PEACOX – Persuasive Advisor for CO2-reducing cross-modal trip planning

Project Reference: 288466

FP7-ICT 2011: 6.6 Low carbon multi-modal mobility and freight transport

Project Duration: 1 Oct 2011 – 30 Sept 2014

D3.2 Eco Driving Model and Emissions Exposure Model

[Trinity College Dublin]

Author(s)

Md. Saniul Alam (TCD)

Dr. Aonghus McNabola (TCD)

Dr. Brian Caulfield (TCD)

Date: 27/01/2013

Dissemination level: (PU, PP, RE, CO): PU

Project Co-Funded by the European Commission within the 7th Framework Program

SEVENTH FRAMEWORK PROGRAMME

Abstract

This project involves the examination of real-time eco-driving data that enables users to make pre-trip and on-route decisions when driving as to the optimal route to take. The basis of this project is to estimate how efficiently drivers are performing in relation to fuel consumption per kilometer. The analysis uses details on the vehicle specification, in terms of fuel efficiency, and relates this to the distance travelled to provide the user with information on the efficiency per KM travelled. Eco-driving involves the training of individuals to change their driving patterns and to adapt to driving conditions. This project examines data collected by TomTom in the Netherlands and measures the emissions saved by providing eco-driving information.

The PEACOX project has set grounds for managing eco-friendly driving issues more efficiently along with their other set targets. One of the aims is to provide information to travellers about safer routes in terms of exposure. Studies showed that a reduction in exposure to particulate matter (PM10) could reduce premature deaths significantly and, could also offer a healthy environment for travelling. Thus, PM10 has been chosen as a generic pollutant whose concentration level indicates the level of exposure in the routes. In order to estimate exposure concentrations, exposure models were built.

To carry out the task, PM10 concentration has been estimated for Dublin and Vienna using a Landuse regression model approach. Routine monitoring PM10 data has been used for building models where explanatory variables included weather, land use, topographic and demographic information. After model validation, a neural network was also used to obtain the best fit model, optimising the relationship between response and explanatory variables. This was necessary to offset the limitation of using the small number of PM10 monitors available. Fourteen emissions maps for different days of the week over the summer and winter seasons were predicted for each city. PM10 concentrations were then transferred to the road network level to highlight the best route in terms of exposure level, or dose for trips.

Table of Contents

1. Eco-driving model

1.1 Introduction

This first section of the deliverable relates to the results of the eco-plus (eco-driving) trial conducted between Trinity College Dublin and TomTom.

In recent years many authors have written about the success of eco-driving and its ability to reduce emissions and how it can be used as a tool to combat climate change. Barkenbus (2010) suggests that eco-driving is the overlooked climate change initiative and that following a policy of eco-driving can result in a 10% reduction in fuel consumption which will have a knock on effect of reducing emissions. A range of studies have shown that the benefits from eco-driving can range from a 5 to 20% reduction in emissions (Stillwater et al, 2012).

Beusen et al (2009) examined 10 cars over a 10-month period after taking a course, which provided them with eco-driving training. The authors found that drivers on average had a 5.8% reduction in fuel usage. However, the study showed that the fuel savings reduced over time and drivers went back to their original habits. Delhomme et al (2013) conducted a survey of French drivers to ascertain their opinions in relation to eco-driving and how they feel about adopting eco-driving styles. The findings show that generally respondents said it would be easy to adapt to the eco-driving styles. The results did show that younger and middle aged drivers said it may be difficult to adapt to the driving styles.

Boriboonsomsin et al (2011) conducted a study of 20 drivers in Southern California using an on-board eco-driving feedback tool. The findings of the study showed modest increases in fuel economy of 6% for urban streets and 1% on motorways. This was attributed increased congestion in the area. Martin et al (2013) conducted a study of 18 drivers in California using on-board feedback for eco-driving. The study took a similar approach to the one reported in this deliverable in that the devices were turned off for the first month and then switched on to give drivers feedback on driving style. Similar to the results found in Boriboonsomsin et al (2011), the authors show that modest improvements in fuel efficiency. In 2012, Martin et al (2012) conducted a longitudinal study of a sample of participants in California. This study surveyed participants over three time intervals to determine if eco-driving behavior would last in the long run using information from an eco-driving website. The study looked at before and after information on how the study worked. The results showed that more than half of the sample improved their eco-driving behavior and that females, those living in smaller households and those with newer cars were more likely to improve eco-driving behavior.

Stillwater and Kurani (2012) employed the theory of planned behavior to examine how driving behaviors change using an on-board eco-driving feedback tool. The findings showed that that setting goals for participants and real-time feedback resulted in drivers increasing their fuel efficiency. Rutty et al (2013) examined the impacts of eco-driving on Calgary's municipal fleet. In the study fifteen drivers in a study to reduce the emissions associated with vehicle idling. The results of the study showed that average vehicle idling was reduced by between 4% and 10% per day. Other road users have been examined to ascertain if eco-driving can be applied to public transport drivers. Sromberg and Karlsson (2013) examined bus drivers in Sweden using in vehicle feedback tools to reduce harsh acceleration. The findings of the study showed that a 6.8% reduction in fuel usage occurred in the study period.

The research presented in this section shows the benefits of eco-driving, while the results are modest; it shows how eco-driving strategies can be successful. While as mentioned these results were modest they these policies can be used in a suite of policies to reduce emissions.

1.2 Methodology

1.2.1 Data Collection

The eco-driving trial started in January 2012 and the results presented in this report track the vehicles up to October 2012. Five different groups were analyzed during the trial period.

Group A: This group had 82 users and these users were provided with on-board active driver feedback for the duration of the trial and access to webfleet online.

Group B: This group had 27 users and these users were provided with on-board active driver feedback for the duration of the trial and from the 1st of July 2012 were given access to webfleet online.

Group C: This group had 27 users and for the first two months had no interventions. Then this group was given both on-board active driver feedback and webfleet online.

Group D: This group had 16 users and was not given any on-board information. This group was given webfleet online from March 2012.

Group E: This group had 15 users and they received no information at all on driving style. This group was used a reference group to compare the other groups.

Table 1 presents a description of the trip characteristics of those participants in each of the groups in the trial. Table 1 presents the average of the sample and the standard deviation of the characteristic. The standard deviation provides an indication of the range in the values recorded. The results show that characteristics such as average daily travel distance and the number of trips are similar, this indicates that the comparisons preformed later in the report are based upon drivers with similar characteristics.

	Group A		Group B		Group C		Group D		Group E	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Number of trips	3.4	2.0	3.2	2.0	3.6	2.0	3.4	1.9	3.4	1.8
Distance (in KM)	56.8	64.6	61.2	72.8	57.3	66.5	69.0	79.7	49.0	63.1
Driving time (in mins)	59	50	62	56	59	51	69	60	55	47
Fuel usage (in liters)	3	3.4	3.5	4.2	3.2	3.7	3.7	4.5	2.8	3.4
Idle time (in mins)	9	8	12	11	13	16	13	15	22	18
Average time spent speeding (in mins)	3	4	4	7	5	6	5	6	3	6

Table 1 Description of data collected

1.3 Eco-driving model

This section of the deliverable presents the model schematic used to analyze the results of the eco-plus trial. Figure 1 shows the model schematic. The model shows that the eco-driving interventions of on-line driver feedback and on-board driver feedback are used to improve driving style. The success of these interventions is then measured by monitoring the reductions in idling time, fuel consumption and speeding. These then all have an over all impact on the reductions in emissions. This model is tested in the next section and the results are presented.

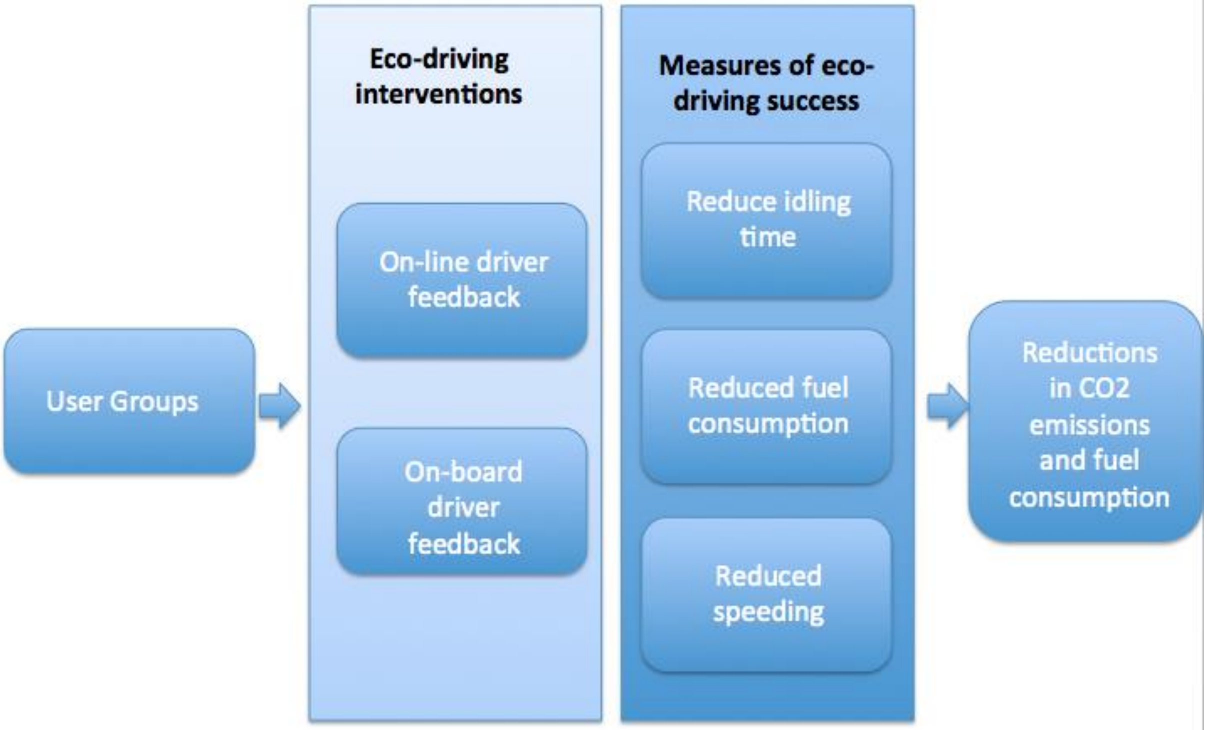


Figure 1 Eco-driving Model

1.4 Analysis and Results

1.4.1 Reductions in CO2

The following section reports the reductions in CO2 emissions from each of the five groups examined in the trial. In order to examine what if any reductions in CO2 occurred the emissions from the first two weeks of driving were averaged and used a baseline to compare subsequent weeks for reductions in emissions. Figures 2 – 6 present the findings for the average reductions in CO_2 emissions per KM for each of the 37 weeks of the trial. The results for each of the groups show that there was a decrease in CO_2 emissions per KM driven. The results from each of the groups are also presented in Tables 2-4 and more discussion of the results is presented in section 4.2.

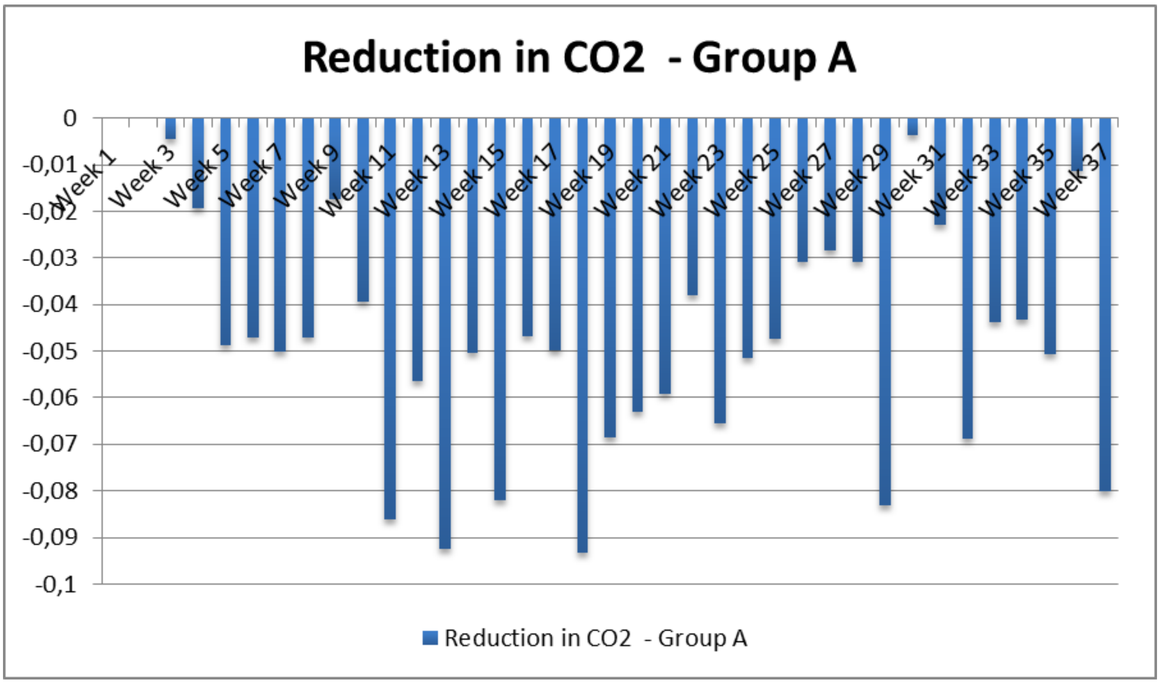


Figure 2 Reduction in CO2 - Group A

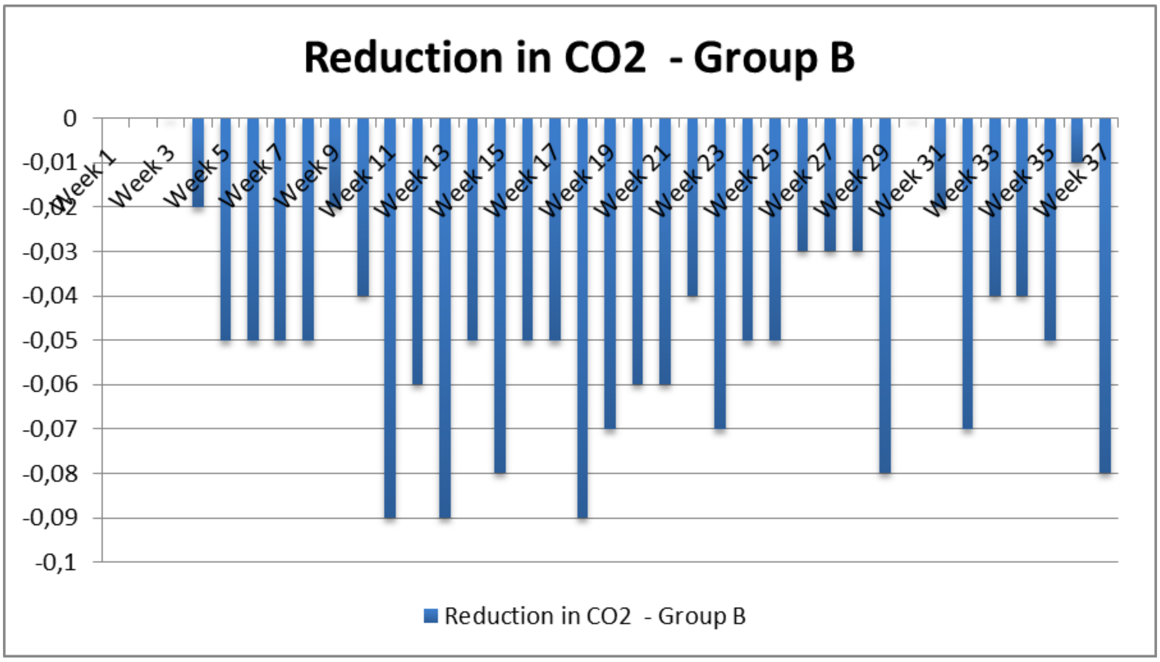


Figure 3 Reduction in CO2 - Group B

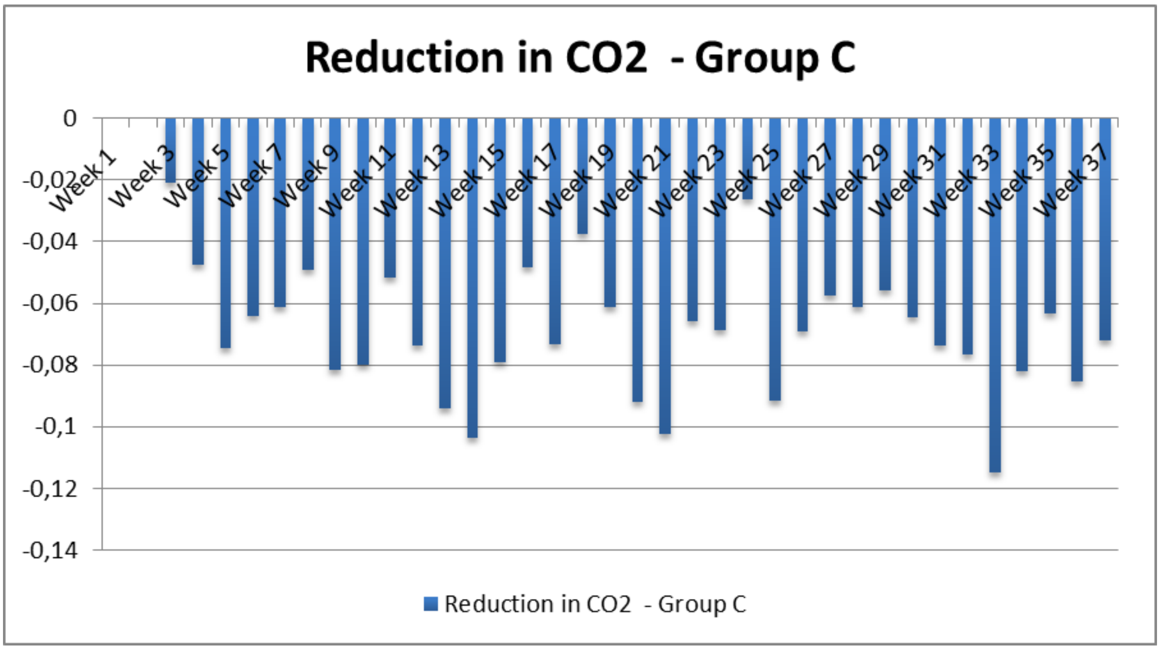


Figure 4 Reductions in CO2 - Group C

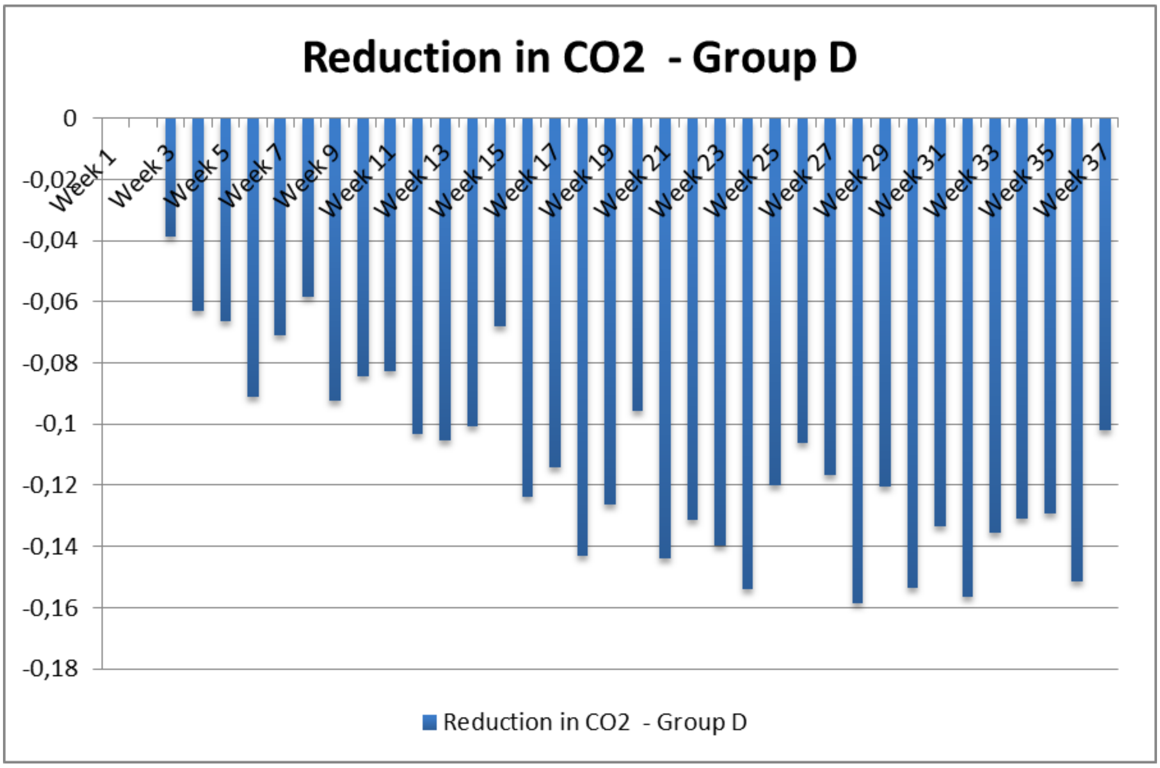


Figure 5 Reductions in CO2 - Group D

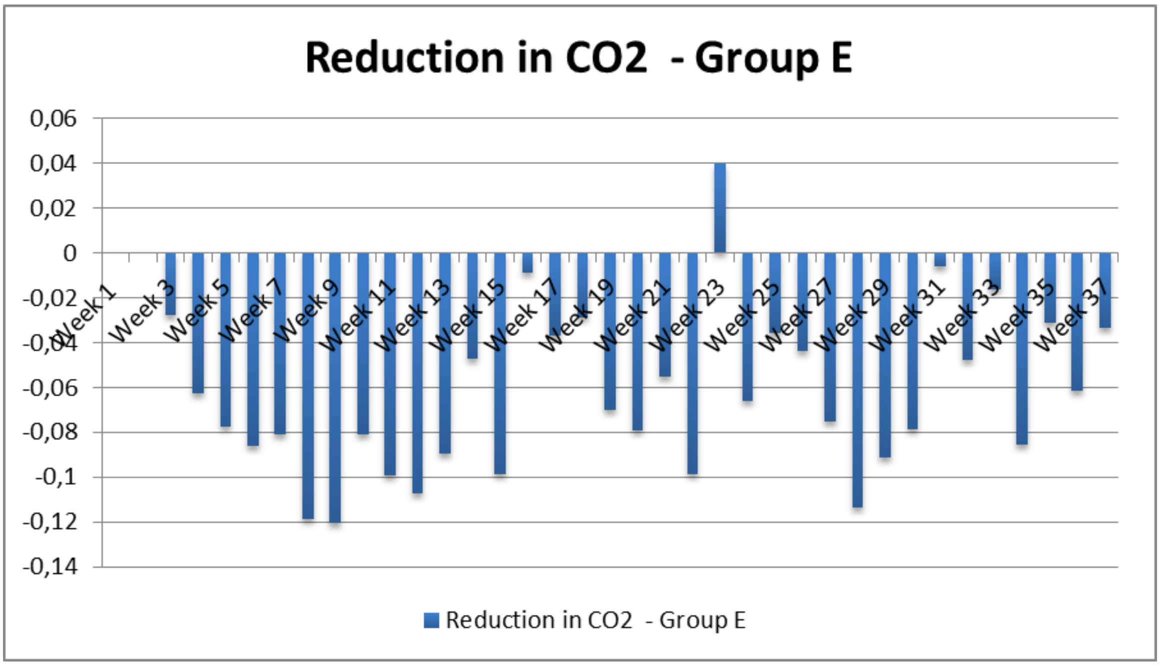


Figure 6 Reductions in CO2 - Group E

Tables 2 – 4 present similar data to that shown in Figures 1-5. The results show the values for emissions per KM in average emissions and the standard deviation (S.D.). The results presented in the table show the first two-week average and then the data is broken down into 5-week periods to show the changes in emissions over time. The results show that for Group A that there is a drop in emissions from week 13 to week 27 and then there is a rebound and the average emissions increases. The results for Group B show that there is a steady drop in the average emissions in the first 12 weeks of the trial. From week 13 to 22 the results show a greater decline in the average CO_2 emissions per KM compared to the other time periods. This time period was when those in Group B were provided with information via webfleet. The results for Group C also show a steady decline in the average CO_2 emissions. This group was provided with on-board information and webfleet from week 8. The results show that from week 8 there was a decrease in average emissions. The findings for Group D show that from week 12, when the participants got access to webfleet, that a decrease in average emissions was experience by those in this group. Group E was used as the reference group in this study, as they were given no extra information on driving performance. As one would expect there was little change in the over all average CO_2 emissions per KM in this group.

Weeks	Group A		Group B		Group C		Group D		Group E	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1 & 2	133	6.2	138	2.1	144	0.6	144	3.1	142	7.2
3-7	135	6.7	135	2.9	137	3.0	137	2.7	138	3.5
8-12	136	10.4	133	3.5	135	2.3	134	2.4	132	2.4
13-17	130	1.2	130	2.9	133	3.7	131	3.1	139	5.5
18-22	131	3.5	130	2.7	134	3.7	128	2.9	138	3.9
23-27	130	3.5	133	2.1	135	3.5	128	2.8	142	6.7
28-32	136	3.7	134	4.6	135	1.2	125	2.5	138	6.2
33-37	136	1.7	133	3.4	132	2.8	128	2.6	141	4.1

Table 2 Overall reductions in CO_2 emissions per KM

In order to ascertain if drivers had different behavior on weekends compared to weekdays the dataset was divided between weekdays and weekends to determine if there was any difference. Table 3 presents the results for the weekends and Table 4 presents the results for weekdays. The findings for Group A shows that on average participants had higher average emissions on weekends. The results from Group B also show a similar trend with higher average emissions on weekends compared to weekdays. These trends are also shown for Groups C, D and E.

Weeks	Group A		Group B		Group C		Group D		Group E	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1 & 2	134	5.4	142	6.8	135	24.3	162	17.3	136	1.0
3-7	138	4.2	143	7.3	147	9.2	142	14.3	147	4.9
8-12	137	4.3	139	5.3	129	11.7	137	4.4	135	2.5
13-17	135	2.1	133	5.0	139	4.4	130	4.5	138	8.7
18-22	137	3.7	129	2.0	139	1.5	129	7.3	144	6.1
23-27	137	4.9	129	6.4	139	5.9	128	1.8	150	23.1
28-32	138	2.9	138	6.4	133	6.6	126	7.6	138	7.7
33-37	136	6	140	5.7	131	4.9	130	1.8	146	8.3

Table 3 Reductions in CO2 emissions - Weekends

Weeks	Group A		Group B		Group C		Group D		Group E	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1 & 2	119	10.0	134	12.1	134	14.6	133	7.4	143	7.6
3-7	133	17.1	135	3.9	137	6.2	138	1.8	136	3.8
8-12	136	14.0	132	3.6	136	4.8	138	1.9	131	3.2
13-17	125	7.8	129	2.1	132	2.9	132	3.1	138	6.1
18-22	123	14.8	124	16.4	134	3.9	127	5.2	137	5.1
23-27	120	17.1	120	23.8	134	3.9	128	4.6	138	5.1
28-32	136	3.2	136	7.4	134	4.6	125	2.8	136	6.7
33-37	135	2.3	139	5.8	132	2.7	127	1.2	139	4.3

Table 4 Reductions in CO2 emissions - Weekdays

1.4.2 Comparison between results

This section of the report presents a comparison between the results found for each of the groups to determine which interventions had the greatest impact upon CO2 emissions per KM. Table 5 presents the findings of a comparison on the average weekly CO2 emissions per KM from groups A-D and compared against group E. This set of results shows how each of the test groups preforms against the control group. The results in Table 4, if positive show that the control group being compared had a reduction in CO2 emissions in the week in question compared to the control group. Whereas a negative result would indicate that that the group being compared to the control group had higher average emissions per KM driven. The results in Table 4 show that on average each of the test groups had a greater reduction in CO2 emissions compared to the control sample. The results show that Group D performed the best with an average reduction in emissions of 6% compared to the control group. Groups A and B also had on average a 4% reduction in CO2 emissions compared to the control group with those in group C having a 3% decrease in emissions.

	Group A	Group B	Group c	Group D
Week 1	7%	1%	-5%	-3%
Week 2	8%	6%	2%	1%
Week 3	8%	3%	1%	2%
Week 4	3%	1%	0%	1%
Week 5	-7%	3%	2%	-1%
Week 6	3%	1%	0%	1%
Week 7	3%	2%	0%	0%
Week 8	-3%	-2%	-5%	-6%
Week 9	-2%	-5%	-2%	-2%
Week 10	-12%	1%	2%	1%
Week 11	3%	4%	-3%	-1%
Week 12	0%	0%	-2%	0%
Week 13	4%	6%	3%	2%
Week 14	7%	6%	8%	7%
Week 15	1%	4%	0%	-3%
Week 16	12%	10%	6%	14%
Week 17	10%	7%	6%	9%
Week 18	12%	13%	3%	14%
Week 19	7%	6%	1%	7%
Week 20	4%	4%	3%	3%
Week 21	4%	6%	7%	11%
Week 22	-1%	-1%	-2%	4%
Week 23	18%	18%	14%	22%
Week 24	5%	4%	-2%	11%
Week 25	14%	7%	8%	10%

Week 26	6%	4%	5%	8%
Week 27	2%	1%	0%	5%
Week 28	-7%	-3%	-4%	6%
Week 29	-3%	5%	-2%	4%
Week 30	3%	-2%	1%	10%
Week 31	7%	8%	10%	15%
Week 32	7%	8%	5%	14%
Week 33	6%	9%	13%	15%
Week 34	1%	1%	2%	6%
Week 35	5%	8%	6%	12%
Week 36	1%	0%	5%	11%
Week 37	7%	11%	6%	8%
Average reduction in CO2 Emissions	4%	4%	3%	6%

Table 5 Groups compared to control group

1.4.3 Minutes spent idling

One of the main bad habits that drivers display when driving is idling. This is when drivers for one reason or another have the car engine running but the car is not moving. In this study participants that were idling for periods of longer than 5 minutes, this data was recorded. Those in participants were that were provided with on-board information would have been provided with information on their idling. Figures 7-11 show the weekly average amount of idling for each of the groups examined in the study. The results show that those in groups A, B and C had lower rates of idling with a weekly average idling time of 8, 8 and 10 minutes respectively. Those in groups D and E were shown to have larger idling times of 12 and 23 minutes per week respectively. The results of this analysis shows that those in groups receiving on-board information and substantially less idling times that those not receiving this information.

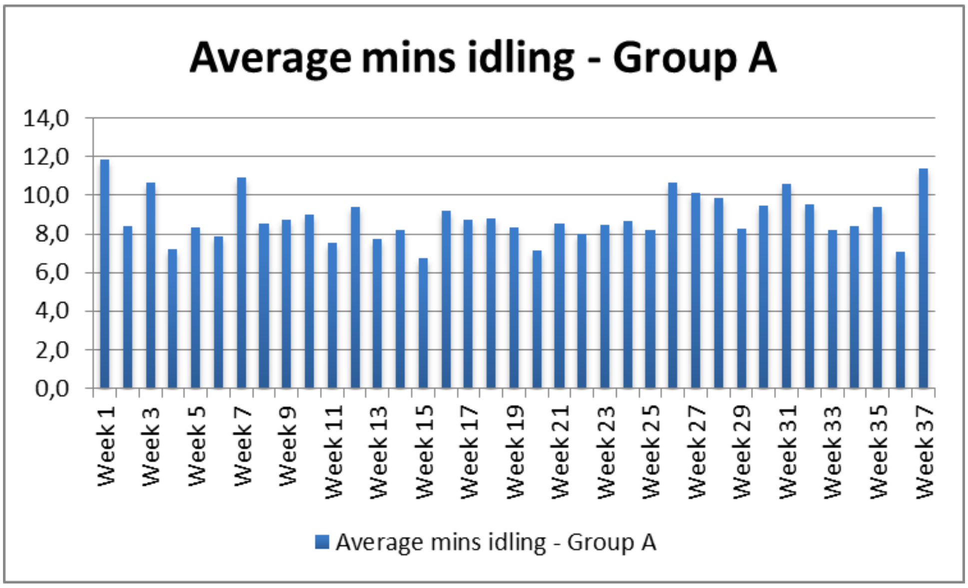


Figure 7 Average mins spent idling – Group A

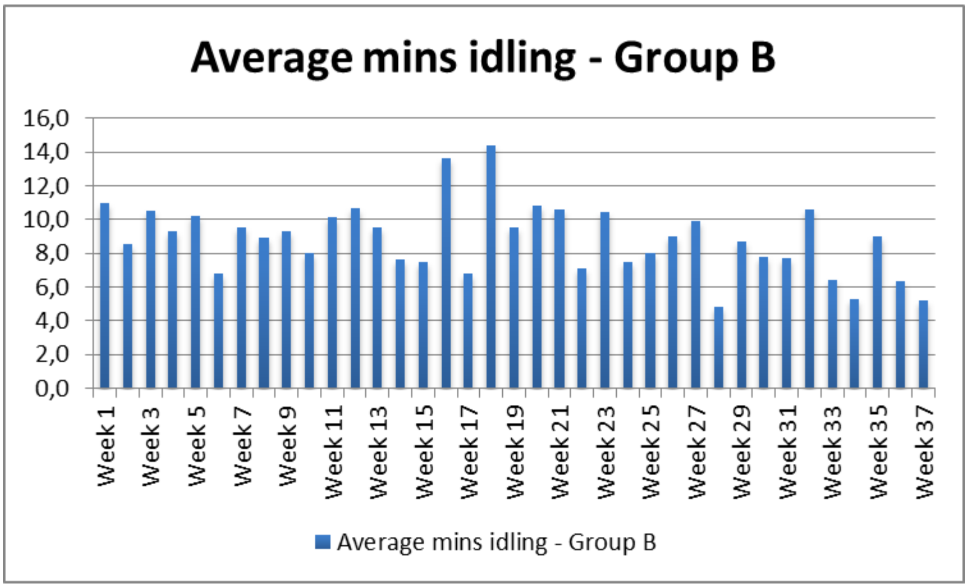


Figure 8 Average mins spent idling - Group B

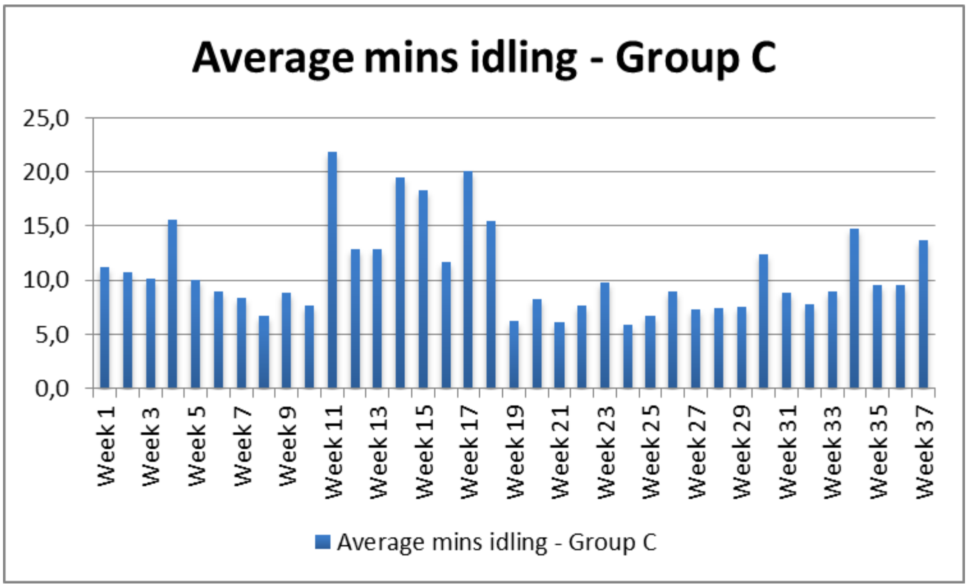


Figure 9 Average mins spent idling - Group C

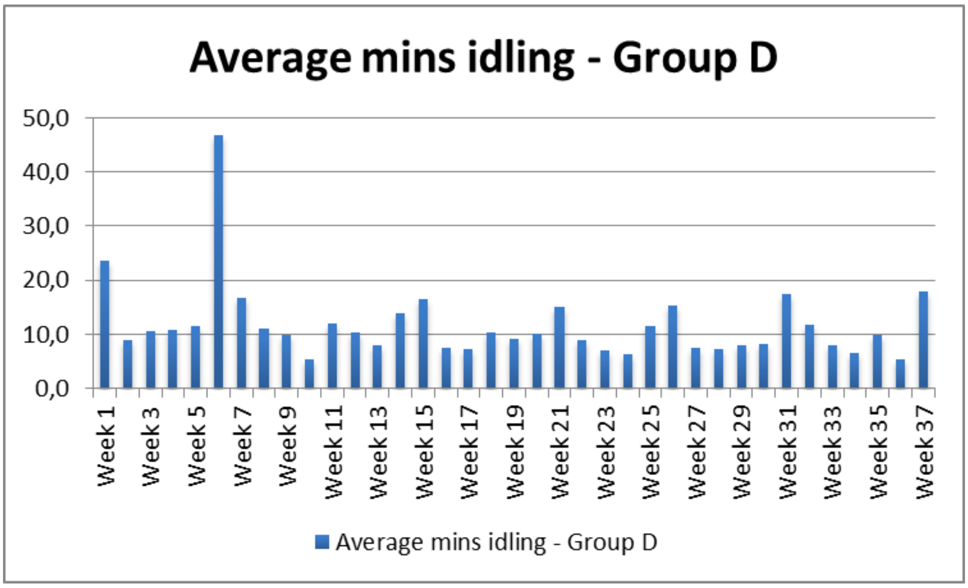


Figure 10 Average mins spent idling - Group D

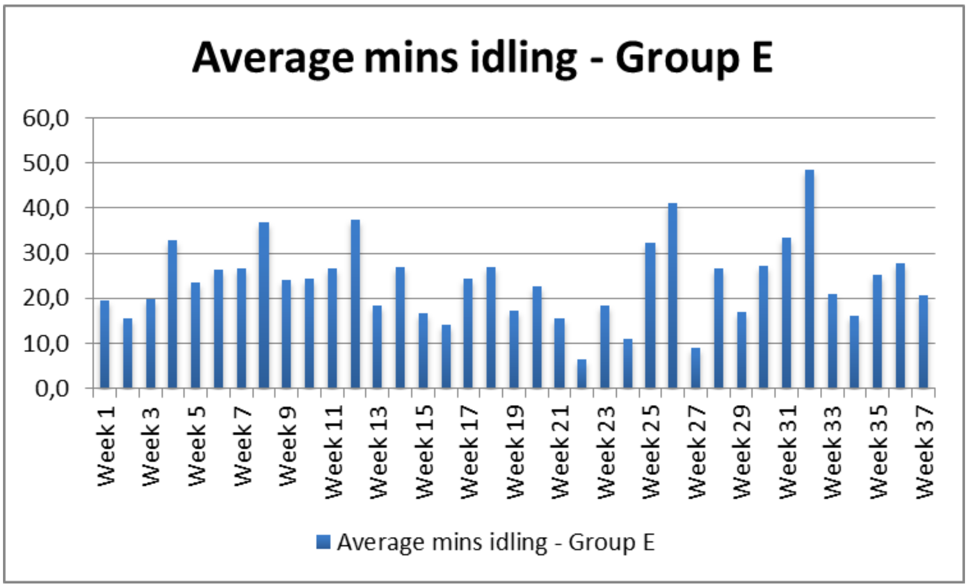


Figure 11 Average mins spent idling - Group E

1.5 Discussion and Conclusions

The results presented in this section of the deliverable show the success of the eco-plus (eco-driving trial) conducted as part of the PEACOX project. The findings of the trial conducted in the Netherlands concur with those presented in the international literature, in that while the savings in emissions and fuel consumption were modest, they do represent a significant reduction in emissions and a reduction in idling time.

2. Emissions Exposure Model

2.1 Introduction

2.1.1 Background of the deliverable

Vehicle emission has an adverse impact on the environment at local and global scales. Many of these pollutants are carcinogenic like benzene, and many cause respiratory problems (e.g. PM2.5. NO_2 and O_3), cardiac admissions (e.g. PM) (Katsouyanni, et al., 2010) . Among them, nitrogen dioxide (NO_2) and particulate matter (PMx) are causes of concern for Europe. Traffic has been considered as a primary source of NO_2 and is also one of the main sources for PM (O'Dwyer, 2011). Investigations noted that if PM10 concentration was reduced to 20 µg/m3 on all days, in Europe, it would lead to a decrease of 15 premature deaths per 1,00,000 inhabitants per year (Katsouyanni, et al., 2010).

The PEACOX project has set grounds for handling eco-friendly driving issues more efficiently along with other set targets. The aim of the third work package of the PEACOX project is to build models which will estimate emissions and exposure levels for travelers. A trip with an origin and a destination may have many possible routes. Thus, the user will be able to choose an option from a given set of options to peruse his/her journey for his/her destination in safer and healthier ways.

2.1.2 Scope of the work package

The exposure model, under the task WP3.4, has been identified as city specific and thus, requires building two different models for two cities (i.e. Dublin and Vienna). Considering the primary aim of work package 3, the objectives of the work pack include:

Objective 1: Selection of appropriate approaches for exposure modelling in the current context.

Objective 2: Apply the model's outcome at route level.

Objective 3: Replicate the model for another city (Vienna).

To carry out the objectives a lead model was developed based on Dublin city where different approaches were tested top obtain a best fit model with the available data. The Vienna model was then developed following the same methodology.

2.2 Selection of exposure modelling approach

2.2.1 Required features for exposure modelling at route level

Air pollution exposure concentration at a particular place is very complex and involves many atmospheric chemical and local physical processes. Difficulties arise in developing models that predict exposure concentrations for all over a city in real-time where the study area is more than 100 square kilometres. Besides, the data availablity for building such a model is often in low resolution e.g. daily PM count. In addition, citywide models require a large number of monitoring stations that have been capturing data for a long period of time in the same resolution. For instance, most of the monitors in Dublin city capture PM_{10} data on a daily average basis. Thus, the highest resolution that model can capture is daily concentration, at least for the Dublin city model. Therefore, the temporal resolution is restricted to daily average. The level of exposure for a person to a particular pollutant, also depends on the exposure duration as well as on the travel time of a person. The exposure model is capable of providing real-time exposure ratings based on the breakdown of daily exposure concentration and real-time travel time information for different routes.

2.2.2 Overview of the methodology

According to the methodology developed (see Figure 12), two different steps are involved in the exposure models predictions. For the first level of analysis, it is necessary to estimate the exposure concentration for every road link. For defining exposure concentration, land use regression (LUR) can be used among the candidate models, which utilises the monitored levels of the pollutant of interest as the dependent variable, and variables such as traffic, topography, and other geographic variables are considered as independent variables in a multivariate regression model (Gilliland et al., 2005; Ryan & LeMasters, 2008). The LUR model is suitable for this research for following reasons: 1) The incorporation of site-specific variables into this method detects small area variations more effectively than other methods of interpolation (Briggs et al., 1997; Gilliland, et al., 2005); 2) the levels of pollution may then be predicted for any location using a regression model (Ryan & LeMasters, 2008). Besides, the predictive performance of the LUR model is no less than that of alternatives such as dispersion modeling (Gulliver, et al., 2011; Hoek, et al., 2008). The landuse regression will be in the form of Equation (1).

$$E = C_0 + A_1X_1 + A_2X_2 + A_2X_3 + \in \quad ; \qquad (1)$$

Where, E = Exposure Concentration; X_1= Traffic data; X_2 = Land use data ; X_3 = Weather data ; $\in$ = Error ; A_1= regressing coefficient.

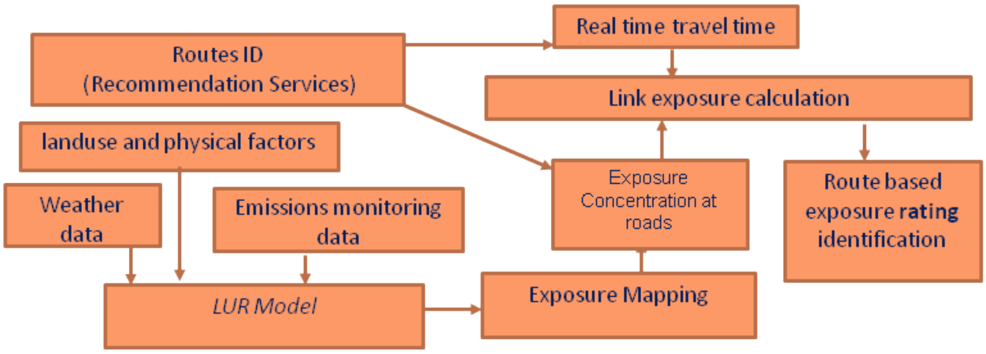


Figure 12 Methodology of the Model

The explanatory power for the LUR model was reported from R-squared .17 to .97 for various types of pollutants. However, R-squared for PM_{10} ranges from .45-.90 for both model calibration and validation (Hoek et al, 2008) in the reference studies. RMSE for PM_{10} by Briggs's et al. (2007, 2011) showed values of 6.7 and 3.3 respectively. With the limitation of the number of Air Quality monitors available in Dublin, it is not expected that high accuracy of the resulting of the models could be obtained. The spatial coverage of the monitors is much lower than the recommended minimums in the areas of interest, 7 for 117 sq.km in Dublin and 13 for 414.6 square km for Vienna.

The LUR model was developed based on the limited number of monitoring stations and the model was used to predict exposure for selected different points of interest in the city.

Those selections were based on grouping analysis. This technique generated a good number of sites to extrapolate exposure concentration across the city area using the kriging technique. The generated exposure maps were then intersected with road centroid lines in a GIS environment.

Exposure level estimation is required for each route in order to integrate the model with other components of the PEACOX project. The original exposure model (Landuse regression model) provides exposure concentration along the route; however, a new derived factor 'Dose' may be required for rating the routes. The dose will indicate the level of exposure and it is the amount of pollutant that someone inhales during travel, and thus, it is a function of exposure concentration of a pollutant, travel time and inhalation rate. At the second level, the technique was determined to calculate the exposure rating by following equation (2):

$$D = \int_{t1}^{t2} C(t).\,\delta(t).\,IR(t,m).\,dt\ ; \qquad (2)$$

Here, D=dose (µg); $\delta(t)$= Time factor (unitless); $IR(t,m)$= Inhalation rate (m^3/hr) based on mode; time in hour; and $C(t)$= µg/m^3

The dose will be calculated for each alternative route. The total value calculated from different modes provides the possible dose for each route. Although the outcome of the modeling will provide dose, the value will be expressed for the users as a band score. The level of concentration will be given in a scale rating where 'A' will indicate excellent travel environment. Similarly, 'B' refers 'Good', 'C' indicates 'Average', 'D' as 'Poor', and 'E' refers 'Unhealthy' conditions. While there will be a number of alternative routes between an origin and destination, dose (µg) for each alternative will be calculated and lower dose will be rated as 'A' and so on.

2.2.3 Selection of the pollutant

PM10 was selected for this task as a generic pollutant. Several reasons exist to justify considering PM10 as a representative pollutant of air quality. PM is a known source pollutant in traffic emissions as well as originating from re-entrained dusts, brake and tyre wear, sea spray, combustions, etc. PM is one of the main pollutants of concern in Ireland as well as in the EU and is monitored routinely by local governments across the EU.

2.2.4 Data Sources

A large quantity of data has been collected in order to build exposure models for Vienna and Dublin. The sources of data include:

- Environmental Protection Agency (EPA, 2012),
- umweltbundesamt (umweltbundesamt, 2013),
- Dublin City Council,
- Met.ie,
- Central Statistics Office (CSO, 2012),
- Trinity College Dublin (internal source),
- Dublinked (Dublinked, 2012),
- Central institute for meteorology and geodynamics, Vienna, Austria (ZAMG, n'd),
- GADM database of Global Administrative Areas (GADM, 2012),
- Geofabrik GmbH (GmbH, 2012),
- European Environment Agency (EEA, 2012a),
- CGIAR (CGIAR, 2012).

2.3 Overview of the data for lead model

The concept of “Lead Model” has been included to avoid duplication of the same type of information that has been generated in the building of Vienna Model. Details of the Dublin exposure model have been included here in sections 3 to 5. This model is then followed by the development of the Vienna Model.

2.3.1 Geographical coverage of the Dublin Model (Lead model)

The geographical coverage of the Dublin city area is approx 115 sq.km. This area is under the jurisdiction of the Dublin city council (Figure 13).

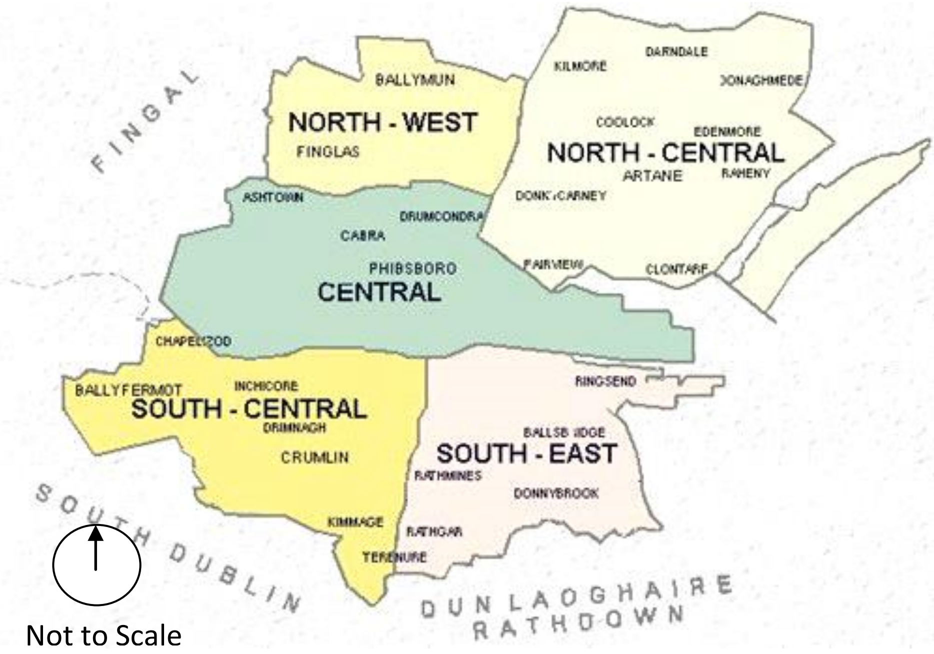


Figure 13 Dublin City

Source: DCC

2.3.2 Air Quality monitoring stations and monitoring data

There are almost 15 Air Quality monitoring stations (both temporary and permanent) in the greater Dublin area (Figure 14). However, only some of them are useful for building the exposure model due to their longer temporal coverage of data.

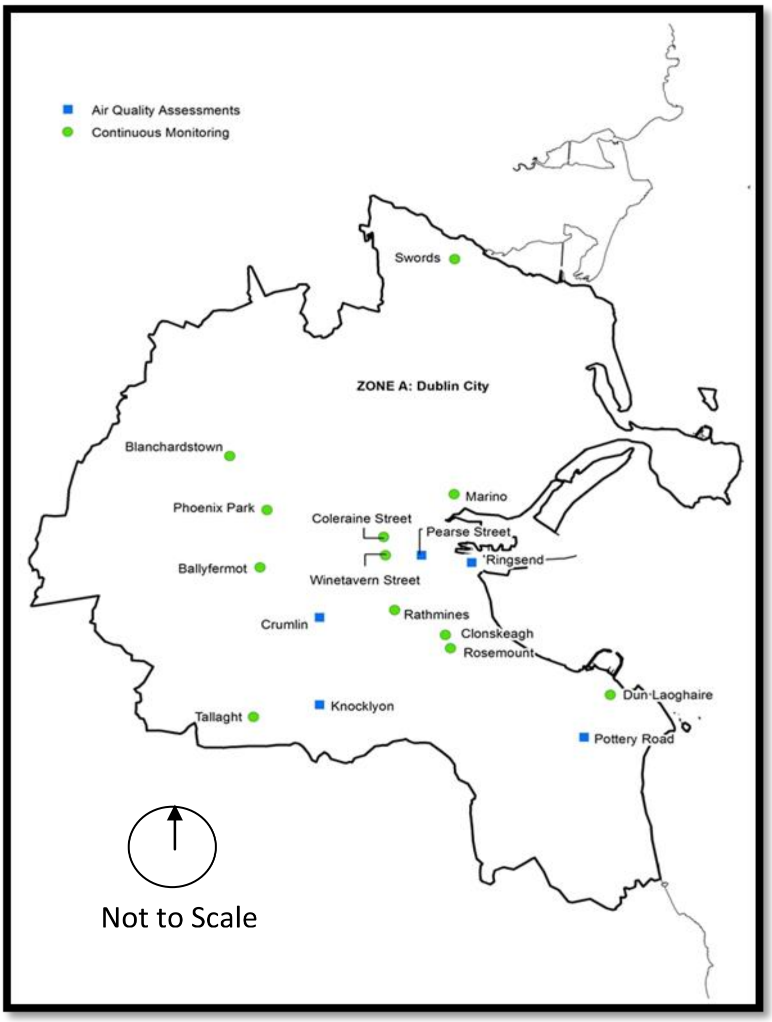


Figure 14 Air Quality Monitoring stations in Dublin City

Source: EPA

The number of Air Quality monitoring stations where PM_{10} data is available for LUR model was seven: Ballyfermott, Coleraine Street, Knocklyon, Marino, PhoenixPark, Winetavern and Ringsend (highlighted in Figure 15). There are also three weather monitoring stations available in this area: Phoenix park, Casement at North and Dublin airport.

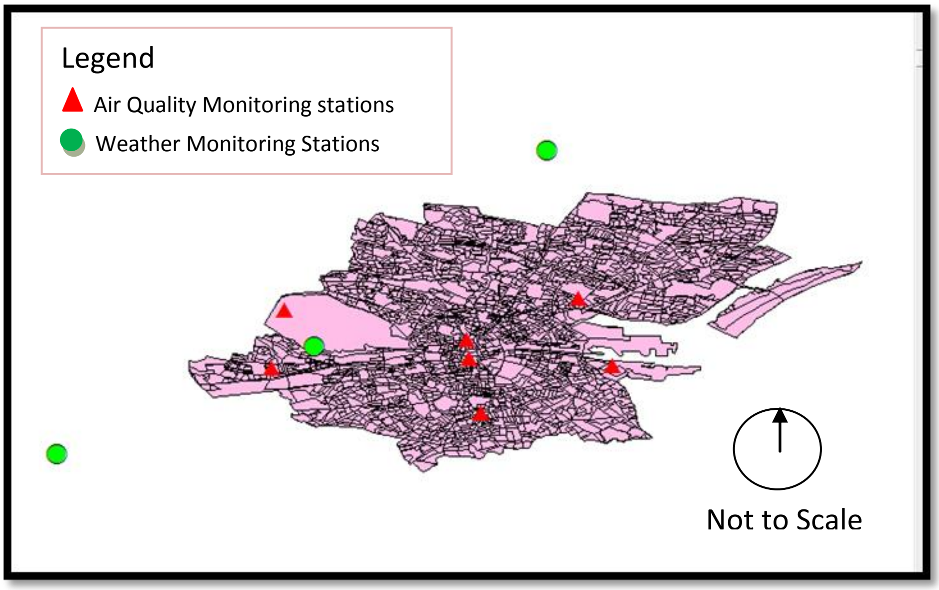


Figure 15 Air Quality and Weather monitoring stations in Dublin City

2.3.3 Response and explanatory variables

The following variables were used to develop the regression equation:

Weather:

- Rainfall (Daily total-2007-2009)
- Wind direction (in Degree-2007-2009)
- Radiation (Daily Average-2007-2009)
- Stability index (Daily-2007-2009)
- Wind speed (Daily Average-2007-2009)
- Temperature (Daily Average-2007-2009)
- Humidity(Daily Average-2007-2009)
- Dew Point (Daily Average-2007-2009)

Physical Parameter:

- Distance from Coast (in km)
- Pollutant's distance from the monitors (in km)
- Pollutant Angle(in Degree)

- Population Number (2007-2009, estimated)
- Housing Stock (2007-2009, estimated)
- Road Length (in km, 2011)
- Digital Elevation Model (DEM)/Altitude (90m at the equator)
- Road types (Motorway and/or with link, Trunk and/or with link, Primary and/or with link, Secondary, Tertiary)
- Land use type (commercial and open space, from land cover data, 2006)

2.3.4 Data management and processing

2.3.4.1 *Data pre-processing and selection of data type:*

ARCGIS, SPSS and Excel softwares were used to extract and process data for modelling.

Landuse:

The considered GIS dataset had a predefined land use category. To use these data for modelling, several land use categories have been merged into groups. The land use data has been recoded based on their impact on exposure concentration. Thus, two categories have been identified that have spatial positive and negative relationships with the exposure concentration.

- Industrial and commercial land use
- Open Space and water body (Predefined categories were: Intertidal flats, Land principally occupied by agriculture with significant areas of natural vegetation, Pastures, Non-irrigated arable land, and Green urban areas)

PM_{10} Data:

As the data was limited in terms of monitoring station numbers, it was decided to use panel data to achieve better results. However the following assessment shows that many years of data were also missing. If we consider the data availability on the monitoring stations for the period 2007-2009, this gives the best available option (2008-2010) to develop the LUR Model.

SL.	Monitoring Sites for Air Quality (PM_{10})	Data Available	Period 2007-2009	Period 2008-2010	Resolution
1	Ballyfermott *	2003-2010	-	-	Daily
2	Blanchardstown	2008-2009	2007	2010	Daily
3	Coleraine Street *	2001-2008	2009	2009,2010	Daily
4	DunLaoghaire	2008-2010	2007	-	Daily
5	Knocklyon	2008	2007,2009	2009,2010	Daily
6	Marino *	2001-2008	2009	2009,2010	Daily
7	PhoenixPark *	1996-98, '01-10	-	-	Daily
8	Rathmines *	1996-98, 2001, '03-05, '07-10	-	-	Daily
9	Tallaght	2008-10	-	-	Daily
10	Winetavern *	2001-2010	-	-	Daily
11	Ringsend *	2009-10	2007,2008	2008	Daily
Interpretation			8 Missing years	8 Missing years	

*Within study area.

Table 6 PM_{10} data availability in Dublin City

The following table (7) outlines the average exposure concentration is high in the central areas (Coleraine Street and Winetavern street) and lowest at peripheries (Phoenix Park) in Dublin as expected.

Station	2007			2008			2009		
	Minimum $\mu g/m^3$	Maximum $\mu g/m^3$	Average $\mu g/m^3$	Minimum $\mu g/m^3$	Maximum $\mu g/m^3$	Average $\mu g/m^3$	Minimum $\mu g/m^3$	Maximum $\mu g/m^3$	Average $\mu g/m^3$
Ballyfermott	2.64	78.47	14.82	2.50	43.19	11.64	1.53	46.10	12.44
Coleraine	4.31	75.28	18.43	4.58	93.47	18.54	-	-	-
Rathmines	1.20	87.92	16.69	1.00	101.30	16.91	2.36	59.58	14.74
Marino	1.67	74.31	13.41	2.50	75.00	12.62	-	-	-
PhoenixPark	1.53	66.19	11.72	1.39	59.44	10.74	2.08	38.89	10.19
Ringsend	-	-	-	-	-	-	5.20	36.52	14.40
Winetavern	3.19	93.47	18.30	1.69	82.36	17.49	1.39	55.83	17.29

Table 7 PM_{10} in different monitoring stations (2007-2009)

Wind Index:

Wind index is a strong determinant of pollutant concentration. The wind index has been calculated (Chen et al., 2010) based on equation (3):

$$\varphi = \frac{(1-\cos(\emptyset-\theta))}{2}; \qquad (3)$$

Where, Wind Index= φ ; $\emptyset$ = Euclidian direction from the nearest major road to monitoring site; θ = Wind direction in respect of true north

Stability Class

Stability class refers to the state of the atmosphere that is resisting or enhancing vertical motion. Different stability states can be categorised based on wind speed and solar radiation. Stability class for Dublin was adopted here as an additional explanatory variable.

Weather data source selection

Weather data from Phoenix Park has been used primarily for model development, except for solar radiation and wind data. Data from Dublin airport station has been used for these latter two weather variables. Missing data in Phoenix Park have been replaced using Dublin airport station's data. The following Table 8 shows the variation of the weather data used in the model.

Weather Variables	2007			2008			2009		
	Minimum	Maximum	Average	Minimum	Maximum	Average	Minimum	Maximum	Average
Temperature (C)	-.16	17.77	10.25	-.74	17.55	9.77	-.90	18.29	9.79
Humidity(%)	61.88	98.67	82.95	63.50	99.42	83.48	62.96	99.29	84.99
Dew Point (C)	-3.35	14.87	7.31	-2.71	14.73	6.92	-4.42	16.44	7.21
Wind speed (m/s)	1.52	13.95	5.66	1.78	14.34	5.94	1.22	14.04	5.68
Radiation (W/m^2)	158.61	1.02	51.20	2.47	113.90	33.15	1.09	123.41	38.20
Rainfall (mm)	.00	51.60	-	.00	58.70	-	.00	38.80	-
Stability Class	4.00	5.00	-	4.00	5.00	-	4.00	5.00	-
Wind Index	.00	1.00	.40	.00	1.00	.40	.00	1.00	.36

Table 8 Weather variables in Dublin Area (2007-2009)

Population and housing stock

There was no data available for housing stock and population for the period of 2007-2009 for the areas of interest. However, the Census data were available for 2011 and 2006 for population and housing stock and thus simple extrapolation has been used for estimating data for 2007, 2008 and 2009. The resolution of the data is at the small area level (lowest census boundary for Irish database).

2.3.4.2 ***Data extraction and sorting:***

PM and weather data have been sorted in Excel software, whereas spatial data has been extracted in a GIS environment. Different overlay data management tools, and spatial analysis tools have been deployed to obtain data. To get information around the Air Quality Monitoring stations buffer operations was used in GIS environment. A buffer in GIS is a zone around a point measured in units of distance. The distance of the buffers for each attribute (e.g. Population, road length) was determined based on relevant literature review and site characteristics. The concept captures the physical properties of the areas that might have an influence on the PM_{10} concentration in the air quality monitoring stations. The following buffer sizes (Figure 16) were considered to extract data from GIS shape files:

- Population Number/ Housing Stock/ Altitude --- 500m
- Road Length ---- 100m, 350m, 750m
- Land use Area according to type ----1000m

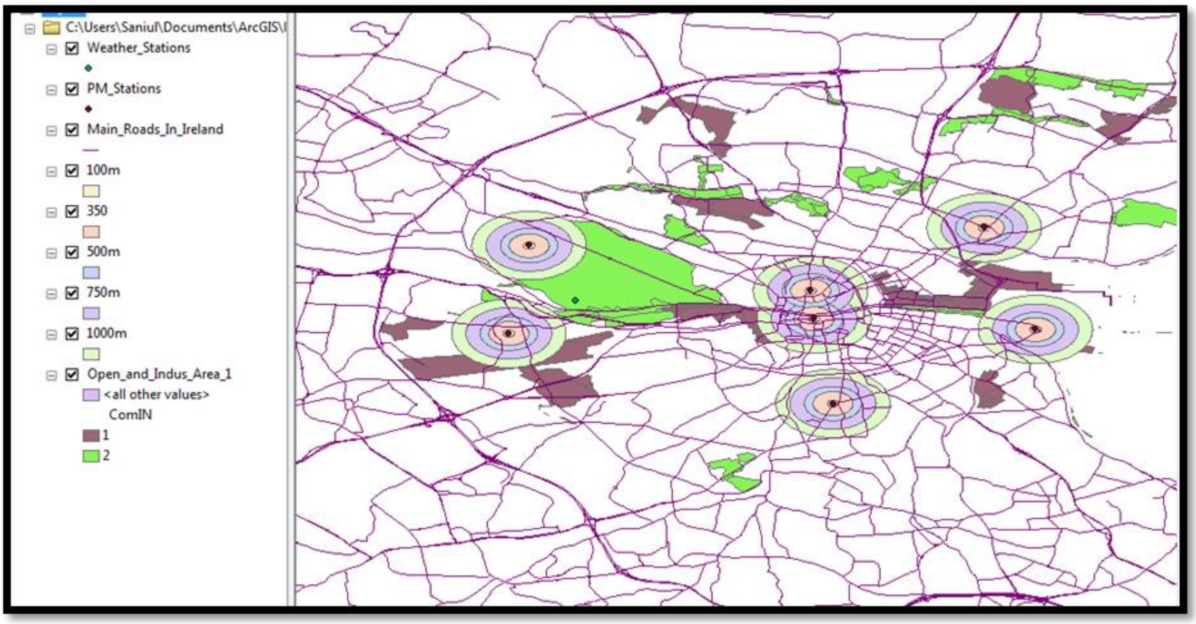


Figure 16 Different buffer sizes around the Air Quality monitors in Dublin

Population and housing stock for buffers have been calculated based on densities of population and housing stock in relevant small areas and the area covered by the buffer boundary. The proportions of buffer area were multiplied by the density of the corresponding small areas to determine population and housing stock for the year 2011 (Table 9). Later, back projection was conducted using a simple growth factor for determining the values for 2007-2009 at each station. However, population and housing stock have been considered constant for the phoenix park area (unpopulated national park).

NAME	Demographic Variables (500m buffer)	
	Housing Stock 2011	Population 2011
Rathmines	5681	9295
Ballyfermot	2825	7184
Ringsend	917	2422
Winetavern Street	9698	18151
Coleraine Street	8663	16484
Phoenix Park	292	1039
Marino	2438	4472

Table 9 Demographic Variables within 500m buffer

The land use and transportation variables around each station are given in tables 10 and 11.

Landuse Variables	Ballyfermott	ColeraineSt	Marino	PhoenixPark	Rathmines	Ringsend	Winetavern
Coast Distance in km	9.50	2.79	4.17	.96	8.20	.20	3.05
Altitude average in meter (500m)	41.85	14.38	27.09	13.47	52.94	4.53	12.15
Commercial area in Sq.km(1000m)	.66	.02	.00	.04	.00	.25	.39
Open Space area in sq. km(1000m)	.08	.29	.00	.29	1.96	.02	.06

Table 10 Values of land use variables around each monitoring station

Transportation variables	Ballyfermott	ColeraineSt	Marino	PhoenixPark	Rathmines	Ringsend	Winetavern
Major Road length in km(100m)	.00	.19	.18	.35	.00	.40	.44
Major Road length in km(350m)	1.70	1.84	1.26	1.43	.00	1.28	4.11
Major Road length in km(750m)	5.17	14.09	4.61	6.27	1.43	4.36	18.76
Minor Road length in km (350m)	3.17	2.65	2.53	2.73	1.10	1.95	4.14
Minor Road length in km (500m)	6.14	6.36	5.04	4.86	2.47	2.86	8.15
Nearest Major Road Distance (km)	.18	.26	.06	.04	.53	.05	.12

Table 11 Values of transportation variables around each monitoring station

2.3.4.3 ***Setting data for modelling***

There were 7673 observations available for building the model. However, due to absence of data in some variables, only 5535 observations were taken into account.

2.4 **LUR modelling for Dublin**

2.4.1 Land use regression models

The following important assumptions were made in the LUR model:

- The physical characteristics of the seven Air Quality monitoring sites are a good representation of the whole area.
- The land use and elevation data have been considered stationary over the years, whereas population and housing stock number have been assumed as having a constant growth rate.

To develop the LUR model the analysis was performed using R – statistical software. Having the limitation of the number of routine monitoring sites and the number of observations, a deviation has been used from the traditional approach of building an LUR model. For model validation common approaches were either leaving one station and carry out the cross validation for (n-1) times, or leaving a certain percent of the data for validation and establish the model using the rest of the data. Here, a few models have been developed at the initial stage and thus, best fitted models were redeveloped again with a certain percentage of the dataset and thus, validated against the rest of the dataset. The best performed model was then selected. The first model (M1) was developed with simple multivariate linear regression (**Fehler! Verweisquelle konnte nicht gefunden werden.**) with all the available explanatory variables.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1.205e+01	2.081e+01	-0.579	0.562519	
temp	2.443e+00	8.263e-01	2.957	0.003117	**
humidi	4.971e-01	1.603e-01	3.101	0.001939	**
dew_point	-3.205e+00	8.390e-01	-3.821	0.000135	***
wind_speed	-1.169e+00	5.489e-02	-21.299	< 2e-16	***
Radiation	1.372e-02	4.240e-03	3.235	0.001225	**
Rainfall	-1.432e-01	2.193e-02	-6.529	7.21e-11	***
Stability_Class	8.593e+00	4.689e-01	18.326	< 2e-16	***
Coast_km	7.265e+01	2.640e+01	2.751	0.005952	**
WindINdex	1.489e+00	3.721e-01	4.001	6.39e-05	***
Altitude	-1.308e+01	4.676e+00	-2.798	0.005167	**
Commer	-3.569e+02	1.356e+02	-2.633	0.008491	**
Open_Area	2.310e+01	8.434e+00	2.739	0.006189	**
mRd350_m	1.316e+02	5.301e+01	2.483	0.013074	*
mRd500_m	-5.744e+01	2.360e+01	-2.434	0.014968	*
Rdj750_m	NA	NA	NA	NA	
NearRdj_km	NA	NA	NA	NA	
Rdj350_km	NA	NA	NA	NA	
Rdj100_km	NA	NA	NA	NA	
Population	-2.575e-03	1.805e-03	-1.426	0.153903	
HS	-4.917e-03	5.574e-03	-0.882	0.377747	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Table 12 Regression Model Output (Model1)

The model shows an adjusted R-squared of 0.35. That means the model can explain 35% variability of the dataset. However, this model provides two insights: some variables have illogically negative correlation (e.g. Commercial area should be an anthropological source of PM_{10}) which might produce unrealistic results if chosen; secondly, four of the variables were not defined because of singularities i.e. an extreme form of multicollinearity/perfect linear relationship exists between variables and these can be replaced. Therefore, this provides an understanding that, there is an option for choosing the best combination of explanatory variables in the model. If negatively correlated variables were retained, the model could produce negative values if other variables (those having a higher impact on the model) are silent or have "0" values (e.g. Open space).

Before choosing the best combination of explanatory variables, another innovation has been used in model M1. Two dummy variables have been selected for the model, namely seasons and days of the week. The following table (13) provides grounds for choosing the first dummy variable. Although there is less variation of average wind speed across the seasons, the other two variables show significant variation. Higher rainfall reduces the PM_{10} concentration in the air, on the other hand, people operate solid fuel heating appliances, which in turn may cause an increase in emitted PM_{10} on cold days. In addition, the traffic is one of the primary sources of PM_{10} and the traffic volume varies according to the days of the week, e.g. weekdays vs. weekends. This provides the logic for choosing the dummy variable for days. Previous studies, (Chen, et al., 2010) for season (only) and (Maynard, A Coull, Gryparis, & Schwartz, 2007) for days (only) used such division in their models. Here, dummy variables have been used to make a better fit of the model with the data set. The model thus, yields an adjusted R-squared 0.37, an improvement of 2% extra explanatory power.

Subsequently, another statistical technique has been deployed as PM_{10} data was not normally distributed, whereas, regression assumes PM_{10} data should follow a normal distribution. To create PM_{10} data as normally distributed, natural log transformation of PM_{10} was considered in the model. Then log-level analysis has been performed, which yielded an adjusted R^2 of 0.43.

	Average Temperature (C)			Average Rainfall (mm)			Average Wind speed (m/s)		
Seasons	2007	2008	2009	2007	2008	2009	2007	2008	2009
Summer*	13.97	14.25	14.56	3.99	3.93	2.80	4.97	5.34	5.12
Winter	9.00	8.01	8.13	1.53	2.26	2.35	5.89	6.17	5.88
Difference	4.97	6.25	6.43	2.47	1.66	.44	-.91	-.83	-.76

*June,July and August (according to met.ie , http://en.wikipedia.org/wiki/Summer)

Table 13 Seasonal environmental data

Following the log-level model, the best logical combination of input variables (**Fehler! Verweisquelle konnte nicht gefunden werden.**) has been chosen for model M2 as below:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1.1283249	0.8525444	-1.323	0.185731	
temp	0.1440752	0.0439355	3.279	0.001047	**
humidi	0.0229599	0.0085121	2.697	0.007011	**
dew_point	-0.1764334	0.0445813	-3.958	7.67e-05	***
wind_speed	-0.0764062	0.0029307	-26.071	< 2e-16	***
Radiation	0.0011457	0.0002272	5.043	4.72e-07	***
Rainfall	-0.0069847	0.0011812	-5.913	3.56e-09	***
Stability_Class	0.3158587	0.0250940	12.587	< 2e-16	***
Coast_km	-0.1824649	0.0165637	-11.016	< 2e-16	***
windINdex	0.0829934	0.0196716	4.219	2.49e-05	***
Altitude	0.0343872	0.0038785	8.866	< 2e-16	***
Rdj350_km	0.1172716	0.0117056	10.018	< 2e-16	***
Open_Area	-0.8736484	0.0530676	-16.463	< 2e-16	***
NearRdj_km	2.8245471	0.1705116	16.565	< 2e-16	***
TestDum12	0.0808790	0.0208837	3.873	0.000109	***
TestDum13	0.1080254	0.0209271	5.162	2.53e-07	***
TestDum14	0.1391088	0.0210412	6.611	4.17e-11	***
TestDum15	0.1227024	0.0210116	5.840	5.53e-09	***
TestDum16	0.0424421	0.0209573	2.025	0.042899	*
TestDum17	-0.0903914	0.0208630	-4.333	1.50e-05	***
TestDum22	0.2600979	0.0167829	15.498	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Table 14 Regression Model Output (M2)

The model could be rewritten in the following form for better understanding:

$$
\begin{aligned}
&Ln(PM10)\\
&= -1.1283 + .26Winter + .081Tuesday + .11Wednesday + .14Thursday + .123Friday\\
&+ .042Saturday - .09Sunday + .14Temperature + .022Humidity - .17Dewpoint - .07Wind\ speed\\
&+ .001Radiation - .006Rainfall + .31Stability\ Class + .08Wind\ Index + .03Altitude - .87Open\ Space\\
&+ .11Major\ road\ \ length\ (within\ 350m\ buffer) + 2.82Nearest\ major\ road\ distance)
\end{aligned}
$$

The model's coefficient for dummy variables shows conformity with the logic for using dummy variables. Winter Mondays get 26% excess PM_{10} over the summer Mondays. On the other hand, coefficients for weekends are comparatively lower than the weekdays, especially on Sunday. For Saturday, the model shows slightly higher PM_{10} than Monday.

Among the other variables, the coastal distance variable shows a negative correlation, which is completely logical in the context. A study (Yin et al., 2005) in Ireland shows that the primary source contributing to PM_{10} is marine aerosol (NaCl). An observation from the past data (2007-2009), monitoring sites (Marino, Ringsend and Dun Laoghaire) showed that there is no significant reduction of PM_{10} being close to the ocean (Figure 17 Average daily PM10 concentration (2007-2009) in monitoring stations).

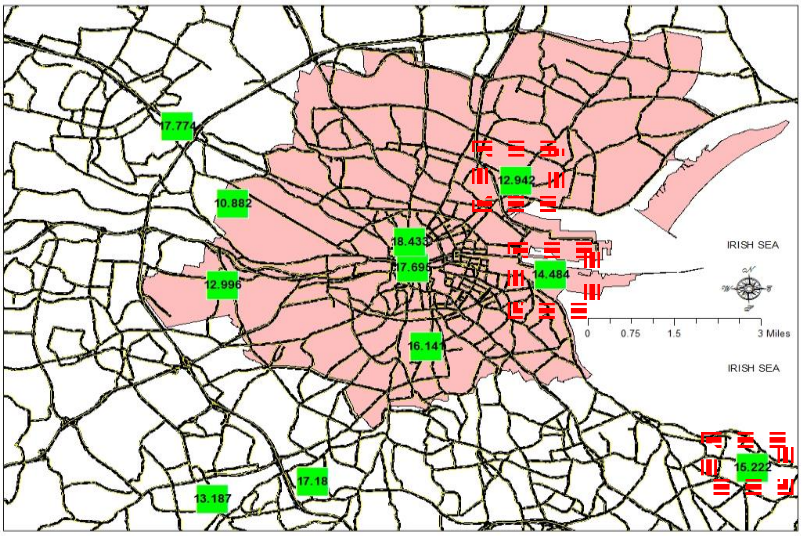


Figure 17 Average daily PM10 concentration (2007-2009) in monitoring stations

Several techniques (Figure 18 Time series data for PM_{10} vs. PM_{10} without outliers Figure 19, Figure 20) have been further used to improve the log-level model. These included:

Model 3: Limiting the PM_{10} concentration data to within two standard deviations. The treatment of PM_{10} in this way has been mentioned as PM_{10}_2SD. Here, standard Deviation; has been calculated for each station and each year.

Model 4: 3-days moving average for PM_{10}_2SD

Model 5: 3-day weighted Moving average (Weight:.5 for the day ' 0', .3 for for day-1, and .2 for day-2) for PM_{10}_2SD.

Model 6: Exponential Moving average with .6 smoothing factor for PM_{10}_2SD

Model 7: Exponential Moving average with .3 smoothing factor for PM_{10}_2SD

The idea here was to capture any long term trends, avoiding short term fluctuations as the resolution of the explanatory variables was either static over the area (weather data), and no land use variable was similar to the resolution of daily PM_{10} data.

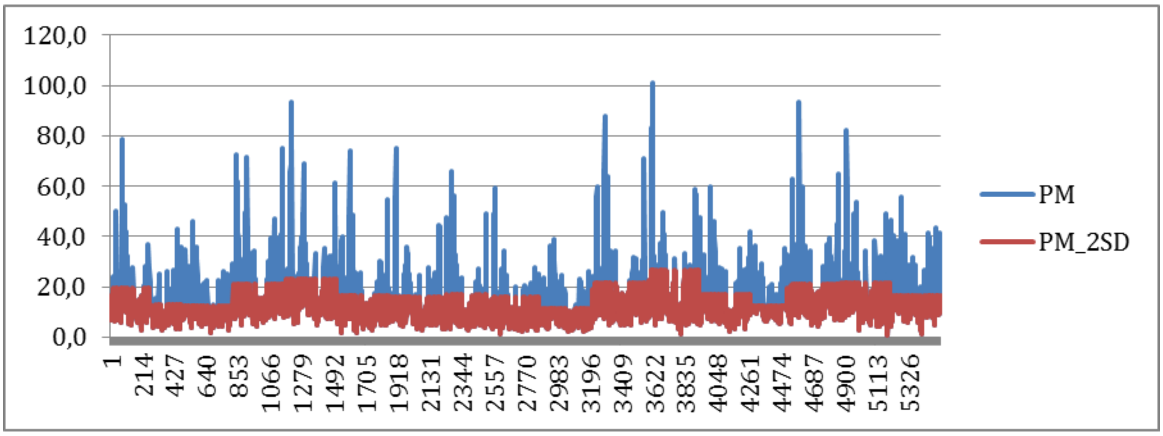


Figure 18 Time series data for PM_{10} vs. PM_{10} without outliers

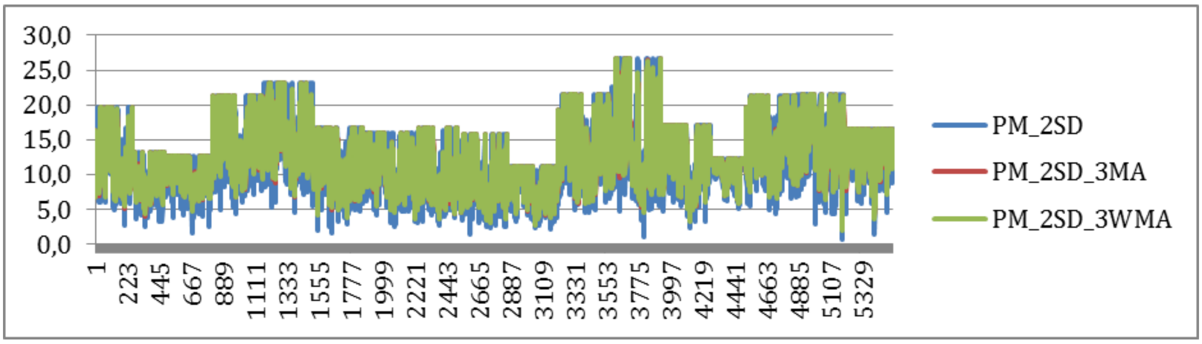


Figure 19 Time series data for PM_{10} without outliers, 3-day moving average and 3-day weighted moving average

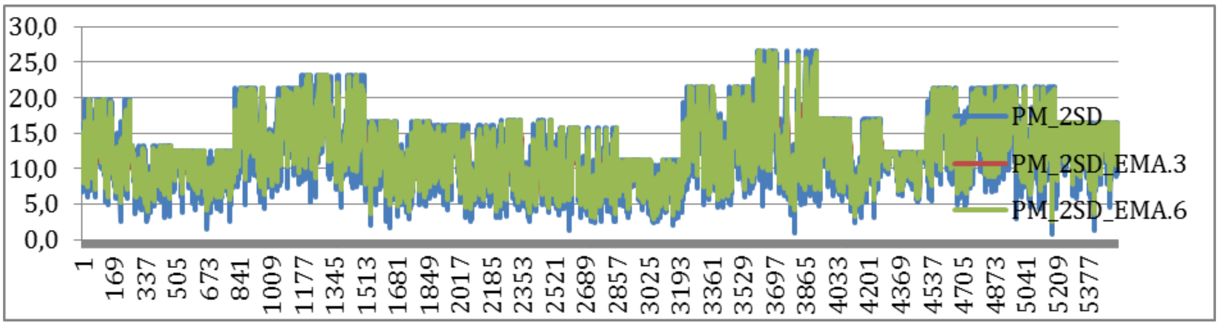


Figure 20 Time series data for PM_{10} without outliers and exponential moving average

The models yielded the following R squared values (Table 15). This means the reduction of short-term fluctuation of PM_{10} data matched better with the low resolution of explanatory variables.

Models	3	4	5	6	7
Adjusted R squared	0.44	0.48	0.49	0.50	0.54

Table 15 Performance of the Models

The target values outlined in Section 2.2 highlighted a desirable R^2 adjusted figure of 50%. Here models 6 and 7 meet or exceed these performance criteria.

2.4.2 Validation of LUR Model

To validate and to choose the best fitted model, cross validation has been performed. Statistical tests like the coefficient of determination (R^2) and Root Mean Squared Error (RMSE) were measured (Table 16) to ensure the calibrated model's efficiency. Using SPSS software, 15% of the total observation was kept for validation and the rest of the 85% data was used to reproduce best performed models: Model 2 (best fitted model before tempering the data), Model-6 and Model-7.

Indicator	Model-2	Model-6	Model-7
R-squared (Model)	0.43	0.50	0.54
R-squared (Validation) *	0.46	0.34	0.30
RMSE*	7.65	9.02	9.23
RMSE (Log)*	2.03	2.19	2.22
Pearson r*	0.67	0.58	0.54
R-sqr*	0.46	0.34	0.30

*validation with 15% data.

Table 16 Model validation

The validation shows that reproduced Model-2 although having a comparatively lower fit, yields a good R squared in the validation process. This model was then further examined using a normality test (Figure 21).

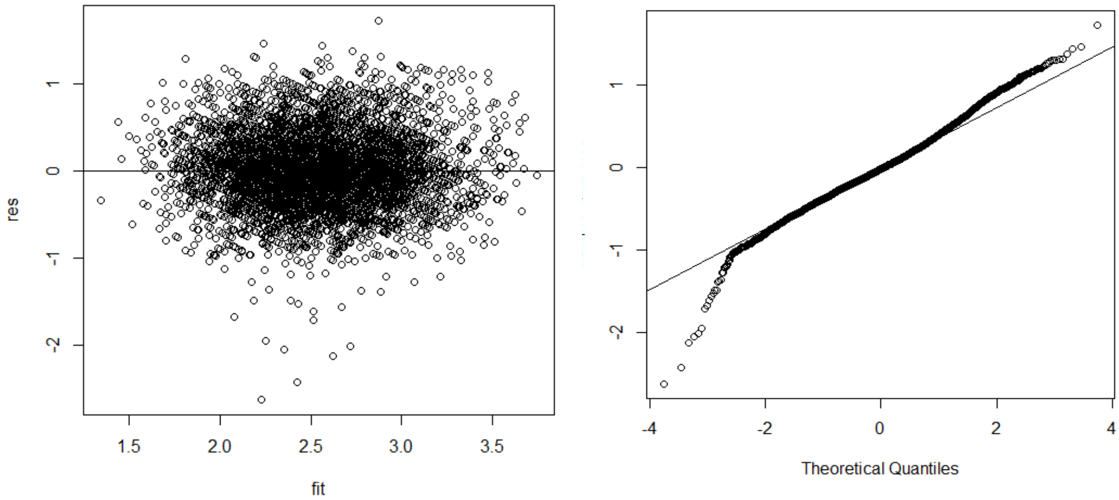
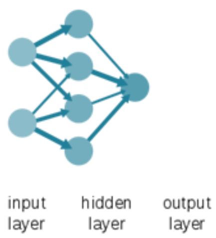


Figure 21 Normality Test (a. Residual vs. fitted value; b. Normal Q-Q plot)

Figure 21 (a) showed an unbiased and homoscedastic relationship between residual and fitted values, while Figure 21 (b) shows the residuals were normally distributed and scattered around the line.

2.4.3 Model Optimisation

The log-level model was further optimised using a neural network (NN) approach in Matlab. As the data was highly nonlinear in nature, a neural network-NN (Figure 22) has been applied in this case.



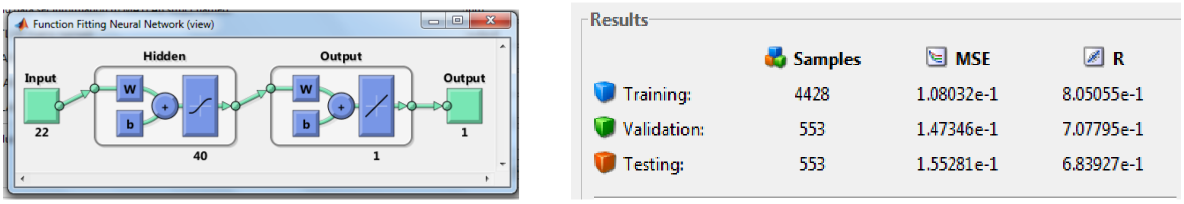
Mathematically, the neuron j can be described as follows

$$u_q = \sum_{p=1}^{n=m} w_{pq} . x_p ; \quad a_q = \varphi (u_q + b_q)$$

Where $x_1, x_2, \ldots\ldots, x_p$ are the input signals, w_{pq} is the connection weight from p in layer l to neuron q in layer l+1, u_q is the linear combiner output due to the input signals, b_q is the bias, φ (….) is the activation faction and a_q is the output signal of the neuron .

Figure 22 Typical Neural Network Structure (Dunne & Ghosh, 2011; Haykin, 1994)

Here, the Levenberg-Marquardt backpropagation technique has been applied. After several iterations, the network architecture (Figure 23a) that performed best was selected. The 22-40-1-1- combination yielded following result (Figure 23b).



Results

	Samples	MSE	R
Training:	4428	1.08032e-1	8.05055e-1
Validation:	553	1.47346e-1	7.07795e-1
Testing:	553	1.55281e-1	6.83927e-1

Figure 23 Neural network (a. Architecture, b. Model performance)

The mean square error was found at validation was 0.14. (RMSE 0.374) in log scale. In normal scale the error was 1.45, a significant improvement from the previous 7.65 RMSE mentioned in Table 16. Figure 24 demonstrates the improved performance obtained using the NN methodology.

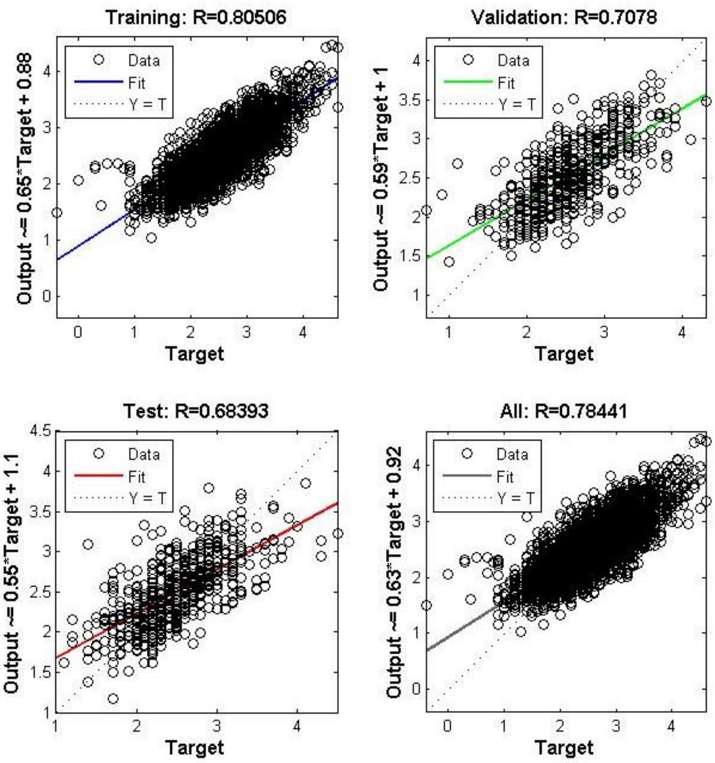


Figure 24 Performance of the Neural Network

2.5 Exposure concentration for Dublin roads

2.5.1 Introduction

To generate exposure concentration at route level, exposure mapping is necessary. This can be obtained using the kriging geo-statistical interpolation technique. However, as the number of the monitoring stations was low, it was necessary to extrapolate data for some unmonitored sites (sample points) using the regression equation. For this reason grouping analysis was performed to understand the physical setting of the study area. This will provide an indication of the sample number and location for each group.

2.5.2 Exposure mapping

2.5.2.1 *Grouping analysis*

Based on the physical characteristics (Population density, Housing Stock density, Major road density, Land cover and Altitude) of the area, the following eight classified areas (Figure 25) have been determined for Dublin city.

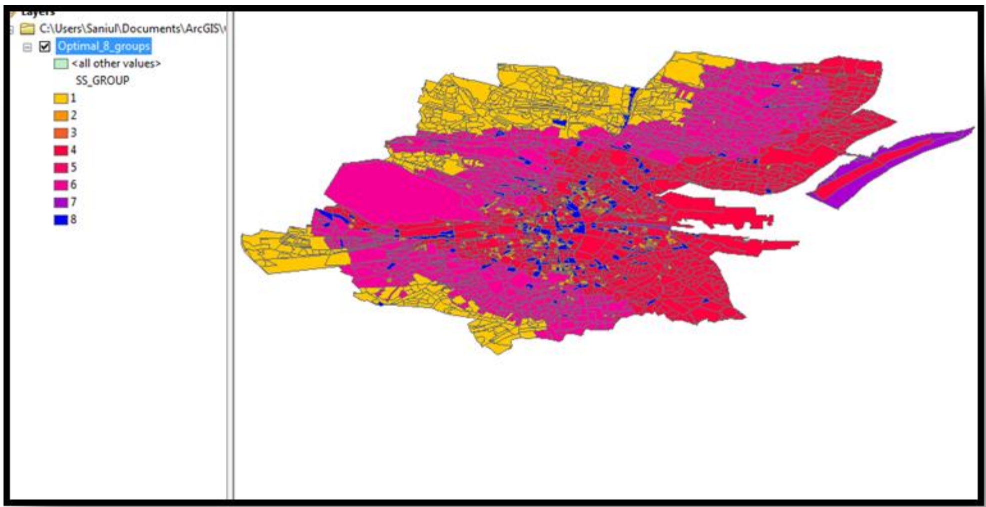


Figure 25 Eight optimal groups for Dublin

Figure 26 (Pseudo F- statistics) shows that the eight groups are the optimum number for this area according to the given criteria.

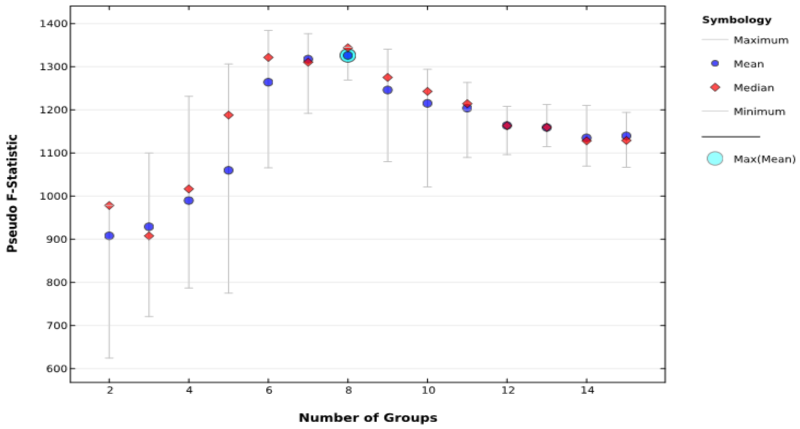


Figure 26 Optimal Group Test

2.5.2.2 *PM_{10} estimations for new sample sites*

Landuse variables like distance from the coast, altitude, major road within 350m, wind index, nearest major road, open space for the relevant buffer sizes were derived from the GIS data. In addition, average weather conditions were calculated based on weather data from 2007 to 2009.

Season	Days	Temperature (C)	Humidity (%)	Dew Point (C)	Wind Speed (m/s)	Wind Direction (degree)	Radiation (w/m^2)	Rainfall (mm)	Stability Class
Summer	Monday	14.42	82.30	11.22	4.90	194.98	63.47	2.44	4.02
	Tuesday	14.72	80.95	11.22	5.29	209.81	73.30	3.20	3.98
	Wednesday	14.42	83.19	11.42	5.23	208.58	69.93	3.82	4.00
	Thursday	13.86	83.85	11.00	5.18	195.17	66.38	3.01	4.00
	Friday	14.04	82.70	10.93	5.19	199.40	66.19	3.42	4.00
	Saturday	14.27	83.65	11.34	5.27	213.83	59.61	6.56	4.03
	Sunday	13.98	82.36	10.82	4.99	197.02	60.79	3.10	4.02
Winter	Monday	8.09	84.13	5.40	6.08	193.59	32.47	1.58	4.06
	Tuesday	8.18	84.46	5.55	5.83	196.65	32.59	2.02	4.07
	Wednesday	8.22	84.71	5.64	5.89	203.19	30.75	2.25	4.06
	Thursday	8.78	83.87	6.05	6.31	206.53	33.03	2.15	4.05
	Friday	8.50	84.25	5.86	5.81	210.37	30.92	1.94	4.16
	Saturday	8.74	83.46	5.90	6.01	205.26	33.18	2.02	4.09
	Sunday	8.27	84.11	5.59	5.91	193.76	31.06	2.28	4.06

Table 17 Average weather condition

Both landuse and environmental data were input to the optimised model in mathlab and PM_{10} data for 32 sites were derived for seven days for each season (example, Figure 27).

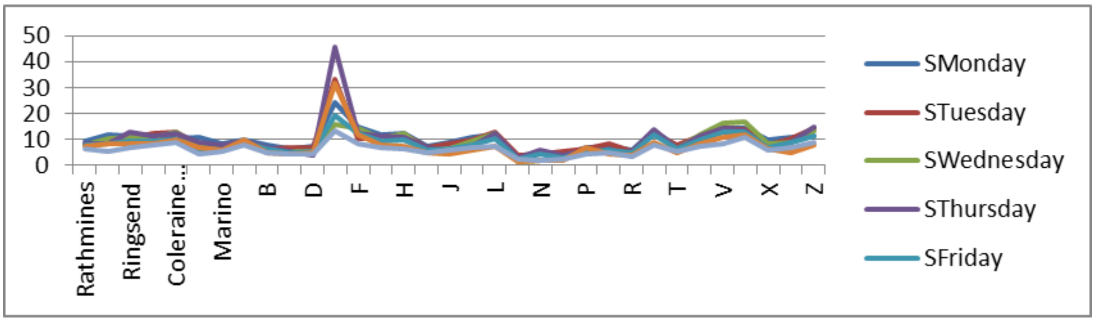


Figure 27 Predicted PM10 over the week in summer

2.5.2.3 *Spatial Auto-correlation*

Moran's I test was conducted which is a measure of global spatial autocorrelation. Negative values indicate negative spatial autocorrelation. Values range from −1 (indicating perfect dispersion) to +1 (perfect correlation). A zero value indicates a random spatial pattern. Most of the cases Moran’s I is near to 0 which indicates the patterns do not appear to be significantly different from random. There is no alarming outcome except for a summer Wednesday. In this case, given the z-score of 2.43, there is less than 5% likelihood that this clustered pattern could be the result of random chance.

	Summer							Winter						
Days	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Moran's I	0.196	0.096	0.316	0.071	0.187	0.084	0.141	0.029	0.049	0.136	0.015	0.025	-0.003	-0.051
z-score*	1.709	1.172	2.428	1.091	1.595	1.121	1.236	0.693	0.643	1.329	0.870	0.752	0.214	-0.136

p-value	0.087	0.241	0.015	0.275	0.111	0.262	0.217	0.488	0.520	0.184	0.384	0.452	0.831	0.892

*Z-scores in which values greater than 1.96 or smaller than −1.96 indicate spatial autocorrelation that is significant at the 5% level.

Table 18 Moran's I test

2.5.2.4 *Kriging*

Kriging was used to develop the exposure concentration map over the city area (Example: Figure 28). The Kriging technique was used to interpolate the value of a random field at an unobserved location from observations of its value at nearby locations.

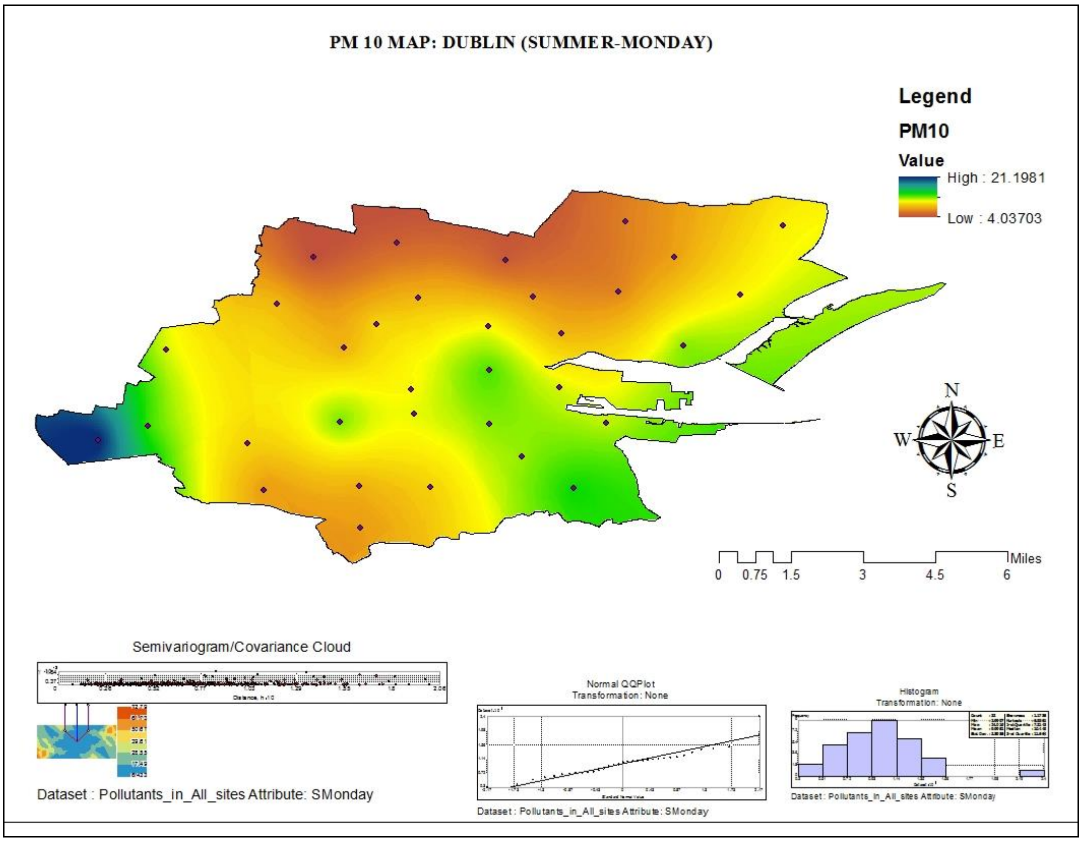


Figure 28 PM_{10} Exposure Map for Dublin area (Summer-Monday)

The exposure map was then overlapped with the road network. In order to obtain the daily exposure concentration on every road. To achieve this, the road layers were first converted to point layers using XTools Pro 9.1 which generated 59262 points on the center line of the roads (Figure 29). PM_{10} concentration for each point has been extracted from the exposure map (Figure 30). Having co-ordinates for each point this output can be used for the route choice purposes, or route level information can also be used for route choice (Figure 32).

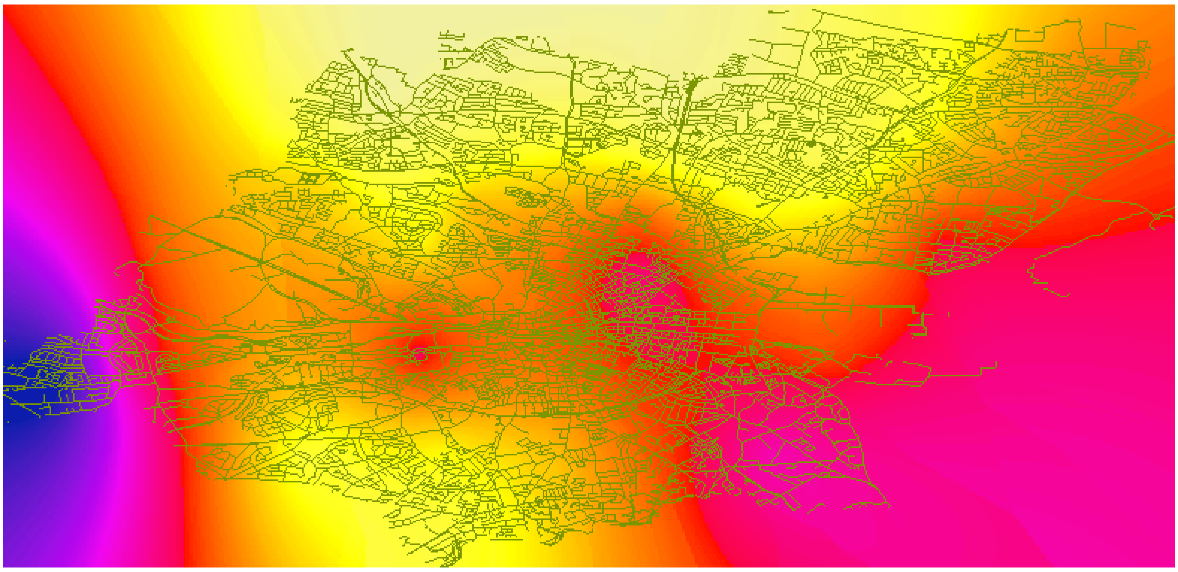


Figure 29 Exposure map with road network (line)

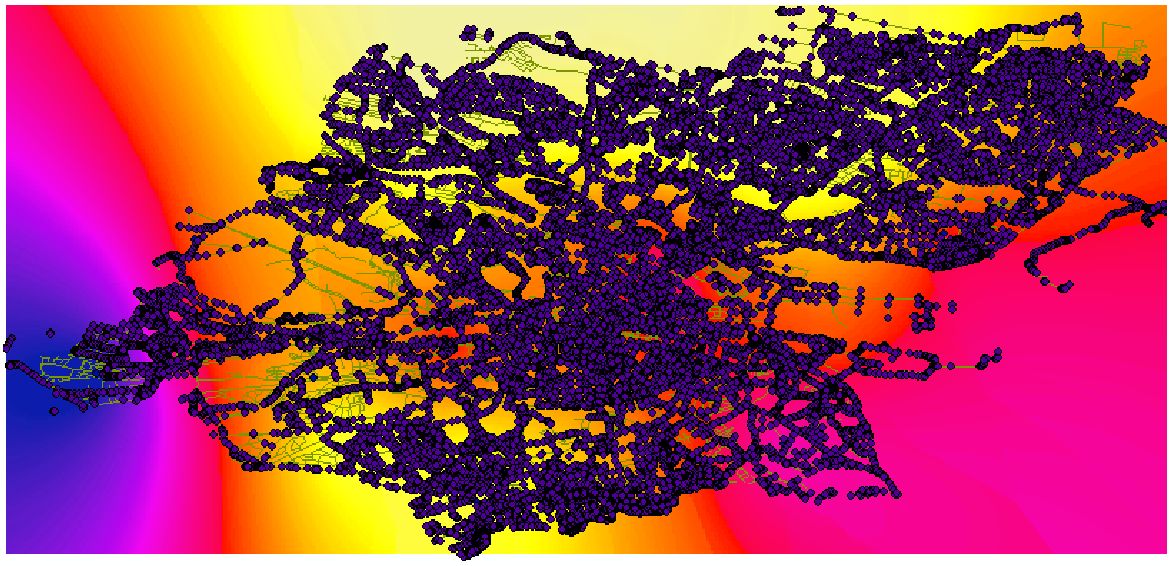


Figure 30 Exposure map with road network (Points)

2.6 Exposure model for Vienna

Vienna is Austria's capital city, with a population of about approximately 1.731 million. The area covered by the Vienna boundary is 414.6 km² which is approximately four times higher than the area of Dublin.

The methodology of the exposure modelling remains the same for Vienna. The dissemination of air quality information is excellent in Vienna. Air Quality monitors in Vienna (Figure 31) provide real time PM_{10} concentration.

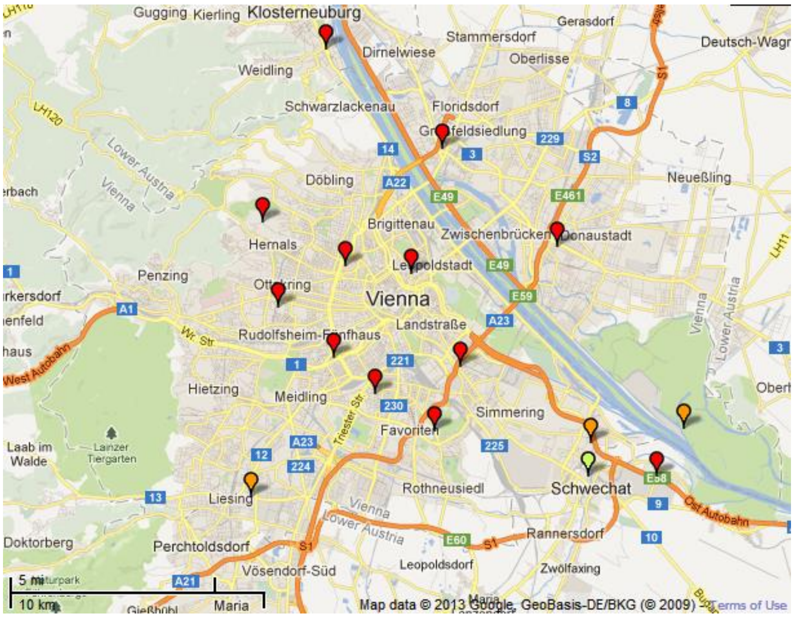


Figure 31 Air Quality monitoring sites in Vienna

Source: Umweltbundesamt

The source for values of the response variable (PM_{10}) is Umweltbundesamt . The latitude and longitude for the PM_{10} monitoring stations have been taken from the Umweltbundesamt website and inputted in the GIS environment. The administrative boundary has been taken from the GADM database of Global Administrative Areas (GADM, 2012). Figure 31 shows a few of the monitors are just outside the Vienna city boundary.

These monitoring sites have also been considered to capture more spatial information about Vienna' exposure concentration dynamics.

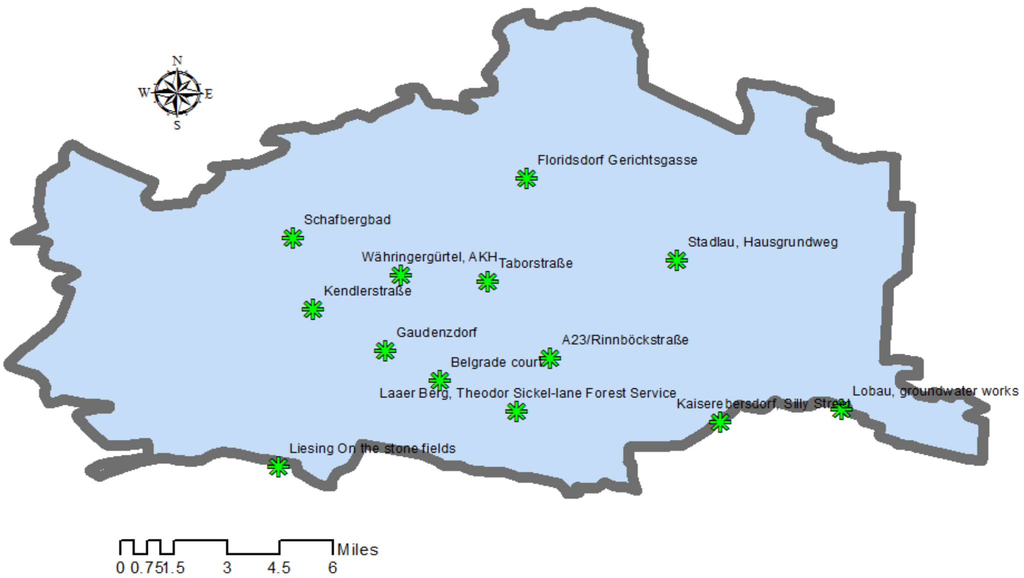


Figure 32 Air Quality monitoring sites in Vienna within the city boundary

Station	2011			2012		
	Minimum µg/m^3	Maximum µg/m^3	Average µg/m^3	Minimum µg/m^3	Maximum µg/m^3	Average µg/m^3
A23/Rinnböckstraße	7.0	148.1	34.44	5.7	98.9	25.98
AKH, Südringweg	4.2	123.7	26.72	4.0	89.6	23.16
Belgradplatz	4.2	145.2	33.87	4.7	99.9	27.33
Floridsdorf Gerichtsgasse	7.9	135.4	31.25	8.0	154.5	27.45
Gaudenzdorf	7.1	136.2	30.51	6.3	106.7	25.57
Kaiser Ebersdorf	3.6	131.1	29.36	3.3	96.3	22.66
Kendlerstraße	6.7	128.3	30.35	4.6	115.4	26.47
Laaerberg	5.9	130.6	27.99	4.2	95.4	23.66
Liesing	4.7	131.7	31.62	4.3	112.1	27.30
Lobau	5.3	125.0	25.99	5.3	87.6	20.28
Schafbergbad	5.2	106.0	24.54	5.7	147.6	21.34
Stadlau	4.2	122.9	28.28	5.0	132.7	24.88
TaborstraBe	5.1	126.4	29.35	5.0	90.9	24.20

Source: Municipal government of Vienna

Table 19 PM_{10} in different monitoring stations (2011-2012) in Vienna

2.6.1 Deviations from the lead model

There is no significant deviation in the modelling process from the one used for Dublin city. The selection of data was performed based on the availability. Explanatory variables used in the Vienna model were similar to Dublin (Table 20). However, the resolution of the dataset was not always the same as the lead model.

Sl no	Variable	Sl no	Variable
1	Latitude+ Longitude	10	Average Population density within 500m (persons/km2)
2	Open area in Sq.km(1000m)	11	Average Altitude in m (1000m)
3	All type of Road length in km(350m)	12	Average Population density within 1000m (persons/km2)
4	All type of Road length in km(750m)	13	Minimum Daily Temperature(C)
5	Major Road length in km(350m)	14	Maximum Daily Temperature(C)
6	Major Road length in km(750m)	15	Wind speed (m/s)
7	Average Altitude in m (500m)	16	Precipitation (mm)
8	Nearest Major Road Distance (km)	17	Average Daily Temperature (C)
9	No. of Building Centroid within 1000m	18	Industrial area in Sq.km(1000m)

Table 20 Explanatory variables for Vienna Model

2.6.2 Model Selection and optimisation

Models for Vienna (Table 21) have been developed followed by the same methodology described in section 4.1.

Location	Air pollutant	PM10 Model for Dublin and Vienna (variables <=.001 Significance)	R2	F	P	SE	DF
	Primary Model	$LnY=27.249077+0.021683\underline{X}_1-0.053404\underline{X}_2-0.045463\underline{X}_3-0.020775\underline{X}_4-0.361544\underline{X}_5-0.110482\underline{X}_6+0.029500\underline{X}_9-0.031520\underline{X}_{12}$	0.34	605.2	< 2.2e-16	0.5045	9325
	With Seasonal Impact	$LnY=27.084983+0.021706\underline{X}_1-0.047978\underline{X}_2-0.045630\underline{X}_3-0.019589\underline{X}_4-0.361544\underline{X}_5-0.110482\underline{X}_6+0.029500\underline{X}_9-0.031520\underline{X}_{12}+0.153344Winter$	0.35	553.8	< 2.2e-16	0.5019	9324
Vienna	Final Model: With Seasonal and Daily Impact:	$LnY=27.055656+0.020783\underline{X}_1-0.047124\underline{X}_2-0.046071\underline{X}_3-0.019863\underline{X}_4-0.361544\underline{X}_5-0.110482\underline{X}_6+0.029500\underline{X}_9-0.031520\underline{X}_{12}+0.147878Winter+0.095848Tuesday+0.129733Wednesday+0.159774Thursday+0.050301Friday-0.013376Saturday-0.083362Sunday$	0.36	356.9	< 2.2e-16	0.4957	9318

The model with high R-squared value was chosen. That equation can be rewritten as below:

Model with variable names	LnY=27.055656+0.020783MindailyTemp-0.047124Max.dailyTemp-0.046071WindSpeed-0.019863DailyPrecipitation-0.361544(Latitude+Longitude)-0.110482OpenAreaSq.km(1000m)+0.029500MajorRoadlengthinkm(350m)-0.031520NearestMajorRoadDistance(km) +0.147878Winter+0.095848Tuesday+0.129733Wednesday+0.159774Thursday+0.050301Friday-0.013376Saturday-0.083362Sunday

Table 21 Panel Data Models for Vienna

Weather parameters that exist in the final model are highly non-linear in nature and have complex interactions among explanatory variables. So, non-parametric regression and neural network fits here. Using of panel data provides an additional benefit of using non-parametric regression, as such the methodology requires a large number of observations to define the structure of the model first and then make estimatations.

Non-parametric regression in the form of locally weighted scatterplot smoothing (LOWESS) has been deployed through XLSTAT 2013. The LOWESS method combines multiple regression models in a *k*-nearest-neighbour-based meta-model. To define the structure of the model, smoothing parameter value (k nearest neighbours: % = 50) and the degree of the local polynomial as 1 is defined). Tricube weight function has been used as kernel function. On the other hand, the log-level models were also optimised using a neural network (NN) approach in Matlab. Here, the Levenberg-Marquardt backpropagation technique has been applied. After several iterations with different number of hidden neurons (10, 25, 30, 40, 50, 60), the network architecture that performed best was selected. The combination for “input-hidden layers- output” was 17-50-1-1 for Vienna yielded consistent satisfactory results. 85% data were devoted to training, 10% for validation and 5% for testing the model, and run the models for consecutive five times. The MATLAB Neural Network Toolbox has been applied to perform this analysis. From the below table, Non-parametric regression was found to consistently yield better results.

Model description			Vienna			Model description			Vienna		
Stage and Performance			Multiple Linear Regression Model	Non-parametric Regression Model	Neural Network	Stage and Performance			Multiple Linear Regression Model	Non-parametric Regression Model	Neural Network
Full Model		R2	0.36	0.51	---						
		MSE	---	0.19	---						
Run 1	Calibration	r'	---	---	0.9	Run 4	Calibration	r'	---	---	0.9
		R2	0.38	0.51				R2	0.37	0.5	
		MSE	---	0.19	0.07			MSE	---	0.19	0.07
	Validation	r'	0.59	0.71	88 (.77)*		Validation	r'	0.6	0.71	.85 (.91)*
		R2	0.35	0.51	---			R2	0.36	0.51	---
		MSE	---	0.16	.08 (.19)*			MSE	---	0.16	.11 (07)*
Run 2	Calibration	r'	---	---	0.91	Run 5	Calibration	r'	---	---	0.92
		R2	0.37	0.51				R2	0.37	0.51	---
		MSE	---	0.19	0.06			MSE	---	0.19	0.06
	Validation	r'	0.6	0.73	.87 (.82)*		Validation	r'	0.57	0.7	0.85 (.86)*
		R2	0.36	0.53	---			R2	0.32	0.49	---
		MSE	---	0.16	.10 (.14)*			MSE	---	0.17	0.1 (.1)*
Run 3	Calibration	r'	---	---	0.87						
		R2	0.37	0.5	---						
		MSE	---	0.19	0.09						
	Validation	r'	0.6	0.71	.76 (.84)*						
		R2	0.36	0.51	---						
		MSE	---	0.15	.17 (.11)*						

* 5% testing data result.

Table 22 Model validation and performance analysis

2.6.3 Estimation of exposure concentration for new points and Interpolation

Similar to Dublin Model, 14 maps (seven days over the summer and winter) were produced for average PM_{10} concentration in Vienna. For this 25 extra points were selected randomly and the prediction was carried out for a total of 38 (25+13 fixed monitoring stations) points in Vienna. Thus, kriging techniques were applied for seven days in summer and winter (example, Figure 33).

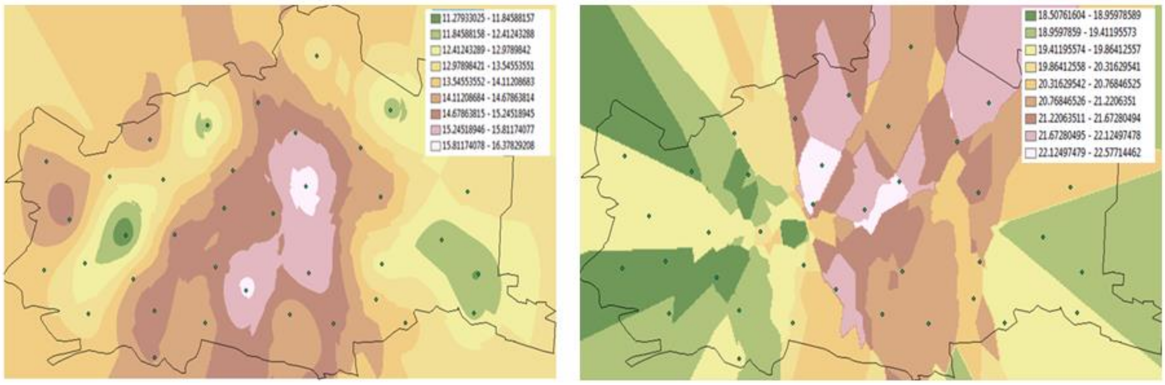


Figure 33 Summer Friday and Winter Monday

The exposure concentration was subsequently extracted to the points on the road links to maps.

2.7 Factor selection & Exposure Rating

As discussed earlier in the section 2.2, the dose for a trip will be identified through equation (2). For this, time factor and mode factors are needed to be identified.

The time factor enhances the estimated average PM_{10} values, and mode factor provides scope to estimate dose for the travellers using different modes. The time factor has been calculated using hourly nitrogen oxide (NO_x) values of five monitoring stations in Dublin, as there is a limitation of availability for hourly PM_{10} data. The daily concentration of NO_x and PM_{10} has a r=.6 Pearson correlation, and that is comparatively better than with other pollutants (PM_{10} vs. NO was r=.51; PM_{10} vs. NO_2 was r=.58). The average hourly NO_x values for each monitoring station was calculated for 2009, and the two peaks were observed in the Figure 34 for all the monitoring stations (the hourly distribution over the day also showed the same tendency as average hourly values). These peaks are consistent with the peak traffic periods mentioned in the emission report (deliverable D3.1), that is, fluctuation of emission concentrations is consistent with traffic level. Then six time factors were estimated from the ratio of average concentration values in different periods, and daily average concentration for each monitor. Time factors (Table 23) for six time periods were calculated taking the average of the ratios in six different periods of all monitors.

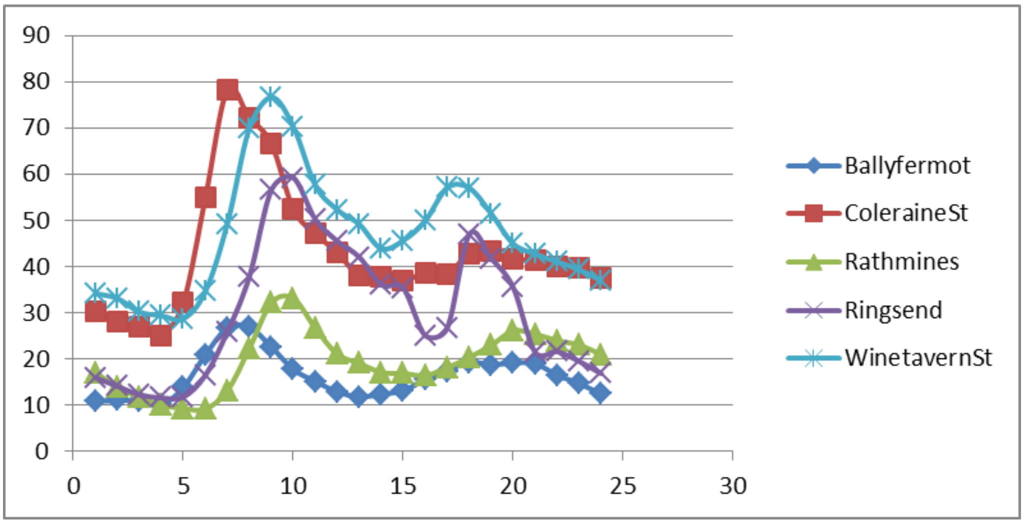


Figure 34 Average NOx concentration in the monitoring stations

Serial No.	Factor Name	Time Period	Time Factor
1	Kick off factor	5am -6.59am	0.74
2	Morning Peak Factor	7am-10.59am	1.35
3	Settling Factor: Noon	11am-13.59pm	0.96
4	Average Traffic Factor	14pm-15.59pm	0.96
5	Evening Peak	16pm-18.59pm	1.16
6	Settling Factor: Night	19pm-21.59pm	0.95
7	Night factor	22pm-4.59am	0.66

Note: Seven Time Factors (e.g. Peak and off-peak hour) have been derived based on the traffic situation in Dublin and assumed to be same in general for Vienna; Inhalitation rate has been taken from the literature (US EPA).

Table 23 List of Time Factors

The mode factor, that is the inhalation factor (Table 24) for each mode has been taken from US EPA 2009 (Exposure Factors Handbook: 2009 Update. EPA/60 0/R-09/052A.)

Serial No	Mode	Inhalation factor
1	Luas (Tram), Dart(Metro), Bus*	0.228 m^3/hr
2	Bike	1.620 m^3/hr
3	Car/Taxi	0.570 m^3/hr
4	Walk	0.720 m^3/hr

* Same as because the body movement is similar to the resting period while using these modes of travel.

Table 24 List of Inhalation factor

All these factors were considered along with the average PM_{10} values to calculate dose for each type of mode users. The total value calculated from different modes provide the possible dose for each route. Although the outcome of the modelling will provide dose, the value will be expressed for the users as a band score as outlined earlier.

2.8 Algorithm for PEACOX application

The algorithm developed for PEACOX was on the Java platform. This final algorithm has been focused on the link exposure calculation in figure 1 in realtime. The input and output were the main concern in this stage. The flowchart in figure 35 illustrates the final algorithm.

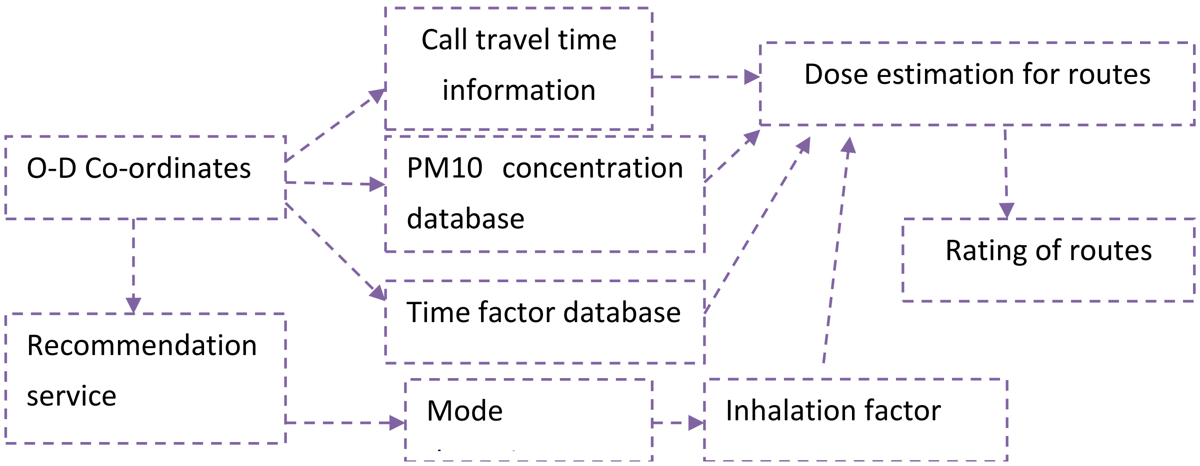


Figure 35 Exposure algorithm flowchart

2.9 Conclusion

The methodologies for trip-by-trip exposure have been developed for the given task of PEACOX project. PM_{10} has been chosen as generic pollutant for exposure modelling. The given task was first carried out for whole city areas by land use regression modelling and thus, the output was used to determine route level exposure. There was a challenge in developing an LUR model with limited spatial coverage of the monitors which was confronted by using advanced modelling techniques. This limitation that was particularly prominent in Dublin was balanced by using data from the additional sites in longer periods. On the other hand, seasonal and daily variation has been added to increase model's performance for both of the cities. The final MLR models in this study yield acceptable explanatory power for either city. The techniques for using exposure information for realtime route level have also been outlined clearly. The final version of the models for PEACOX project has been developed, and delivered in the Java platform with necessary adjustment for instantaneous application.

3. Glossary of Terms

Air Pollutant: Anything emitted to the air which could have a detrimental effect on human health or the environment.

Ambient Air: The air located outside of buildings/ Outdoor air.

Exposure: The amount of contact that a person has with the pollutant.

Land use: The total of arrangements, activities and inputs undertaken in a certain land cover type.

Neural networks/NN: It is often called as statistical black box; is composed of interconnecting artificial neurons that build mathematical models mimicking the properties of biological neurons.

Particulate $Matter_{10}$/PM_{10}: Particulate matter is made up of many different compounds that has size of 10µm.

Routine Monitoring Network: A collection of air monitoring equipment spread across an area whose readings are used to understand the Air Quality in that area.

Solar radiation: Radiation emitted by the Sun.

Regression: In statistics, regression analysis is a technique for estimating the relationships among response and explanatory variables.

Small Area: The smallest area grouping used in the Irish census.

$\mu g/m^3$: 1 $\mu g/m^3$ means that for every cubic meter of air there is 1 micro-gram of the pollutant present.

4. References

Ahn, K. (1998). Microscopic fuel consumption and emission Modeling, Blacksburg, Virginia., Faculty of the Virginia Polytechnic Institute and State University.

Barkenbus, J,. Eco-driving: An overlooked climate change initiative. Energy Policy 38, 2010, 762-769.

Beusen, B., Broekx, S., Denys, T., Beckx, C., Begraeuwe, B., Gijsbers, M., Scheepers, K., Govaerts, L., Torfs, R., Panis, L. Using on-board logging devices to study the longer-term impact of an eco-driving course. Transportation Research Part D. 14, 2009, 514-520.

Boriboonsomsin, K., Barth, M., Vu, A., Evaluation of driving behaviour and attitude towards eco-driving: A Southern California Limited Case Study. *90th Annual Meeting of the Transportation Research Board*, *Washington, D.C*, *January*, 2011.

Briggs, D., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., et al. (1997). Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Science, 11*, 699–718.

Briggs, D.J., (2007). The role of GIS to evaluate traffic-related pollution. Ed. Occup. Environ.

Briggs, D., Hoogh, C.D. and Gulliver, J.(2011) Comparative assessment of GIS based methods and metrics for modeling exposure to air pollution. J. Toxicol. Environ. Health.

CGIAR. (2012). The CGIAR Consortium for Spatial Information from http://www.cgiar-csi.org/

Chen, L., Baili, Z., Kong, S., Han, B., You, Y., Ding, X., et al. (2010). A land use regression for predicting NO_2 and PM_{10} concentrations in different seasons in Tianjin region, China. *journal of Environmental Sciences, 22*(9), 1364-1373.

CSO. (2012). Central Statistics office. from http://www.cso.ie/en/index.html

Delhomme, P., Cristea, M., Francoise, P., Self-reported frequency and perceived difficulty of adopting eco-friendly driving behaviour according to gender, age, and environmental concern. Transportation Research Part D. 20, 2013, 55-58.

Dublinked. (2012). from http://www.dublinked.ie/

Dunne, S., & Ghosh, B. (2011). *Traffic Condition Forecasting: traffic flow predictions employing neural networks in a novel traffic flow regime separation techniqu.* Paper presented at the ITRN2011 31st August – 1st September, University College Cork,Ireland.

EEA. (2012a). Corine Land Cover 2006 seamless vector data & Population density disaggregated with Corine land cover 2000. from http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version-2

EPA. (2012). Environmental Protection Agency, Datasets. 2012, from http://www.epa.ie/

EURADSystem IM. (2012). Air Quality Forecast and Data Assimilation Service for Austria. from http://db.eurad.uni-koeln.de/promote/RLAQS/riu_rlaqs.php?domain=AUT&year=2012&month=01&day=24&force=AUT&mode=1

GADM. (2012). from http://www.gadm.org/

Gilliland, F., Avol, E., Kinney, P., Jerrett, M., Dvonch, T., Lurmann, F., et al. (2005). Air pollution exposure assessment for epidemiologic studies of pregnant women and children: Lessons learned from the Centers for Children's Environmental Health and Disease Prevention Research. *Environ Health Perspect, 113*(1447–1454),

GmbH, G. (2012). OpenStreetMap Shapefiles. from http://www.geofabrik.de/data/shapefiles.html

Gulliver, J., Hoogh, K. d., Fecht, D., Vienneau, D., & Briggs, D. (2011). Comparative assessment of GIS-based methods and metrics for estimating long-term exposures to air pollution. *Atmospheric Environment 45*, 7072-7080.

Haykin, S. (1994). *NN: A Comprehensive Foundation* New York: NY: Macmillan.

Hoek, G., Beelen, R., Hoogh, K. D., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of Landuse regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment, 42*, 7561-7578.

Katsouyanni, K., Gryparis, A., & Samoli, E. (2010). Short-Term Effects of Air Pollution on Health. *Encyclopedia of Environmental Health*, 51–60.

Martin, E., Boriboonsomisn, K., Chan, N., Williams, N., Shaheen, S., Barth, M. Dynamic ecodriving in Northern California: A Study of survey and vehicle operations data from an ecodriving feedback device. . *92nd Annual Meeting of the Transportation Research Board*, *Washington, D.C*, *January*, 2013.

Martin, E., Chan, N., Shaheen, S. Understanding how ecodriving public education can result in reduced fuel use and greenhouse gas emissions. . *91st Annual Meeting of the Transportation Research Board*, *Washington, D.C*, *January*, 2012.

Maynard, D., A Coull, B., Gryparis, A., & Schwartz, J. (2007). Mortality Risk Associated with Short-Term Exposure to Traffic Particles and Sulfates. *Environ Health Perspect, 115*(5), 751-755.

O'Dwyer, M. (2011). *Air Quality in Ireland 2010 :Key Indicators of Ambient Air Quality*.

Pilla F. (2012). *A GIS model for personal exposure to* PM_{10} *for Dublin commuters.* Trinity College Dublin, Ireland.

Rutty, M., Matthews, L., Andrey, J., Del Matto, T. Eco-driver training within the City of Calgary's municipal fleet: Monitoring the impact. Transportation Research Part D. 24, 2013, 44-51.

Ryan, P. H., & LeMasters, G. K. (2008). A Review of Land-use Regression Models for Characterizing Intraurban Air Pollution Exposure. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233947/

Stromberg, H.K., Karlsson, M. Comparative effects of eco-driving initiatives aimed at urban bus drivers – Results from a field trial. Transportation Research Part D. 22, 2013, 28-33.

Stillwater, T., Kurani, K. In-vehicle ecodriving interface: Theory, design and driver responses. *91st Annual Meeting of the Transportation Research Board*, *Washington, D.C*, *January*, 2012.

Stillwater, T., Kurani, K., Mokhtarian, P. Increasing – and decreasing – fuel economy using feedback. A behavioural theory inspired eco-driving experiment. . *91st Annual Meeting of the Transportation Research Board*, *Washington, D.C*, *January*, 2012.

Umweltbundesamt (2013). from http://www.umweltbundesamt.at/

Yin, J., Allen, A. G., M., H. R., S.G., J., Wright, E., Fitzpatrick, M., et al. (2005). Major component composition of urban PM_{10} and $PM_{2.5}$ in Ireland. *Atmospheric Research 78*, 149 – 165.

ZAMG. (n.d). CENTRAL INSTITUTE FOR METEOROLOGY AND GEODYNAMICS. from http://www.zamg.ac.at/cms/de/umwelt/luftqualitaetsvorhersagen