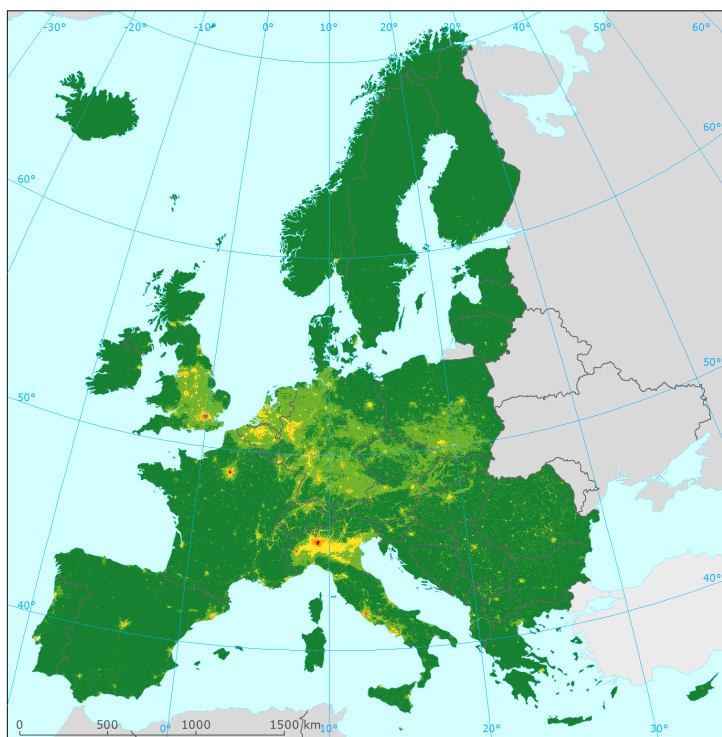


Inclusion of land cover and traffic data in NO₂ mapping methodology



ETC/ACM Technical Paper 2016/12
March 2017

*Jan Horálek, Peter de Smet, Phillipp Schneider,
Pavel Kurfürst, Frank de Leeuw*



The European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM) is a consortium of European institutes under contract of the European Environment Agency
RIVM Aether CHMI CSIC EMISIA INERIS NILU ÖKO-Institut ÖKO-Recherche PBL UAB UBA-V VITO 4Sfera

Front page picture:

Concentration map of NO₂ annual average of 2013, method including land cover and road data and including traffic map layer. Resolution: 1x1 km. Units: $\mu\text{g.m}^{-3}$. (This paper, Map 5.2.)

Author affiliation:

Jan Horálek, Pavel Kurfürst: Czech Hydrometeorological Institute (CHMI), Prague, Czech Republic

Peter de Smet, Frank de Leeuw: National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

Phillipp Schneider: Norwegian Institute of Air Research (NILU), Kjeller, Norway

DISCLAIMER

This ETC/ACM Technical Paper has not been subjected to European Environment Agency (EEA) member country review. It does not represent the formal views of the EEA.

© ETC/ACM, 2017

ETC/ACM Technical Paper 2016/12

European Topic Centre on Air Pollution and Climate Change Mitigation

PO Box 1

3720 BA Bilthoven

The Netherlands

Phone +31 30 2748562

Fax +31 30 2744433

Email etcacm@rivm.nl

Website <http://acm.eionet.europa.eu/>

Contents

1	Introduction	5
2	Methodology	7
2.1	Current method	7
2.2	Land cover and road data inclusion	8
2.3	Traffic map layer inclusion.....	9
2.4	Uncertainty estimates of the concentration maps.....	10
3	Input data.....	13
3.1	Monitoring data	13
3.2	Chemical transport modelling data	13
3.3	Altitude, meteorological data, population density	13
3.4	Land cover	14
3.5	Road type vector data	15
4	Land cover and road type inclusion.....	17
4.1	Background concentration map.....	17
4.2	Population exposure.....	23
5	Traffic map layer inclusion	27
5.1	Traffic map layer creation.....	27
5.2	Inclusion in the concentration map.....	30
5.3	Inclusion in the population exposure	33
6	Conclusion	36
6.1	Land cover and road data inclusion	36
6.2	Traffic map layer inclusion.....	36
6.3	Recommendations	36
	References	39
	Annex Brief review of potential LUR parameters to be used in NO₂ mapping..	41
A.1	Introduction.....	41
A.2	Brief summary of previous work	41
A.3	Frequently used parameters for LUR of NO ₂	42
A.4	Summary and Recommendations	44
A.5	References	44

1 Introduction

As a part of the work carried out by the European Topic Centre on Air Quality and Climate Change Mitigation (ETC/ACM), annual Europe-wide maps of air quality have been produced using geostatistical techniques for many years (Horálek et al., 2016 and references therein). The main species under consideration in previous years have been particulate matter (PM_{10} and $PM_{2.5}$) as well as ozone. Nitrogen dioxide (NO_2) maps have not been produced regularly within the framework of the ETC/ACM until the mapping report Horálek et al. (2016). Until then, NO_2 maps had been produced at irregular intervals in some years¹. Now, the importance of the NO_2 mapping is growing, due to new evidence on NO_2 health impacts. For the NO_2 map included in Horálek et al. (2016) the uncertainty analysis gives poorer results compared to the maps of PM and ozone. Next to this, it concerns only rural and urban background areas not accounting for hot spot location (traffic), although traffic is the most important source of NO_2 .

In order to produce more advanced European-scale maps of annual average NO_2 concentrations on regular basis, it can be helpful to adopt methods from land use regression (LUR) modelling, which often tends to be used more locally for the urban and regional scales (Hoek et al., 2008). However, one limiting factor in adopting such methods is that the parameters with the most explanatory power for NO_2 tend to be related to datasets of traffic volume (e.g. average daily traffic, ADT). Unfortunately such detailed traffic information at the level of each road segment is currently not available at the European scale, so the parameters to be considered for the European-scale NO_2 mapping within the framework of ETC/ACM have to be restricted to datasets that are generally available throughout all of Europe.

One of the primary candidates for such a dataset is the CORINE land cover dataset (CLC), which provides frequently updated land cover and land use information for all of Europe at a spatial resolution of approximately 100 m.

Another candidate of interest is the worldwide road-type Global Road Inventory Project (GRIP) data base (Meijer et al., 2016) of the Netherlands Environmental Assessment Agency (PBL). This GRIP data base is publicly available as a raster version expressing for each 5 arc-minutes grid cell the road lengths for 4 classes from highway to local road. This resolution appeared not to contribute to the necessary improvement of the spatial interpolation results. However, we were able to obtain from PBL its original vector based version, which seemed fit for our purpose. PBL has the intention to release this vector database as open source in a near future. It fulfils as such our principle of using only public accessible data sources, but until then we – as ETC/ACM – are restricted to no further distribution of this vector database.

In the paper, we examine the inclusion of the CLC land cover data and the GRIP road data in the rural and urban background NO_2 mapping (going to 1x1 km resolution). A quick survey (literary review) of land-use regression (LUR) models has been executed first. Then, the most suitable variants has been included in the mapping methodology (i.e. multiple linear regression and the residual kriging) and further tested. The newly created maps have been compared with the routinely produced maps.

¹ Table (A) at http://acm.eionet.europa.eu/databases/interpolated_aq_maps/index_html provides overview for which years ad hoc NO_2 maps are produced at the ETC/ACM.

Next to this, the mapping of the traffic related air quality using measurement data from the traffic stations and available supplementary data has been tested. Different options on how to include such a traffic map layer in the background map and in the exposure estimates has been examined. All the analysis has been based on 2013 data.

Chapter 2 describes the methodology applied. Chapter 3 documents the input data. Chapter 4 presents the analysis and results of the application of land cover and road-type data as supplementary data sources in addition to the current mapping methodology of the rural background and urban background areas, while Chapter 5 shows the same when including additionally a traffic map layer in the mapping methodology.

The Annex offers a brief literary review and our findings on the potential of adopting methods from land use regression (LUR) modelling. The findings of this review were taken into account while designing the most suitable candidates for further analysis, which is presented in the main body of this paper.

2 Methodology

2.1 Current method

The current mapping methodology used to create the NO₂ concentration maps is described in Horálek et al. (2013, 2017). So far, only background NO₂ maps have been constructed, therefore not accounting for traffic.

The mapping method consists of a linear regression model followed by kriging of the residuals from that regression model (residual kriging):

$$\hat{Z}(s_0) = c + a_1 X_1(s_0) + a_2 X_2(s_0) + \dots + a_n X_n(s_0) + \hat{\eta}(s_0), \quad (2.1)$$

where $\hat{Z}(s_0)$ is the estimated concentration at a point s_0 ,
 $X_1(s_0), X_2(s_0), \dots, X_n(s_0)$ are n individual supplementary variables at point s_0 ,
 c, a_1, a_2, \dots, a_n are the $n+1$ parameters of the linear regression model calculated based on the data at the points of measurement,
 $\hat{\eta}(s_0)$ is the spatial interpolation of the residuals of the linear regression model at point s_0 , based on the residuals at the points of measurement.

The spatial interpolation of the regression's residuals is carried out using ordinary kriging, according to

$$\hat{\eta}(s_0) = \sum_{i=1}^N \lambda_i \eta(s_i) \quad \text{with} \quad \sum_{i=1}^N \lambda_i = 1, \quad (2.2)$$

where $\hat{\eta}(s_0)$ is the interpolated value at a point s_0 , derived from the residuals of the linear regression model at the points of measurement $s_i, i = 1, \dots, N$,
 $\eta(s_i)$ are the residuals of the linear regression model at N points of measurement $s_i, i = 1, \dots, N$,
 $\lambda_1, \dots, \lambda_N$ are the estimated weights based on the variogram, which is a measure of a spatial correlation, see Cressie (1993).

Separate map layers are created for the rural and the urban background areas on a grid at 10x10 km resolution. The rural background map layer is based on the rural background stations and the urban background map layer on the urban and the suburban background stations. Subsequently, the rural background map layer and the urban background map layer are merged into one combined final map using a weighting procedure based on the population density grid at 1x1 km resolution, according to

$$\hat{Z}_F(s_0) = (1 - w_U(s_0)) \cdot \hat{Z}_R(s_0) + w_U(s_0) \cdot \hat{Z}_{UB}(s_0) \quad (2.3)$$

where $\hat{Z}_F(s_0)$ is the resulting estimated concentration in a grid cell s_0 for the final map,
 $\hat{Z}_{UB}(s_0)$ is the estimated concentration in a grid cell s_0 for the urban background map layer,
 $\hat{Z}_R(s_0)$ is the estimated concentration in a grid cell s_0 for the rural background map layer,
 $w_U(s_0)$ is the weight representing the ratio of urban areas in a grid cell s_0 .

For further details, see Horálek et al. (2006, 2017).

Population exposure for individual countries and for Europe as a whole is calculated from the air quality maps and population density data, both at 1x1 km resolution. For each concentration class ‘ j ’, the percentage population per country as well as the European-wide total is determined:

$$P_j = \frac{\sum_{i=1}^N I_{ij} p_i}{\sum_{i=1}^N p_i} \cdot 100 \quad (2.4)$$

where P_j is the percentage population living in areas of the j -th concentration class in either the country or in Europe as a whole,
 p_i is the population in the i -th grid cell,
 I_{ij} is the Boolean 0-1 indicator showing whether the concentration in the i -th grid cell is within the j -th concentration class ($I_{ij} = 1$), or not ($I_{ij} = 0$),
 N is the number of grid cells in the country or in Europe as a whole.

In addition, we express per-country and European-wide exposure as the population-weighted concentration, i.e. the average concentration weighted according to the population in a grid cell:

$$\hat{c} = \frac{\sum_{i=1}^N c_i p_i}{\sum_{i=1}^N p_i} \quad (2.5)$$

where \hat{c} is the population-weighted average concentration in the country or in Europe as a whole,
 c_i is the concentration in the i -th grid cell,
 p_i is the population in the i -th grid cell,
 N is the number of grid cells in the country or in Europe as a whole.

2.2 Land cover and road data inclusion

As presented in literature review in Annex, land use regression (LUR) techniques have been widely used in the NO₂ mapping. In principle, these techniques use different supplementary variables in a multiple linear regression. Our goal is to select the most useful of such variables for inclusion in the regression part of Equation 2.1. Only those variables that are available at the European scale are included in the selection.

Annex shows that a diversity of traffic-related parameters explain the largest amount of variability for NO₂ land use regression models. Thus, not only the land cover, but also the available traffic-related, i.e. the road data (see Section 3.5) is investigated for the use in the NO₂ mapping.

The most useful supplementary data are selected by using a stepwise regression and backwards elimination (Horálek et al., 2007). Subsequently, based on the selected set of parameter data, the maps are constructed. Through cross-validation (Section 2.4) these maps are compared with the map as prepared under the current methodology.

The analysis (see Section 4.1) is performed separately for the rural and urban background areas, at the 1x1 km resolution. This means a shift into a finer resolution compared to the current method, in which the separate rural and urban background map layers are constructed at 10x10 km resolution.

The population exposure (see Section 4.2) is calculated on basis of the air quality maps and population density data, both at 1x1 km resolution, which is similar to the current methodology.

2.3 Traffic map layer inclusion

A map of the traffic related air quality, the so-called *traffic map layer*, is constructed according to Equation 2.1, based on the traffic stations. The supplementary data (including the land cover and the road data) are selected using the stepwise regression and backward elimination.

In principle, we intended to distinguish *urban traffic* and *rural traffic* areas, each based on each type of traffic station. However, out of the total pool of 874 traffic stations fulfilling the data coverage criterion, only 19 are classified as rural traffic. This small set of rural traffic stations is too little to allow for statistically sound analysis and it lacks accuracy in spatial distribution and density to serve as source for a representative rural traffic map layer. Contrary to that, the large number of urban traffic stations allows execution of sound statistical analysis and to create a spatially representative urban traffic map layer.

This made us decide to use only the urban traffic stations to construct an *urban traffic map layer*. This traffic map layer should be applied in the urban areas where the air quality is classified as being under direct influence of traffic.

As the component of the traffic map layer is representative for urban traffic areas only, we included it in the urban background map layer, resulting in an adapted urban map layer, according to

$$\hat{Z}_U(s_0) = (1 - w_T(s_0)) \cdot \hat{Z}_{UB}(s_0) + w_T(s_0) \cdot \hat{Z}_T(s_0) \quad (2.6)$$

where $\hat{Z}_U(s_0)$ is the resulting estimated concentration in a grid cell s_0 for the urban map layer,
 $\hat{Z}_{UB}(s_0)$ is the estimated concentration in a grid cell s_0 for the urban background map layer,
 $\hat{Z}_T(s_0)$ is the estimated concentration in a grid cell s_0 for the urban traffic map layer,
 $w_T(s_0)$ is the weight representing the ratio of areas exposed to traffic air quality in a grid cell s_0 .

The weight w_T is estimated for each 1x1 km grid cell, based on the detailed road data, see Section 5.2. The resulting urban map layer is constructed at 1x1 km resolution. Subsequently, it is merged with the rural map layer into the combined final map like in the current methodology (see Eq. 2.3), according to

$$\begin{aligned} \hat{Z}_F(s_0) &= (1 - w_U(s_0)) \cdot \hat{Z}_R(s_0) + w_U(s_0) \cdot \hat{Z}_U(s_0) \\ &= (1 - w_U(s_0)) \cdot \hat{Z}_R(s_0) + w_U(s_0) (1 - w_T(s_0)) \cdot \hat{Z}_{UB}(s_0) + w_U(s_0) w_T(s_0) \cdot \hat{Z}_T(s_0) \end{aligned} \quad (2.7)$$

where $\hat{Z}_F(s_0)$ is the resulting estimated concentration in a grid cell s_0 for the combined final map,
 $\hat{Z}_R(s_0)$ is the estimated concentration in a grid cell s_0 for the rural background map layer,
 $w_U(s_0)$ is the weight representing the ratio of urban areas in a grid cell s_0 .

The population exposure estimate is calculated separately for the areas where the air quality is considered to be directly influenced by traffic and for the background (both rural and urban) areas. The percentage population living in a given concentration class 'j' per country or for Europe as a whole, is calculated according to

$$P_j = \frac{\sum_{i=1}^N I_{Bij}(1 - w_U(i)w_T(i))p_i + \sum_{i=1}^N I_{Tij}w_U(i)w_T(i)p_i}{\sum_{i=1}^N p_i} \cdot 100 \quad (2.8)$$

where P_j is the percentage population living in areas of the j -th concentration class in either the country or in Europe as a whole,
 p_i is the population in the i -th grid cell,
 I_{Bij} is the Boolean 0-1 indicator showing whether the background air quality concentration (estimated by the combined rural/urban background map layer) in the i -th grid cell is within the j -th concentration class ($I_{Bij} = 1$), or not ($I_{Bij} = 0$),
 I_{Tij} is the Boolean 0-1 indicator showing whether the traffic air quality concentration in the i -th grid cell is within the j -th concentration class ($I_{Tij} = 1$), or not ($I_{Tij} = 0$),
 N is the number of grid cells in the country or in Europe as a whole.

The population-weighted concentration is calculated according to

$$\hat{c} = \frac{\sum_{i=1}^N c_{Bi}(1 - w_U(i)w_T(i))p_i + \sum_{i=1}^N c_{Ti}w_U(i)w_T(i)p_i}{\sum_{i=1}^N p_i} = \frac{\sum_{i=1}^N c_i p_i}{\sum_{i=1}^N p_i} \quad (2.9)$$

where c_{Bi} is the concentration in the i -th grid cell for the rural/urban background map layer,
 c_{Ti} is the concentration in the i -th grid cell for the traffic map layer.

One can see that Equation 2.9 turns to Equation 2.5, but calculated based on the final map (see Eq. 2.7).

2.4 Uncertainty estimates of the concentration maps

The uncertainty estimation of the mapping results is based on the ‘leave one out’ *cross-validation* method. It computes the quality of the spatial interpolation for each measurement point from all available information except from the point in question, i.e. it withholds one data point and then makes a prediction at the spatial location of that point. This procedure is repeated for all measurement points in the available set. The results of the cross-validation are expressed by statistical indicators and scatter plots. The main indicators used are *root mean squared error* (RMSE) and *bias*:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{Z}(s_i) - Z(s_i))^2} \quad Bias = \frac{1}{N} \sum_{i=1}^N (\hat{Z}(s_i) - Z(s_i)) \quad (2.10)$$

where $Z(s_i)$ is the observed air quality indicator value at the i^{th} point,
 $\hat{Z}(s_i)$ is the estimated air quality indicator value at the i^{th} point using other information, except the observed indicator value at the i^{th} point,
 N is the number of the observational points.

Next to the RMSE expressed in absolute units, one could express this uncertainty in percentage by relating the RMSE to the mean of the air quality indicator value for all stations:

$$RRMSE = \frac{RMSE}{\frac{1}{N} \sum_{i=1}^N Z(s_i)} \cdot 100 \quad (2.11)$$

where $RRMSE$ is the relative RMSE, expressed in percentage.

Other cross-validation indicators are the coefficient of determination R^2 and the regression equation parameters *slope* and *intercept*, following from the scatter plot between the cross-validation predicted and the observed concentrations.

3 Input data

3.1 Monitoring data

Air quality station monitoring data for the relevant year are extracted from the EEA Air Quality e-Reporting database, EEA (2015). Only data from stations classified by the Air Quality e-Reporting database and/or EBAS of the type *background* and *traffic* for the areas *rural*, *suburban* and *urban* are used. Station type *industrial* is not considered; it represents local scale concentration levels not applicable at the mapping resolution employed. The following pollutant and its indicator is considered:

NO₂ – annual average [$\mu\text{g}\cdot\text{m}^{-3}$], year 2013

Only the stations with annual data coverage of at least 75 percent are used. We excluded the stations outside the EEA map extent *Map_1c* (EEA, 2011).

In total, 377 rural background stations, 1083 urban/suburban background stations and 855 urban/suburban traffic stations are used. Due to the small number of the rural traffic stations (i.e. 19), these stations are further not considered and only the estimation of the urban traffic air quality is discussed in this paper (see Section 2.3).

3.2 Chemical transport modelling data

The chemical dispersion model used in this paper is the EMEP MSC-W (formerly called Unified EMEP) model (version rv4.7), which is an Eulerian model. Simpson et al. (2012, 2013) and https://wiki.met.no/emep/page1/emepmscw_opensource (web site of Norwegian Meteorological Institute) describe the model in more detail. Emissions for the relevant year 2013 (Mareckova et al., 2015) are used and the model is driven by ECMWF meteorology for the relevant year 2013. EMEP (2015) provides details on the EMEP modelling for 2013. The resolution of this model run is $0.1^\circ \times 0.1^\circ$, i.e. circa 10x10 km. Information from this model has been converted to 1x1 km grid resolution: the data representing the EMEP grid cells are imported into *ArcGIS* and transformed into the ETRS89-LAEA5210 projection, subsequently converted into a 100x100 m resolution raster grid and spatially aggregated into the reference EEA 1x1 km grid. The parameter used is the same as for the monitoring data, i.e.

NO₂ – annual average [$\mu\text{g}\cdot\text{m}^{-3}$], year 2013

3.3 Altitude, meteorological data, population density

The *altitude* data field (in m) with an original grid resolution of 15x15 arcseconds comes from U.S. Geological Survey Earth Resources Observation and Science (GTOPO), see Danielson et al. (2011). The data were converted into the ETRS89-LAEA5210 projection, resampled to 100x100 m resolution, shifted to the extent of EEA reference grid, and spatially aggregated into 1x1 km grid resolution.

Next to this, another aggregation has been executed based on the 1x1 km grid cells, i.e. the floating averaging of the circle with *radius of 5 km* around all relevant grid cells. For motivation, see Section 3.4.

The *meteorological parameters* used are *wind speed* (annual average for 2013, in $\text{m}\cdot\text{s}^{-1}$), *surface net solar radiation* (annual average of daily sum for 2013, $\text{MW}\cdot\text{s}\cdot\text{m}^{-2}$), *temperature* (annual average for 2013, $^\circ\text{C}$) and *relative humidity* (annual average for 2013, percentage). The daily data in resolution 15x15 arc-seconds were extracted from the Meteorological Archival and Retrieval System (MARS) of

ECMWF, see ECMWF (2015). For details, see Horálek et al. (2007). The data have been imported into ArcGIS as a point shapefile. Each point represents the centre of a grid cell. The shapefile has been converted into ETRS89-LAEA5210 projection, converted into a 100x100 m resolution raster grid and spatially aggregated into the reference EEA 1x1 km grid.

Population density (in inhabitants.km⁻², census 2011) is based on Geostat 2011 grid dataset (Eurostat, 2014). The dataset is in 1x1 km resolution, in the EEA reference grid. For regions not included in the Geostat 2011 dataset we use as alternative sources JRC (2009) and ORNL (2008) data. For details, see Horálek et al. (2017). Next to the basic resolution of 1x1 km, the floating averaging of the circle with radius 5 km around all individual 1x1 km grid cells has been prepared. For motivation, see Section 3.4.

3.4 Land cover

CORINE Land Cover 2006 – grid 100 x 100 m, Version 17 (12/2013) is used (CLC2006 – 100m, g100_06.zip; EEA, 2013b). The countries missing in this database are Andorra and Greece; the areas missing are Faroe Islands, Jersey and Guernsey. Greece is missing in the CLC2006 but present in the CLC2000. Therefore, we inserted for Greece the CLC2000 data (grid 100 x 100 m, Version 17, 12/2013 EEA, 2013a).

In order to reduce the high number of degrees of freedom in the CORINE Land Cover description, the 44 CLC classes (<http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version-3/corine-land-cover-2006-classes>) have been re-grouped into the 8 more general classes in agreement with the recommendations of Annex, Section A.4, i.e. similarly like in Beelen et al. (2013).

Table 3.1 Definition of general land cover classes, based on CLC2006 classes

Label	General class description	CLC classes grid codes	CLC classes codes	CLC classes description
HDR	High density residential areas	1	111	Continuous urban fabric
LDR	Low density residential areas	2	112	Discontinuous urban fabric
IND	Industry	3, 7, 8	121, 131, 132	Industrial or commercial units, Mineral extraction sites, Dump sites
TRAF	Traffic	4 – 6	122 – 124	Road and rail networks and associated land, Ports, Airports
UGR	Urban green	10 – 11	141 – 142	Artificial, non-agricultural vegetated areas
AGR	Agricultural areas	12 – 22	211 – 244	Agricultural areas
NAT	Natural areas	23 – 34	311 – 335	Forest and semi natural areas
OTH	Other areas	9, 35 – 44	133, 411 – 523	Construction sites, Wetlands, Water bodies

Annex, Section A.4 recommends different buffer sizes to be used. Under the ESCAPE project, buffer sizes of 100, 300, 500, 1000, and 5000 m were used (Beelen et al., 2013). In this paper, for consistency with other supplementary data, the 1x1 km grid resolution is used as the basic grid resolution. Based on this, two aggregations are used, i.e. into 1x1 km grid and into the circle with radius of 5 km. For each general CLC class we spatially aggregated the high land use resolution into the 1x1 km EEA standard grid resolution. The aggregated grid square value represents for each general class the total area of this class as percentage of the total 1x1 km square area. For the floating averaging of the circle with radius 5 km around all relevant grid cells, the aggregated grid square value represents for each general class the total area of this class as percentage of the total area of this circle (which is 8.1 square kilometers; this value is influenced by the 100x100 m resolution of the land cover data).

In the analysis, the first seven classes have been used only, while the class OTHER has been omitted as redundant, as it can be expressed by the other general classes: the percentage of the grid square area attributed to this class can be calculated by subtracting the percentages attributed to other seven classes from 100.

3.5 Road type vector data

GRIP (Meijer et al., 2016) vector road type data base provided by PBL is used. The road types are distributed into 5 classes, from highways to local roads and streets, see Table 3.2.

GRIP stands for Global Road Inventory Project and PBL aims to cover in the end all roads worldwide. PBL intends to maintain and expand the database in coming years and to release it as open source data as soon as they have published about the database and its application. Its primary application is in climate change oriented research.

PBL aimed at a worldwide coverage with a precision of about 250 meters. We were able to use the vector data for the European window on an ‘as is’ basis, which means that no formal quality stamp is given by PBL. The GRIP vector data is not publicly available yet but PBL publication is in preparation. We received the database under embargo and cannot make it available to others. In the analysis, we used the classes 1 – 4 as these road types have high traffic density and intensity and as such do contribute significantly to air quality impacts in the direct vicinity of these road types. Class No. 5 is considered not relevant for the traffic air quality assessment. Due to their low traffic intensity they contribute limitedly to the traffic air quality exposure of the population in its direct vicinity.

Table 3.2 Definition of GRIP type classes

GRIP type class number	Class description
1	Highways
2	Primary roads
3	Secondary roads
4	Tertiary roads
5	Local, residential, urban roads

Based on the GRIP vector data, three characteristics were calculated in ESRI ArcGIS for different road types and their combinations, namely

- *Length of the roads* in km: for the individual classes 1 – 4 and for all classes 1 – 4 together, at all 1x1 grid cells,
- *Percentage of the area influenced by traffic* is represented by buffers around the roads: for the individual classes 1 – 4, for all classes together and for classes 1 – 3 together, at all 1x1 grid cells; a buffer of 75 metres distance at each side from each road vector is taken for the roads of classes 1 and 2, while a buffer of 50 metres is taken for the roads of classes 3 and 4,
- *Monitoring station distance from the nearest road* in km: for the individual classes 1 – 4 and for all classes 1 – 4 together, for all stations.

The parameter “*Percentage of the area influenced by traffic*” is based on the buffers around the roads. The size of the buffers is chosen with a large degree of simplification. According to EEA (1999), the distance of the urban and suburban background stations from the nearest road should be more than 50 meters, supposing this is a distance further not influenced by traffic. According to Su et al. (2015), the impact of highways in an open area comes up to 300 meters, while the impact of major

roadways up to 50 meters. In our setting of size of the buffers, we took into account the urban character of the area the buffers are used for.

During the analysis, we discovered that in the GRIP data the different traffic lanes are considered as separate roads, see the example given in Figure 3.1. This observation disqualified the parameter “*Length of the roads*” for the analysis. Instead, we used the parameter “percentage of the area influenced by traffic” using the buffers around the roads.

Figure 3.1 Road GRIP vector data with different road type classes, example



4 Land cover and road type inclusion

4.1 Background concentration map

First, we selected the suitable set of supplementary data, separately for the rural background and urban background areas. The supplementary data tested on suitability for inclusion in the linear regression model included:

- EMEP model
- altitude (GTOPO 1x1 km grid altitude, and the GTOPO floating average of circle radius 5 km around 1x1km grid cell)
- meteorological parameters
 - wind speed
 - temperature
 - surface net solar radiation
 - relative humidity
- population (1x1 km and radius 5 km)
- road type data
 - percentage of the area influenced by traffic, for individual road type classes 1–4, for all classes 1–4 together and for classes 1–3 together (1x1 km)
 - monitoring station distance from the nearest road of the given class, for individual road type classes 1–4 and for all classes 1–4 together (in km)
- land cover type data
 - HDR (1x1 km and radius 5 km)
 - LDR (1x1 km and radius 5 km)
 - IND (1x1 km and radius 5 km)
 - TRAF (1x1 km and radius 5 km)
 - UGR (1x1 km and radius 5 km)
 - AGR (1x1 km and radius 5 km)
 - NAT (1x1 km and radius 5 km).

Apart from the monitoring station distance from the nearest road, all the parameters are related to 1x1 km grid.

The most useful supplementary data have been selected through a stepwise regression and backwards elimination (Horálek et al., 2007). The set of the supplementary variables has been selected in two variants, i.e. one with the inclusion of the chemical transport model EMEP among the supplementary variables and one without. The reason for the examination of the variant without the use of the EMEP model is the fact that the land use regression techniques mostly do not use the chemical transport models. As we like to gain insight on how the land use contributes in the performance of the regression model independent from the influence of a chemical transport model (CTM), we choose these two variants.

The two selected variants are compared with the current methodology (Horálek et al., 2014), which uses as supplementary data the EMEP model, altitude and wind speed, both for rural and urban background areas. However altitude is not statistically significant for 2013 data at the 10x10 km resolution (see Table 4.2), and is therefore excluded from the linear regression model on that resolution.

For better comparability with the improved variants that are all executed on 1x1 km, the current methodology – which is performed at the 10x10 km resolution – is also performed for both the rural and urban background areas as additional variants at the 1x1 km resolution. The same supplementary variables as at 10x10 km are applied (i.e. EMEP model and wind speed). Next to this, an improved

current version is additionally performed, adding altitude in the rural areas (as is a statistically significant variable at the 1x1 km resolution) and population density in the urban areas (in agreement with Horálek et al., 2017). Altogether, five methods with different variants of the linear regression model are mutually compared and summarised in Table 4.1.

Table 4.1 List of mutually compared mapping methods

Label	Method description	Area type	Grid resolution	CTM	Meteo- rology	Alti- tude	Popu- lation	CLC and road data
(i)	Current method, 10x10 km	Rural background	10x10 km	+	+	+ ^(a)	-	-
		Urban background	10x10 km	+	+	+ ^(a)	-	-
(ii)	Current method, 1x1 km	Rural background	1x1 km	+	+	-	-	-
		Urban background	1x1 km	+	+	-	-	-
(iii)	Improved current method, 1x1 km	Rural background	1x1 km	+	+	+	-	-
		Urban background	1x1 km	+	+	+ ^(a)	+	-
(iv)	Land cover (LC) and road data included	Rural background	1x1 km	+	+	+	+ ^(b)	+
		Urban background	1x1 km	+	+	+	+	+
(v)	LC and road data incl., without CTM	Rural background	1x1 km	-	+	+	+	+
		Urban background	1x1 km	-	+	+	+	+

^(a) Statistically non-significant parameter, see Table 4.2.

^(b) Not selected by the selecting procedure, see Table 4.2.

The selected variables at the method including all variables (iv) are:

rural background areas: *EMEP model*, *wind speed*, altitude (both 1x1 km *GTOPO_1km*, and 5 km radius *GTOPO_5km_rad*), and land cover parameters *LDR_1km* and *NAT_1km*;

urban background areas: *EMEP model*, *wind speed*, altitude (both 1x1 km *GTOPO_1km* and 5 km radius *GTOPO_5km_rad*), *population* (1x1 km), area influenced by traffic of class 1 (1x1 km) *T1buf75m_1km*, and land cover parameters *AGR_1km*, *NAT_1km*, and *LDR_5km_rad*.

At the method excluding the EMEP model (v), the following variables were selected:

rural background areas: *wind speed*, *surface solar radiation*, *temperature*, altitude (both 1x1 km *GTOPO_1km* and 5 km radius *GTOPO_5km_rad*), *population* (1x1 km), and land cover parameters *LDR_5km_rad*, *TRAF_5km_rad* and *NAT_5km_rad*;

urban background areas: *wind speed*, altitude (both 1x1 km *GTOPO_1km* and 5 km radius *GTOPO_5km_rad*), *population* (1x1 km), area influenced by traffic of class 1 roads (1x1 km) *T1buf75m_1km*, and land cover parameters *HDR_5km_rad*, *LDR_5km_rad*, *TRAF_5km_rad* and *NAT_5km_rad*.

It should be noted that – quite surprisingly – out of all road data parameters, only the area influenced by traffic of class 1 (i.e. *T1buf75m_1km*) was selected in the procedure of stepwise regression and backwards elimination.

Table 4.2 presents, next to the supplementary variables ultimately applied, the relevant statistical parameters for both multiple linear regression and the subsequent interpolation by the ordinary kriging of its residuals.

Table 4.2 Parameters of the linear regression models and of the ordinary kriging variograms (nugget, sill, range) of NO₂ annual average for 2013 in rural and urban areas for each of the five methods (i) – (v)

linear regr. model + OK of its residuals	(i) current		(ii) curr. 1k		(iii) impr. curr. 1k		(iv) incl. LC		(v) without CTM	
	rural	urban	rural	urban	rural	urban	rural	urban	rural	urban
	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.	coeff.
c (constant)	5.10	19.30	4.79	19.24	7.95	18.88	5.67	19.23	24.02	26.87
a1 (EMEP model)	0.830	0.565	0.841	0.524	0.765	0.487	0.766	0.452		
a2 (GTOPO_1km)	<i>n. sign.</i>	<i>n. sign.</i>			-0.0025	<i>n. sign.</i>	-0.0098	-0.0188	-0.0083	-0.0229
a3 (GTOPO_5km_rad)							0.0095	0.0167	0.0062	0.0182
a4 (wind speed)	-0.791	-1.701	-0.732	-1.666	-1.172	-1.728	-0.643	-1.94	-1.846	-3.17
a5 (s. solar radiation)									-0.917	
a6 (temperature)									0.377	
a7 (population_1km)						0.00023		0.00016	0.00200	0.00021
a8 (T1buf75m_1km)								10.74		9.47
a9 (LDR_1km)							0.0638			
a10 (AGR_1km)								-0.0273		
a11 (NAT_1km)							-0.0177	-0.0663		
a12 (HDR_5km_rad)										0.0185
a13 (LDR_5km_rad)								0.0064	0.0227	0.0189
a14 (TRAF_5km_rad)									0.0752	0.0342
a15 (NAT_5km_rad)									-0.0068	-0.0057
adjusted R²	0.66	0.48	0.67	0.48	0.69	0.51	0.76	0.57	0.54	0.41
st. err. [µg.m⁻³]	3.48	5.55	3.40	5.52	3.31	5.40	2.89	5.04	4.03	5.89
nugget	12	19	11	19	10	17	3	15	9	17
sill	13	25	13	24	12	23	9	21	15	29
range [km]	230	310	230	310	230	310	30	290	150	290

Note: Dark grey indicates variables not considered in the variant of the linear regression model. Light grey indicates variables not selected in the variant by the selecting procedure.

The land cover parameters show for agricultural areas (*AGR_1km*) and natural areas (*NAT_1km*, *NAT_5km_rad*) a negative dependence, while traffic influences (*T1buf75m_1km*, *TRAF_5km_rad*), high density residential areas (*HDR_5km_rad*) and low density residential areas (*LDR_5km_rad*) show a positive dependence. It means that accounting for the agricultural and nature land cover parameters reduces the NO₂ air concentrations at regression predictions of an area or point, and that the traffic and residential parameters will provide an increase in NO₂ concentrations at the regression predictions.

Quite interesting is the NO₂ dependence on the altitude: at both the rural and the urban background areas there is a negative dependence at the 1x1 km altitude grid (*GTOPO_1km*), while the floating average altitude at the 5 km radius around 1x1 km grid cells (*GTOPO_5km_rad*) shows a positive dependence. On top of that, the negative dependence is slightly stronger. While the variable *GTOPO_1km* shows the altitude of a grid cell, *GTOPO_5km_rad* gives the mean altitude in the circle of 5 km around the grid cell. Together, they show whether the vicinity of a given grid cell is concave (and better aired, leading to lower NO₂ concentrations) or convex (and worse aired, leading to higher NO₂ concentrations).

Furthermore, one can conclude that the best linear regression results (i.e. prior to interpolation) are given by variant (iv), i.e. including land cover and road data, using the EMEP model.

The mapping results of all five methods are mutually compared by means of the ‘leave one out’ cross-validation (Section 2.4). The comparison results are presented in Table 4.3. The best results are marked dark green, the second best light green.

Table 4.3 Comparison of different methods of spatial interpolation showing RMSE, RRMSE, bias, R^2 and linear regression from the cross-validation scatter plots of NO_2 annual mean predicted values, 2013. Units: $\mu\text{g}\cdot\text{m}^{-3}$ except RRMSE and R^2 .

spatial interpolation variant + supplementary data used		rural areas				
		RMSE	RRMSE	bias	R^2	regr. eq.
(i)	current (EMEP, wind speed; 10x10 km)	3.4	35.8%	0.1	0.699	$y = 0.677x + 3.2$
(ii)	current 1km (EMEP, wind speed; 1x1 km)	3.4	35.2%	0.1	0.681	$y = 0.694x + 2.0$
(iii)	impr. current 1km (EMEP, altitude, wind speed; 1x1 km)	3.2	33.8%	0.1	0.706	$y = 0.725x + 2.7$
(iv)	including LC (EMEP, altitude, w. sp., land cover; 1x1 km)	2.8	29.2%	0.1	0.782	$y = 0.810x + 1.9$
(v)	without CTM (alt., w.sp., s.s. rad., temp., pop., LC; 1x1 km)	3.3	34.6%	0.2	0.698	$y = 0.760x + 2.5$

spatial interpolation variant + supplementary data used		urban background areas				
		RMSE	RRMSE	bias	R^2	regr. eq.
(i)	current (EMEP, wind speed; 10x10 km)	5.1	23.9%	0.0	0.557	$y = 0.572x + 9.2$
(ii)	current 1km (EMEP, wind speed; 1x1 km)	5.1	23.6%	0.0	0.568	$y = 0.586x + 8.9$
(iii)	impr. current 1km (EMEP, wind speed, population; 1x1 km)	4.8	22.7%	0.0	0.603	$y = 0.624x + 8.1$
(iv)	including LC (EMEP, alt., w. sp., pop., road, LC; 1x1 km)	4.6	21.3%	0.0	0.645	$y = 0.670x + 7.1$
(v)	without CTM (alt., w.sp., pop., road, land cover; 1x1 km)	4.7	22.1%	0.0	0.624	$y = 0.659x + 7.4$

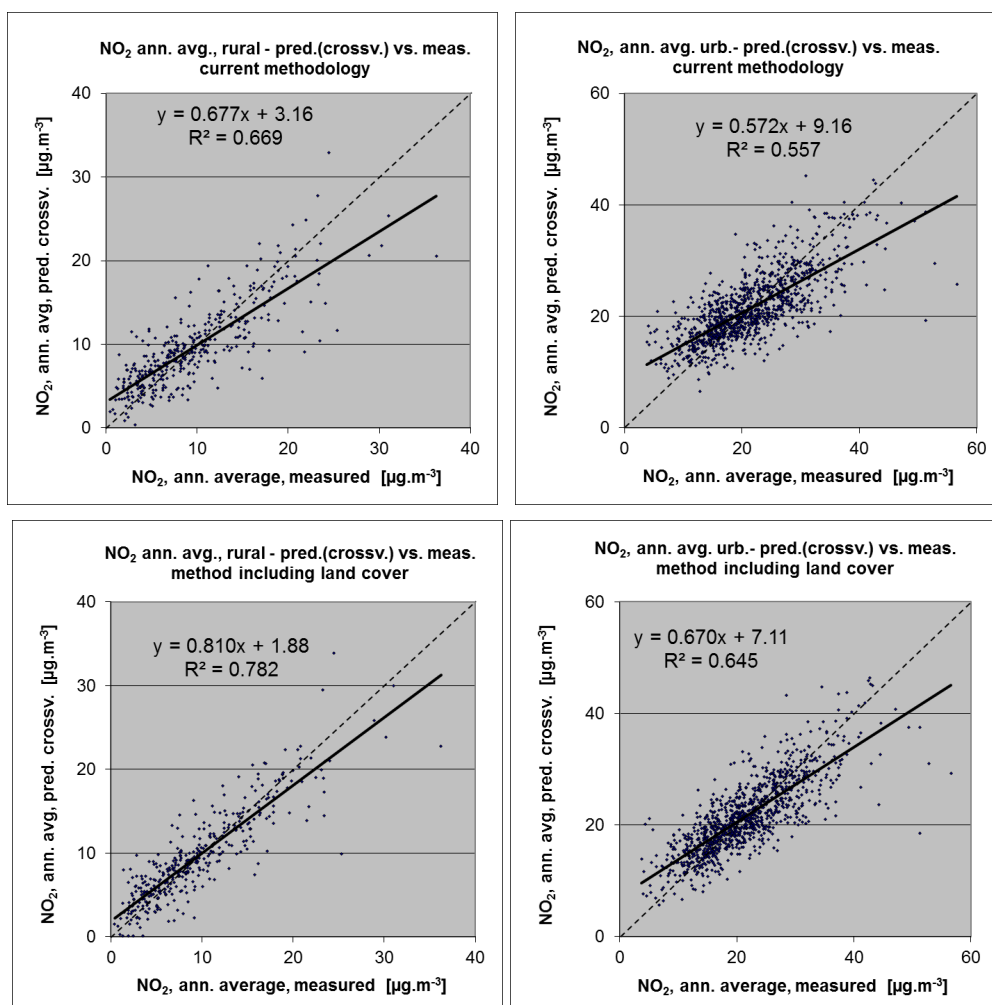
It can be seen that the best results are given by the variant (iv), i.e. current method including land cover and road data, using the EMEP model, at both rural and urban background areas. Compared to the current methodology (i), one can see the improvement of the relative RMSE from 36 % to 29 % for rural areas and from 24 % to 21 % for urban background areas.

The second best results are given by the improved current method (iii) for rural background areas and by the variant (v), i.e. current method including land cover and road data but without the use of the EMEP model for urban background areas.

For better illustration of the results presented in Table 4.3, the cross-validation scatter plots are presented in Figure 4.1 for the current method (i) and for the method including the land cover and road data (iv) at both rural background and urban background areas. One can see the improvement of the R^2 of the cross-validation scatter plot for method (iv) compared to (i) going from 0.67 to 0.78 at rural areas and from 0.56 to 0.65 at urban background areas. The regression equation also proves to provide a clear improvement at both the slope and intercept.

One can conclude that the inclusion of the land cover and road data provides clear improvement on the NO_2 mapping methodology. Therefore, it is recommended to implement these supplementary data sources in the routine methodology. When introducing this, it is recommended to also move the application of the 1x1 km resolution from the combined final merging process-step to the early process-step of creation of the separate rural and urban background map layers.

Figure 4.1 Correlation between cross-validated predicted and measurement values for NO₂ annual average 2013 for rural background (left) and urban background (right) areas, for current method (i) (top) and method including LC and road data (iv) (bottom)

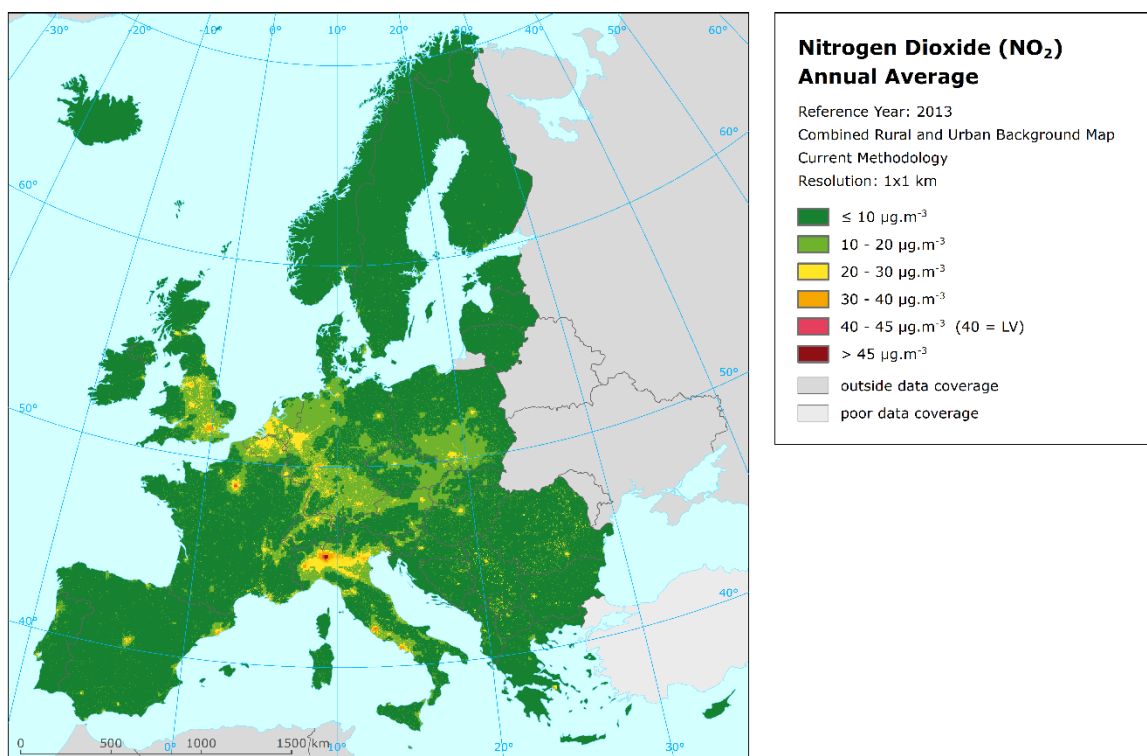


Map 4.1 presents the combined final map as a result of the merging of the rural background and urban background map layers created by means of the current methodology (i) using the 10x10 km grid resolution in the rural and urban map layers creation.

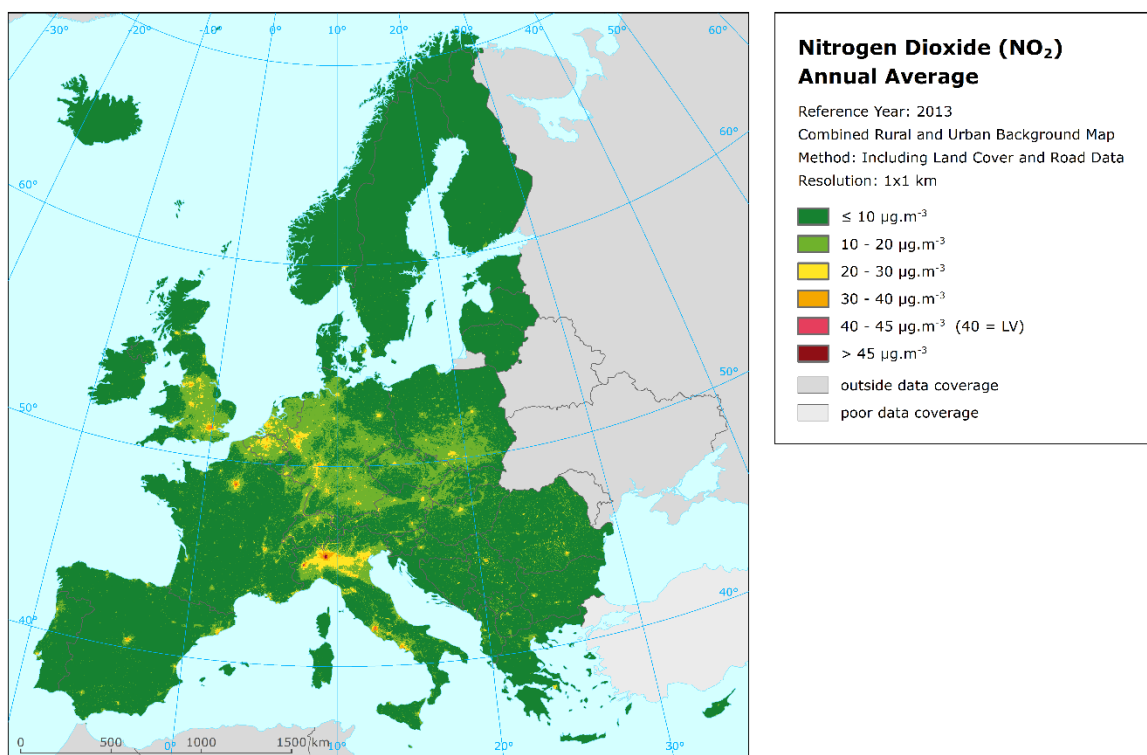
Map 4.2 gives the same map created by means of the method (iv) including the land cover and road data on 1x1 km grid resolution. In the limited areas with lacking land cover data, i.e. Andorra, Jersey, Guernsey, and Faroes (Section 3.3), the improved current version (iii) is used instead.

Map 4.3 shows that in most of Europe, the improved method (iv) provides in general a slight decrease or increase in concentrations of -2 to 2 $\mu\text{g}/\text{m}^3$ compared to the current method (i), with larger areas showing slight decrease. Only in scattered areas of northern and southern Spain, north-eastern France, central Germany, Po Valley in Italy and in Serbia and Romania it leads to somewhat more increased concentrations. Contrary to that, scattered areas of southern Norway, southern Finland, central and southern Greece, central Spain, the Netherlands, the Alps and the Pyrenees show more reduced concentrations if the improved method (iv) is used.

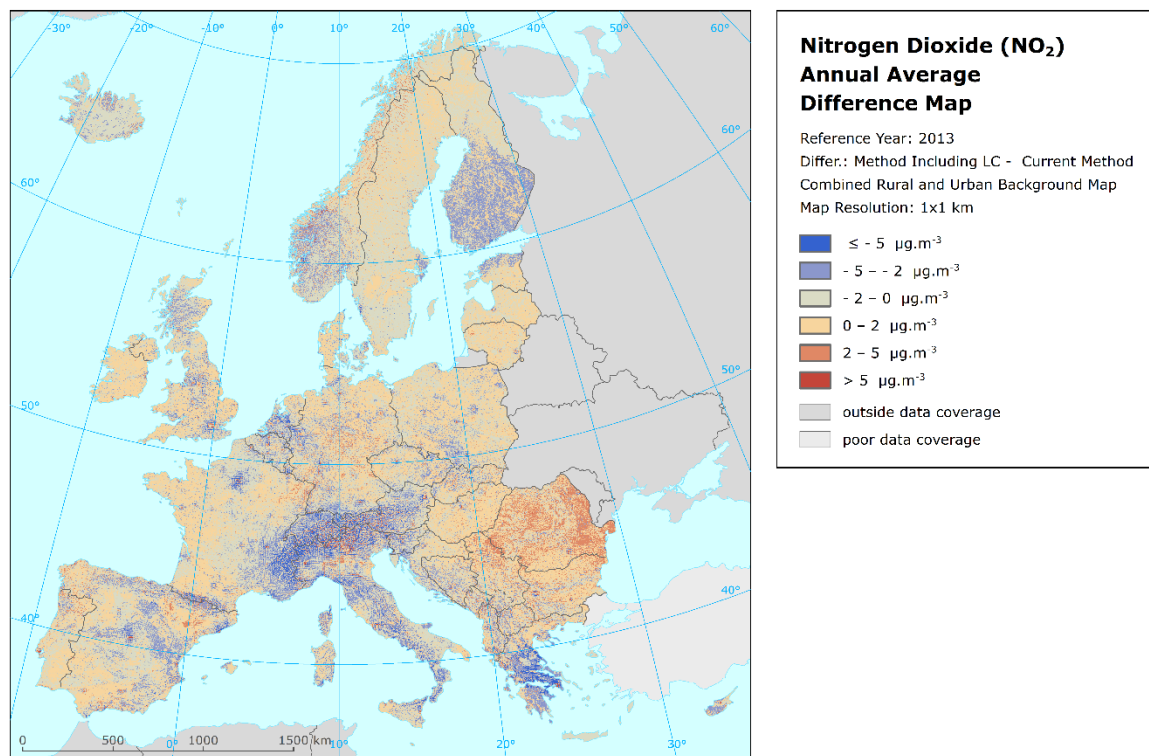
Map 4.1 Rural and urban background map of NO₂ annual average for 2013, current method (i)



Map 4.2 Rural and urban background map of NO₂ annual average for 2013, method (iv) including land cover and road data



Map 4.3 **Difference rural and urban background map, NO₂ annual average for 2013, difference between method including land cover and road data (iv) and current method (i)**



4.2 Population exposure

Based on the concentration maps presented in Section 4.1, population exposure tables have been calculated, based on Equation 2.5.

Table 4.4 gives, for the current method (i) as presented in Map 4.1, the population frequency distribution for a limited number of exposure classes, as well as the population-weighted concentration for individual countries and for Europe as a whole. This population exposure of NO₂ annual average for 2013 based on the current method has been used in EEA (2016) for the NO₂ health impact assessment.

Table 4.5 offers the same, for the method including the land cover and road data (iv) as presented in Map 4.2.

At both the overall European and the EU-28 population-weighted annual average NO₂ concentrations the method (iv) including the land cover and road data provides an estimate for 2013 being 0.6 µg.m⁻³ lower. Taking into account a large ratio of European population living in the urban areas (being 77%, while 16% in the mixed and 7% in the rural areas), the most likely explanation for this is a more realistic estimation of the population exposure in urban background areas due to taking into account the land cover data source.

However, it should be noted that the population exposure is based on the maps representing the background areas only, not the traffic areas. The inclusion of the traffic map layer is further discussed in Chapter 5.

Table 4.4 Population exposure and population-weighted concentrations, NO₂ annual average for 2013, current method (i)

Country		Population [inhbs . 1000]	NO ₂ annual average, exposed population [%]						Population-weighted conc. [µg.m ⁻³]
			< LV				> LV		
			< 10 µg.m ⁻³	10 - 20 µg.m ⁻³	20 - 30 µg.m ⁻³	30 - 40 µg.m ⁻³	40 - 45 µg.m ⁻³	> 45 µg.m ⁻³	
Albania	AL	2 899	18.6	67.6	13.8				15.9
Andorra	AD	76	1.8	98.2					14.3
Austria	AT	8 452	8.9	43.3	47.9				19.3
Belgium	BE	11 162	1.0	17.4	68.4	13.2			23.6
Bosnia & Herzegovina	BA	3 836	24.9	43.3	31.8				15.7
Bulgaria	BG	7 285	15.2	55.9	28.9				16.5
Croatia	HR	4 262	22.6	51.2	26.2				15.8
Cyprus	CY	866	100.0						6.9
Czech Republic	CZ	10 516	6.4	72.2	21.4				17.1
Denmark	DK	5 603	25.5	60.7	13.9				13.0
Estonia	EE	1 320	30.0	70.0					10.8
Finland	FI	5 427	46.0	54.0					9.4
France (metropolitan)	FR	63 652	18.1	44.7	27.8	5.2	4.1		18.7
Germany	DE	80 524	3.8	41.2	52.9	2.1			20.4
Greece	GR	11 004	42.1	24.9	15.2	17.8			14.6
Hungary	HU	9 909	9.4	72.6	18.1				16.8
Iceland	IS	322	13	87.4					14.3
Ireland	IE	4 591	39	53.8	6.8				11.6
Italy	IT	59 685	3.8	32.2	38.4	20.5	1.7	3.4	24.5
Latvia	LV	2 024	27.3	54.1	18.6				13.7
Liechtenstein	LI	37	0.1	9.8	90.0				22.7
Lithuania	LT	2 972	29.4	70.6					11.5
Luxembourg	LU	537	0.5	21.1	78.3				23.4
Macedonia, FYROM of	MK	2 062	3.7	37.8	58.5				20.8
Malta	MT	421	3.1	96.9					12.0
Monaco	MC	38			100.0				23.2
Montenegro	ME	621	21.4	22.8	55.8				17.2
Netherlands	NL	16 780	0.5	32.2	67.4				21.3
Norway	NO	5 051	30.5	44.0	25.5				14.4
Poland	PL	38 063	15.4	62.6	22.0				16.1
Portugal (excl. Az., Mad.)	PT	9 977	21.5	62.8	15.7				14.0
Romania	RO	20 020	17.5	32.3	50.1				17.9
San Marino	SM	34	5.3	94.7					15.4
Serbia (incl. Kosovo*)	RS	8 997	14.4	16.7	68.9				20.0
Slovakia	SK	5 411	6.5	93.5					16.0
Slovenia	SI	2 059	17.2	39.5	43.4				17.6
Spain (excl. Canarias)	ES	44 623	9.3	61.1	18.6	11.0			18.0
Sweden	SE	9 556	27.1	72.9					11.5
Switzerland	CH	8 039	3.1	17.5	77.6	1.9			22.4
United Kingdom (& dep.)	UK	63 905	4.2	25.2	58.4	12.2			22.8
Total		532 614	11.5	43.6	37.6	6.3	0.7	0.4	19.2
			98.9				1.1		
EU-28		500 603	11.2	44.4	36.6	6.7	0.7	0.4	19.3
			98.9				1.1		
Kosovo*	KS	1 816	13.2	17.9	68.8				19.3
Serbia (excl. Kosovo*)	RS	7 182	14.7	16.4	68.9				20.2

*) under the UN Security Council Resolution 1244/99

Note: Turkey is not included in the calculation due to the lack of air quality data.

Table 4.5 Population exposure and population-weighted concentrations, NO₂ annual average for 2013, method (iv) using land cover and road data

Country		Population [inhbs . 1000]	NO ₂ annual average, exposed population [%]						Population- weighted conc. [µg.m ⁻³]
			< LV				> LV		
			< 10 µg.m ⁻³	10 - 20 µg.m ⁻³	20 - 30 µg.m ⁻³	30 - 40 µg.m ⁻³	40 - 45 µg.m ⁻³	> 45 µg.m ⁻³	
Albania	AL	2 899	17.6	47.6	34.8				17.1
Andorra	AD	76	6.0	94.0					14.0
Austria	AT	8 452	8.9	45.8	41.0	4.4			19.5
Belgium	BE	11 162	0.9	29.3	53.1	16.7			23.4
Bosnia & Herzegovina	BA	3 836	27.9	58.2	13.9				14.1
Bulgaria	BG	7 285	16.3	58.4	25.3				16.3
Croatia	HR	4 262	24.7	50.5	24.7				15.2
Cyprus	CY	866	48.2	51.8					9.4
Czech Republic	CZ	10 516	4.8	73.4	20.4	1.4			16.8
Denmark	DK	5 603	46.1	37.6	16.2				12.1
Estonia	EE	1 320	54.1	45.9					9.3
Finland	FI	5 427	61.6	38.3	0.1				8.7
France (metropolitan)	FR	63 652	22.6	45.2	19.8	9.1	3.0	0.2	17.7
Germany	DE	80 524	3.6	51.5	39.6	5.2	0.0		19.6
Greece	GR	11 004	39.8	24.8	26.6	8.9			15.2
Hungary	HU	9 909	7.8	73.8	18.1	0.3			16.5
Iceland	IS	322	24	75.6					12.9
Ireland	IE	4 591	50	37.3	13.2				11.3
Italy	IT	59 685	7.2	33.8	36.0	15.8	3.5	3.6	23.3
Latvia	LV	2 024	34.8	40.6	24.5				13.2
Liechtenstein	LI	37	1.1	33.9	65.0				20.4
Lithuania	LT	2 972	36.1	61.7	2.2				11.8
Luxembourg	LU	537	0.9	40.1	59.0				20.9
Macedonia, FYROM of	MK	2 062	3.4	44.1	50.7	1.8			20.2
Malta	MT	421	15.5	84.5					13.2
Monaco	MC	38			100.0				26.0
Montenegro	ME	621	20.8	48.2	31.0				16.3
Netherlands	NL	16 780	0.8	39.1	55.9	4.2			21.4
Norway	NO	5 051	36.7	44.5	18.7	0.1			13.3
Poland	PL	38 063	18.5	59.8	20.1	1.6			15.5
Portugal (excl. Az., Mad.)	PT	9 977	27.2	50.9	20.5	1.4			14.8
Romania	RO	20 020	14.1	47.0	34.6	4.4			18.1
San Marino	SM	34	28.0	72.0					11.4
Serbia (incl. Kosovo*)	RS	8 997	14.2	43.9	37.3	4.5			18.5
Slovakia	SK	5 411	3.9	84.5	11.6				16.0
Slovenia	SI	2 059	20.0	50.3	29.3	0.4			16.4
Spain (excl. Canarias)	ES	44 623	13.9	50.6	22.3	12.7	0.4		18.5
Sweden	SE	9 556	44.9	54.5	0.6				10.2
Switzerland	CH	8 039	4.1	40.9	50.7	4.3			20.6
United Kingdom (& dep.)	UK	63 905	5.5	40.2	41.2	10.7	2.4		21.4
Total		532 614	13.9	46.9	30.4	7.2	1.1	0.4	18.6
			98.5				1.5		
EU-28		500 603	13.2	54.4	24.9	6.1	0.8	0.5	18.7
			98.7				1.3		
Kosovo*	KS	1 816	0.0	3.4	16.5	27.6	52.4	0.0	16.7
Serbia (excl. Kosovo*)	RS	7 182	0.0	3.5	38.1	51.0	7.4	0.0	18.9

*) under the UN Security Council Resolution 1244/99

Note 1: Turkey is not included in the calculation due to the lack of air quality data.

Note 2: The percentage value "0.0" indicates an exposed population exists, but it is small and estimated less than 0.05 %. Empty cells mean: no population in exposure.

5 Traffic map layer inclusion

So far, only background NO₂ maps (representing both rural and urban background areas) have been constructed not including the monitoring information from for traffic stations, although traffic is the most important source of NO₂. In this chapter, the estimation of the traffic related air quality using measurement data from the traffic stations and available supplementary data is executed. Furthermore, we examined how to incorporate such a traffic map layer with the background map and the exposure estimates. We discuss in this chapter only the interpolation methodological variants applying the 1x1 km grid resolution in all process steps.

The traffic map layer is based on urban and suburban traffic stations only and applies as such in urban areas only. Section 2.3 provides the details on the reasons behind it: the lack of rural traffic stations prevents us from even considering the construction of an interpolated rural traffic map layer. Although we use mostly a description *traffic map layer* in the paper, note it means *urban traffic map layer* in fact.

5.1 Traffic map layer creation

Similarly to the rural background and the urban background areas, a map of traffic related to air quality in urban areas can be constructed, based on measurement data of the urban and suburban traffic stations and suitable supplementary data, using Equation 2.1.

First, the suitable set of supplementary variables for application in the linear regression model has been selected by the stepwise regression and backwards elimination from the same pool of variables as was used at the rural and urban background areas, see Section 4.1.

The selection has been executed in four variants: two variants with land cover and road data, i.e. one with the inclusion of the chemical transport model EMEP and one without, are compared with another set of two variants without land cover and road data, i.e. one variant without inclusion of altitude and population density averaged in 5 km radius and one with these variables. For simplicity, we call these two variants “current” and “alternative current” methods. Table 5.1 summarises these four methodical variants that are mutually compared on their interpolation prediction performance. The variant numbering running from (ii) to (v) is chosen to match with its comparable variants type examined in Chapter 4. The variant (i) of Chapter 4 (i.e. current method on 10x10 km resolution) is not applied, as earlier no traffic map layer in the 10x10 km resolution was constructed.

Table 5.1 List of mutually compared mapping methods

Label	Method description	Area type	Grid resol.	CTM	Meteo- rology	Alti- tude	Popu- lation	CLC and road data
(ii)	Current method, 1x1 km	Urban traffic	1x1 km	+	+	+	(^a)	-
(iii)	Alternative current method	Urban traffic	1x1 km	+	+	+	(^a)	-
(iv)	Land cover (LC) and road data included	Urban traffic	1x1 km	+	+	+	(^a)	+
(v)	Land cover (LC) and road data included, without CTM	Urban traffic	1x1 km	-	+	+	+	+

(^a) Not selected by the selecting procedure, see Table 5.2.

Based on the selection procedure, the following variables have been selected for the examined method variants:

The selected variables at the variant without land cover, road data and variables averaged in 5 km radius (ii) are: *EMEP model*, *wind speed*, *surface solar radiation*.

At the variant without land cover and road data (iii), the selected variables are: *EMEP model*, *surface solar radiation*, altitude (both 1x1 km *GTOPO_1km* and 5 km radius *GTOPO_5km_rad*).

The selected variables at the variant including all variables (iv) are: *EMEP model*, *wind speed*, altitude (both 1x1 km *GTOPO_1km* and 5 km radius *GTOPO_5km_rad*), area influenced by traffic of class 2 (1x1 km) *T2buf75m_1km*, and land cover parameter *LDR_5km_rad*.

At the variant excluding the EMEP model (v), the following variables were selected: *wind speed*, *surface solar radiation*, *temperature*, *relative humidity*, altitude (both 1x1 km *GTOPO_1km* and 5 km radius *GTOPO_5km_rad*), population in 5 km radius *population_5km_rad*, and land cover parameter *LDR_5km_rad*,

Table 5.2 presents the parameters for the multiple linear regression and subsequent ordinary kriging of its residuals.

Table 5.2 Parameters of the linear regression models and of the ordinary kriging variograms (nugget, sill, range) of NO₂ annual average for 2013 in traffic areas for each of the four methods (ii) – (v)

linear regr. model + OK of its residuals	(ii) current 1k urban traffic	(iii) altern. curr. urban traffic	(iv) incl. LC urban traffic	(v) without CTM urban traffic
	coeff.	coeff.	coeff.	coeff.
c (constant)	42.58	36.16	27.39	-109.71
a1 (EMEP model)	0.654	0.685	0.550	
a2 (GTOPO_1km)		-0.0172	-0.0254	-0.0204
a3 (GTOPO_5km_rad)		0.0168	0.0213	0.0176
a4 (wind speed)	-1.278		-1.529	1.628
a5 (s. solar radiation)	-1.100	-1.019		-3.414
a6 (temperature)				-1.550
a7 (relative humidity)				1.206
a8 (population_5km_rad)				0.000017
a8 (T2buf75m_1km)			7.263	
a10 (LDR_5km_rad)			0.0208	0.0248
adjusted R²	0.33	0.33	0.36	0.31
st. err. [µg.m⁻³]	10.80	10.80	10.53	10.93
nugget	82	83	74	70
sill	120	120	120	124
range [km]	260	260	340	370

Note: Dark grey indicates variables not considered in the variant of the linear regression model. Light grey indicates variables not selected in the variant by the selecting procedure.

To statistically compare the interpolation performance, all four methods are mutually compared by means of the ‘leave one out’ cross-validation (Section 2.4). The comparison results are presented in Table 5.3. The best results are marked dark green.

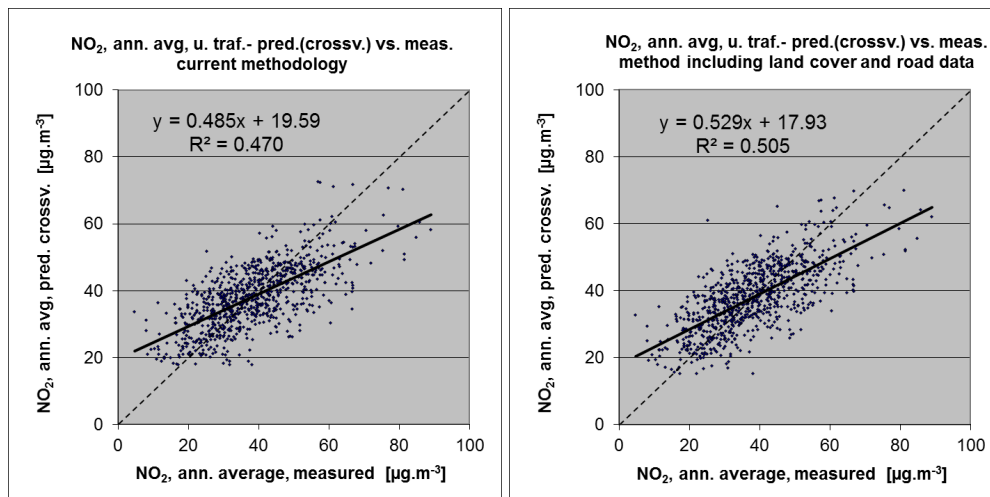
Table 5.3 Comparison of different methods of spatial interpolation showing RMSE, RRMSE, bias, R^2 and linear regression from the cross-validation scatter plots of NO_2 annual mean predicted values, 2013, for urban traffic areas. Units: $\mu\text{g.m}^{-3}$ except RRMSE and R^2 .

spatial interpolation variant + supplementary data used		urban traffic areas				
		RMSE	RRMSE	bias	R^2	regr. eq.
(ii)	current 1km (EMEP, wind speed, s. solar radiation; 1x1 km)	9.5	25.1%	0.0	0.470	$y = 0.485x + 19.6$
(iii)	alternative current (EMEP, alt., s. solar rad.; 1x1 km)	9.5	25.1%	0.1	0.470	$y = 0.483x + 19.7$
(iv)	including LC (EMEP, alt., w. sp., road data, LC; 1x1 km)	9.2	24.3%	0.1	0.505	$y = 0.529x + 17.9$
(v)	without CTM (alt., w.sp., s.s.r., r.h., temp., pop., LC; 1x1 km)	9.2	24.3%	0.1	0.504	$y = 0.534x + 17.7$

It can be seen that all the four methods give quite similar results. The best results are given by the methods (iv) and (v), i.e. including land cover and road data, either using the EMEP model or without the use of the EMEP model. The relative RMSE of the best method is 24 % and R^2 of the cross-validation scatter plot is 0.51. From this one can conclude that the traffic map layer gives quite reliable estimates for the urban traffic air quality. Therefore, the traffic map layer will be further integrated with the rural and urban background map layer as developed in Chapter 4.

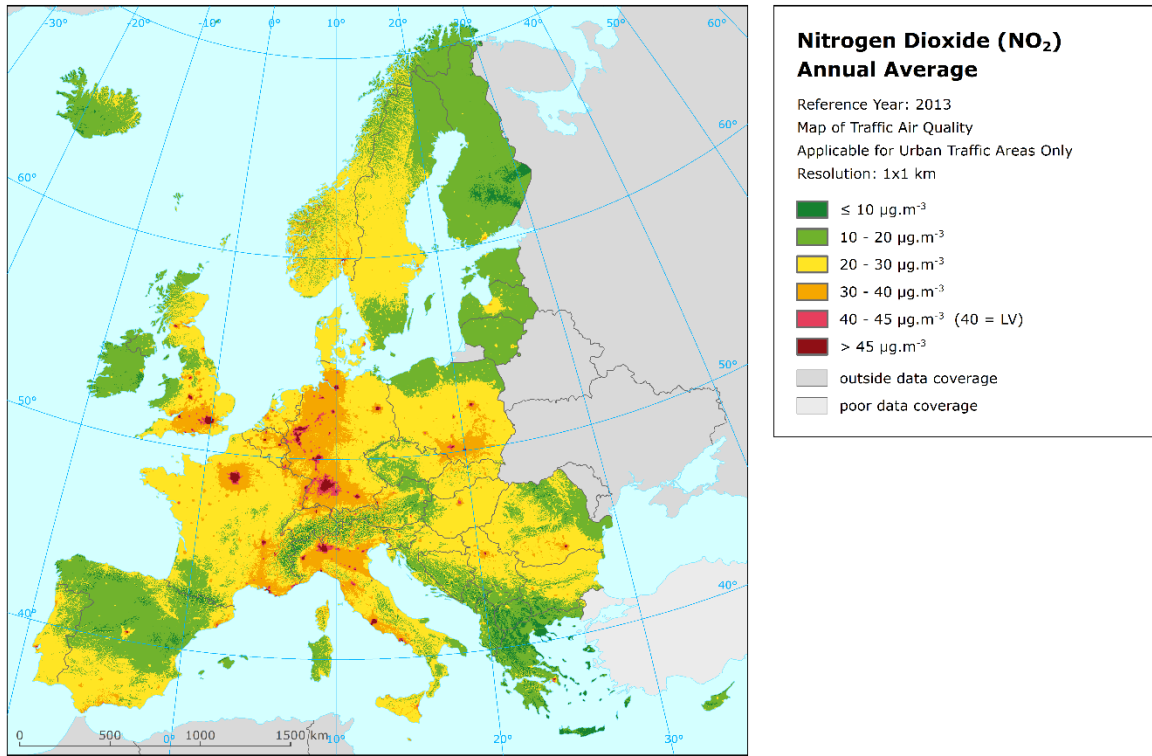
For better illustration of the results presented in Table 5.3, the cross-validation scatter plots are presented in Figure 5.1 for the current method (ii) and for the method including the land cover and road data, using the EMEP model (iv).

Figure 5.1 Correlation between cross-validated predicted and measurement values for NO_2 annual average 2013 for urban traffic areas, for current method (ii) (left) and method including LC and road data (iv) (right)



Map 5.1 presents the urban traffic map layer prepared on basis of the urban/suburban traffic stations by the method including land cover and road data (iv). It should be noted that at limited areas the urban traffic map layer provides lower concentrations values than the urban background map layer (Section 4.1, variant (iv)). At these cases we substitute the estimated urban traffic map layer value with the value of this urban background map layer.

Map 5.1 Concentration map of NO₂ annual average for 2013, urban traffic air quality, method (iv) including land cover and road data. Applicable for urban traffic areas only.



5.2 Inclusion in the concentration map

The traffic map layer introduced in Section 5.1 should be integrated with the rural and urban background map as developed in Section 4.1. This should be done on basis of Equations 2.6 and 2.7 introduced in Section 2.3. As the traffic map layer represents the urban traffic areas, it is incorporated with the urban background map layer. Such urban layer is subsequently merged with the rural map layer into the combined final map. The crucial factor in this traffic map layer inclusion is the weight of the traffic map layer $w_T(i)$ (see Equations 2.6 and 2.7). In this chapter we estimate the weight factor $w_T(i)$ for each 1x1 km grid cell i , based on the detailed road type data (see Section 3.5).

We assume the area influenced by traffic is represented by buffers around the roads. A buffer of 75 metres is considered for the road type classes 1 and 2, while a buffer of 50 metres is taken for the road type classes 3 and 4 (Section 3.5). Leading from this, the weight factor $w_T(i)$ of Equations 2.6 and 2.7 is dependent on the percentage of the area that is influenced by traffic in the 1x1 km grid cell i . That is the total area in a cell that falls within the buffer area of the group of road type classes considered. Two variants of groups of road type weight factors are considered: a $w_{T1}(i)$ for all classes 1 – 3 together, and a $w_{T2}(i)$ for classes 1 – 4 together, according to

$$w_{T1}(i) = T123buf_1km(i) / 2 \quad (5.1)$$

$$w_{T2}(i) = T1234buf_1km(i) / 2 \quad (5.2)$$

where $w_{T1}(i)$, $w_{T2}(i)$ is the traffic weight factor in the two respective variants, of Equations 2.6 and 2.7, for grid cell i ,
 $T123buf_1km(i)$ is the percentage of area influenced by urban traffic in 1x1 km grid cell i , for all classes 1 – 3,
 $T1234buf_1km(i)$ is the percentage of area influenced by urban traffic in 1x1 km grid cell i , for classes 1 – 4.

The reason for dividing the weight factor by 2 is that the buffer represents the area *influenced* by traffic, whereas it is assumed that near the road the concentration is at the level of the traffic map layer itself, while close to the edge of the buffer it is at the level of the urban background map layer. The division is aimed to compensate the flat buffer concentration value for its ‘dilution’ effect occurring when going from the road kerb towards the buffer edge. It should be noted that using this, a linear decrease in concentration from road axis to the background is assumed, while in fact a Gaussian decay might be more realistic. This simplification is somewhat compensated by the fact we assume the traffic stations are located in the zero distance from the road kerb, which is not true in some cases.

Examining the w_{T1} and w_{T2} variants of the weight, the mean percentage of area influenced by traffic for individual road type classes 1 – 4 were compared for the urban/suburban background and the urban/suburban traffic stations, see Table 5.4. The reason for this comparison is the fact that the weights w_{T1} and w_{T2} are based on the percentages of the area influenced by traffic for different road classes.

Table 5.4 Mean percentage of area influenced by traffic for individual road classes 1–4 in grid cells with urban/suburban background and urban/suburban traffic stations with enough NO₂ annual average data for 2013

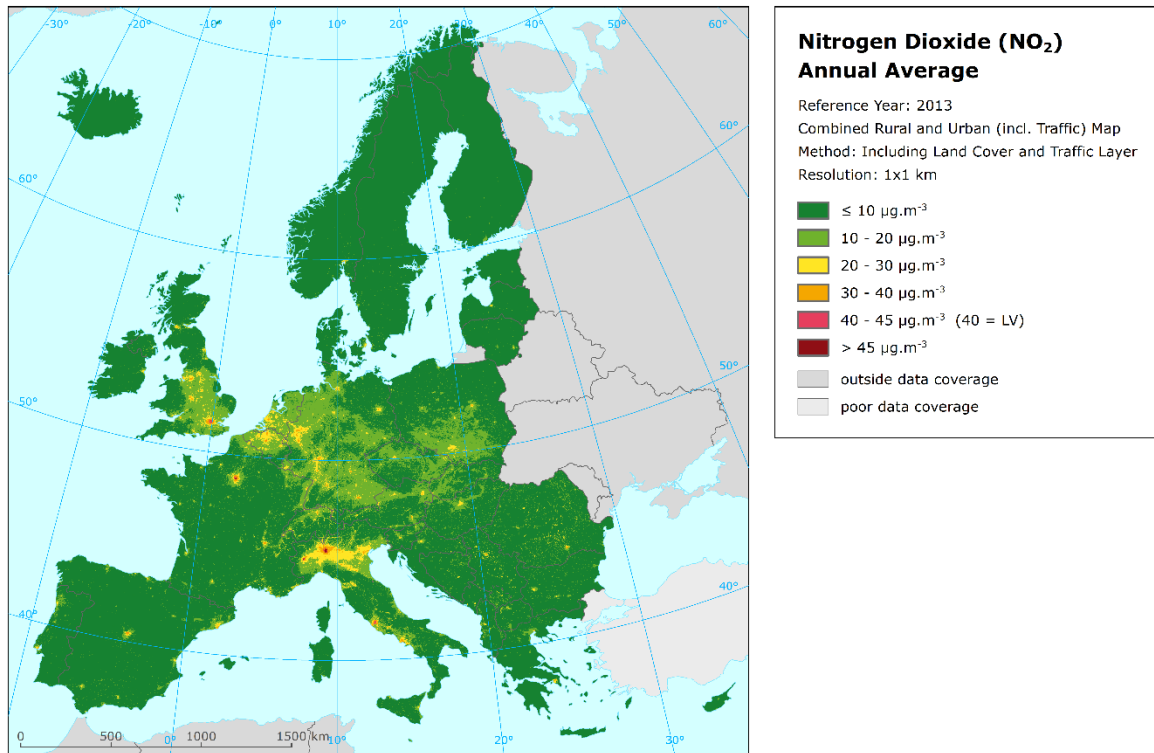
	class 1	class 2	class 3	class 4
	buffer 75 m	buffer 75 m	buffer 50 m	buffer 50 m
urban/suburban background stations (USB)	0.029	0.080	0.079	0.086
urban/suburban traffic stations (T)	0.047	0.100	0.105	0.083
ratio T/USB	1.63	1.26	1.33	0.97

It can be assumed that the mean percentage of the area influenced by traffic should be higher for the urban/suburban traffic stations compared to the urban/suburban background stations. One can see that this is true for the buffers around the roads of classes 1–3, but not for the roads of class 4. Leading from this, one can conclude that the weight w_{T1} (i.e. using the buffers around the roads of classes 1–3 only) is more realistic compared to the weight w_{T2} (i.e. using the buffers around the roads of all classes 1–4). Based on this, the weight w_{T1} is used for the final map creation.

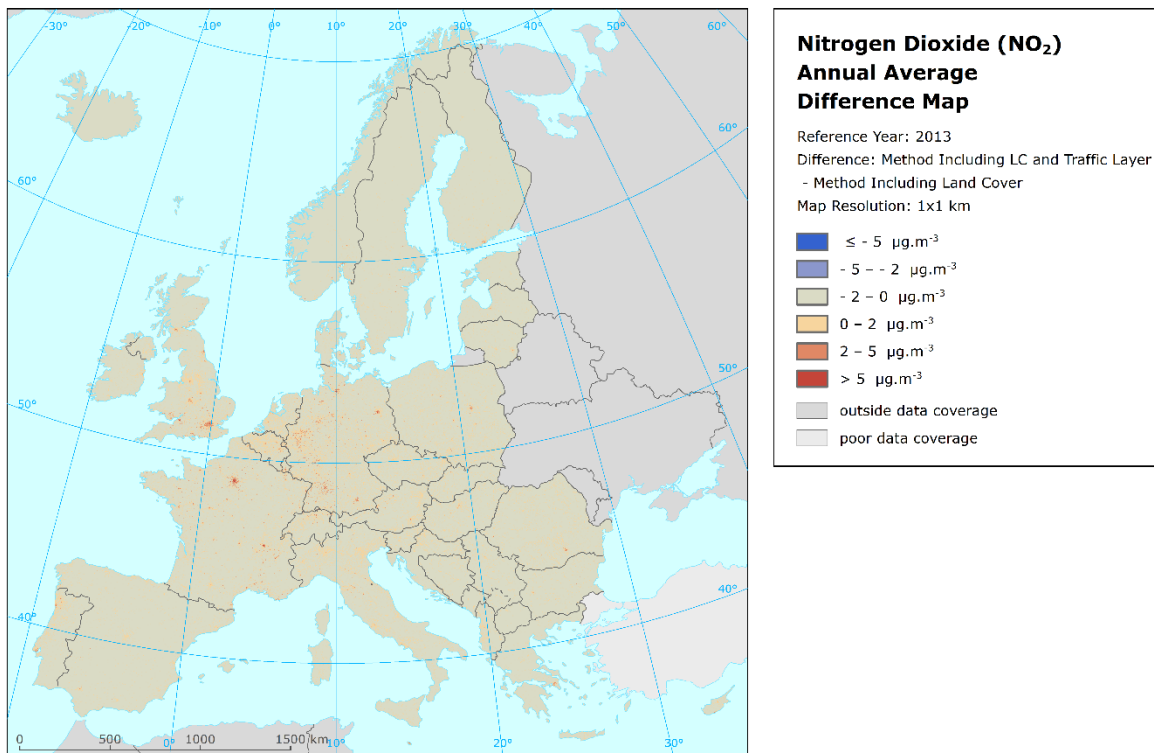
Map 5.2 presents the NO₂ concentration map created by including the traffic map layer (as presented in Map 5.1) in the rural and urban background map (as presented in Map 4.2), using Equations 2.6 and 2.7 with the weight w_{T1} .

Map 5.3 shows the difference between Map 5.2 and Map 4.2. One can see the differences mainly in large cities with high NO₂ concentration in urban traffic areas (Map 5.1).

Map 5.2 Concentration map of NO₂ annual average for 2013, including traffic map layer, method (iv) including land cover and road data, weight w_{T1} used.



Map 5.3 Difference map for NO₂ annual average for 2013, difference between map including traffic map layer and background map without traffic map layer



5.3 Inclusion in the population exposure

Next to the inclusion of the traffic map layer in the concentration map, it is also used for the population exposure estimate, using Equations 2.8 and 2.9. These equations are used in order to guarantee the concentrations used for the population exposure are not smoothed inside the 1x1 km grid cells, which would lead into the underestimation of the population exposure. It should be noted that in this approach, it is supposed that the population inside the 1x1 km grid cell is distributed evenly, which is not the fact in general. However, this shortcoming seems to be of smaller magnitude than the omitting of the population living nearby the traffic hotspots.

Table 5.5 presents the population exposure for NO₂ annual average based on the background concentration map as presented in Map 4.2, with inclusion of the traffic map layer as presented in Map 5.1, using Equations 2.8 and 2.9 with the weight w_{TI} (see Section 5.2).

As can be seen, the overall European population-weighted concentration calculated by this approach is only slightly higher compared to the population exposure calculated based on the background map without inclusion of the traffic map layer, see Table 4.5. The main difference is in the population exposure related to high concentration classes. According to Table 5.4, it is estimated that in 2013 about 3 % of the European population lived in areas above 40 $\mu\text{g.m}^{-3}$, i.e. above the LV according the Ambient Air Quality Directive (EU, 2008).

It can be concluded that Map 5.2 and Table 5.5 give the most realistic results. *The method including the traffic map layer in the mapping and population exposure estimates can be recommended for the further use.*

Table 5.6 shows the overview of the population-weighted concentration, based on several methods presented in this paper. The method presented in Table 5.5 (i.e. with the most realistic results) is marked by orange.

Table 5.5 Population exposure and population-weighted concentrations, NO₂ annual average for 2013, method (iv) including land cover and road data, traffic map layer included using weight w_{T1}

Country		Population [inhbs . 1000]	NO ₂ annual average, exposed population [%]						Population weighted conc. [µg.m ⁻³]
			< 10 µg.m ⁻³	< LV		> LV			
			10 - 20 µg.m ⁻³	20 - 30 µg.m ⁻³	30 - 40 µg.m ⁻³	40 - 45 µg.m ⁻³	> 45 µg.m ⁻³		
Albania	AL	2 896	17.5	45.8	34.0	2.7			17.5
Andorra	AD	73	5.6	93.9	0.4				14.0
Austria	AT	8 507	8.8	44.0	38.2	7.9	0.7	0.4	20.2
Belgium	BE	11 204	0.9	28.4	51.3	17.5	0.8	1.1	24.0
Bosnia & Herzegovina	BA	3 831	27.9	55.9	15.4	0.8			14.4
Bulgaria	BG	7 246	16.1	54.8	25.9	2.8	0.3		17.0
Croatia	HR	4 247	24.6	47.3	26.4	1.7	0.0		15.8
Cyprus	CY	858	46.2	46.8	5.7	1.2			10.5
Czech Republic	CZ	10 512	4.8	70.6	21.6	2.8	0.2		17.2
Denmark	DK	5 627	45.3	36.3	17.2	0.7	0.4	0.1	12.6
Estonia	EE	1 316	52.8	42.4	4.1	0.6			10.0
Finland	FI	5 451	60.1	36.3	2.2	1.3			9.3
France (metropolitan)	FR	63 989	22.4	42.4	19.3	10.1	3.1	2.6	19.0
Germany	DE	80 767	3.6	49.3	37.0	6.9	1.3	1.9	20.7
Greece	GR	10 927	38.3	25.1	24.6	9.3	1.8	0.9	16.0
Hungary	HU	9 877	7.8	70.8	19.3	1.2	0.9	0.0	17.0
Iceland	IS	326	23.7	70.9	5.3	0.0			13.4
Ireland	IE	4 606	48.5	36.4	13.4	1.7			11.8
Italy	IT	60 783	7.1	32.7	35.3	16.4	3.9	4.5	23.8
Latvia	LV	2 001	34.2	38.8	25.0	1.6	0.4		13.8
Liechtenstein	LI	37	1.1	33.6	63.9	1.3	0.0		20.6
Lithuania	LT	2 943	35.6	57.6	5.9	0.8			12.4
Luxembourg	LU	550	0.9	37.1	51.2	4.1	2.8	3.8	23.0
Macedonia, FYROM of	MK	2 066	3.4	44.0	49.3	3.4			20.3
Malta	MT	425	14.5	75.3	1.3	9			15.2
Monaco	MC	38			78		21.5		29.7
Montenegro	ME	622	20.8	46.6	32.6	0.0			16.4
Netherlands	NL	16 829	0.8	37.4	54.3	6.7	0.7	0.1	22.0
Norway	NO	5 108	36.1	41.1	18.0	3.2	0.6	1.0	14.4
Poland	PL	38 018	18.4	57.8	20.4	2.7	0.4	0.4	16.0
Portugal (excl. Az., Mad.)	PT	9 922	26.8	47.6	20.8	3.4	0.9	0.4	15.7
Romania	RO	19 947	14.0	45.4	33.8	5.7	0.1	1.0	18.6
San Marino	SM	33	27.2	68.3	3.9	1			12.0
Serbia (incl. Kosovo*)	RS	7 147	14.2	42.8	36.8	5.4	0.4	0.4	18.8
Slovakia	SK	5 416	3.9	81.6	12.9	1.6			16.4
Slovenia	SI	2 061	19.9	48.5	29.2	1.9	0.4		16.8
Spain (excl. Canarias)	ES	44 397	13.6	48.0	23.6	13.3	1.1	0.4	19.1
Sweden	SE	9 645	44.1	51.4	2.7	1.9	0.0		10.9
Switzerland	CH	8 140	4.1	38.9	46.9	7.8	1.3	1.0	21.6
United Kingdom (& dep.)	UK	64 351	5.4	37.9	39.5	12.2	2.9	2.1	22.5
Total		532 738	13.7	44.8	29.8	8.5	1.6	1.6	19.4
			96.8				3.2		
EU-28		502 424	13.5	44.8	29.5	8.8	1.7	1.6	19.5
			96.7				3.3		

Kosovo*	KS	1 821	15.5	55.2	29.3				16.7
Serbia (excl. Kosovo*)	RS	7 147	13.9	39.8	38.6	6.8	0.5	0.5	19.4

*) under the UN Security Council Resolution 1244/99

Note 1: Turkey is not included in the calculation due to the lack of air quality data.

Note 2: Empty cells mean: no population in exposure.

Table 5.6 Population-weighted concentrations, NO₂ annual average for 2013, using different methods; the method including traffic map layer with the preferred weight w_{T1} marked by orange

Country		Population [inhbs . 1000]	NO ₂ annual average, Population weighted conc. [µg.m ⁻³]				
			(i) curr. 10x10 no traffic layer	(ii) curr. 1x1 no traffic layer	(iv) incl. LC 1x1 no traffic layer	(iv) incl. LC 1x1 tr. layer, wT1	(iv) incl. LC 1x1 tr. layer, wT2
Albania	AL	2 896	15.9	16.0	17.1	17.5	17.5
Andorra	AD	73	14.3	14.4	14.0	14.0	14.0
Austria	AT	8 507	19.3	19.3	19.5	20.2	20.4
Belgium	BE	11 204	23.6	23.8	23.4	24.0	24.3
Bosnia & Herzegovina	BA	3 831	15.7	15.6	14.1	14.4	14.6
Bulgaria	BG	7 246	16.5	16.4	16.3	17.0	17.5
Croatia	HR	4 247	15.8	15.6	15.2	15.8	16.0
Cyprus	CY	858	6.9	6.8	9.4	10.5	10.9
Czech Republic	CZ	10 512	17.1	16.9	16.8	17.2	17.6
Denmark	DK	5 627	13.0	13.1	12.1	12.6	13.1
Estonia	EE	1 316	10.8	10.7	9.3	10.0	10.3
Finland	FI	5 451	9.4	9.2	8.7	9.3	9.7
France (metropolitan)	FR	63 989	18.7	18.6	17.7	19.0	19.6
Germany	DE	80 767	20.4	20.3	19.6	20.7	21.2
Greece	GR	10 927	14.6	14.7	15.2	16.0	16.6
Hungary	HU	9 877	16.8	16.5	16.5	17.0	17.3
Iceland	IS	326	14.3	15.0	12.9	13.4	13.6
Ireland	IE	4 606	11.6	11.9	11.3	11.8	12.0
Italy	IT	60 783	24.5	24.3	23.3	23.8	24.2
Latvia	LV	2 001	13.7	14.0	13.2	13.8	14.2
Liechtenstein	LI	37	22.7	21.8	20.4	20.6	20.6
Lithuania	LT	2 943	11.5	11.7	11.8	12.4	12.5
Luxembourg	LU	550	23.4	23.0	20.9	23.0	23.2
Macedonia, FYROM of	MK	2 066	20.8	20.9	20.2	20.3	20.4
Malta	MT	425	12.0	11.8	13.2	15.2	15.9
Monaco	MC	38	23.2	22.8	26.0	29.7	29.8
Montenegro	ME	622	17.2	17.0	16.3	16.4	16.4
Netherlands	NL	16 829	21.3	21.4	21.4	22.0	22.5
Norway	NO	5 108	14.4	14.2	13.3	14.4	14.7
Poland	PL	38 018	16.1	16.0	15.5	16.0	16.6
Portugal (excl. Az., Mad.)	PT	9 922	14.0	14.5	14.8	15.7	16.0
Romania	RO	19 947	17.9	17.7	18.1	18.6	18.8
San Marino	SM	33	15.4	14.8	11.4	12.0	12.8
Serbia (incl. Kosovo*)	RS	7 147	20.0	19.8	18.5	18.8	19.0
Slovakia	SK	5 416	16.0	15.8	16.0	16.4	16.6
Slovenia	SI	2 061	17.6	17.6	16.4	16.8	16.9
Spain (excl. Canarias)	ES	44 397	18.0	18.1	18.5	19.1	19.4
Sweden	SE	9 645	11.5	11.3	10.2	10.9	11.4
Switzerland	CH	8 140	22.4	22.0	20.6	21.6	22.1
United Kingdom (& dep.)	UK	64 351	22.8	22.6	21.4	22.5	23.0
Total		532 738	19.2	19.2	18.6	19.4	19.8
EU-28		502 424	19.3	19.2	18.7	19.5	19.9
Kosovo*	KS	1 821	19.3	19.3	16.7	16.7	16.7
Serbia (excl. Kosovo*)	RS	7 147	20.2	20.0	18.9	19.4	19.6

*) under the UN Security Council Resolution 1244/99

Note: Turkey is not included in the calculation due to the lack of air quality data.

6 Conclusion

The paper examines the potential improvement of the NO₂ mapping using the land cover and traffic related data. At first, the inclusion of land cover data and available road data in the rural and urban background NO₂ mapping has been tested. Next to this, the mapping of the traffic related air quality using measurement data from the traffic stations and available supplementary data has been examined, as well as the inclusion of such a traffic map layer in the background map and in the exposure estimates.

6.1 Land cover and road data inclusion

The inclusion of CLC land cover data and GRIP road data in the rural and urban background NO₂ mapping has been examined. The most useful supplementary data have been selected through a stepwise regression and backwards elimination in two variants, i.e. one with the use of the chemical transport model EMEP and one without its use. The best results are given by the variant including land cover and road data, using the EMEP model, both for the rural and the urban background areas. Compared to the current methodology, one can see the improvement of the relative RMSE from 36 % to 29 % for the rural background areas and from 24 % to 21 % for the urban background areas.

One can conclude that the inclusion of the land cover and road data brings clear improvement of the NO₂ mapping methodology. Therefore, it is recommended to implement these supplementary data sources in the routine methodology. When introducing this, it is recommended to also move the application of the 1x1 km resolution from the combined final merging process-step to the early process-step of creation of the separate rural and urban background map layers.

6.2 Traffic map layer inclusion

Based on the urban and suburban traffic stations and available supplementary data, traffic map layer has been constructed. This map layer applies as such on urban areas only, since an interpolated rural traffic map layer cannot be constructed due to the lack of rural traffic stations. Several variants of the traffic map layers have been prepared and mutually compared using cross-validation. The relative RMSE of the best method is 24 % and R² of the cross-validation scatter plot is 0.51. From this one can conclude that the traffic map layer gives quite reliable estimates for the urban traffic air quality.

The best traffic map layer has been incorporated with the background map and the exposure estimates, using the road data showing the area influenced by urban traffic. Two options how to include the traffic map layer in the rural and urban background map and population exposure estimates has been examined. The option using the buffers around the roads of classes 1–3 has been selected as preferred.

The selected method including the traffic map layer in the mapping and the population exposure estimates can be recommended for further use.

6.3 Recommendations

As stated above, it is recommended to implement the inclusion of the land cover and road data, as well as the inclusion of the traffic map layer as presented in this paper in the routine methodology for creation of NO₂ maps and exposure estimates.

Next to this, it is recommended to use this improved method as a starting point when examining the potential further improvements of the NO₂ mapping (e.g. by including the satellite data or by using the logarithmical transformation of the measurement and modelled data).

Based on the improvements seen for NO₂, it is recommended to examine whether a similar approach improves the mapping of other pollutants, namely the PM mapping.

References

- Beelen R, Hoek G, Vienneau D, Eeftens M et al. (2013), Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmospheric Environment* 72, 10–23.
- Cressie N (1993). *Statistics for spatial data*. Wiley series, New York.
- Danielson JJ, Gesch DB (2011). Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073. <https://lta.cr.usgs.gov/GMTED2010>
- EC (2008). Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. OJ L 152, 11.06.2008, 1-44.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF>
- ECMWF: Meteorological Archival and Retrieval System (MARS). <http://www.ecmwf.int/>
- EEA (1999). Criteria for EUROAIRNET. The EEA Air Quality Monitoring and Information Network Technical Report 12/1999. <http://www.eea.europa.eu/publications/TEC12>
- EEA (2011). Guide for EEA map layout. EEA operational guidelines. August 2011, version 4. http://www.eionet.europa.eu/gis/docs/GISguide_v4_EEA_Layout_for_map_production.pdf
- EEA (2013a). Corine land cover 2000 (CLC2000) raster data. 100x100m gridded version 17 (12/2013). <http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-raster-3>
- EEA (2013b). Corine land cover 2006 (CLC2006) raster data. 100x100m gridded version 17 (12/2013) <http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3>
- EEA (2015). Air Quality e-Reporting. Air quality database. <http://www.eea.europa.eu/data-and-maps/data/aqereporting>
- EEA (2016). Air quality in Europe – 2016 Report. EEA Report 28/2016. <http://www.eea.europa.eu/publications/air-quality-in-europe-2016>
- Eurostat (2014). GEOSTAT 2011 grid dataset. Population distribution dataset. <http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography>
- Eurostat (2015). Total population for European states for 2013. <http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&language=en&pcode=tps00001&tableSelecti on=1&footnotes=yes&labeling=labels&plugin=1>
- EMEP (2015). Transboundary particular matter, photo-oxidants, acidifying and eutrophying components. Status Report 1/2015, August 2015, 228 pp. http://emep.int/publ/reports/2015/EMEP_Status_Report_1_2015.pdf
- Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, Briggs D (2008), A review of land-use regression models to assess spatial variation of outdoor air pollution, *Atmospheric Environment* 42, 7561–7578.
- Horálek J, Denby B, de Smet P, de Leeuw F, Kurfürst P, Swart R, van Noije T (2007). Spatial mapping of air quality for European scale assessment. ETC/ACC Technical paper 2006/6. http://acm.eionet.europa.eu/reports/ETCACC_TechPaper_2006_6_Spat_AQ

Horálek J, de Smet P, de Leeuw F, Coňková M, Denby B, Kurfürst P (2010). Methodological improvements on interpolating European air quality maps. ETC/ACC Technical Paper 2009/16. http://acm.eionet.europa.eu/reports/ETCACC_TP_2009_16_Improv_SpatAQmapping

Horálek J, Kurfürst P, de Smet P (2014). Additional 2011 European air quality maps. ETC/ACM Technical Paper 2014/5. http://acm.eionet.europa.eu/reports/ETCACM_TP_2014_5_add_2011_aqmaps

Horálek J, de Smet P, Kurfürst P, de Leeuw F, Benešová N (2017). European air quality maps for 2014. ETC/ACM Technical Paper 2016/6. http://acm.eionet.europa.eu/reports/ETCACM_TP_2016_6_AQMaps2014

JRC (2009). Population density disaggregated with Corine land cover 2000. 100x100 m grid resolution, EEA version popu01clcv5.tif of 24 Sep 2009. <http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2>

Su JG, Apte JS, Lipsitt J, Jerrett M (2015). Populations Potentially Exposed to Traffic-Related Air Pollution in Seven World Cities. *Environment International* 78, 82–89.

Mareckova K, Wankmüller R, Pinterits M, Ullrich B (2015). Inventory Review 2015. Review of emission data reported under the LRTAP Convention and NEC Directive. Stage 1 and 2 review & Status of gridded and LPS data. EEA/CEIP Technical Report 1/2015. www.ceip.at/fileadmin/inhalte/emep/pdf/2015/DP0146_InventoryReport_2015_forWeb.pdf

Meijer JR, Huijbregts MAJ, Schotten CGJ, Schipper AM (2016). Mapping the global road network (submitted). Global Road Inventory Project (GRIP) of PBL. http://geoservice.pbl.nl/website/flexviewer/index.html?config=PBL_GRIP.xml¢er=556597,6800125&scale=5000000

NMI (2016). EMEP/MSC-W modelled air concentrations and depositions. Yearly, monthly, daily and hourly gridded data. http://thredds.met.no/thredds/catalog/data/EMEP/2016_Reporting/catalog.html

ORNL (2008). ORNL LandScan high resolution global population data set. http://www.ornl.gov/sci/landscan/landscan_documentation.shtml

Simpson D, Benedictow A, Berge H, Bergström R, Emberson LD, Fagerli H, Hayman GD, Gauss M, Jonson JE, Jenkin ME, Nyíri A, Richter C, Semeena VS, Tsyro S, Tuovinen J-P, Valdebenito A, Wind P (2012). The EMEP MSC-W chemical transport model – technical description. *Atmospheric Chemistry and Physics*, 12, 7825–7865, doi:10.5194/acp-12-7825-2012. <http://www.atmos-chem-phys.net/12/7825/2012/acp-12-7825-2012.html>

UN (2015). World Population Prospects, the 2015 Revision. United Nations. Department of Economic and Social Affairs, Population Division. <http://esa.un.org/unpd/wpp/Download/Standard/Population/>

Annex Brief review of potential land use regression parameters to be used in NO₂ mapping

When reading this Annex one has to bear in mind that the literature review has been executed prior to the actual analysis presented in this paper. The provisional conclusions from the review were overtaken by those from the actual application of land use and road type data in the spatial interpolation. The improvements are larger than originally from the literature review would have been expected.

A.1 Introduction

As part of the work carried out by the European Topic Centre on Air Quality and Climate Change Mitigation (ETC/ACM), annual Europe-wide maps of air quality have been produced using geostatistical techniques for many years [Horálek *et al.*, 2016 and references therein]. The main species under consideration in previous years have been particulate matter (PM₁₀ and PM_{2.5}) as well as ozone. Nitrogen dioxide (NO₂) has so far not been produced operationally within the framework of the ETC/ACM, although NO₂ maps have been produced at irregular intervals in some years.

In order to produce European-scale maps of annual average NO₂ concentrations, it can be helpful to adopt methods from land use regression (LUR) modelling, which often tend to be used more locally for the urban and regional scale [Hoek *et al.*, 2008]. However, one limiting factor in adopting such methods is that the parameters with the most explanatory power for NO₂ tend to be related to datasets of traffic volume (e.g. average daily traffic, ADT). Unfortunately such detailed traffic information at the level of each road segment is currently not available at the European scale, so the parameters to be considered for the European-scale NO₂ mapping within the framework of the ETC/ACM work have to be restricted to datasets that are generally available throughout all of Europe. One of the primary candidates for such a dataset is the CORINE land cover dataset (CLC), which provides frequently updated land cover and land use information for all of Europe at a spatial resolution of approximately 100 m.

Here we will briefly review the current literature of land use regression techniques with the main goal of investigating which land use parameters in the strict sense of the word (i.e. *not* traffic-related parameters such as distance to major roads/freeways, traffic intensity or length of road types without traffic intensity data) have been used in recent studies and which exhibit some predictive power with the goal of mapping NO₂ at the European scale at a spatial resolution of approximately 1 km as it is planned within the ETC/ACM.

A.2 Brief summary of previous work

This general summary of previous studies on land use regression is intentionally kept very short in order to not duplicate existing work. Please see the paper by Hoek *et al.* [2008] for a comprehensive review of land use regression studies for mapping urban- and regional-scale air quality.

Land-use regression techniques for mapping urban air quality were first introduced by Briggs *et al.* [1997], which, perhaps more accurately, referred to the technique as regression mapping [Hoek *et al.*, 2008]. However, as the large body of existing literature at this point mostly refers to the method under the term *land use regression* techniques, we will continue to use this term here as well, even though most often other parameters besides strictly land use and land cover datasets are the primary predictors in these models.

LUR techniques have seen widespread use in recent years, particularly in epidemiological studies in Europe and North America. They usually combine air quality observations made at between 20 and several hundred sites throughout the study area with spatially exhaustive information from predictor variables that give spatial information about air pollution and thus provide the model with realistic spatial patterns. The observations are often made using passive sampling over several weeks, but limited studies have also been carried out using the routine air quality monitoring networks [Stedman *et al.*, 1997], however air quality monitoring stations tend to be quite sparse and often do not exhibit sufficient spatial density to be used for urban-scale LUR modeling. LUR methods are generally applied for modeling annual mean concentration in urban areas but other areas of application, e.g. regional and continental scales, have been identified [Beelen *et al.*, 2009]. In urban areas, LUR techniques have similar levels of accuracy as geostatistical techniques and dispersion models [Hoek *et al.*, 2008], although the latter have the significant advantage that they are based on deterministic physical and chemical processes.

A.3 Frequently used parameters for LUR of NO₂

The predictor variables that are typically used for LUR models are related to traffic, population density, physical geography (e.g. altitude), climate and land use [Hoek *et al.*, 2008]. The latter, namely “true” land use and land cover classes such as vegetation, urban areas, water, etc., are only one of the many parameters used and generally do not exhibit much predictive power in LUR models for NO₂.

Based on the review study carried out by [Hoek *et al.*, 2008], Table A.1 shows the parameters that were considered in previous studies and that at the same time are either already available or are possibly relatively easy to derive from existing datasets at an approximately 1 km x 1 km scale over all of Europe.

Table A.2 Summary of parameters used in land-use regression models of NO₂, modified after Hoek *et al.* (2008). Parameters highlighted in bold are what could be considered “true” land-use/land-cover parameters as they could be extracted from the CORINE land cover database.

Parameter	Used in	Possible Europe-wide dataset
Length of major roads	[Briggs <i>et al.</i> , 1997], [Gilbert <i>et al.</i> , 2005], [Morgenstern <i>et al.</i> , 2007], [Madsen <i>et al.</i> , 2007], [Jerrett <i>et al.</i> , 2007], [Henderson <i>et al.</i> , 2007], [Wheeler <i>et al.</i> , 2008]	OpenStreetMap
Distance from major roads	[Briggs <i>et al.</i> , 1997], [Gonzales <i>et al.</i> , 2005], [Gilbert <i>et al.</i> , 2005], [Rosenlund <i>et al.</i> , 2008]	OpenStreetMap
Built up land	[Briggs <i>et al.</i> , 1997]	CORINE
Land cover factor	[Briggs <i>et al.</i> , 1997], [Morgenstern <i>et al.</i> , 2007], [Aguilera <i>et al.</i> , 2008]	CORINE
Altitude	[Briggs <i>et al.</i> , 1997], [Briggs <i>et al.</i> , 2000], [Gonzales <i>et al.</i> , 2005], [Smith <i>et al.</i> , 2006], [Briggs <i>et al.</i> , 2005], [Rosenlund <i>et al.</i> , 2008], [Madsen <i>et al.</i> , 2007], [Henderson <i>et al.</i> , 2007], [Aguilera <i>et al.</i> , 2008]	SRTM, EU-DEM, or others

High density housing	[Briggs et al., 2000]	OpenStreetMap
Urban land cover	[Stedman et al., 1997], [Beelen et al., 2007]	CORINE
Length of minor roads	[Gilbert et al., 2005]	OpenStreetMap
Open Space	[Gilbert et al., 2005], [Sahsuvaroglu et al., 2006]	CORINE
Population Density	[Gilbert et al., 2005], [Smith et al., 2006], [Beelen et al., 2007], [Rosenlund et al., 2008], [Henderson et al., 2007]	GPW
Distance to Sea	[Briggs et al., 2005]	OpenStreetMap
Non-residential urban land cover	[Briggs et al., 2005], [Henderson et al., 2007]	CORINE
High-density residential land cover	[Briggs et al., 2005]	CORINE
Agriculture land cover	[Briggs et al., 2005]	CORINE
Building density	[Hochadel et al., 2006]	OpenStreetMap
Road length	[Ross et al., 2006]	OpenStreetMap
Industrial land use	[Sahsuvaroglu et al., 2006], [Jerrett et al., 2007]	CORINE
Length rural roads	[Morgenstern et al., 2007]	OpenStreetMap
Length medium traffic roads	[Madsen et al., 2007]	OpenStreetMap
Length small roads	[Madsen et al., 2007]	OpenStreetMap
Household density	[Jerrett et al., 2007]	OpenStreetMap
Road type	[Aguilera et al., 2008]	OpenStreetMap

More recently, the European Study of Cohorts for Air Pollution Effects (ESCAPE) project [Eeftens et al., 2011, 2012; Beelen et al., 2013; de Hoogh et al., 2013] developed land-use regression models for various pollutants in several cities throughout Europe. For example, multi-city models for NO₂ and PM_{2.5} were developed [Wang et al., 2014]. In terms of actual land use/land cover parameters, the final model for NO₂ used the parameter “natural areas and green vegetation within a 5000 m radius”, which exhibited a partial R² of 0.55. For the actual PM_{2.5} model, no land-use/land cover parameters were used, only traffic-related parameters, but for the model for PM_{2.5} absorbance, the same parameter (“natural areas and green vegetation within a 5000 m radius”) explained a partial R² of 0.69 [Wang et al., 2014].

In terms of single-city models, ESCAPE evaluated the use of the following land cover variables, all derived from the CORINE land cover dataset [Beelen et al., 2013]: High density residential land (HDRES), low density residential land (LDRES), sum of high density and low density residential land (HLDRES), Industry (INDUSTRY), Port (PORT), Urban green (URBGREEN), semi-natural and forested areas (NATURAL), sum of urban green and semi-natural and forested areas (GREEN). Each of these variables was calculated using buffer sizes of 100, 300, 500, 1000, and 5000 m. The Europe-wide CORINE dataset was further complemented by local sources of land cover and land use.

The land cover variables that were actually used in the single-city LUR NO₂ models of ESCAPE [Beelen et al., 2013] were: URBGREEN_500 (Helsinki), PORT_5000 and GREEN_1000 (Copenhagen), NATURAL_300 (Bradford), INDUSTRY_5000, HDRES_500 and NATURAL_1000 (Manchester), HLDRES_5000 (London), INDUSTRY_5000 (Ruhr area), HLDRES_500 (Munich), NATURAL_5000 and INDUSTRY_1000 (Vorarlberg), NATURAL_5000 (Paris), HDRES_5000 (Lyon), INDUSTRY_5000 (Gyor), HLDRES_500 (Basel), NATURAL_5000 (Geneva), NATURAL_5000 (Turin), HLDRES_5000 (Verona), INDUSTRY_5000 and URBGREEN_1000

(Rome), LDRES_5000 (Bilbao), HDRES_300 (Barcelona), NATURAL_5000 (Catalonia), HDRES_100 (Huelva), INDUSTRY_300 and NATURAL_1000 (Athens), PORT_1000 (Heraklion).

It should be noted that while some of the land use variables derived from CORINE were used as part of the LUR models in ESCAPE, the authors of the study [Beelen *et al.*, 2013] state themselves in their conclusions that “it is especially important to have accurate local traffic intensity data as predictor variables available” [Beelen *et al.*, 2013].

Recently, [Marcon *et al.*, 2015] used a Buildings data layer with 5000 m buffer and an Industry layer with 1000 m buffer to model NO₂ in the Veneto region of Italy. Even more recently, [Gaeta *et al.*, 2016] followed the ESCAPE approach to map NO₂ around Ciampino airport near Rome. They used LDRES1000 as the only land use variable in their model of NO₂.

High-resolution mapping of NO₂ in the Bergen (Norway) region was carried out by [Denby, 2015] using techniques similar to land use regression. However, no actual land cover/land use parameters were considered, and the final model included only ADT multiplied by road length and shipping emissions. The former variable alone accounted already for 66% of the variability, with the addition of shipping emission increasing the model performance to explain 73% of the variability. The inclusion of other variables, such as population density, elevation, dispersion model output, was not able to significantly improve the model performance over this level. This once again highlights the importance of using traffic information in regression models of NO₂.

A.4 Summary and Recommendations

A very brief review of recent publications of land use regression techniques for NO₂ mapping revealed that it is primarily traffic-related parameters that explain the largest amount of variability for NO₂ land-use regression models, and, out of those, primarily variables related to traffic volume (e.g., ADT). While some of such parameters, such as road length or distance from major roads, could potentially be derived from Europe-wide datasets on roads (or even data sources like OpenStreetMap), the most valuable parameters such as traffic volume/intensity are not available on a Europe-wide basis.

For the work to be carried out within ETC/ACM in 2016 it is primarily envisaged to use the CORINE Land Cover dataset as an additional data source. From such information it is feasible to derive some useful parameters that have been used in the past, such as built-up land [Briggs *et al.*, 1997], a land cover factor [Briggs *et al.*, 1997; Morgenstern *et al.*, 2007; Aguilera *et al.*, 2008], urban land cover [Stedman *et al.*, 1997; Beelen *et al.*, 2007], open space areas [Gilbert *et al.*, 2005; Sahsuvaroglu *et al.*, 2006], non-residential urban land cover [Briggs *et al.*, 2005; Henderson *et al.*, 2007], high-density residential land cover [Briggs *et al.*, 2005], agriculture land cover [Briggs *et al.*, 2005], and industrial land use [Sahsuvaroglu *et al.*, 2006; Jerrett *et al.*, 2007].

While the mapping of NO₂ without adequate information on traffic volume/intensity is very challenging, the most promising approach probably is to follow the previous work carried out within the ESCAPE project [Eeftens *et al.*, 2011, 2012; Beelen *et al.*, 2013; de Hoogh *et al.*, 2013; Wang *et al.*, 2014]. They developed both single-city and multi-city land use regression models throughout all of Europe. The use of the similar land cover/land use parameters as applied in the LUR NO₂ models of ESCAPE in different buffer sizes might be a good starting point for an analysis examining a possible additional explanatory variables inclusion in the mapping of NO₂ within the framework of the ETC/ACM. These variables are entirely available from the CORINE land cover database and the various buffer layers could be derived.

A.5 References

Aguilera, I., J. Sunyer, R. Fernández-Patier, G. Hoek, A. Aguirre-Alfaro, K. Meliefste, M. T. Bomboi-Mingarro, M. J. Nieuwenhuijsen, D. Herce-Garraleta, and B. Brunekreef (2008), Estimation of

Outdoor NO_x, NO₂, and BTEX Exposure in a Cohort of Pregnant Women Using Land Use Regression Modeling, *Environ. Sci. Technol.*, 42(3), 815–821, doi:10.1021/es0715492.

Beelen, R., G. Hoek, P. Fischer, P. A. van den Brandt, and B. Brunekreef (2007), Estimated long-term outdoor air pollution concentrations in a cohort study, *Atmos. Environ.*, 41(7), 1343–1358, doi:10.1016/j.atmosenv.2006.10.020.

Beelen, R., G. Hoek, E. Pebesma, D. Vienneau, K. de Hoogh, and D. J. Briggs (2009), Mapping of background air pollution at a fine spatial scale across the European Union, *Sci. Total Environ.*, 407(6), 1852–1867, doi:10.1016/j.scitotenv.2008.11.048.

Beelen, R. et al. (2013), Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project, *Atmos. Environ.*, 72(2), 10–23, doi:10.1016/j.atmosenv.2013.02.037.

Briggs, D. J., S. Collins, P. Elliott, P. Fischer, S. Kingham, E. Lebre, K. Pryl, H. VanReeuwijk, K. Smallbone, and A. Van der Veen (1997), Mapping urban air pollution using GIS: a regression-based approach, *Int. J. Geogr. Inf. Sci.*, 11(7), 699–718, doi:10.1080/136588197242158.

Briggs, D. J., C. De Hoogh, J. Gulliver, J. Wills, P. Elliott, S. Kingham, and K. Smallbone (2000), A regression-based method for mapping traffic-related air pollution: Application and testing in four contrasting urban environments, *Sci. Total Environ.*, 253(1-3), 151–167, doi:10.1016/S0048-9697(00)00429-0.

Briggs, D. J., A. Aaheim, C. Dore, G. Hoek, M. Petrakis, and G. Shaddick (2005), *Air pollution modelling for support to policy on health and environmental risks in Europe*, APMOSPHERE Final Report, Imperial College, London, UK.

Denby, B. R. (2015), *Mapping of NO₂ concentrations in Bergen (2012 - 2014)*, Norwegian Meteorological Institute.

Eeftens, M., R. Beelen, P. Fischer, B. Brunekreef, K. Meliefste, and G. Hoek (2011), Stability of measured and modelled spatial contrasts in NO₂ over time., *Occup. Environ. Med.*, 68(10), 765–770, doi:10.1136/oem.2010.061135.

Eeftens, M. et al. (2012), Development of land use regression models for PM_{2.5}, PM_{2.5} absorbance, PM₁₀ and PM_{coarse} in 20 European study areas; results of the ESCAPE project, *Environ. Sci. Technol.*, 46(20), 11195–11205.

Gaeta, A. et al. (2016), Development of nitrogen dioxide and volatile organic compounds land use regression models to estimate air pollution exposure near an Italian airport, *Atmos. Environ.*, 131, 254–262, doi:10.1016/j.atmosenv.2016.01.052.

Gilbert, N. L., M. S. Goldberg, B. Beckerman, J. R. Brook, and M. Jerrett (2005), Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model, *J. Air Waste Manag. Assoc.*, 55(8), 1059–1063, doi:10.1080/10473289.2005.10464708.

Gonzales, M., C. Qualls, E. Hudgens, and L. Neas (2005), Characterization of a spatial gradient of nitrogen dioxide across a United States-Mexico border city during winter, *Sci. Total Environ.*, 337(1-3), 163–173, doi:10.1016/j.scitotenv.2004.07.010.

Henderson, S. B., B. Beckerman, M. Jerrett, and M. Brauer (2007), Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter, *Environ. Sci. Technol.*, 41(7), 2422–2428, doi:10.1021/es0606780.

Hochadel, M., J. Heinrich, U. Gehring, V. Morgenstern, T. Kuhlbusch, E. Link, H. E. Wichmann, and U. Krämer (2006), Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information, *Atmos. Environ.*, 40(3), 542–553, doi:10.1016/j.atmosenv.2005.09.067.

Hoek, G., R. Beelen, K. de Hoogh, D. Vienneau, J. Gulliver, P. Fischer, and D. Briggs (2008), A review of land-use regression models to assess spatial variation of outdoor air pollution, *Atmos. Environ.*, 42(33), 7561–7578, doi:10.1016/j.atmosenv.2008.05.057.

de Hoogh, K. et al. (2013), Development of land use regression models for particle composition in twenty study areas in Europe., *Environ. Sci. Technol.*, 47(11), 5778–86, doi:10.1021/es400156t.

Horálek, J., P. de Smet, P. Kurfürst, F. De Leeuw, and N. Benešová (2016), *European air quality maps of PM and ozone for 2013 and their uncertainty*, ETC/ACM Technical Paper 2015/5.

Jerrett, M., M. A. Arain, P. Kanaroglou, B. Beckerman, D. Crouse, N. L. Gilbert, J. R. Brook, N. Finkelstein, and M. M. Finkelstein (2007), Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada., *J. Toxicol. Environ. Health. A*, 70(3-4), 200–12, doi:10.1080/15287390600883018.

Madsen, C., K. C. L. Carlsen, G. Hoek, B. Oftedal, P. Nafstad, K. Meliefste, R. Jacobsen, W. Nystad, K. H. Carlsen, and B. Brunekreef (2007), Modeling the intra-urban variability of outdoor traffic pollution in Oslo, Norway-A GA2LEN project, *Atmos. Environ.*, 41(35), 7500–7511, doi:10.1016/j.atmosenv.2007.05.039.

Marcon, A., K. de Hoogh, J. Gulliver, R. Beelen, and A. L. Hansell (2015), Development and transferability of a nitrogen dioxide land use regression model within the Veneto region of Italy, *Atmos. Environ.*, 122, 696–704, doi:10.1016/j.atmosenv.2015.10.010.

Morgenstern, V., a Zutavern, J. Cyrys, I. Brockow, U. Gehring, S. Koletzko, C. P. Bauer, D. Reinhardt, H.-E. Wichmann, and J. Heinrich (2007), Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children., *Occup. Environ. Med.*, 64(1), 8–16, doi:10.1136/oem.2006.028241.

Rosenlund, M., F. Forastiere, M. Stafoggia, D. Porta, M. Perucci, A. Ranzi, F. Nussio, and C. a Perucci (2008), Comparison of regression models with land-use and emissions data to predict the spatial distribution of traffic-related air pollution in Rome., *J. Expo. Sci. Environ. Epidemiol.*, 18(2), 192–199, doi:10.1038/sj.jes.7500571.

Ross, Z., P. B. English, R. Scalf, R. Gunier, S. Smorodinsky, S. Wall, and M. Jerrett (2006), Nitrogen dioxide prediction in Southern California using land use regression modeling: potential for environmental health analyses, *J. Expo. Sci. Environ. Epidemiol.*, 16, 106–114, doi:10.1038/sj.jea.7500442.

Sahsuvaroglu, T., A. Arain, P. Kanaroglou, N. Finkelstein, B. Newbold, M. Jerrett, B. Beckerman, J. Brook, M. Finkelstein, and N. L. Gilbert (2006), A land use regression model for predicting ambient concentrations of nitrogen dioxide in Hamilton, Ontario, Canada., *J. Air Waste Manag. Assoc.*, 56(8), 1059–1069, doi:10.1080/10473289.2006.10464542.

Smith, L., S. Mukerjee, M. Gonzales, C. Stallings, L. Neas, G. Norris, and H. Özkaynak (2006), Use of GIS and ancillary variables to predict volatile organic compound and nitrogen dioxide levels at unmonitored locations, *Atmos. Environ.*, 40(20), 3773–3787, doi:10.1016/j.atmosenv.2006.02.036.

Stedman, J. R., K. J. Vincent, G. W. Campbell, J. W. L. Goodwin, and C. E. H. Downing (1997), New high resolution maps of estimated background ambient NO_x and NO₂ concentrations in the U.K., *Atmos. Environ.*, 31(21), 3591–3602, doi:10.1016/S1352-2310(97)00159-3.

Wang, M. et al. (2014), Performance of multi-city land use regression models for nitrogen dioxide and fine particles, *Environ. Health Perspect.*, 122(8), 843–849, doi:10.1289/ehp.1307271.

Wheeler, A. J., M. Smith-Doiron, X. Xu, N. L. Gilbert, and J. R. Brook (2008), Intra-urban variability of air pollution in Windsor, Ontario-Measurement and modeling for human exposure assessment, *Environ. Res.*, 106(1), 7–16, doi:10.1016/j.envres.2007.09.004.