}$	0.433	**0.463**	0.513	**0.6**	0.531	**0.631**	0.518	**0.622**	0.476	**0.56**
$\boldsymbol{com}$	0.471	**0.45**	0.429	**0.443**	0.45	**0.513**	0.453	**0.528**	0.38	**0.436**

pop	0.4	**0.451**	0.482	**0.578**	0.528	**0.637**	0.522	**0.63**	0.481	**0.569**
$agri$	-0.53	**-0.541**	-0.519	**-0.563**	-0.588	**-0.629**	-0.522	**-0.606**	-0.508	**-0.599**

Supplementary material

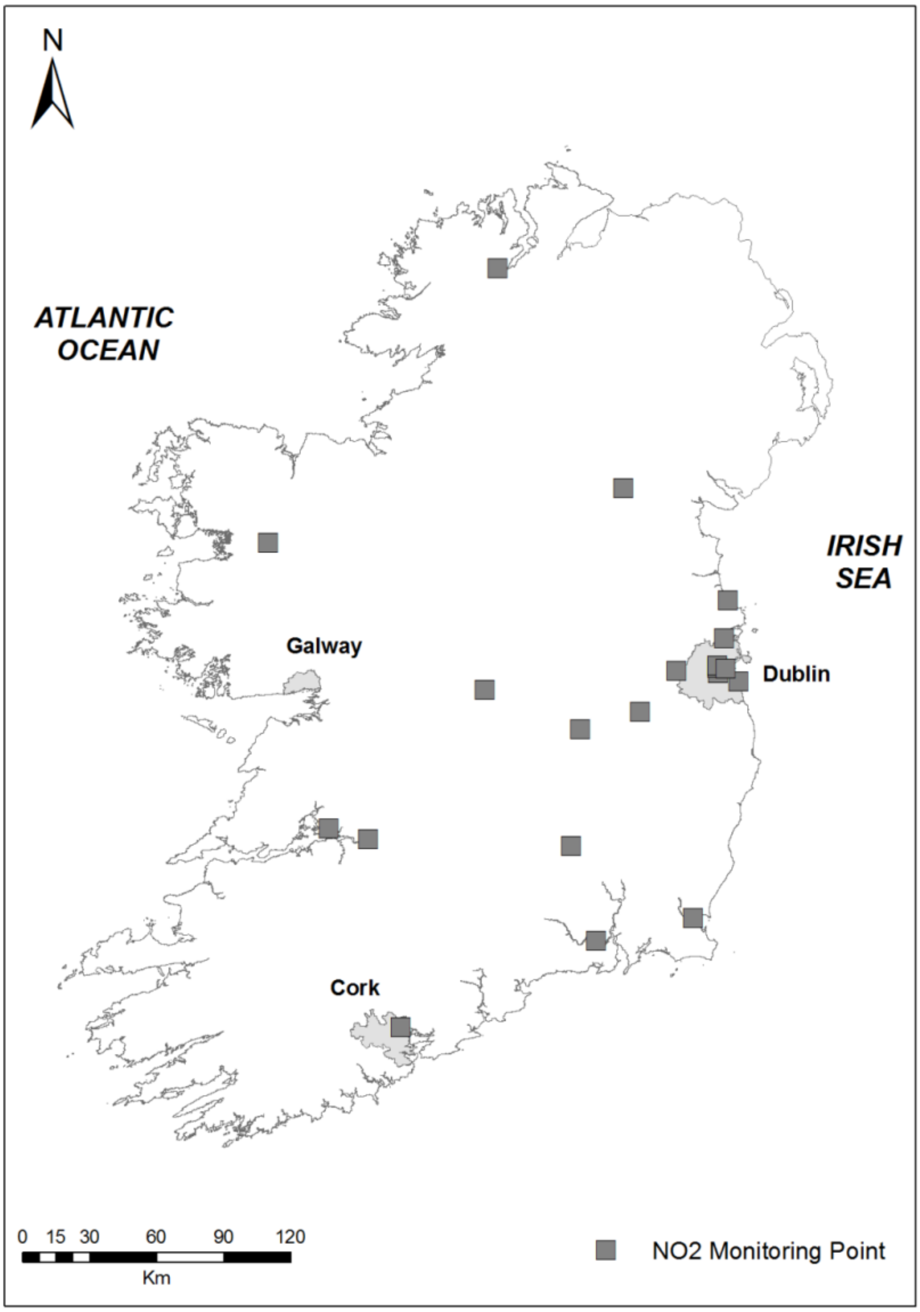

Fig. 1S.

Figure captions

Figure 1. Overview of methodology

Figure 2. Regression of average NO_2 concentration in September against annual mean concentration

Figure 3. Sectors of 4km radius around the monitoring site in Kilkenny

Figure 4. Determination of directional NO_2 sectors at Kilkenny monitoring site

Figure 5. Wind sector data for subset at Rathmines, raw data and data corrected using s_f factors from Winetavern Street

Fig. 1S. NO_2 monitoring sites