Spatial surrogates for the disaggregation of CORINAIR emission inventories

Joachim Maes*, Jo Vliegen, Karen Van de Vel, Stijn Janssen, Felix Deutsch, Koen De Ridder, Clemens Mensink

VITO, Flemish Institute for Technological Research, Environmental Modelling Unit, Boeretang 200, 2400 Mol, Belgium

Article info

Article history:
Received 20 May 2008
Received in revised form 28 October 2008
Accepted 19 November 2008

Keywords:
Atmospheric emissions
CORINE land cover
EMEP
SNAP
Spatial disaggregation

Abstract

CORINAIR atmospheric emission inventories are frequently used input data for air quality models with a domain situated in Europe. In CORINAIR emission inventories, sources are broken down over 11 major source categories. This paper presents spatial surrogates for the disaggregation of CORINAIR atmospheric emission inventories for input of air pollutants and particulate matter to grid or polygon based air quality model domains inside Europe. The basis for the disaggregation model was the CLC2000 land cover data to which statistical weights were added. Weights were population census data for residential emissions, employment statistics for agricultural and industrial area emissions, livestock statistics for ammonia emissions and annual aircraft movements for emissions realized by air transport. Additional road and off-road network information was used to disaggregate emissions realized by traffic. A comparison of top down produced emission estimates with spatially resolved national emission data for The Netherlands and the United Kingdom gave confidence in the present spatial surrogates as a tool for the top down production of atmospheric emission maps. Explained variance at a spatial resolution of 5 km was >70% for CO, NMVOC and NOx, >60% for PM10 and almost 50% for SO2.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Atmospheric emissions of air pollutants are essential input into air quality models. Annual quantities of anthropogenic emissions to the air are mostly compiled by regional or national environmental agencies and categorized by source category for a particular geographic area and specific time interval. Emission inventories serve as policy instruments to evaluate existing regulations and are often used to report to supranational authorities, such as the United Nations or the European Union. Once stored in quality assured databanks, international emission inventories are extremely useful tools to provide input into air quality models of which the model domain crosses administrative boundaries. This is especially evident in the EU where the competence for emission inventories is fragmented over different national or regional agencies. This results in different standards, methods and source categories being used when compiling emission inventories, even within single countries. Collecting emissions from regional and national emission inventories for usage in air quality models with a transboundary model domain is therefore a challenging task. On top of this, regional emission inventories report commonly in a local language, which further complicates the compilation of bottom up based emission input data for particular air quality models.

An alternative for the bottom up compilation of emissions is the top down disaggregation of emission totals reported by supranational or global inventories. In Europe, a commonly used supranational emission inventory is the CORINAIR/EMEP database. CORINAIR (Core Inventory of Air Emissions) is a project performed since 1995 by the European Topic Centre on Air Emissions, with the aim to collect, maintain, manage and publish information on emissions into the air, by means of a European air emission inventory and database system (European Environmental Agency, 2007).

In this paper spatial surrogates are presented to disaggregate CORINAIR based emission totals onto a spatially resolved emission inventory for input into air quality model domains situated in Europe. A spatial surrogate or substitute variable is a value greater than zero and less than or equal to one that specifies the fraction of the emissions realized in a particular country that should be allocated to a particular grid cell of the air quality model domain of interest (Eyth and Habisak, 2003). Typically, some type of geographic attribute is used to weight the attributes into grid cells in a manner more specific than a simple uniform distribution.
Examples of such substitute variables are population maps, spatially resolved industrial production data or mapped traffic census data.

CORINAIR data have been used previously in several studies to produce spatially and temporally resolved emission inventories (Lenhart and Friedrich, 1995; Friedrich and Reis, 2004; Monforti and Pederzoli, 2005; Poupkou et al., 2007; Borge et al., 2008). The focus of this paper essentially lies on the integration of several European, publicly available data sets so as to disaggregate CORINAIR based emission data. In particular, satellite derived CORINE land cover data in combination with EUROSTAT statistics were applied as spatial surrogate variables for disaggregating non-point emission sources over the European countries and seas.

This paper is essentially structured into two parts. Firstly, the methodology of a top down approach for spatial allocation of CORINAIR emissions is presented in detail and for each source category, spatial surrogates are proposed. Next, the top down methodology to generate emissions is illustrated for two cases. CORINAIR emissions are disaggregated to a grid situated in the United Kingdom covering London and to the map with the municipalities of the Netherlands. Each time, the top down disaggregated emissions were compared with estimates provided by the respective national emission inventories.

2. Material and methods

2.1. CORINAIR/EMEP emission inventory

In CORINAIR emission inventories, sources are broken down over 11 SNAP categories. SNAP stands for selected nomenclature for sources of air pollution. CORINAIR guidebooks give the European continental, national and local authorities a set of standard reference tools and methods to estimate pollutants production in a given area and to report it under the SNAP nomenclature (Monforti and Pederzoli, 2005; European Environmental Agency, 2007). A commonly used CORINAIR emission inventory that covers the European continent is the EMEP emissions database. EMEP is the scientific programme of the Convention on Long-range Transboundary Air Pollution. Through its data portal, officially reported scientific programme of the Convention on Long-range Transboundary Air Pollution. Through its data portal, EMEP provides corrected and gapfilled expert emissions for air quality modeling, both on a country basis as national totals as well as gridded on a 50 km resolution map (Vestreng et al., 2007). The geographic area of EMEP covers emissions realized in all European countries and seas. The database can be consulted at http://www.emep-emissions.at/ceip/.

2.2. Spatial disaggregation methodology

The MIMS Spatial Allocator (University of North Carolina, Eyth and Habisak, 2003) was used to generate spatial surrogate variables based on the SNAP based national emission totals reported by CORINAIR emission inventories. Spatial surrogates are created for regularly spaced air quality model grids as well as for any modeling polygon shapes such as municipalities.

Three types of surrogates can be created with Spatial Allocator: polygon-based, line-based, and point-based. Polygon-based surrogates use attribute information that is based on area (e.g., population in a census tract). The surrogate value is calculated as the ratio of the attribute value in the intersection of the country and the grid cell to the total value of the attribute in the country. Examples of polygon-based weight attributes are population, number of households or land use. For line-based surrogates, the length of the linear weight feature (e.g., railroad, river, road) replaces area in the above explanation. For point-based surrogates, instead of using area, the software will allocate a value of 1 if the weight point falls within the region of interest or 0 otherwise. In some cases, no special weight attribute is desired. Instead, the surrogate is based purely on the area of the polygon, the length of the polylines, or the count of the points (Eyth et al., 2006).

2.3. Spatial surrogates for the disaggregation of SNAP structured emission totals

For each SNAP source category, point source emissions are allocated on the air quality domain of interest (Fig. 1). Next, remaining non-point emissions were spatially distributed using quantitative spatial surrogate data.

![Conceptual model of the disaggregation procedure. In a first step point source emissions are allocated on the air quality model domain. Next, remaining non-point emissions were spatially distributed using quantitative spatial surrogate data.](image-url)
quantitative spatial surrogate data (Fig. 1). Table 1 lists all the data sources that were used in this paper for spatial disaggregation of CORINAIR emissions.

### 2.3.1. Allocation of point source emissions

The European Pollutant Emission Register (EPER) served as data source to distribute or rescale emissions of large point sources over a model domain of interest (Table 1). The database contains point source emission data for about 12,000 facilities occurring in the EU-25 for the years 2001 and 2004 (European Commission, 2000). Facilities are characterized by longitude and latitude allowing simple mapping on a grid.

### Table 1

<table>
<thead>
<tr>
<th>Data source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUROSTAT</td>
<td>EUROSTAT publishes economic, monetary, trade statistics, business, social, regional, agricultural, environmental and energy statistics. The web site offers free access to download all data. In this paper three data sets were used. Employment in persons at NUTS 3 level (2001), animal populations at NUTS 2 level (data for 2005, gapfilling using data for previous years); air traffic data by airport (data for 2006, gapfilling using data for previous years). URL: <a href="http://epp.eurostat.ec.europa.eu">http://epp.eurostat.ec.europa.eu</a>.</td>
</tr>
<tr>
<td>EPER</td>
<td>EPER is the European Pollutant Emission Register. EPER contains data on the main pollutant emissions to air and water reported by about 12,000 large and medium-sized industrial facilities in the EU-25 Member States. Data are available for 2001 and 2004. URL: <a href="http://eper.ec.europa.eu/">http://eper.ec.europa.eu/</a>.</td>
</tr>
<tr>
<td>CLC2000</td>
<td>The CORINE Land Cover 2000 data (CLC2000) is a map of the European environmental landscape based on interpretation of satellite images with land cover types in 44 standard classes. The map was created in GIS ARC/INFO format at an original scale of 1:100,000. The resolution of the raster data is 250 \times 250 m. The European Environmental Agency owns CLC2000 and grants free access to the data. Derived from this data set is the population density disaggregated with CLC2000 (Gallelo and Peedell, 2001). The owner of this data set is the Joint Research Centre of the European Commission. URL: <a href="http://www.eea.europa.eu">http://www.eea.europa.eu</a>.</td>
</tr>
<tr>
<td>TREMOVE</td>
<td>TREMOVE is a policy assessment model to study the effects of different transport and environment policies on the emissions of the transport sector. The model covers passenger and freight transport in the EU-25 and covers the period 1995–2030. National emissions are estimated using the COPERT methodology. TREMOVE is owned by the European Commission. The database is available in MS Excel and MS Access. URL: <a href="http://www.tremove.org">http://www.tremove.org</a>. The 2000 Census of Motor Traffic on Main International Traffic Arteries is a compilation of road traffic data on main international roads in Europe. The census shows the average annual daily traffic on the E-Roads of 37 European countries. The data can be purchased at UNECE Transport Division. URL: <a href="http://www.unece.org/trans/">http://www.unece.org/trans/</a>.</td>
</tr>
<tr>
<td>UN Traffic Census</td>
<td>The Global Land Cover 2000 database categorizes land cover using 22 different classes at a spatial resolution of 1 km. The data are made available by the Joint Research Centre of the European Commission. URL: <a href="http://www-gem.jrc.it/egl2000">http://www-gem.jrc.it/egl2000</a>.</td>
</tr>
<tr>
<td>CLC2000</td>
<td>The Global Land Cover 2000 database categorizes land cover using 22 different classes at a spatial resolution of 1 km. The data are made available by the Joint Research Centre of the European Commission. URL: <a href="http://www-gem.jrc.it/egl2000">http://www-gem.jrc.it/egl2000</a>.</td>
</tr>
<tr>
<td>ESRI data and maps</td>
<td>ESRI data and maps is a set of map data that is included with ArcGIS software. It was used to have spatial information of inland waterways and railways of Europe. The data are not publicly available. URL: <a href="http://www.esri.com/data">http://www.esri.com/data</a>.</td>
</tr>
<tr>
<td>RRG GIS database</td>
<td>The RRG GIS database (Raumforschung, Raumplanung und Geoinformation) is a geodatabase covering all trans-European Transport Networks. The database was used to map sea shipping routes in Europe. The data are not publicly available. URL: <a href="http://www.brg.de">http://www.brg.de</a>.</td>
</tr>
</tbody>
</table>

### 2.3.2. Allocation of non-point source emissions

All point source emissions reported under the EPER decision were summed per SNAP source category and per country and subsequently divided by the national emission total reported by the EMEP data portal. This ratio corresponds to the proportion of point source emissions of the total emission. Remaining emissions were considered non-point sources and were distributed using sector specific spatial surrogates. The ratio between emissions reported under EPER and emissions reported under CLRTAP was in few cases >1. Then all CORINAIR emissions were assumed to be point sources and distributed using the EPER emission data as weight attributes. For sectors without industrial point sources or for non-EU member states which do not report to EPER, emission totals were assumed to be area sources.

Non-point sources were disaggregated using specific spatial substitutes and tested against two sets of bottom up produced emission inventories. In general, line sources are distributed using spatial information of transport networks while area based emission totals are disaggregated using the CORINE data set in combination with additional attributes adding weight to the different land use classes.

### 2.3.2.1. Disaggregation of area sources reported under CORINAIR source categories S1, S3, S4 and S5

The SNAP source categories S1, S3 and S4 report emissions from combustion in energy production and transformation industries, combustion in manufacturing industry and industrial production processes, respectively. Remaining non-point sources for this sector were allocated using the CLC2000 land cover class industrial and commercial units (Table 1) in combination with statistical activity data provided by EUROSTAT on the number of employees in industry in European NUTS 3 statistical areas (Table 1). EUROSTAT is the statistical bureau of the European Union. The Nomenclature of Territorial Units for Statistics (NUTS) is a geocode standard for referencing the administrative divisions of European countries for statistical purposes. EUROSTAT data with low spatial resolution (NUTS 3 level) were proportionally disaggregated on a high resolution land cover map (250 \times 250 m²) spreading the numbers of employees in the industry over the land cover class industrial and commercial units. For countries that are not covered by the CLC2000 data and for which no EUROSTAT information is available, emissions were disaggregated using the CLC2000 land cover class artificial surfaces and associated areas (Table 1). Emissions released during the extraction and distribution of fossil fuels and geothermal energy are reported under S5 and were disaggregated using the CLC land class ports.

### 2.3.2.2. Disaggregation of area sources reported under CORINAIR source categories S2 and S6

Sectors S2 and S6 report emissions realized by non-industrial combustion plants and by the application of solvents, respectively. Non-point sources for these two categories were spatially distributed using the population density disaggregated with the CLC2000 data (Gallelo and Peedell, 2001). For countries that are not covered by the CORINE data set, the GPWv3 population data set was used (Table 1).

### 2.3.2.3. Disaggregation of area sources reported under CORINAIR source categories S8 and S10

Source categories S8 and S10 include emissions from agricultural practices. S8 contains emissions from other non-mobile sources as well, including emissions realized by civil aviation, railways, national navigation, military transport and gardening. So as to estimate for the share of agriculture in S8, we used 50.4% as a cut-off value. This value was calculated by the contribution of off-road agricultural vehicles relative to the other sources using the NFR source categories that feed into S8. NFR
codes have been developed for reporting under the United Nations Economic Commission for Europe and are more refined (European Environmental Agency, 2007). The cut-off level is an European average.

Emissions caused by agricultural production processes, predominantly NH\textsubscript{3}, are reported under S10. Ammonia emissions caused by agricultural practices are distributed using a map combining the CLC2000 land cover classes pasture and complex cultivation patterns with EUROSTAT animal population numbers on a NUTS 2 geographic level. Agricultural emissions of other pollutants are disaggregated using a dasymetric map combining the GLC2000 land cover class cultivated and managed areas (Table 1) with EUROSTAT statistical data on the number of employees in agriculture on NUTS 3 geographic level. For countries outside the EUROSTAT statistical area, emission totals are disaggregated using the GLC2000 land cover class cultivated and managed areas without employment statistics.

2.3.2.4. Disaggregation of area sources reported under CORINAIR source category S9. Emissions reported under S9 are coming from waste treatment and disposal. Non-point emissions for this sector are spatially disaggregated using the CORINE population data set (Table 1). The emission total caused by road traffic reported under S7 is allocated over urban roads, non-urban roads and motorways. These data are country specific, except for those countries that are not part of the TREMOVE database (Table 1). In this case, the average was used. Next, sub-totals were disaggregated using different spatial surrogates. Urban road emissions were distributed over an intersection map between the CORINE population data set and the CLC2000 land cover class continuous urban fabric. Similarly, non-urban road emissions were distributed using an intersection map between the CORINE population data set and the discontinuous urban fabric land cover class. Motorway emissions were disaggregated using the UN traffic census for Europe with AADT data (annual average daily traffic) per road segment for the year 2000 realized on the E-roads in Europe (Table 1).

The TREMOVE emission model also estimates national emission totals of three different off-road traffic networks: railways, inland shipping and air transport. For the first two networks, traffic network data were used for spatial disaggregation of emissions (ESRI data and maps, Table 1). Air traffic emissions are distributed over the CLC2000 land cover class Airports. To this spatial feature, attribute information was added using the EUROSTAT statistics on the number of flights per airport (Table 1).

Emissions from international marine shipping are in the EMEP emission database allocated to different European seas and oceans. So emission totals are available for the Baltic sea, the North Sea, the Mediterranean Sea, the Black Sea and the part of the Atlantic Ocean that is covered by the EMEP model domain. These emission totals were distributed over a map with the European marine shipping routes. The data were derived from the RRG GIS database (Table 1).

2.3.2.6. Disaggregation of natural sources reported under CORINAIR source category S11. Data for natural sources are summarized under sector S11. These emissions were not considered in this paper.

### Table 2
Regression diagnostics.

<table>
<thead>
<tr>
<th>UK London area</th>
<th>Emission totals 2005 (Gg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>NH\textsubscript{3}</td>
</tr>
<tr>
<td>Total UK emission EMEP</td>
<td>2408</td>
</tr>
<tr>
<td>Disaggregated total EMEP</td>
<td>456</td>
</tr>
<tr>
<td>Aggregated total NAEI</td>
<td>404</td>
</tr>
<tr>
<td>Difference (%)</td>
<td>113%</td>
</tr>
<tr>
<td>Explained variance (%)</td>
<td></td>
</tr>
<tr>
<td>Spatial resolution</td>
<td>CO (R^2)</td>
</tr>
<tr>
<td>1 × 1 km\textsuperscript{2}</td>
<td>20</td>
</tr>
<tr>
<td>2 × 2 km\textsuperscript{2}</td>
<td>29</td>
</tr>
<tr>
<td>3 × 3 km\textsuperscript{2}</td>
<td>37</td>
</tr>
<tr>
<td>4 × 4 km\textsuperscript{2}</td>
<td>35</td>
</tr>
<tr>
<td>5 × 5 km\textsuperscript{2}</td>
<td>41</td>
</tr>
<tr>
<td>6 × 6 km\textsuperscript{2}</td>
<td>58</td>
</tr>
<tr>
<td>7 × 7 km\textsuperscript{2}</td>
<td>67</td>
</tr>
<tr>
<td>8 × 8 km\textsuperscript{2}</td>
<td>61</td>
</tr>
<tr>
<td>9 × 9 km\textsuperscript{2}</td>
<td>64</td>
</tr>
<tr>
<td>10 × 10 km\textsuperscript{2}</td>
<td>68</td>
</tr>
<tr>
<td>The Netherlands Emission totals 2004 (Gg)</td>
<td>CO</td>
</tr>
<tr>
<td>Explained variance (%)</td>
<td></td>
</tr>
<tr>
<td>Spatial resolution</td>
<td>CO (R^2)</td>
</tr>
<tr>
<td>Municipalities (polygons)</td>
<td>74</td>
</tr>
</tbody>
</table>

CORINAIR/EMEP emission totals for the UK (2005) and for the Netherlands (2004) were spatially disaggregated on a grid based and polygon based map, respectively. For the UK, the national total emission and the proportion of the total that was attributed to the grid map are compared to the aggregated total based on the NAEI 1 km resolution emission map for 6 pollutants. For the Netherlands, emission totals were disaggregated on a polygon based map of the municipalities. The explained variance based on least squares regression \(R^2\) and on least trimmed squares regression \(R_{\text{LTS}}^2\) is used a measure for the reproduction of nationally compiled emissions by the disaggregation procedure.
2.4. Comparison of the spatial disaggregation of CORINAIR emission totals with national emission inventories

The performance of the spatial surrogates that are proposed in this paper to disaggregate national emission totals was evaluated by comparing the top down methodology with two high-resolution emission maps. As an illustration of the methodology, two test areas were selected, one in the United Kingdom and one in The Netherlands. Both countries have well developed emission inventories that are publicly available via the Internet and both report in English. For the UK, we selected a domain of 100 \times 100 \text{ km}^2 that included the London metropolitan area (Fig. 3). EMEP expert emission data for Great Britain for 2005 were disaggregated onto regular grids of identically sized cells with different spatial resolution (from 1 km to 10 km). Note that the selected domain does not cover the entire country. However, by using spatial surrogate data that cover Europe, and hence Great Britain, the methodology calculates first what portion of the national emission totals should be attributed to the model domain of interest and subsequently distributes this portion over the grid cells.

The result of this exercise was compared with a spatially resolved emission inventory provided by the National Atmospheric Emission Inventory on a 1 km resolution (NAEI, 2007).

In a second test case, emissions were distributed over the Netherlands on a polygon based map consisting of the municipalities. The difference with the case for the UK is that now the national total is distributed, rather than a part of it, and that emission totals are allocated to irregularly shaped polygons instead of regularly shaped grid cells. Again, the downscaled emission totals for the Netherlands reported to LRTAP for 2004 and available in the EMEP emission database were compared with the data provided by the national emission registration administration for each municipality in the Netherlands (The Netherlands Pollutant Release and Transfer Register, 2007).

Performance of the spatial surrogates was tested by calculating the explained variance that results from a least squares regression $R^2$ as well as a least trimmed squares regression $R_{LTS}^2$ between the national emission data and the downscaled emission data. $R^2$ measures the proportion of the total variation that is explained by the regression line. Least squares regression minimises the sum of squared residuals. In this context, a residual is the difference between the emission per unit area reported by the national emission inventory and the emission estimated by the disaggregation procedure. In a regression analysis there may be influential data points of which the presence or absence significantly influences the explained variance $R^2$, based on ordinary least squares regression. Typically, grid cells that contain a large point emission source are considered outliers relative to the average emission per grid cell. To avoid that large point sources artificially increased the explained variance, a robust regression method based on least trimmed squares was used as well yielding a robust $R_{LTS}^2$. $R_{LTS}^2$ may be used as a measure for the explained variance for both point and non-point emissions while arguably, $R^2$ should be interpreted as a measure for the explained variance when comparing non-point sources only. Explained variances based on ordinary least squares regression $R^2$ were calculated in MS Excel. Robust explained variances $R_{LTS}^2$ were calculated using the robust regression procedure in SAS.

3. Results

3.1. Grid based emission maps for the London area

National emission totals for the United Kingdom reported by the EMEP emission database for 2005 were disaggregated over a 100 \times 100 \text{ km}^2 domain covering the greater London area. Only 4% of the UK national ammonia emission is realized in the selected domain above London (13 Gg out of 317 Gg, Table 2). For other pollutants, on average 17% of the 2005 national emission total could be attributed to the selected domain. The disaggregated emissions were subsequently compared to aggregated emission maps based on the NAEI emission inventory on a 1 km resolution. For all pollutants but NH$_3$ the national emission total resulted in an overestimation of the emission total for the entire grid domain (Table 2). For CO, NMVOC and NO$_x$ 13% more emissions were assigned to the domain than when summing up the 1 km resolution emissions provided by NAEI for the same domain. This is illustrated here for NMVOC. The UK reported a national emission total of 977 Gg NMVOC in 2005. From this total, 166 Gg was allocated to the domain by the disaggregation procedure. This compares to an aggregated 147 Gg based on the 1 km resolution emission maps provided by the NAEI. This difference increased for PM$_{10}$ (25%) and for SO$_2$ (57%). In contrast, the total NH$_3$ emission was underestimated (Table 2). Disaggregating the national NH$_3$ emission yielded a total of 13 Gg compared to an aggregated total of 17 Gg based on the 1 km resolution emission maps by NAEI.

The disaggregation procedure was run 10 times. Each time, the spatial resolution of the domain was linearly changed from 1 km to 10 km. As expected, the explained variance, based on a linear regression between the NAEI data and the downscaled emission data increased as the spatial resolution of the grid became coarser. This was especially evident for CO, NO$_x$, PM$_{10}$ and NMVOC emissions. For these pollutants, explained variance varied from 20% for the 1 km resolution map to almost 90% for the 10 km resolution map.

A scatterplot between the NAEI CO emission data and the CO emission totals disaggregated on the 5 km resolution map demonstrates the accuracy of the disaggregation model (Fig. 2). The corresponding $R^2$ was 39% based on LS regression and 87% based on LTS regression (Table 2). Fig. 2 illustrates well the different approach of an ordinary least squares estimator and a robust one to assess the explained variance of a linear regression. In Fig. 2 a number of data points are vertically outlying. Vertical outliers affect considerably $R^2$ based on least squares regression but these
data points are neglected by the robust LTS estimator when performing a regression analysis. More important for application in air quality models is that many large point sources are well predicted by the disaggregation model. Statistically, such influential points are known as good leverage points. These data points are outlying on the X axis but they are well fitting to the model. Obviously, the accurate mapping of large point sources is of crucial importance given their importance for local air quality. In least squares regression good leverage points tend to increase considerably the explained variance while in robust regression, their weight is diminished. Another remark is that based on the position of data points relative to the bissectrice identifying a 1–1 relationship between the data (Fig. 2) it is clear that gridded emissions with a lower than average value for CO were overestimated while those with higher than average emissions were underestimated.

The interpretation of the results for SO₂ using \( R^2 \) differed strongly depending on the estimator used. Explained variance based on LS regression yielded higher values than for the LTS regression. Again, the spatial distribution of large point sources was responsible for this difference. Only 4 large point sources contributed to 74% of the total emission of SO₂ over the domain, affecting the value for explained variance.

For ammonia, spatial allocation yielded poor results. Possible reasons for the low reproduction of ammonia emissions are discussed further.

CO, NMVOC and NOx emissions mapped at a 2 km spatial resolution were compared between the two inventories (Fig. 3). The variance \( R^2_{\text{LTS}} \) explained by the disaggregation model was 53% for CO, 77% for NMVOC and 53% for NOx. The major pattern that emerged from the two emission maps is the relative high emissions in grid cells covering the London city and lower emissions in the surrounding area. Emissions related to major highways were reproduced less accurately by the downscaling procedure if compared to the national emission maps.

3.2. Polygon based emission maps for Dutch municipalities

The results obtained for a spatial allocation of Dutch emission totals over the 467 municipalities of the Netherlands were in line with the results for the grid based domain over London. The explained variance was high for CO, NMVOC, NOx and PM10, while SO₂ and NH₃ were less well reproduced in comparison with the national emission for each municipality (Table 2). For CO, the disaggregated data were plotted against the national emission

**Fig. 3.** Gridded emission maps for CO, NMVOC and NOₓ. Difference between disaggregated CORINAIR/EMEP emission totals for the UK using spatial surrogates and aggregated NAEI emissions. The surface area of the maps is 10⁴ km², the spatial resolution is 2 × 2 km². The lower UK map depicts the position of the grid (Transverse Mercator projection; British National Grid).
inventory data. The explained variance corresponding to this regression was 69% for the LS regression between the Dutch emission data and the disaggregated data and 73% for a LTS regression between the two data sets. CO emission totals for municipalities with large point source emissions were reproduced well by the disaggregation procedure while emissions assigned to communes with lower than average CO were somewhat underestimated (Fig. 4).

Emission maps for both the Dutch data and the downscaled data are presented in Fig. 5 for NH₃, SO₂ and PM₁₀.

4. Discussion

4.1. Strengths

Maps of the spatial distribution of atmospheric emissions are key input to any air quality assessment (Bush et al., 2008). Many spatially resolved emission data have been produced in Europe but often they cover single states or regions (e.g. De Kluizenaar et al., 2001; Poupkou et al., 2007; Borge et al., 2008), target single pollutants (e.g. Hutchings et al., 2001) or focus on single source categories (e.g. Panis et al., 2001; Schrooten et al., 2008). This paper presents spatial surrogates for the disaggregation of CORINAIR atmospheric emission inventories for input of classical air pollutants and particulate matter to any grid or polygon based air quality model domain inside Europe. This flexibility can be attributed to the combination of spatial and statistical databases that cover EU member states as well as other European countries.

Another asset is that we presented useful indications of the uncertainty of the disaggregation model relative to two national emission inventories. It is important to notice that the methodology for spatial disaggregation of emission data does not affect the national emission totals. However, the methodology takes into account surrogate data in order to produce emissions on a particular air quality domain that intersects geographically with the country for which emission totals are available. As a result, there was a difference between the NAEI aggregated emission total and the EMEP disaggregated emission total for the selected domain with, in general, an overestimation of the emission total. The explained variance of the emissions that were assigned per grid cell relative to the NAEI emission data increased when the spatial resolution became coarser. For CO, NMVOC and NOₓ, high explained variances (>70%) were obtained at a spatial resolution of 5 km. At this resolution, explained variance for PM₁₀ was >60%. A comparison for a polygon based domain in the Netherlands yielded similar results. The comparison with national emission data gives confidence in the present spatial surrogates as a tool for atmospheric emission maps, at least for the mapping of these 4 pollutants.

Air quality models require both spatially and temporally resolved emission data. While this paper focuses on spatial disaggregation of emission totals, we refer to Monforti and Pederzoli (2005) which present a tool to temporally disaggregate SNAP based emission inventories.

4.2. Limitations

Clearly, the provision of EU wide spatial proxy data introduces a number of limitations. Emission maps generated for SO₂ and in particular for NH₃ appeared to be less reliable than for other pollutants as suggested by the explained variance based on regression between national emission data and downscaled data. For SO₂, explained variance based on ordinary least squares regression was consistently higher than based on a robust regression analysis. This is due to the fact that few large point sources dominate the spatial emission pattern of SO₂. Large point sources tend to increase $R^2$ based on ordinary least squares regression. In 2004, the UK reported 683 Gg SO₂ emissions from large point sources to the EPER database while the total SO₂ emission for that year is 883 Gg. In 2005 the NAEI reports a total of 565 Gg point sources on a national SO₂ emission of 706 Gg. Therefore, the spatial surrogates for non-point sources of SO₂ are of less importance.

Mapping NH₃ emissions over the city of London yielded poor results in terms of explained variance while for the Netherlands, on a coarser spatial resolution, better results were presented with a maximum explained variance of 65%. In both the UK and the Netherlands, 90% of the NH₃ emissions are realized in the agriculture. For the UK, less than 5% of the ammonia was realized in the domain over London. This domain consisted largely of artificial land cover types while the major farming areas of Britain are situated elsewhere. Possibly, the emission domain was therefore probably not suitable for verifying NH₃ emissions. Ammonia emissions were disaggregated according to the distribution of agricultural land, defined as pasture and complex cultivation patterns, in combination with statistics on animal populations on a NUTS 2 level. Such methodology has been used as well for ammonia emissions from agriculture in Denmark (Hutchings et al., 2001). Still, the accuracy of spatial ammonia maps will improve if the methodology includes better statistics on the distribution of ammonia emission source such as grazing, housing, storage and landspreading emissions from cattle, rather than using the distribution of source items as such (Dragosits et al., 1998). Further, emissions from application of fertilizer were not taken into account.

Also the spatial allocation of traffic emissions needs further improvement. Visual mapping of NOₓ and CO on a 2 km resolution map for the greater London area showed that emissions attached to roads were poorly reproduced by the spatial surrogates for SNAP sector S7. The methodology applied here discriminated between urban, rural and highway transports. Emissions from urban and rural driving were disaggregated using CLC2000 land cover data rather than attributing these emissions to detailed road network. Clearly, the assignment of emissions from road transport can be improved considerably if traffic volumes for the road network are available as well. Such data were not used in this study for urban conditions.
and rural driving. For highway driving, UN traffic census data were
used based on annual average daily traffic counts. These data
clearly underestimate the traffic density in and around London.

5. Conclusion

The major benefit of the spatial substitutes that are presented
in this paper to disaggregate CORINAIR emission inventories is
the flexibility to generate emissions for air quality grids
anywhere in the EMEP domain regardless whether or not the
domains cross national borders and without the need for a local
or regional emission inventory. Data uncertainty decreased
considerably for coarser spatial grids and at a spatial resolution
of 5 km, local emission data were reproduced well by the
disaggregation procedure. For finer spatial detail, it is more
appropriate to use local or regional emission inventories,
provided that they are available.

Acknowledgements

Mapped emissions for the UK are made freely available on the
Mapped emissions for the Netherlands are made freely available on the Netherlands Pollutant Release & Transfer Register web site at [http://www.emissieregistratie.nl](http://www.emissieregistratie.nl). The emissions were prepared with the Spatial Allocator, a software tool developed by the University of North Carolina at Chapel Hill with support from the US EPA ([http://www.ie.unc.edu/cempd/projects/mims/spatial/](http://www.ie.unc.edu/cempd/projects/mims/spatial/)).

References


