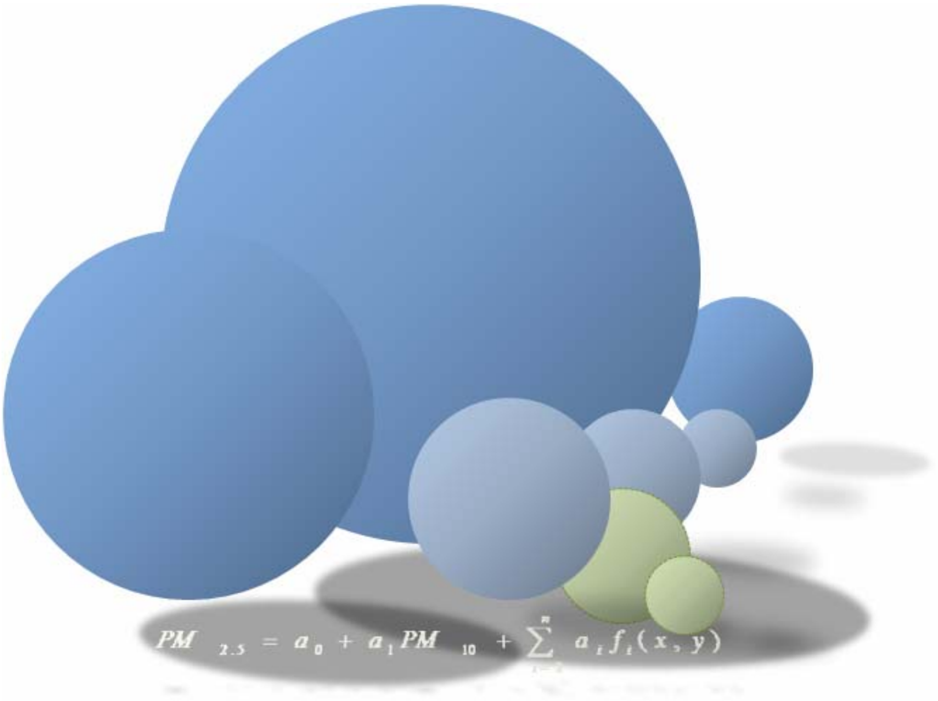
Mapping annual mean $PM_{2.5}$ concentrations in Europe: application of pseudo $PM_{2.5}$ station data



ETC/ACM Technical Paper 2011/5

August 2011

Bruce Rolstad Denby, Jan Horálek,
Peter de Smet,Frank de Leeuw

European Topic Centre on Air Pollution and Climate Change Mitigation

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 UBA-V ÖKO AEAT EMISIA CHMI NILU INERIS PBL CSIC

Front page picture:
Design, Bruce Rolstad Denby

Author affiliation:
Bruce Rolstad Denby: Norwegian Institute of Air Research (NILU), Kjeller, Norway
Jan Horálek: Czech Hydrometeorological Institute (CHMI), Praha, Czech Republic
Peter de Smet, Frank de Leeuw: The National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

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 Technical paper 2011/5
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/

Summary

The European topic centre for air pollution and climate change mitigation (ETC/ACM) has been developing mapping methods and operationally providing maps of European wide air quality for a number of years, as part of its tasks for the European Environment Agency (EEA). These maps are used to present the current state of air quality in Europe and to assess the population exposure throughout Europe and its Member States. To create these maps, monitoring data from the AirBase database is combined with other spatially resolved supplementary data (e.g. EMEP model outputs, altitude, population, meteorology) by first applying multiple linear regression to the supplementary data and there after applying residual kriging. Currently maps for PM_{10} and ozone indicators are operationally produced.

There is also a need to produce European wide maps of $PM_{2.5}$, as this is considered to be a relevant indicator for health impact. This is reflected in the current European Air Quality Directive (2008/50/EC), which introduced the Average Exposure Indicator and a target/limit value for $PM_{2.5}$. Since $PM_{2.5}$ was not a legislated pollutant with a defined target value prior to the 2008 Air Quality Directive, monitoring of $PM_{2.5}$ has not been as extensively pursued in Europe as it has for PM_{10}. For example, in the year 2008, 203 stations reporting annual mean $PM_{2.5}$ were available in Airbase for all background stations. For the same year 1286 annual mean PM_{10} concentrations were available at all background stations. This shortage of monitoring data makes spatial assessment very uncertain and for this reason $PM_{2.5}$ maps have not been operationally produced.

To improve this situation the possibility of creating 'pseudo' $PM_{2.5}$ stations, based largely on the available PM_{10} station data, was investigated in a previous ETC/ACM report. The concept is that annual mean $PM_{2.5}$ concentrations could be derived from measured PM_{10} and other supplementary data. These pseudo data could then be used to improve the spatial assessment of $PM_{2.5}$ in Europe.

This report describes the use of multiple linear regression to derive pseudo $PM_{2.5}$ annual mean concentrations from measured PM_{10} annual mean concentrations and applies these data to derive European wide maps of annual mean $PM_{2.5}$. Maps are made for the years 2007 and 2008. It is found that using multiple linear regression provides estimates of annual mean $PM_{2.5}$ close to, or within, the data quality objectives set out in the European Air Quality Directive (2008/50/EC), when taking into account the less stringent requirements for data below the lower assessment threshold. However, the pseudo $PM_{2.5}$ data does not completely satisfy the more stringent data quality objectives for fixed measurements alone, and so the pseudo $PM_{2.5}$ data cannot be used as a direct replacement for monitoring data. The pseudo $PM_{2.5}$ data derived is considered to be of sufficient quality for its use in the map making application carried out here.

Maps are created and their uncertainty is assessed using cross-validation methods. In particular two types of maps are made for all of Europe. The first of these maps uses both the measured and the pseudo $PM_{2.5}$ data to generate the maps. The second uses only the measured $PM_{2.5}$ data to generate the maps. There are significant differences between the two maps. Both maps are compared with the available measured $PM_{2.5}$ data and the map that includes the pseudo $PM_{2.5}$ stations is shown to provide a significant statistical improvement on the maps made using only measured $PM_{2.5}$ data.

In addition, the population exposure for all the countries in Europe is assessed, using the population weighted concentrations and number of people exposed above the target value threshold of 25 μgm^{-3}. Though population weighted concentrations were similar for both maps the number of people exposed to levels of $PM_{2.5}$ above the target value was significantly different with the use of the pseudo $PM_{2.5}$ data, reducing the exceedance exposure by a third. The inclusion of these pseudo data resulted in lower estimates of exposure rates and is considered to be of higher quality than the results achieved using measured $PM_{2.5}$ alone.

The aggregated uncertainty of the exposure calculations is assessed per country. The relative uncertainty of the population weighted concentration varied between 5 – 25%, depending largely on the size of the country (degree of aggregation). A preliminary assessment is also carried out of the exceedance exposure per country. The relative uncertainty is found to be larger than that found for the population weighted concentration, due to the sensitivity of the exceedance exposure to small changes in concentrations. A number of countries are found to have a significant number of their population above the target value.

As a result of this study it is recommended to implement the pseudo $PM_{2.5}$ station data and the production of $PM_{2.5}$ maps operationally at EEA, at least until sufficient numbers of $PM_{2.5}$ stations become available. In addition it is recommended to develop, and operationally put in place, methods for determining the uncertainty of the exposure calculations.

Contents

1. Introduction

The European topic centre for air pollution and climate change mitigation (ETC/ACM) has been developing methods and providing maps of European wide air quality for a number of years (e.g. Horálek et al., 2007; 2010; Fiala et al., 2009, de Smet et al., 2010). To create these maps, monitoring data from AirBase is combined with other spatially resolved supplementary data (e.g. EMEP model outputs, altitude, population, meteorology) by first applying multiple linear regression to the supplementary data and there after applying residual kriging. Currently maps for PM_{10} and ozone indicators are operationally produced.

There is also a need to produce European wide maps of $PM_{2.5}$, as this is considered to be a relevant indicator to health impact. However, there is currently a shortage of monitoring data for $PM_{2.5}$ in AirBase compared to PM_{10}. This makes spatial assessment very uncertain and for this reason $PM_{2.5}$ maps have not been operationally produced. To improve this situation the possibility of creating 'pseudo' $PM_{2.5}$ stations, based largely on the available PM_{10} station data, has been investigated in two previous studies. Initially this was done by determining a $PM_{2.5}/PM_{10}$ concentration ratio, dependent on location and station type (de Leeuw and Horálek, 2009). In a follow up study (Denby et al., 2011) the concept that the annual mean $PM_{2.5}$ concentrations could be derived from PM_{10} and other supplementary data within a useful uncertainty was further investigated. These data could then be used to improve the spatial assessment of $PM_{2.5}$ in Europe. In the report from Denby et al. (2011) two methods were investigated, those being multiple linear regression and artificial neural networks. The two methods gave similar results and it was recommended to proceed further with the creation of pseudo $PM_{2.5}$ stations by using multiple linear regression, since this method was transparent and reproducible.

The aim of this report is to assess the application of pseudo $PM_{2.5}$ stations for mapping and to generate a robust method for creating $PM_{2.5}$ maps operationally. In this report we further apply the results of the Denby et al. (2011) study and use the pseudo $PM_{2.5}$ stations to map, at 10 km resolution, annual mean $PM_{2.5}$ concentrations for all of Europe in the years 2007 and 2008. The methodology followed for the mapping is similar to the normal operational procedure used to map annual mean PM_{10} concentrations (e.g. Horálek et al., 2010).

2. Datasets

2.1 AirBase air quality data

The air quality datasets used in the study are the PM_{10} and $PM_{2.5}$ annual mean concentrations for Europe taken from AirBase (AirBase, 2011) for the years 2007 and 2008. The geographical positions and altitude of the stations are also used as supplementary data. These data are used to create the pseudo $PM_{2.5}$ concentrations and the observed $PM_{2.5}$ data are also used in the mapping procedure. All stations labelled as 'background' are used in the analysis, these included 'urban', 'suburban', 'rural' and 'unknown' station types. Data coverage of > 75% is required.

For the years 2007/2008, annual mean $PM_{2.5}$ concentrations have been reported for 147/203 background stations (57/93 urban, 41/46 suburban, 49/64 rural). For the same years and station categories 1165/1286 PM_{10} annual mean background concentrations have been reported. The number of stations used in the analysis varies. For the calculation of pseudo $PM_{2.5}$ regression parameters, only stations with concurrent measurements of $PM_{2.5}$ and PM_{10} are used (129/178) and for the mapping regression and interpolation the sum of the available $PM_{2.5}$ background stations and the non-concurrent PM_{10} background stations are used (1118/1303).

2.2 Population

Population density is used as supplementary data. This is provided on a 10 x 10 km^2 grid resolution. (Source EEA, popugrid01v4_1grid, official version 4.1, Jan. 2008; Owner: JRC). For countries and regions not included in this database, we use as alternative source the ORNL LandScan (2002) Global Population Dataset. These data are used both to create the pseudo $PM_{2.5}$ concentrations as well as for the mapping of $PM_{2.5}$ concentrations.

2.3 Meteorological data

Both climatological data (years 1961 – 1990; New et al., 2002) (resolution 10 x 10 minutes) and current meteorological fields for the study years (2007/2008) taken from ECMWF re-analysis (15 x 15 minutes) (European Centre for Medium-range Weather Forecasts; http://www.ecmwf.int/) are used as supplementary data sources. These data are used both to create the pseudo $PM_{2.5}$ concentrations as well as for the mapping of $PM_{2.5}$ concentrations.

The relevant climatological meteorological data (*clim*) available for use, and identified in Denby et al. (2011), include: wind speed (*wind*), relative humidity (*RH*), percentage sunshine duration (*sun*), temperature (*temp*) and precipitation (*prec*). The current meteorological field data (*meteo*) include: wind speed (*wind*), relative humidity (*RH*), accumulated surface solar radiation (*sun*) and temperature (*temp*).

2.4 Altitude data

We use the European covering altitude data field (in meters) of GTOPO30, original grid resolution of 30 x 30 arc seconds (just smaller than 1x1 km). These data are spatially aggregated into the 10x10 km EEA standard grid resolutions. For details see Horálek et al. (2007).

2.5 Unified EMEP model data

Annual mean $PM_{2.5}$ concentration fields for the years 2007/2008 taken from the Unified EMEP model (Simpson et al., 2003; EMEP, 2009; EMEP, 2010; http://www.emep.int/OpenSource/index.html) are used for the mapping procedure.

3. Methodology

The methodology for carrying out the mapping is divided into two steps:

1. Create a set of pseudo $PM_{2.5}$ station data based on existing PM_{10} station data;

2. Use the derived pseudo $PM_{2.5}$ and the available observed $PM_{2.5}$ data to carry out the mapping.

In both these steps testing and assessment of the available supplementary data is carried out. In principle the two steps are kept separate, as the supplementary data suitable for determining the pseudo $PM_{2.5}$ station data may not be the same as for the spatial mapping of the $PM_{2.5}$ concentrations. However, it is expected that the two steps make use of similar sets of supplementary data.

3.1 Creation of pseudo $PM_{2.5}$ concentrations

As previously stated multiple linear regression (MLR) of the observed $PM_{2.5}$ concentrations with the observed PM_{10} concentrations and other supplementary data is carried out to produce pseudo $PM_{2.5}$ stations. In the previous study (Denby et al., 2011) the following recommendation and selection of supplementary data was made based on the 2004-2007 datasets:

1. The Artificial Neural Network gave slightly better results than MLR but it was recommended to proceed further with the creation of pseudo $PM_{2.5}$ stations using MLR, since this method was transparent and reproducible.

2. The selected supplementary data sources for the MLR were latitude, longitude, population density and climatological sunshine duration.

3. Both climatological sunshine duration and yearly meteorological accumulated solar radiation gave similar results. The choice of climatological data over meteorological data was intended to simplify the calculation (as the climatological data is static from year to year).

4. The small number of available stations did not warrant the separation of the data into rural and (sub)urban station types.

The exploratory assessment made in the previous study will be recalculated for the present 2007 and 2008 data. These more recent years provide an increased number of observational $PM_{2.5}$ station data that supports the robustness of the statistical analyses and its conclusions. The following tests are carried out:

1. The two years (2007/2008) are assessed separately for a variety of supplementary data sources;
2. The rural and (sub)urban stations are assessed separately for each of the two years (2007/2008).

The statistical indicators used to assess the results are given as:

1. Root mean square error (RMSE).
2. Correlation (r^2).

3. Fractional bias (FB). For MLR the bias is 0, but this indicator was also used for the kriging application where it is not always 0.
4. Fraction of station predictions within ± 25% of the observed $PM_{2.5}$ (FAC25%). In the European AQ Directive 95% of the data are required to be within this range to fulfil the monitoring (fixed measurement) quality objectives.
5. Number of station predictions that fulfil the European AQ Directive (2008/50/EC; EC, 2008) assessment quality objective criteria (ECQO). This is similar to FAC25% except that for concentrations below the lower assessment threshold for $PM_{2.5}$ (12 $\mu g/m^3$) the allowable error is ± 100% as monitoring is not required below this level.

Selection of a suitable set of supplementary data is based on the following criteria:

1. Consistency between the two years
2. Consistency between statistical indicators (particularly RMSE and correlation)
3. Small differences in the statistical indicators are not considered significant
4. Small improvements in RMSE need to be reflected in the other statistics
5. Consistency between the years in the regression coefficient sign

PM_{10} and other supplementary data are provided as regression variables and regression coefficients are determined to give the best fit. It is important to note that since the technique minimizes the mean square error then the root mean square error (RMSE), which is one of the indicators used to assess the performance of MLR, will always remain the same or decrease with any additional parameter. For this reason it is also important to assess the performance of MLR using the other metrics.

3.2 Mapping of $PM_{2.5}$

Similar to mapping of PM_{10}, mapping of $PM_{2.5}$ is a two step procedure; the overall mapping procedure for PM_{10} is described in previous ETC/ACM and EEA reports (e.g. Horálek et al., 2007; 2010; Fiala et al., 2009, de Smet et al., 2010). The two steps are:

1. EMEP model fields and other spatially distributed supplementary data are used in a multiple linear regression with $PM_{2.5}$ station data (both observed and pseudo) to determine suitable regression coefficients for producing a background field for residual kriging;

2. This background field is subtracted from the set of $PM_{2.5}$ stations (both observed and pseudo) and the residual is interpolated, using ordinary kriging. In addition, kriging of the $PM_{2.5}$ stations (both observed and pseudo) without the subtraction of any background fields is also carried out.

The mapping is carried out separately for the (sub)urban and the rural stations and the maps combined afterwards, using a weighted addition based on population density (see e.g. Horálek et al., 2010).

There are a number of variations possible and, as in the creation of pseudo $PM_{2.5}$ data, the suitability of supplementary data must be assessed. In previous work on PM_{10} (Horálek et al., 2010) the following supplementary data was used for the rural mapping: EMEP data, altitude, wind speed and accumulated solar radiation. For the urban mapping of PM_{10} only the EMEP model data was found to be useful. Log-normal transformation of the concentration data was also applied and found to be superior than when no transformation was applied (see also Denby et al., 2008).

The assessment of the mapping in this study will also follow similar lines to that carried out in the previous PM_{10} mapping. That being

1. Use of logarithmic concentration transformation;
2. Separate interpolation for rural and urban areas/stations;
3. Testing and assessment of suitable spatially distributed supplementary data;
4. Comparison with log-normal kriging, i.e. without using the residual.

In addition, for this study the following assessment will take place to determine the impact on the interpolation when including pseudo $PM_{2.5}$ stations:

5. Comparison of interpolation at observed $PM_{2.5}$ stations only, with and without the addition of pseudo $PM_{2.5}$ stations.

4. Exploration and selection of supplementary data for the estimation of pseudo $PM_{2.5}$ stations

A number of tests are carried out using a variety of combinations of supplementary data as well as logarithmic transformations of some of these data. The results of these are tabulated in Appendix 1 Tables A1.1 – A1.8 for the two years of data. The spatial distribution of both the observed and the pseudo rural and (sub)urban station is presented in Figures A1.1 and A1.2 for 2007 and 2008.

The following is a summary concerning the data exploration:

- The supplementary data tested included latitutude (*lat*), longitude (*lon*), population (*pop*), altitude (*alt*), sunshine duration climatological (*sun(clim)*), accumulated annual radiation (*sun(meteo)*), annual wind speed (*wind(meteo)*). Additional tests were carried out using logarithmic transformations that are not shown in the results tables. The use of logarithmically transformed population density, sunshine and elevation was assessed. These did not provide improved results compared to the non-transformed values.
- The use of logarithmically transformed concentrations was assessed. This did not improve the results.
- As expected, there is a high correlation between latitude and sunshine duration (Table A1.9) and a high correlation between PM_{10} and $PM_{2.5}$.

The following conclusions are made concerning the generation of pseudo $PM_{2.5}$ data using MLR:

- The best performing and selected set of supplementary data is selection number 7 (**PM10 + lat + lon + pop + sun(meteo)**). This agrees with the previous assessment (Denby et al., 2011), with the exception that the meteorological accumulated solar radiation has been used and not the climatological sunshine duration. See Table 1 below for a summary of these statistics and Figure 1 for the resulting scatter plots.
- When the selected supplementary dataset is applied separately to the urban and rural station data the resulting statistics are slightly improved, though this is to be expected for any split of the data (Table A1.4 and A1.8). Due to the limited number of data available the splitting into the urban and rural data sets is not recommended.
- For both years the ECQO indicator is very close to, or satisfies, the required quality objectives for assessment provided in the European AQ Directive, which is ECQO > 0.95. This indicates that the pseudo $PM_{2.5}$ stations provide a satisfactory replacement for assessment applications in the AQ Directive.
- The FAC25% indicator, however, does not satisfy the AQ Directive quality objectives for monitoring (FAC25% > 0.95) and so the pseudo $PM_{2.5}$ cannot be seen as a direct replacement for monitoring of $PM_{2.5}$.
- An indicative value for the uncertainty of the methodology is the normalised RMSE, normalised with the average concentration. This indicates that the pseudo PM2.5 data has an uncertainty (one standard deviation) of approximately 18% and 15% for the years 2007 and 2008 respectively.
- The regression slope parameters (Table 2) vary to some degree between the two years but are consistent in sign. The high correlation between latitude and sunshine duration (Table A1.9) makes the individual coefficients for these two parameters less robust from year to

year. Even so, there was a significant improvement with the inclusion of both these parameters, rather than just one, and so these have both been retained.

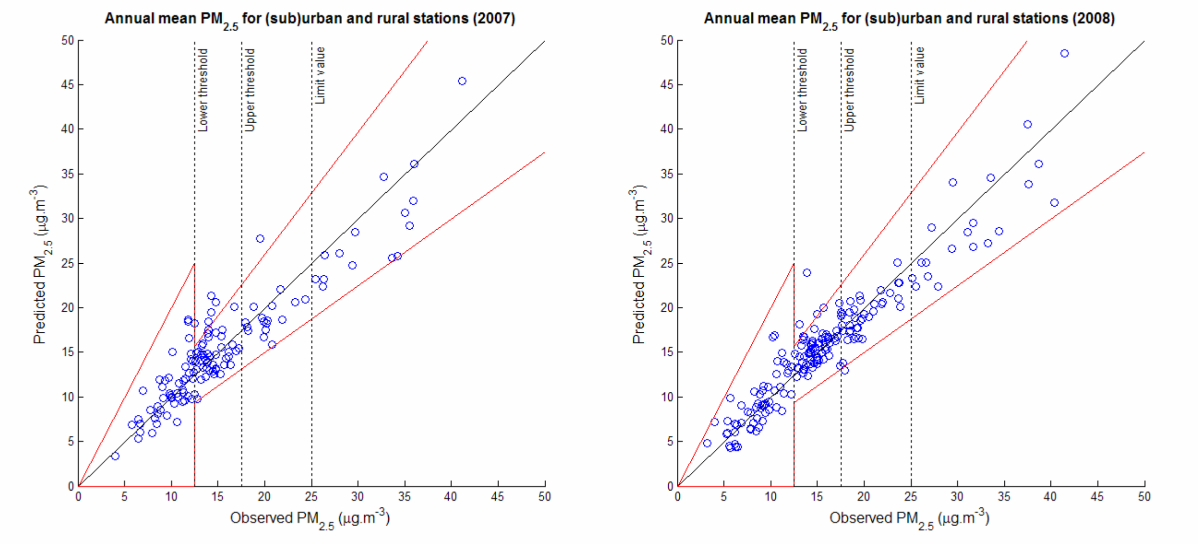


*Figure 1. Scatter plots of the pseudo ('predicted') $PM_{2.5}$ generation for the selected supplementary data (**PM10 + lat + lon + pop + sun (meteo)**) and the observed $PM_{2.5}$ concentrations. Also shown in the plot is the ECQO envelope (red lines) showing the requirements as laid out in the European AQ Directive for monitoring data.*

*Table 1. Statistical assessment of the generation of pseudo $PM_{2.5}$ concentrations using the chosen MLR variables of **PM10 + lat + lon + pop + sun (meteo)** as shown in Figure 1.*

MLR of all background stations using: PM10 + lat + lon + pop + sun(meteo)							
Year	**Number of stations**	**Mean (µg/m³)**	**FB**	**RMSE (µg/m³)**	**R^2**	**FAC25%**	**ECQO**
2007	129	15.48	0.000	2.828	0.843	0.853	0.953
2008	178	15.86	0.000	2.404	0.894	0.899	0.966

*Table 2. Regression coefficients determined for the chosen MLR variables of **PM10 + lat + lon + pop + sun (meteo)** as shown in Figure 1 and Table 1.*

Regression parameters	**Coefficients (2007)**	**Coefficients (2008)**
Zero Offset (µg/m³)	27.02	52.06
Lat (degrees)	-0.382	-0.63
Lon (degrees)	0.196	0.12
Population (inhabitant per 10 x 10 km²)	-2.187e-06	-9.077e-07
Integrated radiation (W.m².hour)	-0.809e-06	-1.96e-06
PM10 (µg/m³)	0.571	0.628

5. Exploration and selection of supplementary data for the residual kriging mapping regression

As previously stated the first step in the mapping procedure is to carry out multiple linear regression of the $PM_{2.5}$ data with other supplementary data to provide a background map for the eventual residual kriging. For this assessment we use a set of $PM_{2.5}$ data that consists of both the observed $PM_{2.5}$ concentrations and, at stations where PM_{10} is measured but not $PM_{2.5}$, the pseudo $PM_{2.5}$ predictions. Results are shown for the urban and the rural stations separately, as this is the method to be used for the $PM_{2.5}$ mapping. The aim is to identify useful supplementary data for their eventual use in the residual kriging. The choice made in this section is indicative only as the best mapping regression result does not necessarily provide the best residual kriging result.

In previous mapping studies the relevant rural background data for PM_{10} was found to be **EMEP + Altitude + Wind speed (*wind(meteo)*) + Solar radiation (*sun(meteo)*)** and for the urban background data only the **EMEP model** was found to improve the results. In addition log-normal transformation of the concentrations has been shown to improve the regression results. Logarithmic transformation of the altitude was not shown to improve the results for PM_{10}. There is no guarantee that the same supplementary data is suitable for $PM_{2.5}$ but it is expected to be similar.

The MLR is carried out for the two years 2007 and 2008 separately. Note that in the testing the natural logarithm of the population was used as a regression parameter rather than just the population as this was found to provide consistently improved results.

The following conclusions are made concerning these tests:

- For rural background mapping regression there is no clear 'best' set of variables. Either one of the choices 7, 8, 10 or 11 provides similar results (Tables A2.1 and A2.3) which include various combinations of the parameters **EMEP model, accumulated solar radiation (*sun(meteo)*), wind speed (*wind(meteo)*), altitude and population.** Choice 7 has the best scoring statistical results and this is shown in Figure 2 and in Table 3.
- For the urban background mapping there is some improvement with the inclusion of more parameters than the **EMEP model**. Similarly to the rural background results the choices 7, 8, 10 and 11 all give similar results. (Tables A2.2 and A2.4). The results are generally poorer than for the rural background, with lower correlations and higher RMSE. As with the rural case Choice 7 is shown in Figures 3 and in Tables 4.

The residual kriging cross-validation assessment will be based on an exploration of the more promising of the above results.

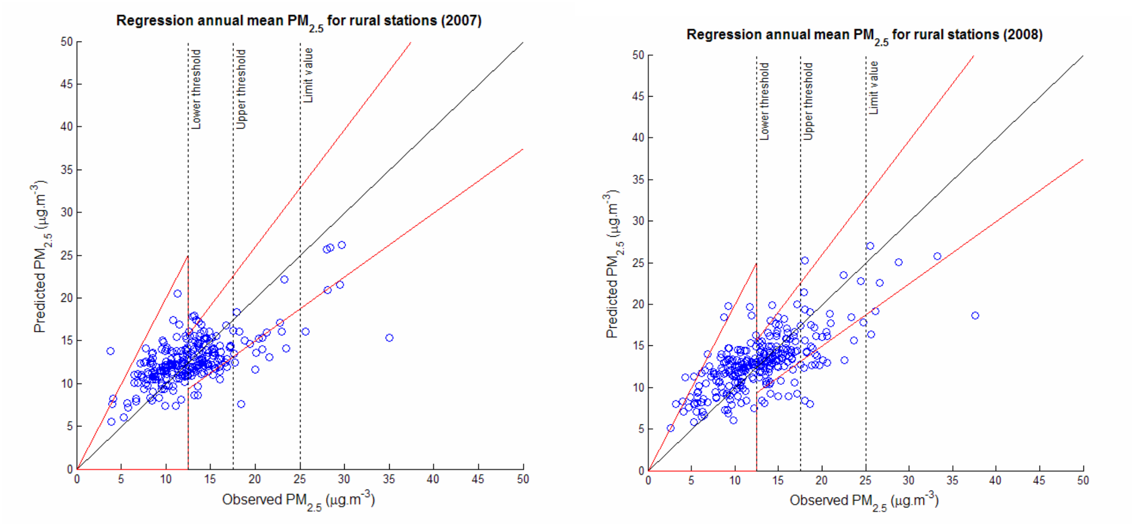


Figure 2. Results of the mapping regression for rural stations using choice 7 in Tables A2.1 and A2.3. Left 2007 and right 2008.

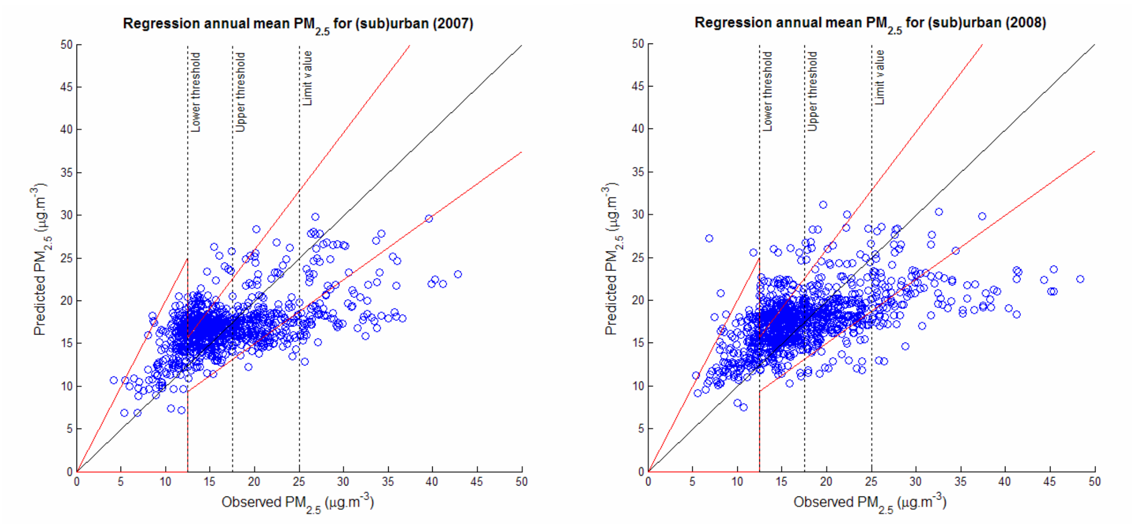


Figure 3. Results of the mapping regression for (sub)urban stations using choice 11 in Tables A2.2 and A2.4. Left 2007 and right 2008.

Table 3. Statistical results of the MLR variables for Choice 7 ***EMEP + sun(meteo) + wind(meteo) + alt + log(pop)*** *as shown in Figure 2 for the rural background stations. Comparison is against both pseudo and observed $PM_{2.5}$ concentrations.*

MLR of rural background stations using: EMEP + sun(meteo) + wind(meteo) + alt + log(pop)							
Year	**Number of stations**	**Mean (µg/m³)**	**FB**	**RMSE (µg/m³)**	**R^2**	**FAC25%**	**ECQO**
2007	253	12.73	0.001	3.415	0.438	0.715	0.901
2008	278	13.02	0.004	3.577	0.477	0.676	0.867

Table 4. Statistical results of the MLR variables Choice 7 ***EMEP + sun(meteo) + wind(meteo) + alt + log(pop)*** *as shown in Figure 3 for the (sub)urban background stations. Comparison is against both pseudo and observed $PM_{2.5}$ concentrations.*

MLR of (sub)urban background stations using: EMEP + sun(meteo) + wind(meteo) + alt + log(pop)							
Year	**Number of stations**	**Mean (μg/m³)**	**FB**	**RMSE (μg/m³)**	**R^2**	**FAC25%**	**ECQO**
2007	928	16.92	-0.002	4.844	0.292	0.635	0.749
2008	1025	17.65	-0.001	5.227	0.294	0.66	0.747

6 Kriging and residual kriging assessment

6.1 Exploratory assessment using both pseudo and measured $PM_{2.5}$

In this section cross-validation is carried out for both log-normal kriging and residual kriging (using a log-normal transformation) using both the pseudo and measured $PM_{2.5}$ data. The residual kriging is based on a range of choices (and some additional combinations to further explore the significance of a number of supplementary parameters) as outlined in Section 5 for the background subtraction field. The rural and (sub)urban stations are treated separately. Exploratory results are tabulated in Appendix 3.1 and 3.2. The variogram parameters used for the interpolations are given in Appendix 3.3.

The following conclusions are made concerning the interpolation

- For the rural kriging the statistical indicators for the residual kriging are better than for the log-normal kriging. Choice 8, combination of **EMEP + wind(meteo) + alt + log(pop),** provides the best results. These are shown in Figure 4 and Table 5. The addition of accumulated solar radiation does not improve the results to any significant degree. The above combination is chosen for map making.

- For urban interpolation the log-normal kriging performs better than the residual kriging. These are shown in Figure 5 and Table 6. Of the residual kriging the best supplementary data combination is found to be just the EMEP model. This is consistent with PM_{10} mapping previously carried out and indicates that none of the supplementary data provides useful additional information for the urban mapping. The log-normal kriging is chosen for map making.

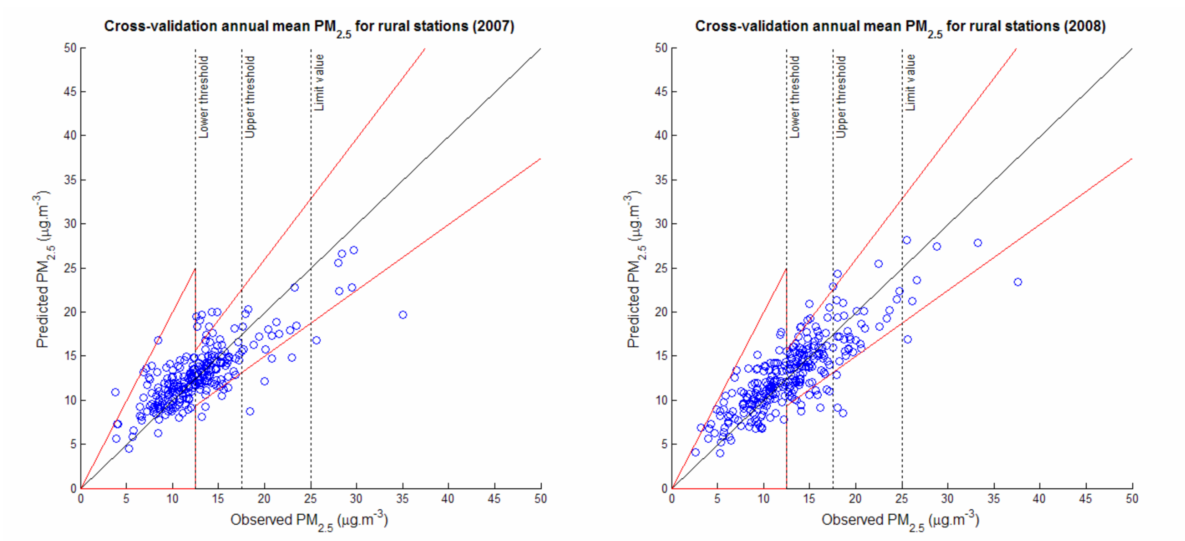


Figure 4. Results of the residual kriging for rural stations using choice 8 (EMEP model+ wind(meteo) alt + log(pop)) in Tables A3.1 and A3.2. Left 2007 and right 2008.

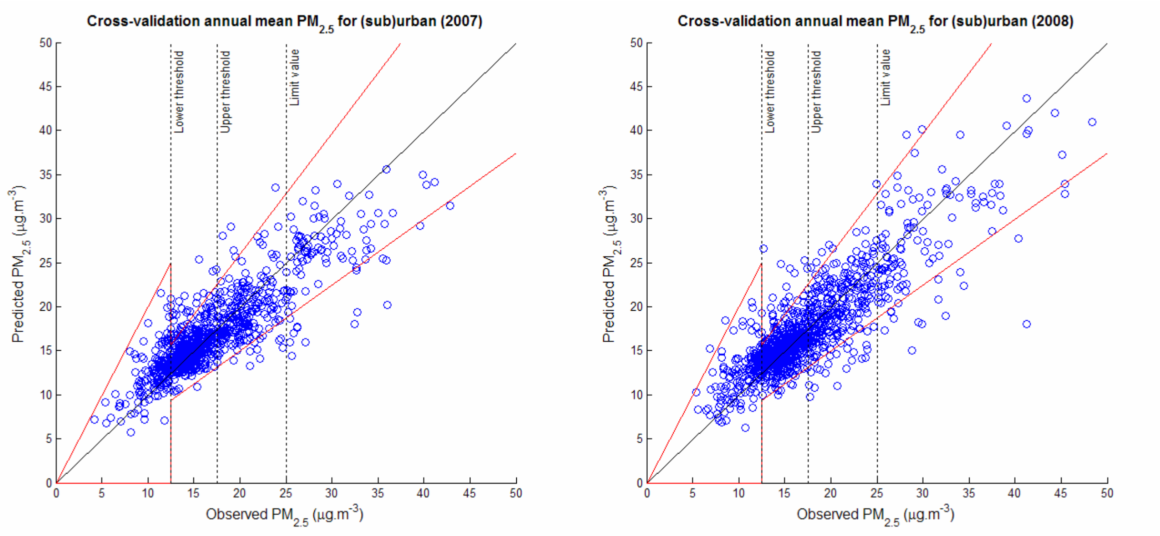


Figure 5 Results of the log-normal kriging for (sub)urban stations using choice 0 in Tables A3.3 and A3.4. Left 2007 and right 2008.

*Table 5. Statistical results of the residual kriging cross-validation using **EMEP + wind(meteo) + alt + log(pop)** as supplementary data for the mapping regression, shown in Figure 4 for the rural background stations. Comparison is against both pseudo and observed $PM_{2.5}$ concentrations.*

Residual kriging of rural background stations using: EMEP + wind(meteo) + alt + log(pop)							
Year	**Number of stations**	**Mean (µg/m³)**	**FB**	**RMSE (µg/m³)**	**R^2**	**FAC25%**	**ECQO**
2007	253	12.73	0.005	2.836	0.613	0.794	0.909
2008	278	13.03	0.006	2.833	0.676	0.763	0.924

Table 6. Statistical results of the log-normal kriging cross-validation as shown in Figure 5 for the (sub)urban background stations. Comparison is against both pseudo and observed $PM_{2.5}$ concentrations.

Log-normal kriging of (sub)urban background stations							
Year	**Number of stations**	**Mean (µg/m³)**	**FB**	**RMSE (µg/m³)**	**R^2**	**FAC25%**	**ECQO**
2007	928	16.92	0.001	2.977	0.733	0.88	0.93
2008	1025	17.65	-0.001	3.231	0.731	0.861	0.911

6.2 Impact of pseudo $PM_{2.5}$ data on the interpolation

As an additional assessment, to see the impact of the pseudo $PM_{2.5}$ stations on the interpolation, the recommended interpolation methods provided in Section 6.1 are also assessed using cross-validation only at sites where $PM_{2.5}$ is measured. The intention is to determine if any improvement is obtained by including pseudo $PM_{2.5}$ stations in the interpolation. This is done by carrying out two cross-validation interpolations:

a) Using both pseudo and observed $PM_{2.5}$ concentrations in the interpolation but assessing the results only at the measured $PM_{2.5}$ sites.

b) Using only observed $PM_{2.5}$ concentrations in the interpolation and assessing these results only at the measured $PM_{2.5}$ sites.

Any difference in the statistical indicators is the result of the addition of the pseudo $PM_{2.5}$ stations. The results are shown in Tables 7 and 8. The following conclusions are made based on these results:

- Inclusion of the pseudo $PM_{2.5}$ data significantly improves the RMSE (by around 12%) and correlation (by 13%) for the rural interpolation when the interpolation is assessed at only the measured $PM_{2.5}$ sites. The FAC25% and the ECQO indicators are not necessarily improved with the inclusion of the pseudo station data.

- The inclusion of $PM_{2.5}$ data significantly improves the RMSE (by around 10%) and correlation (by up to 24%) for the (sub)urban interpolation when the interpolation is assessed at only the observed $PM_{2.5}$ sites. Both the FAC25% and the ECQO indicators are also improved with the inclusion of the pseudo $PM_{2.5}$ station data.

Table 7. Statistical results of the residual kriging cross-validation for the rural background stations. The two cases, a and b, indicate the use of both pseudo and observed concentrations (a) or just observed concentrations (b). Both years 2007 and 2008 are shown separately. Best statistical results are highlighted with red, then orange.

RURAL BACKGROUND stations 2007 at observed $PM_{2.5}$ sites only: n = 49, observed mean = 12.21 μg/m^3						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
8a	(using pseudo + obs) EMEP + wind(meteo) + alt + log(pop)	0.023	3.347	0.731	0.714	0.918
8b	(using obs only) EMEP + wind(meteo) + alt + log(pop)	-0.002	3.776	0.654	0.735	0.918
RURAL BACKGROUND stations 2008 at observed $PM_{2.5}$ sites only: n = 63, observed mean = 11.69 μg/m^3						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
8a	(using pseudo + obs) EMEP + wind(meteo) + alt + log(pop)	-0.019	3.485	0.717	0.619	0.905
8b	(using obs only) EMEP + wind(meteo) + alt + log(pop)	-0.005	4.005	0.625	0.571	0.937

Table 8. Statistical results of the log-normal kriging cross-validation for the (sub)urban background stations. The two cases, a and b, indicate the use of both pseudo and observed concentrations (a) or just observed concentrations (b). Both years 2007 and 2008 are shown separately. Best statistical results are highlighted with red, then orange.

URBAN BACKGROUND stations 2007 at observed $PM_{2.5}$ sites only: n = 98, observed mean = 17.26 μg/m^3						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
0a	(using pseudo + obs) lognormal kriging	-0.041	4.09	0.714	0.786	0.867
0b	(using obs only) lognormal kriging	0.006	4.642	0.574	0.776	0.837
URBAN BACKGROUND stations 2008 at observed $PM_{2.5}$ sites only: n = 137, observed mean = 17.76 μg/m^3						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
0a	(using pseudo + obs) lognormal kriging	-0.038	3.56	0.744	0.854	0.905
0b	(using obs only) lognormal kriging	0.001	3.986	0.654	0.774	0.861

7 Maps of $PM_{2.5}$

This section presents the European maps of annual mean $PM_{2.5}$, using the recommended supplementary data and methods outlined in the previous section. The following maps are presented for both 2007 and 2008:

1. Combined rural and urban maps using both measured and pseudo $PM_{2.5}$ data;
2. Combined rural and urban maps using only measured $PM_{2.5}$ data;
3. Difference between the two maps (measured and pseudo – measured only);
4. Uncertainty maps based on the kriging variance, for the maps using both measured and pseudo $PM_{2.5}$ data.

In all cases urban maps are created using log-normal kriging and rural maps are created using residual kriging after logarithmic transformation of the concentration data and using the supplementary data **log(EMEP) + wind(meteo) + alt + log(pop)** for the spatial regression. The urban and rural maps are combined using population weighting similar to recent European scale air pollutants mapping (de Smet et al., 2011).

The variogram parameters used in the kriging (Table A3.5) have been slightly adjusted to be consistent between the two years and for the two different types of maps (with and without pseudo measurements). These new variogram parameters are provided in Appendix 4, Table A4.1 and A4.2.

7.1 Maps for 2007

In Figure 6 the positions and concentrations of the pseudo and observed $PM_{2.5}$ measurement sites are shown. In Figure 7 the map of annual mean $PM_{2.5}$ using both pseudo and observed $PM_{2.5}$ is given and in Figure 8 the same map but using only observed $PM_{2.5}$ is provided. In Figure 9 the difference between these two maps is shown. In Figure 10 the uncertainty, based on the kriging variance, of Figure 7 is shown.

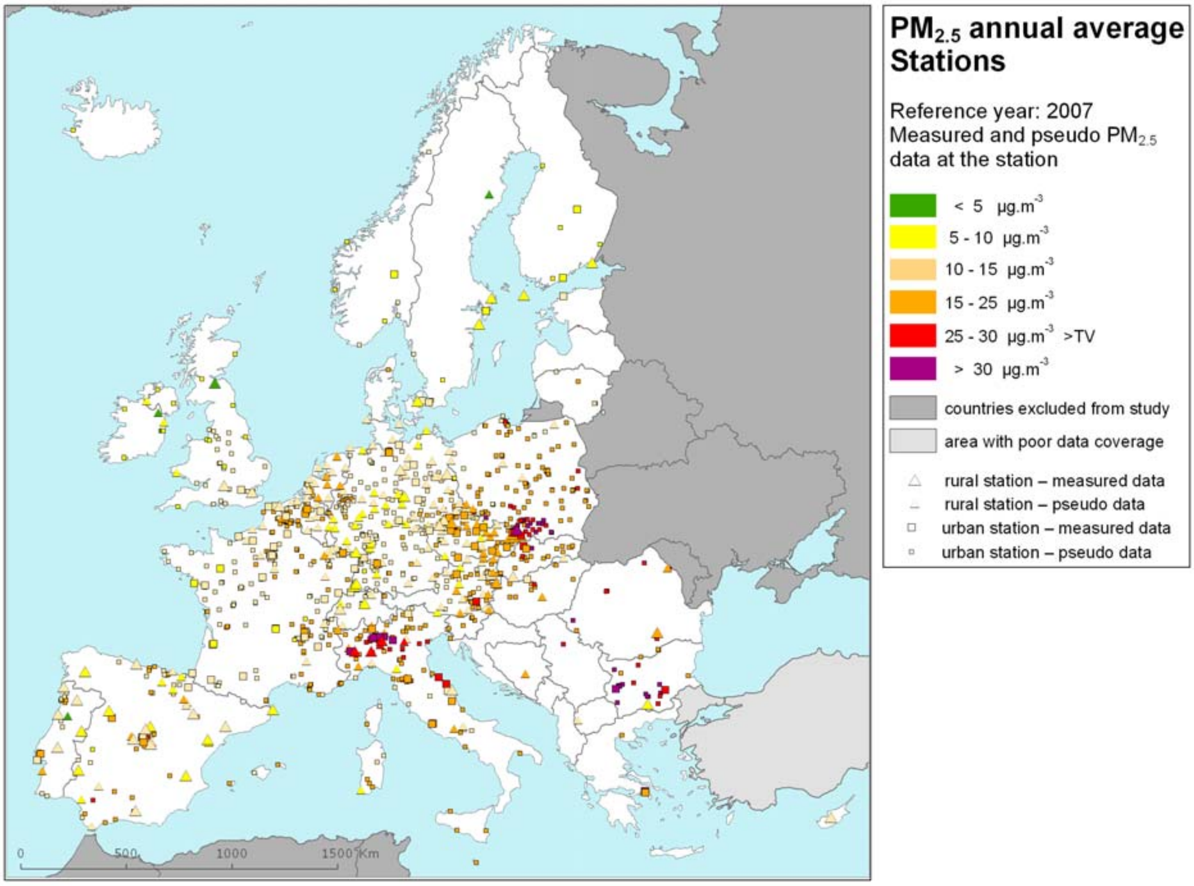


Figure 6. Positions and concentrations of the observed and pseudo annual mean $PM_{2.5}$ concentrations used in the map making for 2007.

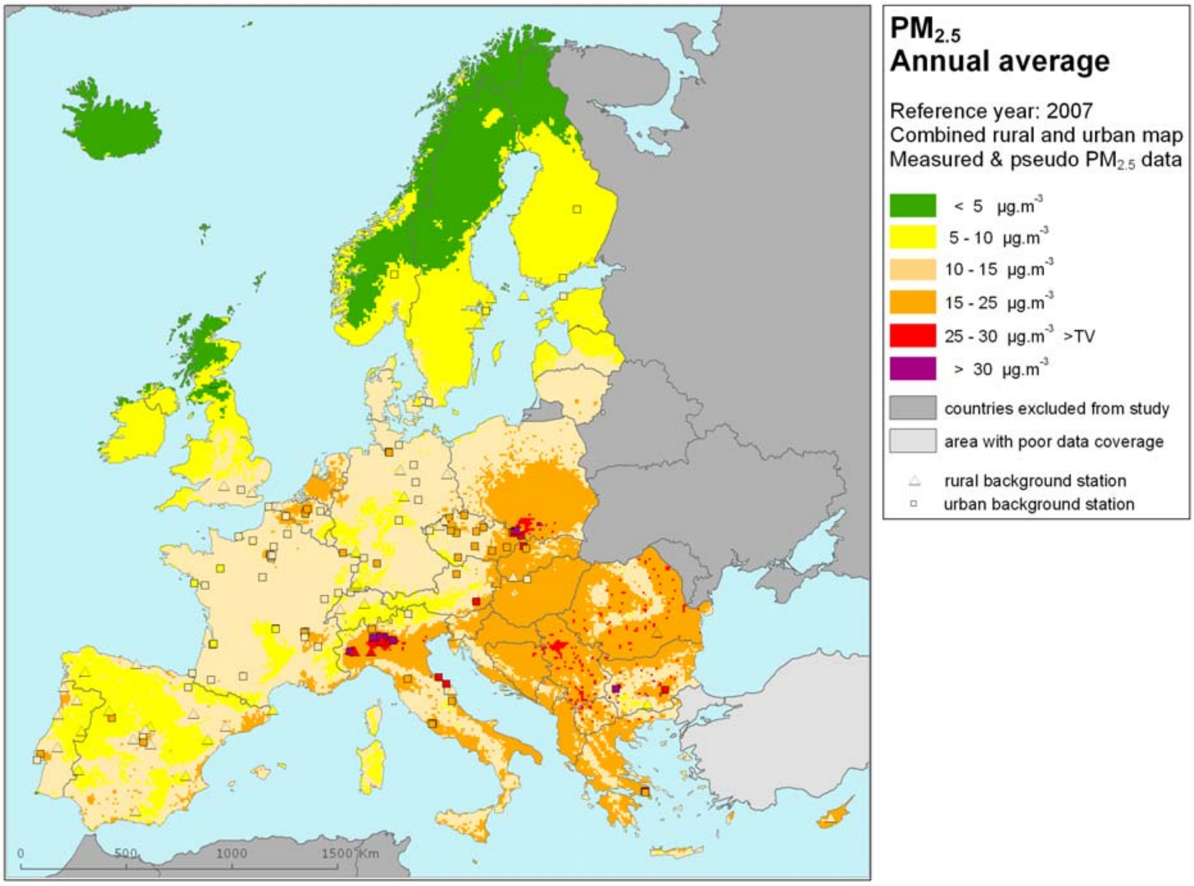


Figure 7. Combined rural and urban map for 2007 made using both observed and pseudo $PM_{2.5}$ data. Also shown are the observed $PM_{2.5}$ concentrations.

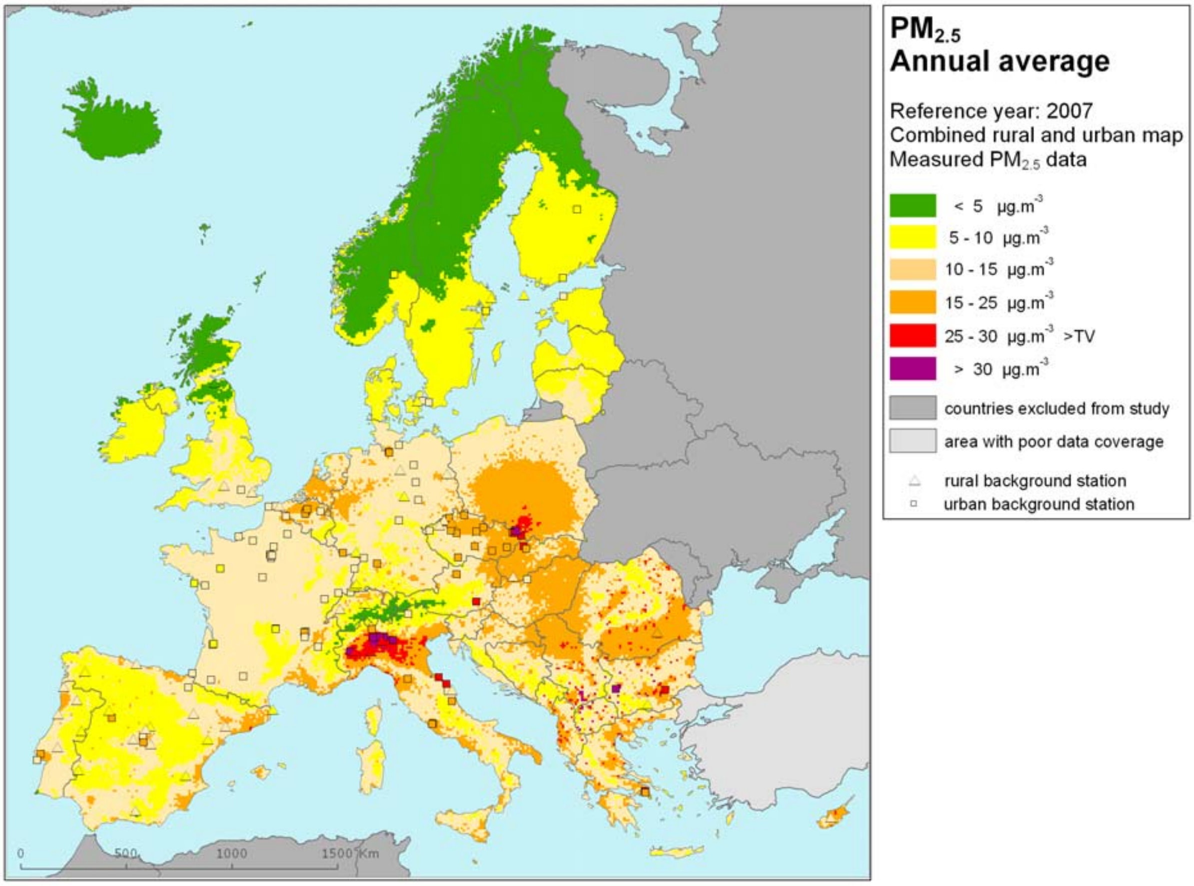


Figure 8. Combined rural and urban map for 2007 made using only observed $PM_{2.5}$ data. Also shown are the observed $PM_{2.5}$ concentrations.

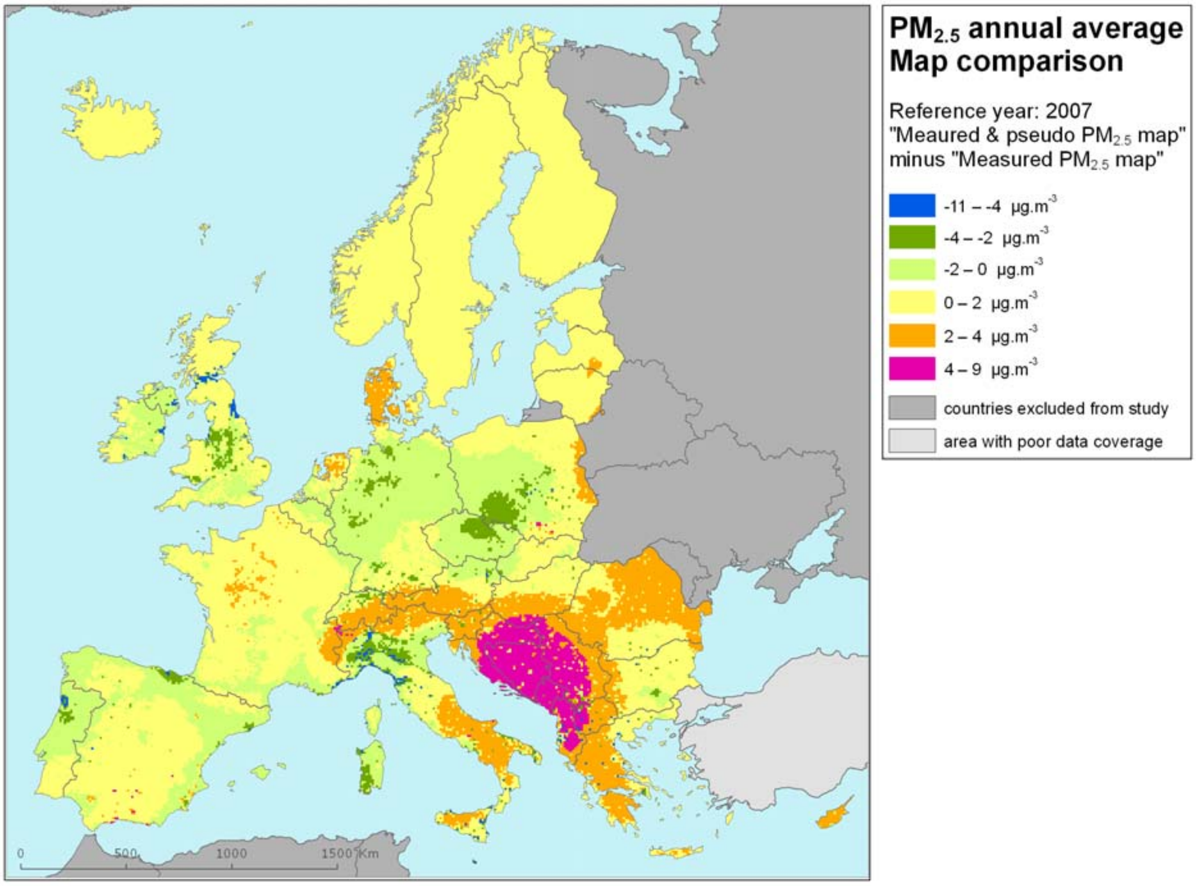


Figure 9. Difference map of the combined rural and urban maps for 2007 that are shown in Figures 7 and 8.

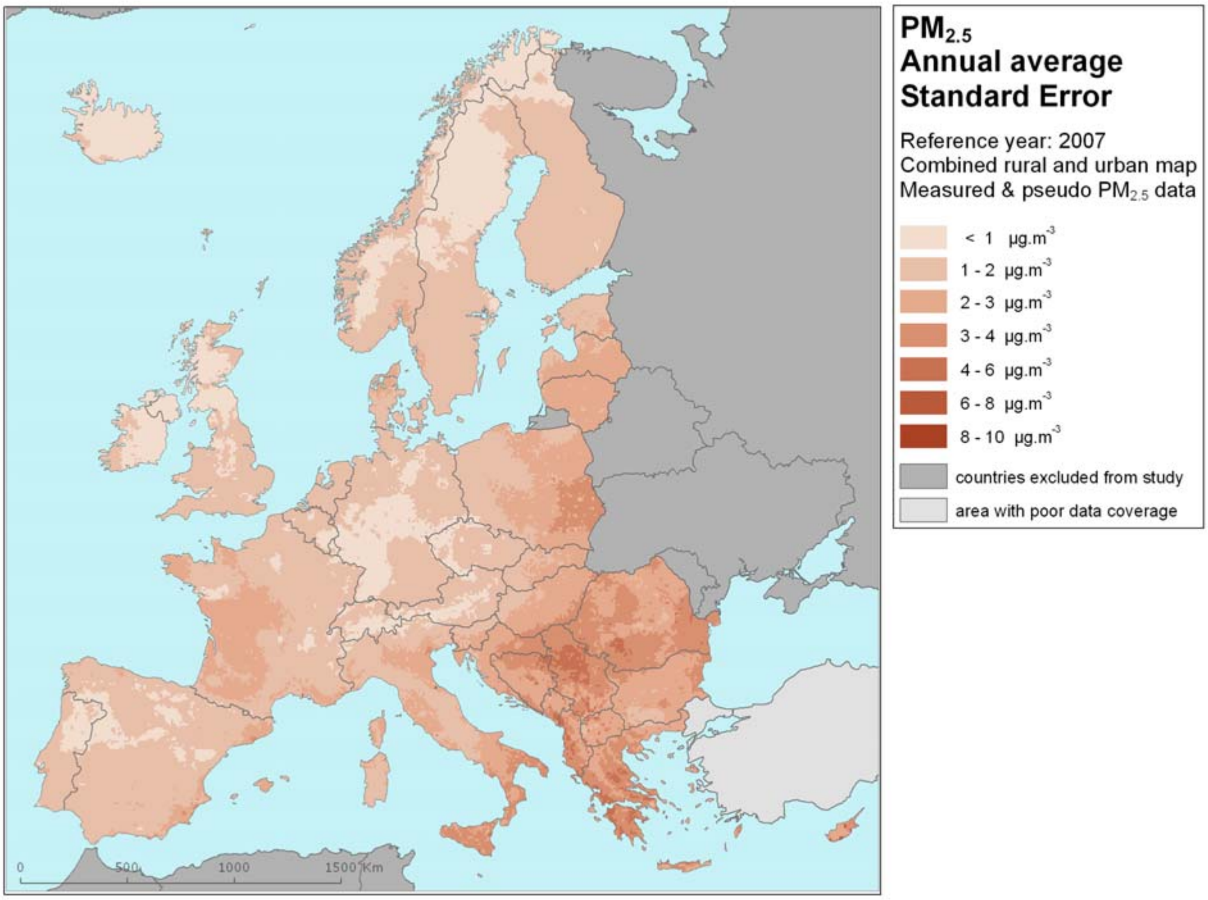


Figure 10. Combined rural and urban map uncertainty map for 2007 made using both observed and pseudo $PM_{2.5}$ data, as in Figure 7. Uncertainty maps are based on the residual and log-normal kriging variance and indicate the interpolation uncertainty as one standard deviation.

7.2 Maps for 2008

In Figure 11 the positions and concentrations of the pseudo and observed $PM_{2.5}$ measurement sites are shown. In Figure 12 the map of annual mean $PM_{2.5}$ using both pseudo and observed $PM_{2.5}$ is given and in Figure 13 the same map but using only observed $PM_{2.5}$ is provided. In Figure 14 the difference between these two maps is shown. In Figure 15 the uncertainty, based on the kriging variance, of Figure 11 is shown.

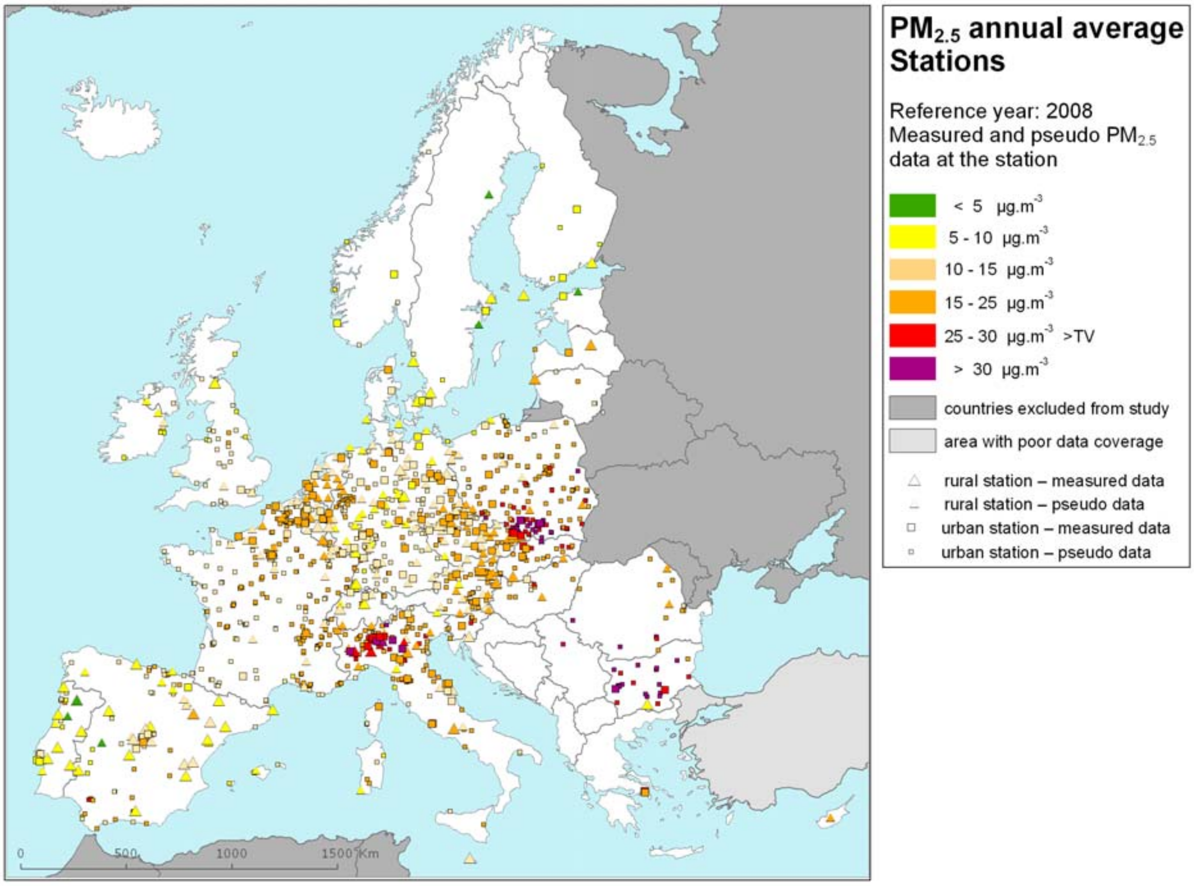


Figure 11. Positions and concentrations of the observed and pseudo annual mean $PM_{2.5}$ concentrations used in the map making for 2008.

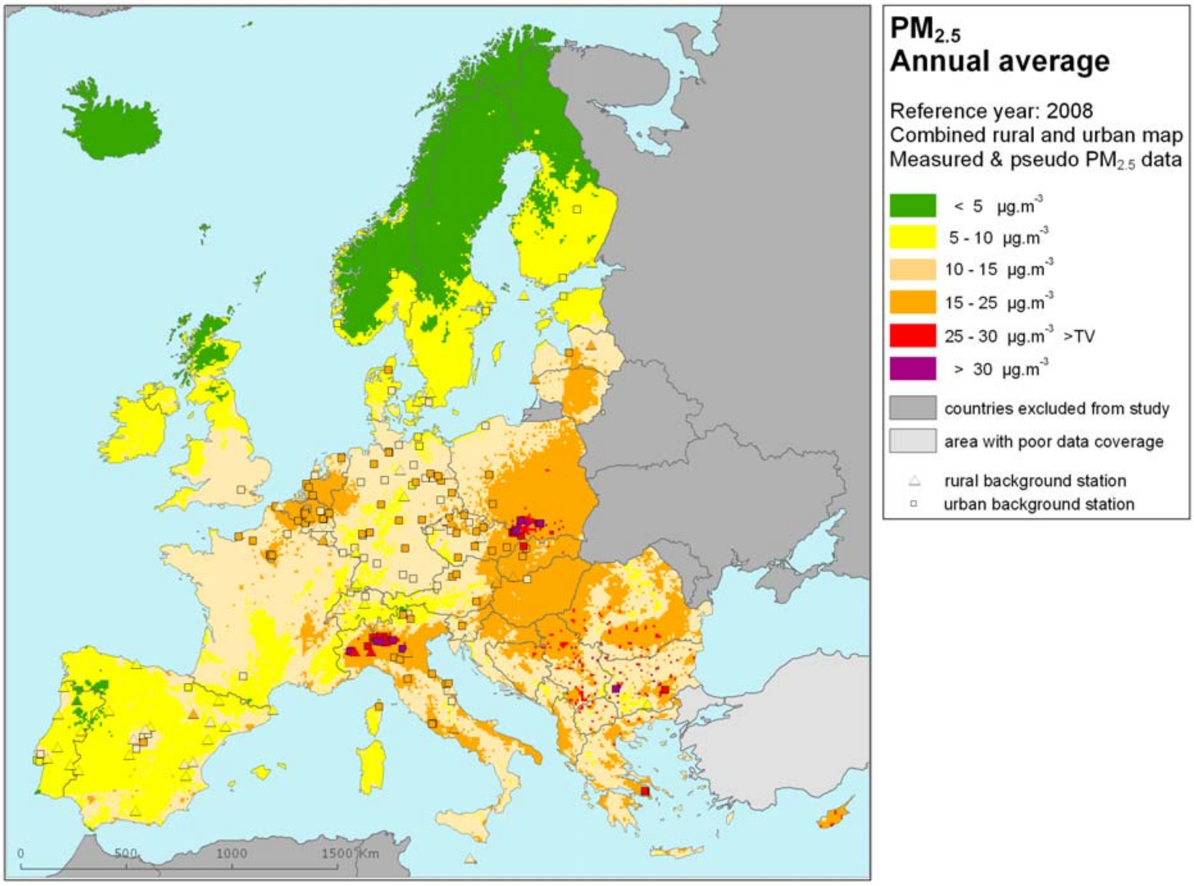


Figure 12. Combined rural and urban map for 2008 made using both observed and pseudo $PM_{2.5}$ data. Also shown are the observed $PM_{2.5}$ concentrations.

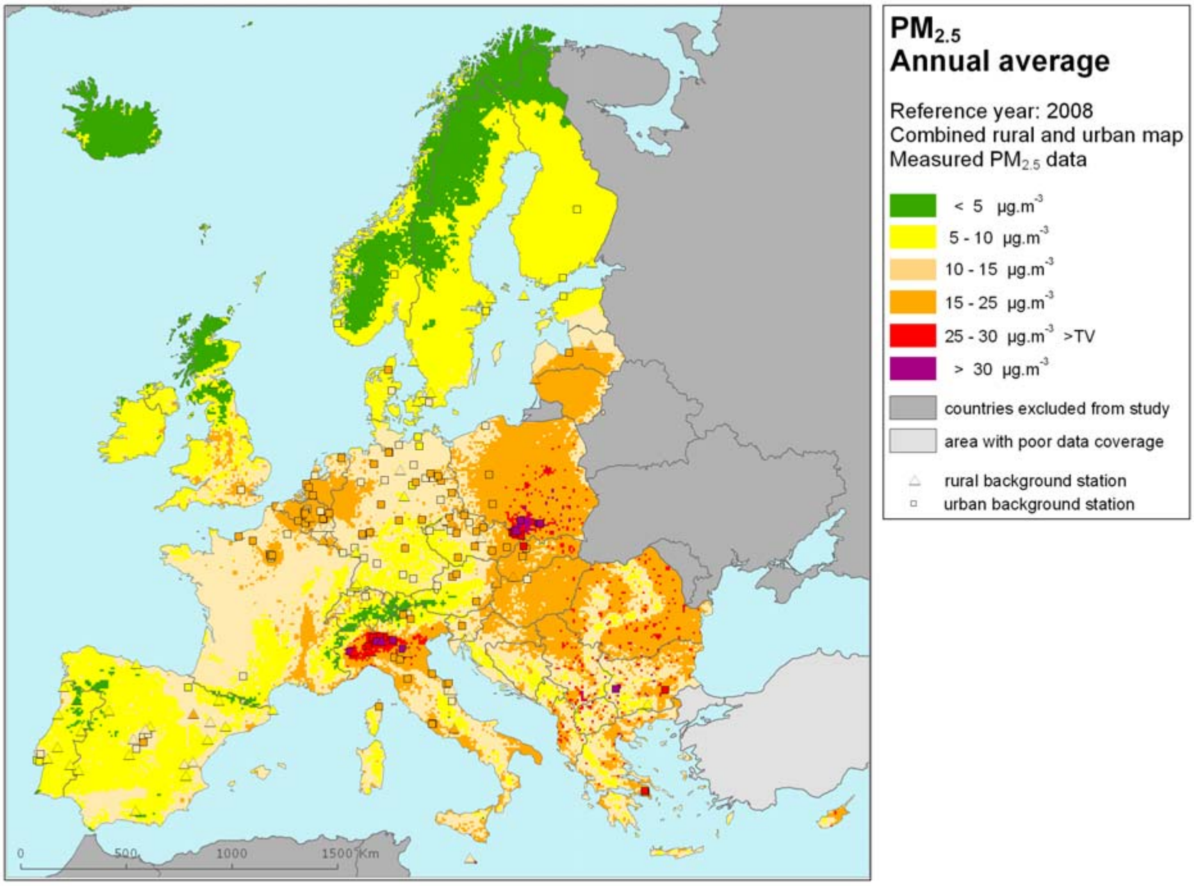


Figure 13. Combined rural and urban map for 2008 made using only observed $PM_{2.5}$ data. Also shown are the observed $PM_{2.5}$ concentrations.

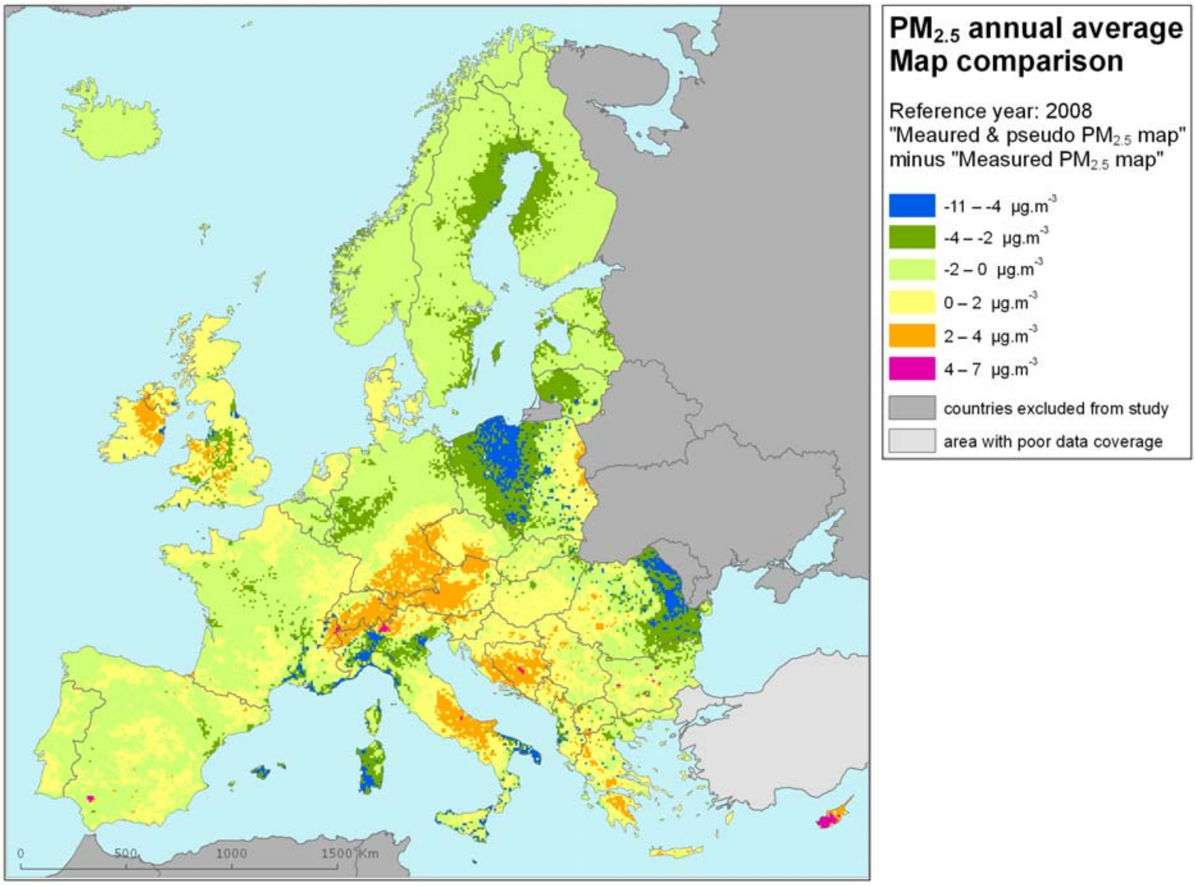


Figure 14. Difference map of the combined rural and urban maps for 2008 that are shown in Figures 11 and 12.

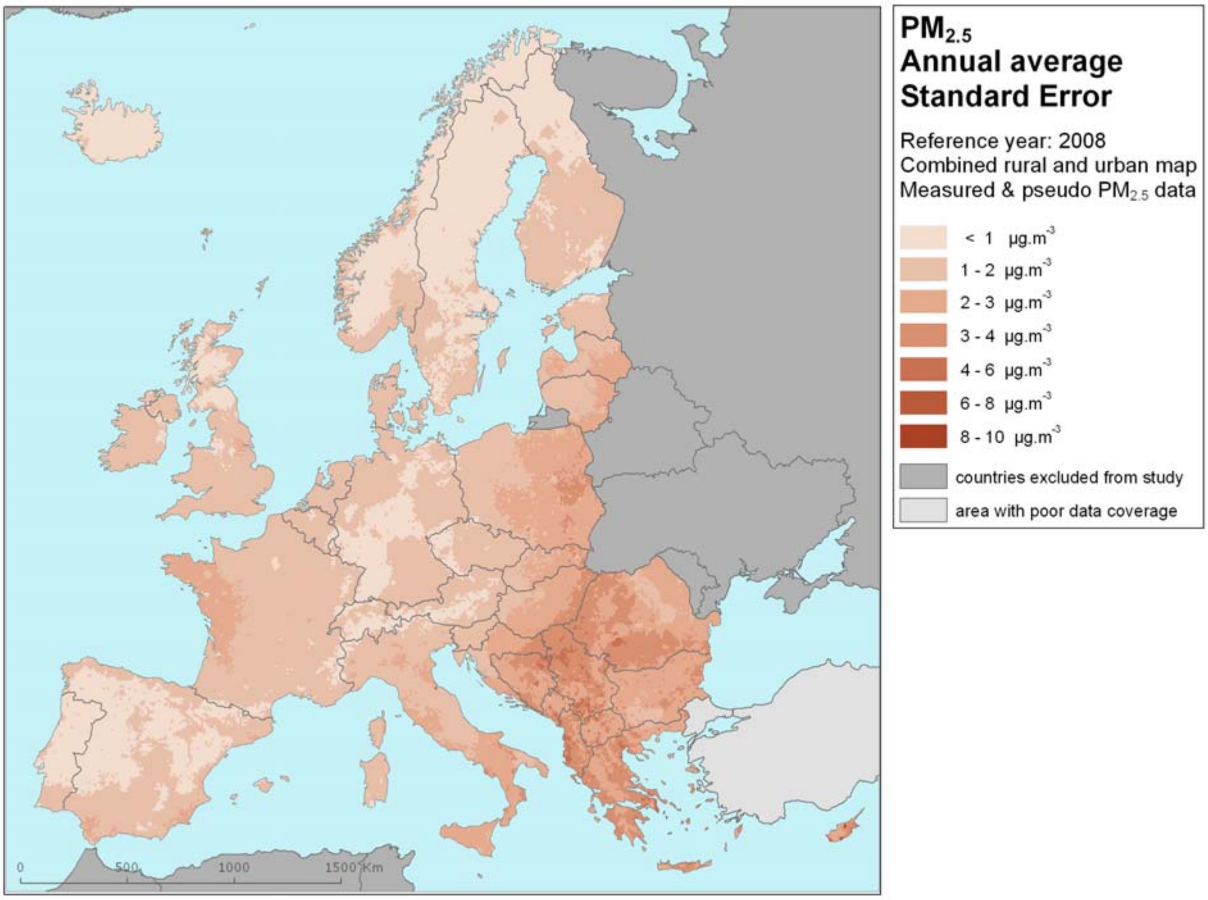


Figure 15. Combined rural and urban map uncertainty map for 2008 made using both observed and pseudo PM$_{2.5}$ data, as in Figure 7. Uncertainty maps are based on the residual and log-normal kriging variance and indicate the interpolation uncertainty as one standard deviation.

7.3 Discussion concerning the maps

There are some significant differences between the two maps presented in Sections 7.1 and 7.2 for both years. These differences are the direct result of including pseudo PM$_{2.5}$ data in the map making. The two years studied will show different results, depending on the number and spatial distribution of pseudo and observed PM$_{2.5}$ concentrations.

Common to both years are the following features:

- For large areas of Europe (e.g. Spain, France, Germany, Scandinavia, UK) there are only small differences (± 2 μg/m^3) in the rural background concentrations. This may be attributable to the higher correlation of the supplementary data used for the residual kriging baseline (Table 3). It may also reflect the higher coverage of existing PM$_{2.5}$ stations in these areas.
- In central and northern Poland the inclusion of pseudo PM$_{2.5}$ stations reduces the concentrations by more than 2 μg/m^3. In this area there are no rural measurements of PM$_{2.5}$

available. The use of pseudo $PM_{2.5}$ stations reduces the impact of stations in the more polluted south of Poland.

- The extent of high concentrations of $PM_{2.5}$ (> 25 μg/m^3) in the Po Valley area is significantly reduced with the introduction of pseudo $PM_{2.5}$ stations.

For 2007 the following features are noted:

- The inclusion of pseudo $PM_{2.5}$ stations leads to a significant increase in the $PM_{2.5}$ concentrations in the Balkan region. This area is highly uncertain as there are virtually no monitoring stations available in this region

For 2008 the following features are noted:

- The inclusion of pseudo $PM_{2.5}$ stations leads to a significant increase in the area of Cyprus.

The uncertainty maps in Figures 10 and 15 show significant uncertainty in some regions of Europe. In areas such as Eastern Europe these uncertainties can be as high as 6 μg/m^3. In areas with high station density, e.g. Germany, these uncertainties are lower at 1 or 2 μg/m^3. In general we see uncertainties in the range 15-30%, similar to the cross-validation RMSE. It should be noted that when using the logarithmic transformation of the data the uncertainty estimate made during the kriging, which uses a single variogram for all of Europe, is proportional in nature. This means that areas with high concentrations will have higher absolute uncertainties compared to areas with low concentrations.

8 Population exposure

We are interested to see the impact the pseudo $PM_{2.5}$ stations have on the population weighted concentrations for all of Europe (relevant for health impact assessment) and on the number of people exposed above the AQ Directive target value of 25 μgm^{-3}. This target value is to be met by 2010 and it is set as a limit value to be met 2015. Country based exposure tables are calculated, based on the maps shown in Section 7, for both the years 2007 and 2008 and for the two maps produced, i.e. with (Figures 7 and 12) and without (Figures 9 and 13) pseudo $PM_{2.5}$ stations. These four tables are provided in Appendix 5 and are summarised for all of Europe (Table 9 and Figures 16 and 17) here.

Exposure is calculated by overlaying the 10x10 km^2 gridded concentration maps with a 10x10 km^2 population density map. Population weighted concentrations are then calculated based on these two maps as well as the number of people exposed between predefined concentration levels. We are particularly interested in the number of people that are exposed to concentration levels above the target or limit value, the 'exceedance exposure'. This is the same exposure methodology that has been applied in previous ETC/ACM exposure calculations (e.g. Horálek et al., 2007). In Horálek et al. (2010) this method was updated to include the use of 1x1 km^2 population maps, but this method has not been applied for this particular study.

8.1 European summary exposure table

In Table 9 we produce a summary of the results contained in detailed tables in Appendix 5. Table 9 includes the population in all of Europe that is exposed above the target value and the population weighted concentration for all of Europe. This table indicates the following:

- That the population weighted concentration for all of Europe is very similar for both 2007 and 2008, at around 15 – 16 (μgm^{-3}), with and without the use of pseudo $PM_{2.5}$ stations.
- That the population weighted concentration for all of Europe is only slightly lower when including the pseudo $PM_{2.5}$ stations at 0.5 μgm^{-3} (2007) and 1.3 μgm^{-3} (2008).
- That the population exposed above the target value is significantly lower, by a factor of around one third, when pseudo $PM_{2.5}$ stations are used. The inclusion of pseudo $PM_{2.5}$ stations significantly reduces the estimated exceedance exposure for both years.

This last point is significant as it indicates that the inclusion of pseudo $PM_{2.5}$ stations can have a significant impact when calculating the exposure of the population above the target value set in the AQ Directive *Table 9. European exposure tables showing the population, the percentage of population exposed in particular concentration ranges and the population weighted average concentration based on the 10 x 10 km^2 maps. Included are the results for both 2007 and 2008 where the concentration maps are made with and without pseudo $PM_{2.5}$ stations. The target value of 25 μgm^{-3} is to be met by 2010, which is set as limit value to be met by 2015.*

Calculation year and type	Population UN 2007	Population exposed above the target value of 25 $\mu g/m^3$	Percentage of population exposed above the target value of 25 $\mu g/m^3$	Pop. weighted concentration ($\mu g/m^3$)
2007 with pseudo PM2.5	516,188,303	32,003,675	**6.2**	**15.5**
2007 without pseudo PM2.5	516,188,303	48,521,700	**9.4**	**16.0**
2008 with pseudo PM2.5	534,112,380	33,114,968	**6.2**	**15.6**
2008 without pseudo PM2.5	534,112,380	57,150,025	**10.7**	**16.9**

2008/50/EC (EC, 2008).

8.2 Sensitivity of the national population weighted concentrations to the use of pseudo $PM_{2.5}$ stations

To clarify the results further we compare the calculated population weighted concentrations, with and without pseudo $PM_{2.5}$, in Figures 16 and 17 for the years 2007 and 2008 respectively. We note the following points concerning these.

- For most countries there is little difference between the two methods, with or without pseudo $PM_{2.5}$ stations. Generally the differences are within ± 2 μgm^{-3}. The differences can be both positive and negative but on average for all countries the population weighted concentration is higher by 0.5 and 1.3 μgm^{-3}, for 2007 and 2008, with the inclusion of pseudo $PM_{2.5}$ stations.

- One exception is the small country of Malta in 2008 where the population weighted averaged concentrations was significantly less with the inclusion of pseudo $PM_{2.5}$ data. This country has only one PM_{10} station and this was only operating in 2007.

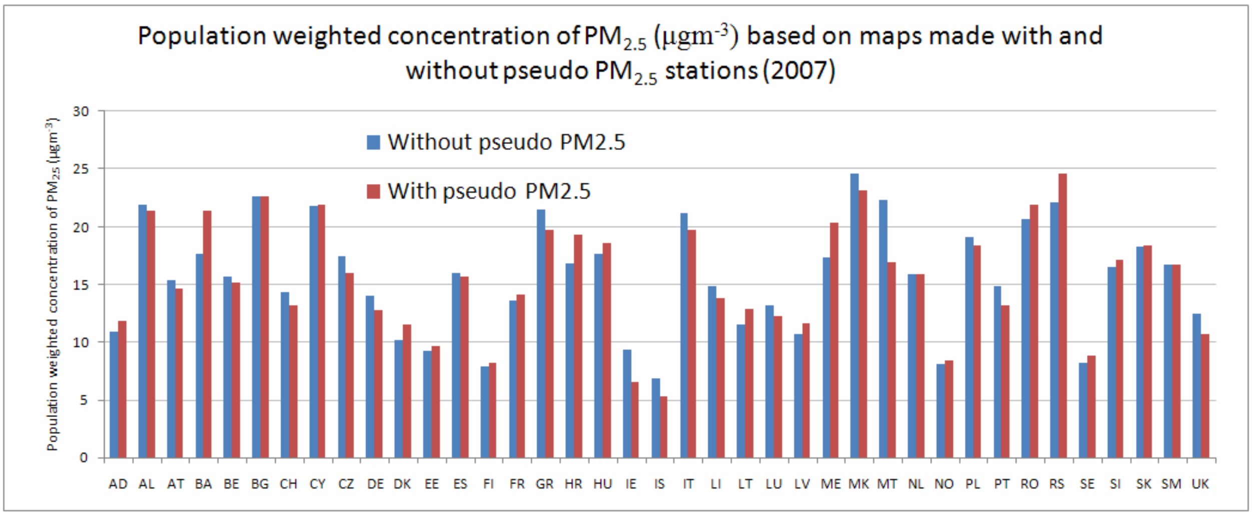


Figure 16. Estimated population weighted concentration (μgm^{-3}) for the year 2007, shown country for country. In blue the calculation without the use of pseudo $PM_{2.5}$, in red the calculation with pseudo $PM_{2.5}$.

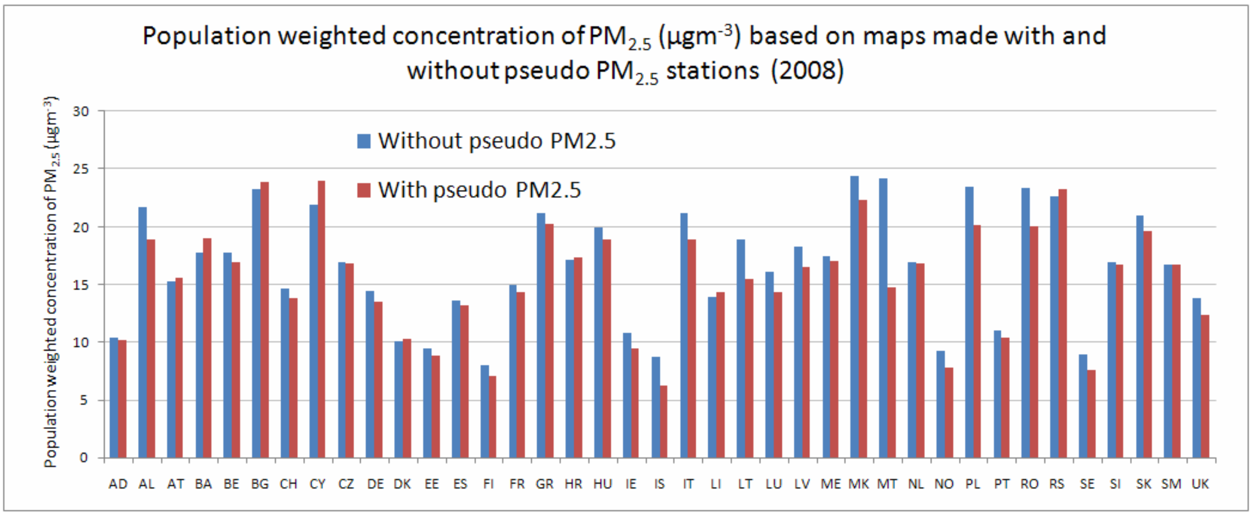


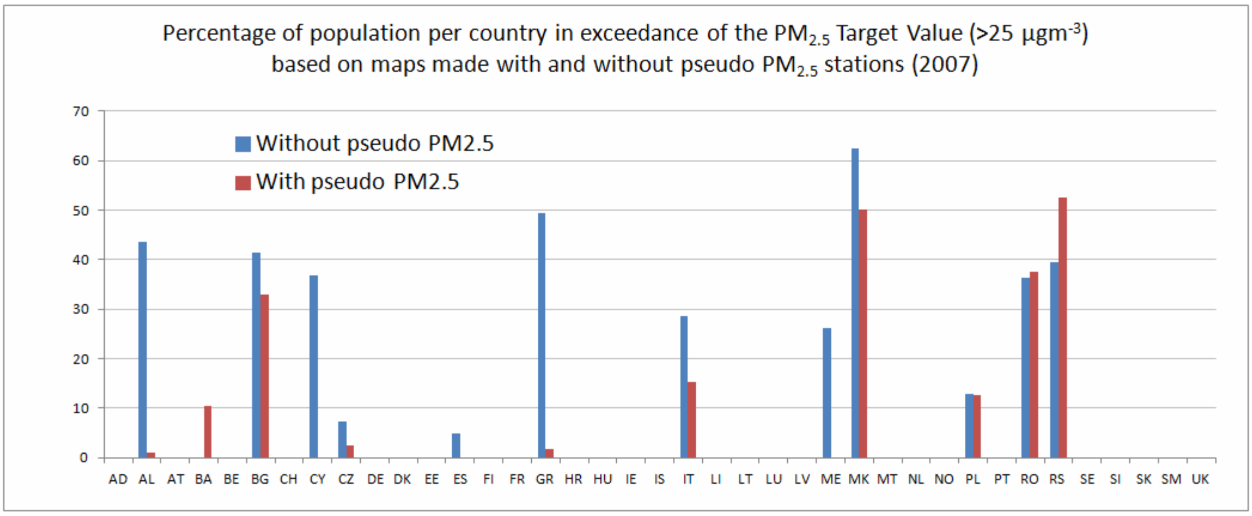
Figure 17. Estimated population weighted concentration (μgm^{-3}) for the year 2008, shown country for country. In blue the calculation without the use of pseudo $PM_{2.5}$, in red the calculation with pseudo $PM_{2.5}$.

8.2 Sensitivity of the national exceedance exposure calculations to the use of pseudo $PM_{2.5}$ stations

In Figures 18 and 19 we show the population exposed (as percentage and as absolute values) above the target value, country by country, both with and without pseudo $PM_{2.5}$. We note the following points concerning these.

- For both years a large proportion of the difference between the two methods seen in Europe is the result of the changed exposure levels in Italy. This reflects the lower concentrations calculated for the Po Valley area, previously commented on in Section 7, with the inclusion of pseudo $PM_{2.5}$ stations.

- In addition, the exposure estimates are significantly less for some other countries when including pseudo $PM_{2.5}$ stations. These include Greece and Albania in 2007 and Romania, Poland and Albania in 2008.



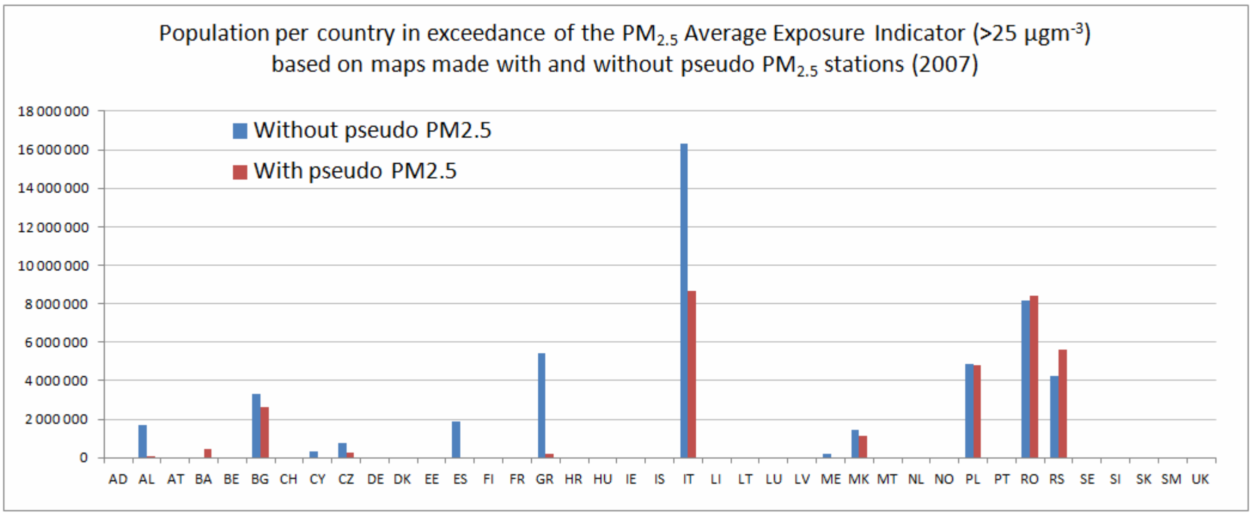
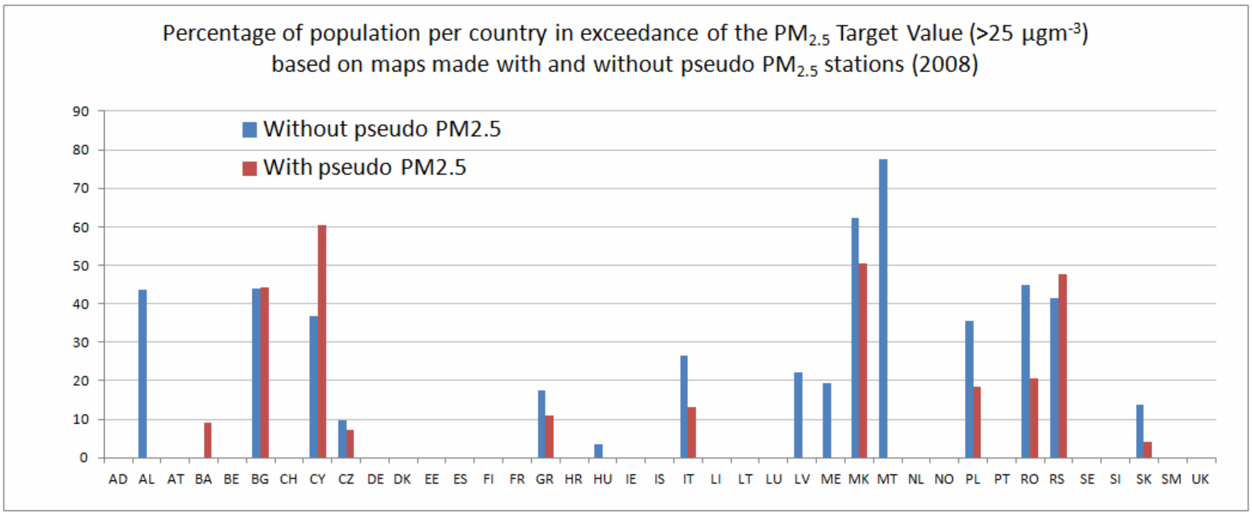


Figure 18. Estimated population exposed, as % (top) and total population (bottom), above the target value (25 μgm^{-3}) for the year 2007, shown country for country. In blue the calculation without the use of pseudo $PM_{2.5}$, in red the calculation with pseudo $PM_{2.5}$.



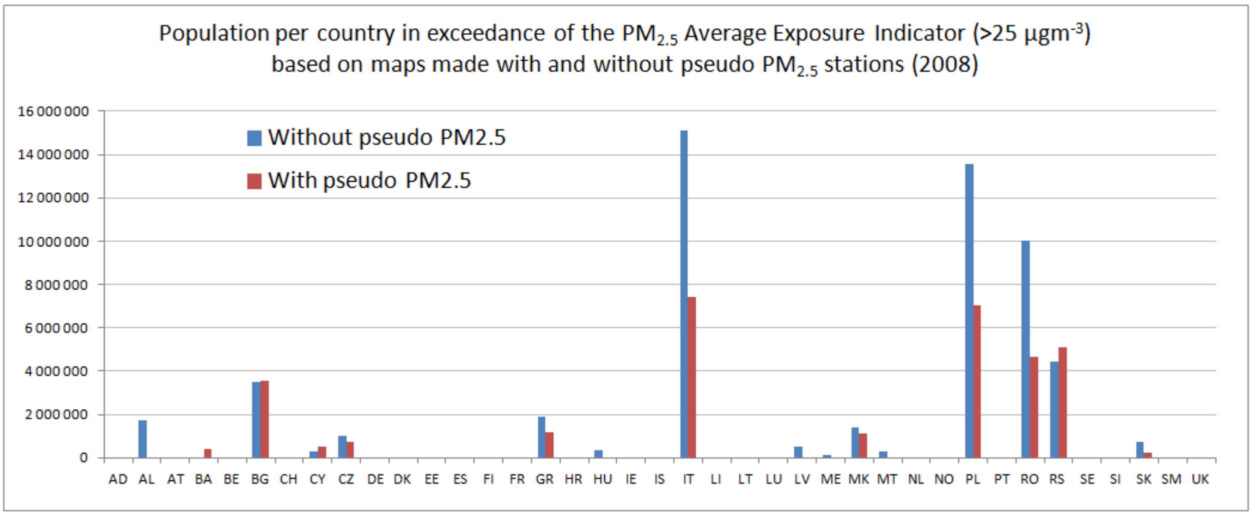


Figure 19. Estimated population exposed, as % (top) and as total population (bottom), above the target value (25 μgm^{-3}) for the year 2008, shown country for country. In blue the calculation without the use of pseudo $PM_{2.5}$, in red the calculation with pseudo $PM_{2.5}$.

8.3 Uncertainty estimates of the exposure

The results show that variations in annual mean concentrations resulting from the use of two different methods, in this case the use or not of pseudo $PM_{2.5}$ stations in the spatial interpolation, can have significant impact on threshold estimates of exposure, especially when considering exposures on national scale. For a number of countries (Italy, Greece, Poland, Romania, Albania) that have significant areas where the concentrations are near the target value threshold level (25 μgm^{-3}) this can lead to large differences in estimated exceedances.

With this high sensitivity also comes high uncertainty. Uncertainties in both of the concentration maps, estimated from the cross-validation RMSE in Tables 5 and 6 are around 20-25%. This estimated uncertainty is on a grid to grid basis and will have less impact on the national and European scale since a significant part of the uncorrelated uncertainty is removed through aggregation (averaging or summation over a country). A method for calculating the aggregated uncertainty for the population weighted concentration, based on the kriging variance as the most appropriate indicator of spatial uncertainty, is described in Appendix 6. In its application we assume that the uncertainty in the population density is negligible compared to the uncertainty of the concentration maps. Because there is a degree of correlation, specified by the variogram used for the kriging interpolation, the aggregated uncertainty will depend on the area over which the uncertainty is calculated. In general this leads to lower uncertainties for larger areas. The reduction of uncertainty through aggregation will depend on the partial sill values (indicating the amount of spatial variance) as well as on the range of the variogram (indicating correlation distance). Shorter correlation distances will lead to a quicker reduction in the uncertainty with increasing spatial aggregation.

We assess the aggregated uncertainty of the population weighted concentration per country for both the years 2007 and 2008, using the pseudo $PM_{2.5}$ concentration and uncertainty maps (Figures 8, 11, 12 and 15). The method is described in Appendix 6 and the variogram parameters needed for the calculation are given in Appendix 3, Table A3.5. The relative uncertainties range from 5 – 25% dependent on the country, with smaller countries having the largest uncertainty (similar to the single grid uncertainty).

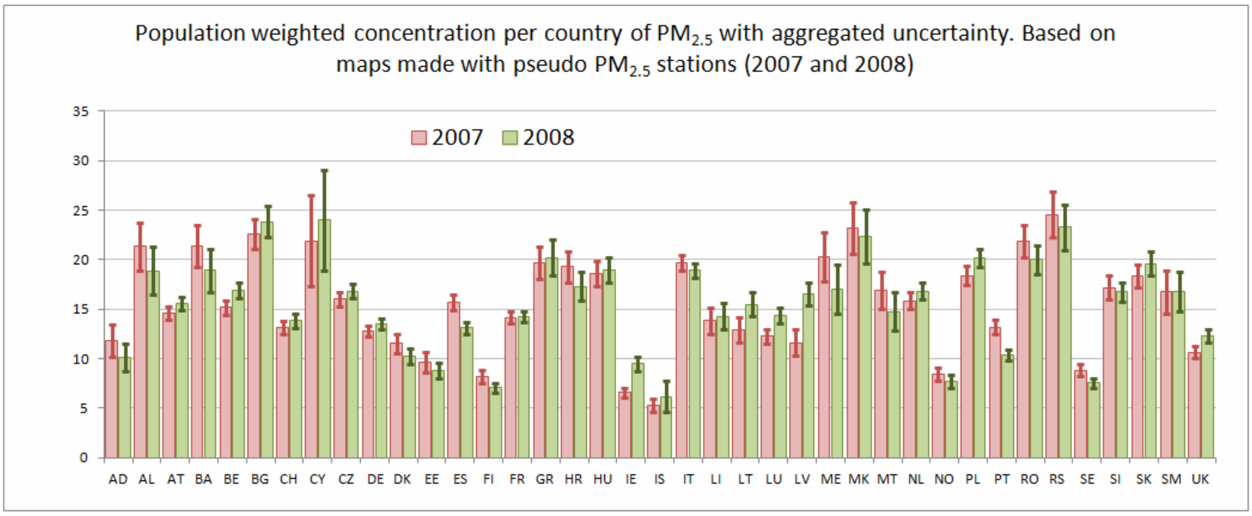


Figure 20. Estimated population weighted concentration of $PM_{2.5}$ (μgm^{-3}) for the year 2007 and 2008, shown country for country using the pseudo $PM_{2.5}$ maps (as in Figure 16 and 17). The error bars indicate the standard deviation (± SD) of the aggregated uncertainty as described in the text and Appendix 6.

The uncertainty of the exceedance exposure is more complex as it is dependent on threshold levels and only areas close to these threshold levels will be sensitive to any uncertainty. Currently a methodology for calculating the exceedance exposure uncertainty is not implemented, but an estimate can be made following the approach outlined in Denby et al. (2008). In that paper it was discussed that it is the correlated uncertainty (bias) that has most impact on the exceedance levels, rather than the uncorrelated uncertainty since this quickly reduces with aggregation. As an estimate of the correlated uncertainty per country we use the calculated population weighted aggregated uncertainty shown in Figure 20. We perturb the calculated concentration fields in each country by this aggregated uncertainty (± SD) and recalculate the exceedance exposure based on these perturbations. This provides an indication of the uncertainty of the exceedance exposure per country by providing a high and low estimate. The results are shown (Figure 21) for both the years 2007 and 2008, where the pseudo $PM_{2.5}$ maps have been applied. This method is indicative only as a thorough uncertainty assessment, e.g. using Monte Carlo methods, would need to be carried out on the $PM_{2.5}$ concentration and uncertainty maps.

In general the relative uncertainty for the exceedance exposure is significantly larger than that for the population weighted concentrations. For several countries, Italy, Poland, Romania, Bulgaria and Serbia the number of people exposed above the target value is significantly higher than the uncertainty. For other countries such as Albania, Bosnia and Herzegovina, Cyprus and Greece the uncertainty is larger than the number of people exposed above the target value.

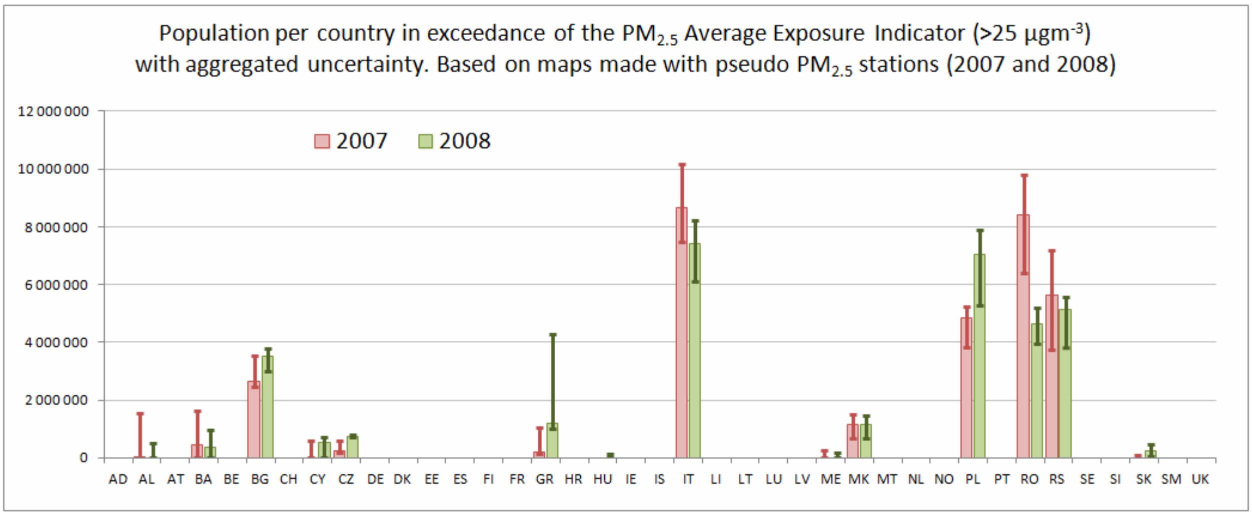


Figure 21. Estimated population exposed to $PM_{2.5}$ concentrations, as total population, above the target value (25 μgm^{-3}) for the year 2007 and 2008, shown country for country. The error bars indicate low and high estimates based on the perturbation (±SD) taken from the population weighted aggregated uncertainty given in Figure 20. The best estimate is given as bars and is the same as that shown in Figures 18 and 19. The pseudo $PM_{2.5}$ map is used for the calculation.

9 Discussion and conclusion

In this study we have presented a method for estimating 'pseudo' $PM_{2.5}$ annual mean concentrations, based on observed PM_{10} concentrations and a set of supplementary data (latitude, longitude, population and accumulated solar radiation). The method uses multiple linear regression of the PM_{10} and supplementary data with available and concurrent $PM_{2.5}$ observations. A set of pseudo $PM_{2.5}$ stations and concentrations are then derived using the established regression coefficients at existing PM_{10} station sites. This increases the number of available sites for interpolation of $PM_{2.5}$ by a factor of approximately seven. The comparison of the pseudo $PM_{2.5}$ with observed $PM_{2.5}$ indicates that the method satisfies the assessment quality objects as laid out in the European AQ Directive when the different requirements for assessment at different concentration levels are taken into account. As such they are an acceptable methodology for determining $PM_{2.5}$ concentrations for Directive purposes. However, the pseudo $PM_{2.5}$ data does not satisfy the quality objectives for fixed monitoring alone. The indicative uncertainty (normalised root mean square error) of the pseudo $PM_{2.5}$ data is approximately 15 – 18%.

The derived pseudo $PM_{2.5}$ stations have then been used to compensate for the insufficient number of existing $PM_{2.5}$ stations to enhance the spatial interpolation of these data, using log-normal and residual kriging as the interpolation method. This has been carried out to produce maps of annual mean $PM_{2.5}$ concentration for all of Europe. The rural and (sub)urban stations have been interpolated separately. For the rural interpolation residual kriging was applied after subtracting a background spatial regression based on supplementary data from the EMEP model, wind speed, altitude and population. Logarithmic transformation of the concentrations and population was used. For the (sub)urban interpolation log-normal kriging was found to provide the best interpolation method. The two interpolations are combined, based on a population weighting of the two fields, at 10 x 10 km^2 resolution.

The interpolation was made using two sets of data. The first set included all pseudo and all measured $PM_{2.5}$ stations. The second data set carried out the interpolation using measured $PM_{2.5}$ data only. Cross-validation of the interpolations was made at the observed $PM_{2.5}$ sites only. This showed that inclusion of the pseudo data improved the cross-validation statistical indicators of RMSE and correlation by 12/13% for the rural interpolation and by 10/24% for the (sub)urban interpolation, for the years 2007/2008. This improvement demonstrates quantitatively the improvement in the $PM_{2.5}$ maps with the introduction of the pseudo $PM_{2.5}$ data.

In addition to the statistical assessment carried out, the resulting 10 x 10 km^2 resolution maps made using both pseudo and measured $PM_{2.5}$ are compared to maps made using only measured $PM_{2.5}$. There are a number of areas where clear differences are observed in both years. The differences are mostly in areas where poor coverage of $PM_{2.5}$ is available but where pseudo $PM_{2.5}$ data has a significant impact on the results, or where both pseudo and observed $PM_{2.5}$ stations are not available. However, the improved spatial coverage, the established statistical quality of the pseudo $PM_{2.5}$ stations and the improved interpolation established at existing $PM_{2.5}$ monitoring sites indicate that the maps made using pseudo data provide a better assessment of the spatial distribution of $PM_{2.5}$ in Europe.

Estimation of the uncertainty in the derived maps is obtained from the kriging variance (spatial uncertainty) and from the cross-validation RMSE (indicative uncertainty). The relative cross-validation RMSE is approximately 20%, considering both years and both the (sub)urban and rural maps. The uncertainty maps, based on the kriging variance, indicate some areas where uncertainty in the predictions is large (> 6 μgm^{-3}) but these correspond to areas with high concentrations (> 20 μgm^{-3}) and generally poor station coverage. Areas where station density is highest show lower uncertainties. In general the kriging uncertainties are in the range 15 – 30%. The kriging uncertainty is in addition to the pseudo $PM_{2.5}$ uncertainty of 15 – 18%. However, the pseudo $PM_{2.5}$ uncertainty will also result in an increased spatial variability of the pseudo measurement data used for the mapping. This will automatically be included in the empirical variogram that is used to estimate the kriging variance. As a result a simple addition of the kriging variance and the pseudo $PM_{2.5}$ variance would lead to an overestimation of the uncertainty. We therefore consider the kriging variance to be representative of the $PM_{2.5}$ mapping uncertainty.

The resulting maps have also been used to estimate European wide $PM_{2.5}$ exposure assessment for the years 2007 and 2008. The results indicate that both years show quite similar distributions of exposure levels and target value exceedances (> 25 μgm^{-3}) on both the national and overall European scale. When applying the maps made both with, and without, pseudo stations the resulting population weighted concentrations per country, and Europe wide, were similar (± 2 μgm^{-3}) for almost all countries. However, there were clear differences, particularly in Italy, Greece, Poland, Romania and Albania when the population in exceedance of the target value was assessed. For Europe as a whole the interpolation without pseudo $PM_{2.5}$ stations indicated that 48.5/57.1 million inhabitants were in exceedance of this value for the years 2007/2008. When pseudo data was included in the interpolation the number of exceedances was estimated to be less, i.e. 32.0/33.1 million inhabitants. These are significant differences and reflect the sensitivity of threshold exceedance calculations to small changes in concentrations in densely populated areas close to the threshold value.

An indicative assessment of the uncertainty of the population weighted concentration and exceedance exposure was provided per country. This was based on the aggregated uncertainty derived from the kriging variance. For the population weighted exposure the aggregated uncertainty was generally highest for smaller countries, since increased aggregation reduces the uncorrelated uncertainty. Values between 5 – 25% were estimated. Based on this the uncertainty of the population exposed above the target value was also estimated per country. This indicated a fairly high uncertainty in countries where exceedance exposure exists, but for a number of countries the uncertainty was still significantly less than the exceedance exposure calculated. The current uncertainty assessment of aggregated exceedance exposure is indicative and will require refinement in the future.

The need for pseudo $PM_{2.5}$ stations is the result of a shortage of $PM_{2.5}$ measurements in Europe. This is a direct result of the fact that no target or limit values were defined for $PM_{2.5}$ in the former air quality Directives. The new Directive 2008/50/EC introduced for $PM_{2.5}$ a target value to be met by 2010 and a limit value by 2015 with extended conditional directives on implementation of $PM_{2.5}$ measurements to address these issues. Until a significant increase in $PM_{2.5}$ stations occurs it is recommended, based on this report, to continue using pseudo $PM_{2.5}$ data to provide improved maps of $PM_{2.5}$ annual mean concentrations for Europe.

References

AirBase (2011), European air quality database, http://airbase.eionet.europa.eu/

de Leeuw, F. and J. Horálek (2009) Assessment of the health impacts of exposure to PM2.5 at a European level ETC/ACC Technical Paper 2009/1.
http://acm.eionet.europa.eu/reports/ETCACC_TP_2009_1_European_PM2.5_HIA

Denby, B., G. Gola, F.de Leeuw, P. de Smet, J. Horálek (2011). Calculation of pseudo PM2.5 annual mean concentrations in Europe based on annual mean PM10 concentrations and other supplementary data. ETC/ACC Technical Paper 2010/9. URL:
http://acm.eionet.europa.eu/reports/ETCACC_TP_2010_9_pseudo_PM2.5_stations

Denby B., M. Schaap, A. Segers, P. Builtjes and J. Horálek (2008). Comparison of two data assimilation methods for assessing PM10 exceedances on the European scale. Atmos. Environ. 42, 7122-7134.

Denby B. (2010). Uncertainty when aggregating spatial data, ETC/ACM task 5.3.3.1 discussion paper, Versionv2.1. In preparation. Available on request.

de Smet, P., Horálek, J., Conková, M., Kurfürst, de Leeuw, F. and Denby, B. (2011). European air quality maps of ozone and PM10 for 2008 and their uncertainty analysis. ETC/ACC Technical Paper 2010/10.
http://acm.eionet.europa.eu/reports/ETCACC_TP_2010_10_spatAQmaps_2008

de Smet, P., Horálek, J., Conková, M., Kurfürst, de Leeuw, F. and Denby, B. (2010). European air quality maps of ozone and PM10 for 2007 and their uncertainty analysis. ETC/ACC Technical paper 2009/9.
http://acm.eionet.europa.eu/reports/ETCACC_TP_2009_9_spatAQmaps_2007

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. Official Journal L 152, 11.6.2008, p. 1–44

EMEP 2009. EMEP Status Report 4/09 "Transboundary Particulate Matter in Europe: Status Report 2009" Joint CCC & MSC-W & CEIP & CIAM Report. URL: http://tarantula.nilu.no/projects/ccc/reports/emep4-2009.pdf

EMEP 2010. EMEP Status Report 4/10 "Transboundary Particulate Matter in Europe: Status Report 2010" Joint CCC & MSC-W & CEIP & CIAM Report. URL: http://tarantula.nilu.no/projects/ccc/reports/emep4-2010.pdf

Fiala, J., B. Denby, J. Horálek, P. Kurfürst, P. de Smet, F. de Leeuw and A.Lükewille (2009). Spatial assessment of PM10 and ozone concentrations in Europe (2005). Technical report No 1/2009.
http://www.eea.europa.eu/publications/spatial-assessment-of-PM10-and-ozone-concentrations-in-europe-2005-1

Horálek, J.; Denby, B.; de Smet, P. A. M.; de Leeuw, F. A. A. M.; Kurfürst, P.; Swart, R.; van Noije, T., (2007). Spatial mapping of air quality for European scale assessment. ETC/ACC Technical paper No 2006/6.
http://acm.eionet.europa.eu/reports/ETCACC_TechnPaper_2006_6_Spat_AQ

Horálek, J., de Smet, P. A. M., de Leeuw, F. A. A. M., 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

JRC population density data (2008). EEA version popugrid01v4_1grid.zip of 15 jan 2008.
http://www.eea.europa.eu/data-and-maps/data/ds_resolveuid/30B341A4-A07B-4306-B8CBD7F0A05CEB36

New, M., Lister, D., Hulme, M., Makin, I. (2002). A high-resolution data set of surface climate over global land areas. Climate Research, Vol. 21, pp. 1-25. http://www.cru.uea.ac.uk/cru/data/hrg.htm/; www.cru.uea.ac.uk/cru/data/TH

ORNL (2002). ORNL LandScan 2002 Global Population dataset. http://www.ornl.gov/sci/landscan/landscanCommon/landscan02_release.html

Simpson D, Fagerli H, Jonson JE, Tsyro S, Wind P, Tuovinen J-P (2003). Transboundary acidification and eutrophication and ground level ozone in Europe: Unified EMEP model description. EMEP Status Report 1/03 Part I. MNP, Oslo, Norway. http://www.emep.int/publ/reports/2003/emep_report_1_part1_2003.pdf

Appendix 1 Tables of results for pseudo PM2.5 exploration

This appendix provides the tables on which the selection of the supplementary data sources for pseudo $PM_{2.5}$ generation took place. Best scoring statistical indicators are shown as red, the next as orange and the next as green. A total 'indicative' colour coding is given in the column 'Supplementary data' to indicate the total best scoring supplementary data set. These are to some extent subjective and are intended to be indicative to help the reader locate the appropriate values in the table.

Note also that the number of stations used depends on the number of concurrent $PM_{2.5}$ and PM_{10} measurements sites. As a result the number of stations used for the exploratory work will always be less than the total number of stations available.

A1.1 Year 2007

Table A1.1. Results of the exploratory study using MLR for all the background stations in the year 2007.

All BACKGROUND stations 2007: n = 129, observed mean = 15.48 μg/m³						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
1	PM10	0.000	3.329	0.783	0.798	0.891
2	PM10 + lat + lon	0.000	2.929	0.832	0.845	0.946
3	PM10 + pop	0.000	3.256	0.792	0.837	0.930
4	PM10 + lat + lon + pop	0.000	2.892	0.836	0.853	0.953
5	PM10 + lat + lon + pop + alt	0.000	2.888	0.836	0.860	0.953
6	PM10 + lat + lon + sun (meteo)	0.000	2.874	0.838	0.814	0.922
7	PM10 + lat + lon + pop + sun (meteo)	0.000	2.828	0.843	0.853	0.953
8	PM10 + lat + lon + pop + sun (clim)	0.000	2.765	0.850	0.837	0.946
9	PM10 + lat + lon + pop + sun (meteo) + wind (meteo)	0.000	2.804	0.846	0.837	0.946

Table A1.2. As in Table A1.1 but for rural stations only.

RURAL BACKGROUND stations 2007: n = 44, observed mean = 11.91 μg/m³						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
1	PM10	0.000	2.667	0.776	0.886	0.977
2	PM10 + lat + lon	0.000	2.408	0.817	0.841	0.955
3	PM10 + pop	0.000	2.666	0.776	0.864	0.955
4	PM10 + lat + lon + pop	0.000	2.407	0.817	0.841	0.955
5	PM10 + lat + lon + pop + alt	0.000	2.406	0.817	0.841	0.955
6	PM10 + lat + lon + sun (meteo)	0.000	2.298	0.833	0.841	0.955
7	PM10 + lat + lon + pop + sun (meteo)	0.000	2.273	0.837	0.864	0.955
8	PM10 + lat + lon + pop + sun (clim)	0.000	2.248	0.841	0.818	0.977
9	PM10 + lat + lon + pop + sun (meteo) + wind (meteo)	0.000	2.267	0.838	0.841	0.950

Table A1.3. As in Table A1.1 but for urban stations only.

(SUB)URBAN BACKGROUND stations 2007: n = 85, observed mean = 17.33 µg/m³						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
1	PM10	0.000	3.602	0.745	0.788	0.859
2	PM10 + lat + lon	0.000	3.021	0.821	0.894	0.965
3	PM10 + pop	0.000	3.471	0.764	0.824	0.918
4	PM10 + lat + lon + pop	0.000	2.980	0.826	0.882	0.953
5	PM10 + lat + lon + pop + alt	0.000	2.972	0.827	0.882	0.953
6	PM10 + lat + lon + sun (meteo)	0.000	2.968	0.827	0.859	0.929
7	PM10 + lat + lon + pop + sun (meteo)	0.000	2.926	0.832	0.871	0.941
8	PM10 + lat + lon + pop + sun (clim)	0.000	2.859	0.840	0.894	0.953
9	PM10 + lat + lon + pop + sun (meteo) + wind (meteo)	0.000	2.873	0.838	0.859	0.941

Table A1.4. As in Table A1.1 but where the regression has been carried out separately for the rural and (sub)urban stations using the selected supplementary data. 7a = one single regression, as in Table A1.1 row 7: 7b = separate urban and rural regression.

(SUB)URBAN + RURAL BACKGROUND stations 2007 using: PM10 + lat + lon + pop + sun (meteo) **n = 129, observed mean = 15.48 µg/m³**						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
7a	PM10 + lat + lon + pop + sun (meteo)	0.000	2.828	0.843	0.853	0.953
7b	PM10 + lat + lon + pop + sun (meteo)	0.000	2.721	0.855	0.868	0.946

A1.2 Year 2008

Table A1.5. Results of the exploratory study using MLR for all the background stations in the year 2008.

All BACKGROUND stations 2008: n = 178, observed mean = 15.86 µg/m³						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
1	PM10	0.000	3.091	0.825	0.820	0.944
2	PM10 + lat + lon	0.000	2.858	0.850	0.831	0.955
3	PM10 + pop	0.000	3.090	0.825	0.831	0.955
4	PM10 + lat + lon + pop	0.000	2.857	0.851	0.831	0.955
5	PM10 + lat + lon + pop + alt	0.000	2.856	0.851	0.831	0.961
6	PM10 + lat + lon + sun (meteo)	0.000	2.410	0.894	0.882	0.955
7	PM10 + lat + lon + pop + sun (meteo)	0.000	2.404	0.894	0.899	0.966
8	PM10 + lat + lon + pop + sun (clim)	0.000	2.584	0.878	0.871	0.961
9	PM10 + lat + lon + pop + sun (meteo) + wind (meteo)	0.000	2.392	0.895	0.893	0.966

Table A1.6. As in Table A1.5 but for rural stations only.

RURAL BACKGROUND stations 2008: n = 55, observed mean = 11.25 µg/m³						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
1	PM10	0.000	3.333	0.724	0.618	0.855
2	PM10 + lat + lon	0.000	3.059	0.768	0.673	0.891
3	PM10 + pop	0.000	3.194	0.747	0.636	0.855
4	PM10 + lat + lon + pop	0.000	2.991	0.778	0.691	0.891
5	PM10 + lat + lon + pop + alt	0.000	2.984	0.779	0.709	0.891
6	PM10 + lat + lon + sun (meteo)	0.000	2.583	0.834	0.745	0.964
7	PM10 + lat + lon + pop + sun (meteo)	0.000	2.563	0.837	0.800	0.964
8	PM10 + lat + lon + pop + sun (clim)	0.000	2.808	0.804	0.745	0.927
9	PM10 + lat + lon + pop + sun (meteo) + wind (meteo)	0.000	2.537	0.840	0.836	0.964

Table A1.7. As in Table A1.5 but for urban stations only.

(SUB)URBAN BACKGROUND stations 2008: **n = 123, observed mean = 17.92 μg/m³**						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
1	PM10	0.000	2.831	0.831	0.878	0.959
2	PM10 + lat + lon	0.000	2.718	0.844	0.878	0.959
3	PM10 + pop	0.000	2.794	0.835	0.894	0.967
4	PM10 + lat + lon + pop	0.000	2.699	0.846	0.886	0.967
5	PM10 + lat + lon + pop + alt	0.000	2.699	0.846	0.886	0.967
6	PM10 + lat + lon + sun (meteo)	0.000	2.309	0.887	0.943	0.959
7	PM10 + lat + lon + pop + sun (meteo)	0.000	2.283	0.890	0.951	0.976
8	PM10 + lat + lon + pop + sun (clim)	0.000	2.423	0.876	0.943	0.976
9	PM10 + lat + lon + pop + sun (meteo) + wind (meteo)	0.000	2.277	0.890	0.951	0.976

Table A1.8. As in Table A1.5 but where the regression has been carried out separately for the rural and (sub)urban stations using the selected supplementary data. 7a = one single regression, as in Table A1.6 row 7; 7b = separate urban and rural regression.

(SUB)URBAN + RURAL BACKGROUND stations 2008 using: PM10 + lat + lon + pop + sun (meteo) **n = 178, observed mean = 15.86 μg/m³**						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
7a	PM10 + lat + lon + pop + sun (meteo)	0.000	2.404	0.894	0.899	0.966
7b	PM10 + lat + lon + pop + sun (meteo)	0.000	2.373	0.897	0.904	0.972

A1.3 Correlation matrix for the selected supplementary data

The following table indicates the correlation between the various supplementary data parameters. There is a high correlation between PM_{10} and $PM_{2.5}$ and also a high negative correlation between latitude and accumulated radition (sun).

Table A1.9. Correlation matrix of the supplementary data used in the selected regression

Parameter	**PM25**	**Lat**	**Lon**	**Pop**	**Sun(meteo)**	**PM10**
2007						
PM25	1	-0.13	0.42	0.047	0.079	0.88
Lat	-0.13	1	0.44	0.049	-0.92	-0.22
Lon	0.42	0.44	1	-0.067	-0.38	0.24
Population	0.047	0.049	-0.067	1	-0.089	0.16
Sun(meteo)	0.079	-0.92	-0.38	-0.089	1	0.17
PM10	0.88	-0.22	0.24	0.16	0.17	1
2008						
PM25	1	-0.015	0.42	0.13	-0.12	0.90
Lat	-0.015	1	0.48	0.039	-0.92	-0.16
Lon	0.42	0.48	1	0.0072	-0.43	0.31
Pop	0.13	0.039	0.0072	1	-0.073	0.16
Sun	-0.12	-0.92	-0.43	-0.073	1	0.083
PM10	0.90	-0.16	0.31	0.16	0.083	1

A1.4 Spatial distribution of measurement sites

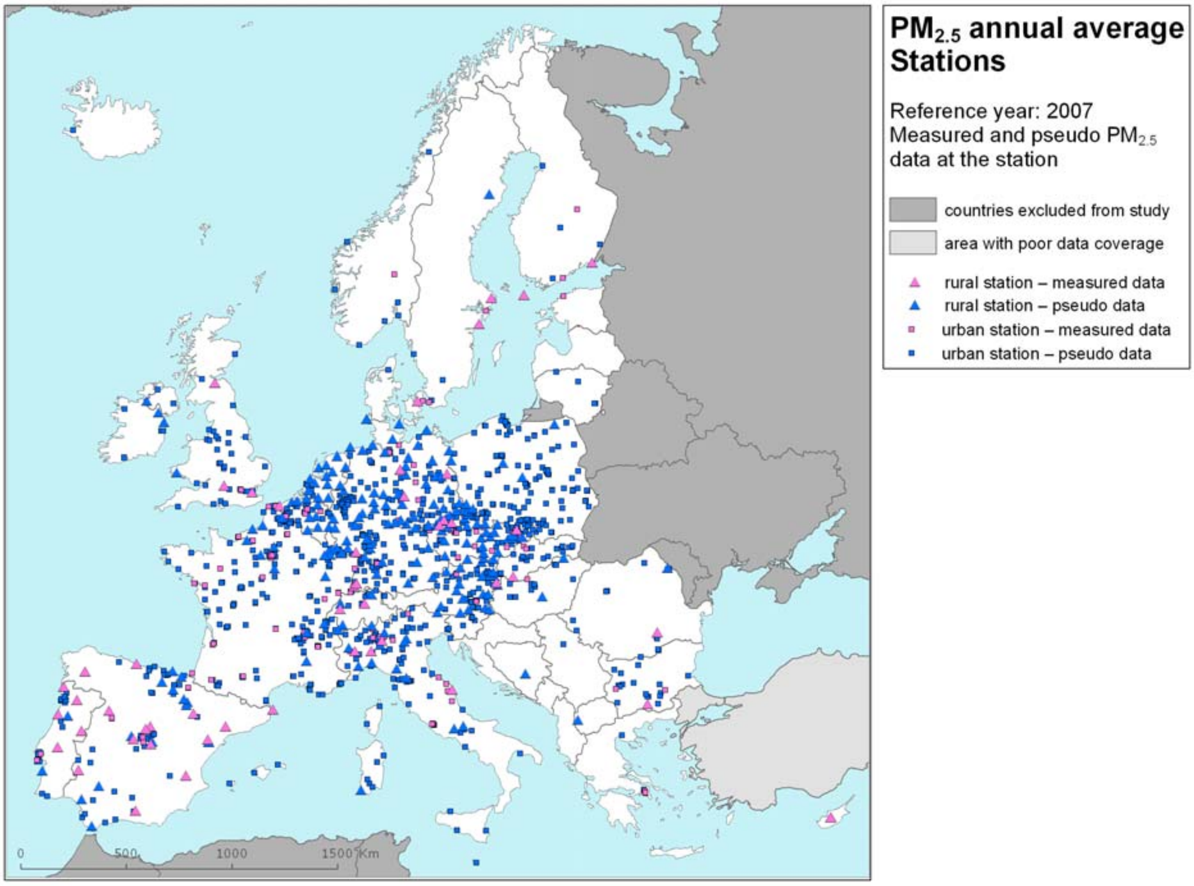


Figure A1.1. Spatial distribution of observed $PM_{2.5}$ and pseudo $PM_{2.5}$ sites (based on positions of PM_{10} only sites), year 2007.

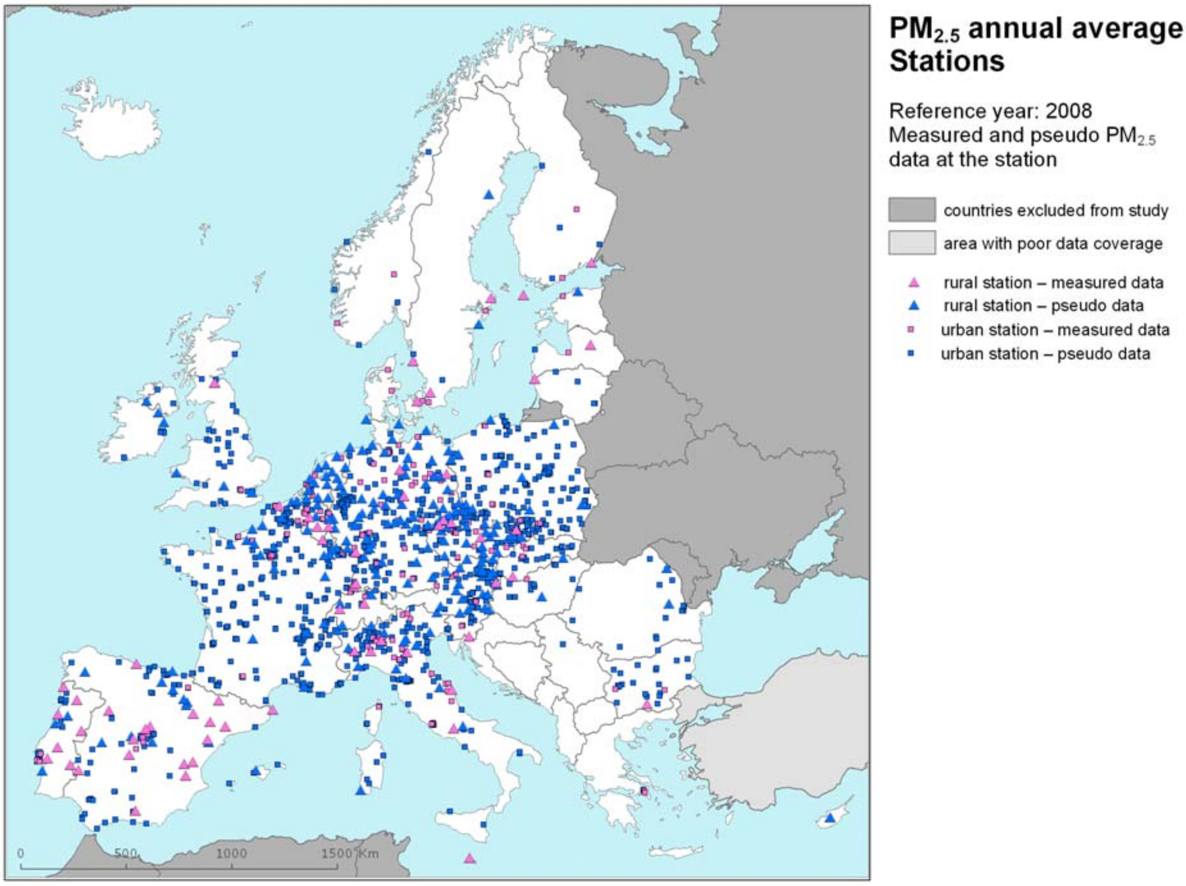


Figure A1.2. Spatial distribution of observed $PM_{2.5}$ and pseudo $PM_{2.5}$ sites (based on positions of PM_{10} only sites), year 2008.

Appendix 2 Tables of results for mapping regression exploration

In this appendix the tables on which the exploration of the supplementary data sources for $PM_{2.5}$ mapping regression are provided. Best scoring statistical indicators are shown as red, the next as orange and the next as green. These are to some extent subjective and are intended to be indicative to help the reader locate the appropriate values in the table. No 'best' supplementary data set is indicated as this selection occurs only after the interpolation (Appendix 3).

A2.1 Year 2007

The following tables show the results of the exploratory activity for the year 2007.

Table A2.1. Results of the exploratory study using MLR for the rural background stations only in the year 2007.

RURAL BACKGROUND stations 2007: n = 253, observed mean = 12.73 μg/m³						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
1	EMEP	0	3.885	0.272	0.632	0.858
2	EMEP + alt	0	3.865	0.28	0.632	0.862
3	EMEP + log(pop)	0	3.837	0.29	0.652	0.858
4	EMEP + sun(meteo)	0	3.826	0.294	0.632	0.866
5	EMEP + wind(meteo)	0	3.676	0.35	0.656	0.862
6	EMEP + sun(meteo) + wind(meteo)	0	3.676	0.349	0.664	0.862
7	EMEP + sun(meteo) + wind(meteo) + alt + log(pop)	0.001	3.415	0.438	0.715	0.901
8	EMEP + wind(meteo) + alt + log(pop)	0.001	3.423	0.436	0.688	0.893
9	EMEP + sun(meteo) + alt + log(pop)	0	3.693	0.342	0.656	0.877
10	EMEP + sun(meteo) + wind(meteo) + log(pop)	0	3.586	0.382	0.688	0.889
11	EMEP + sun(meteo) + wind(meteo) + alt	0.001	3.434	0.431	0.68	0.877

Table A2.2. Results of the exploratory study using MLR for the urban background stations only in the year 2007.

URBAN BACKGROUND stations 2007: n = 928, observed mean = 16.92 μg/m³						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
1	EMEP	-0.002	5.201	0.184	0.6	0.745
2	EMEP + alt	-0.002	5.045	0.233	0.605	0.73
3	EMEP + log(pop)	-0.002	5.182	0.19	0.605	0.744
4	EMEP + sun(meteo)	-0.002	5.115	0.211	0.612	0.747
5	EMEP + wind(meteo)	-0.002	4.855	0.289	0.621	0.741
6	EMEP + sun(meteo) + wind(meteo)	-0.002	4.852	0.29	0.62	0.741
7	EMEP + sun(meteo) + wind(meteo) + alt + log(pop)	-0.002	4.844	0.292	0.635	0.749
8	EMEP + wind(meteo) + alt + log(pop)	-0.002	4.848	0.291	0.63	0.745
9	EMEP + sun(meteo) + alt + log(pop)	-0.002	4.991	0.249	0.621	0.739
10	EMEP + sun(meteo) + wind(meteo) + log(pop)	-0.002	4.851	0.29	0.633	0.751
11	EMEP + sun(meteo) + wind(meteo) + alt	-0.002	4.844	0.292	0.623	0.738

A2.2 Year 2008

Table A2.3. Results of the exploratory study using MLR for the rural background stations only in the year 2008.

RURAL BACKGROUND stations 2008: n = 278, observed mean = 13.02 μg/m^3						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
1	EMEP	0.003	3.981	0.352	0.612	0.853
2	EMEP + alt	0.003	3.977	0.353	0.615	0.853
3	EMEP + log(pop)	0.003	3.913	0.374	0.608	0.827
4	EMEP + sun(meteo)	0.003	3.982	0.351	0.608	0.853
5	EMEP + wind(meteo)	0.004	3.786	0.414	0.64	0.849
6	EMEP + sun(meteo) + wind(meteo)	0.004	3.738	0.429	0.658	0.863
7	EMEP + sun(meteo) + wind(meteo) + alt + log(pop)	0.004	3.577	0.477	0.676	0.867
8	EMEP + wind(meteo) + alt + log(pop)	0.004	3.613	0.467	0.673	0.863
9	EMEP + sun(meteo) + alt + log(pop)	0.003	3.91	0.375	0.612	0.827
10	EMEP + sun(meteo) + wind(meteo) + log(pop)	0.003	3.659	0.452	0.673	0.874
11	EMEP + sun(meteo) + wind(meteo) + alt	0.005	3.614	0.467	0.647	0.863

Table A2.4. Results of the exploratory study using MLR for the urban background stations only in the year 2008.

URBAN BACKGROUND stations 2008: n = 1025, observed mean = 17.65 μg/m^3						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
1	EMEP	-0.002	5.529	0.211	0.612	0.728
2	EMEP + alt	-0.002	5.318	0.27	0.638	0.739
3	EMEP + log(pop)	-0.002	5.519	0.213	0.622	0.734
4	EMEP + sun(meteo)	-0.002	5.528	0.211	0.61	0.725
5	EMEP + wind(meteo)	-0.001	5.309	0.272	0.647	0.74
6	EMEP + sun(meteo) + wind(meteo)	-0.001	5.277	0.281	0.655	0.742
7	EMEP + sun(meteo) + wind(meteo) + alt + log(pop)	-0.001	5.227	0.294	0.66	0.747
8	EMEP + wind(meteo) + alt + log(pop)	-0.001	5.246	0.289	0.656	0.745
9	EMEP + sun(meteo) + alt + log(pop)	-0.002	5.317	0.27	0.635	0.737
10	EMEP + sun(meteo) + wind(meteo) + log(pop)	-0.001	5.276	0.281	0.651	0.74
11	EMEP + sun(meteo) + wind(meteo) + alt	-0.001	5.228	0.294	0.657	0.745

Appendix 3 Tables of results for kriging exploration

In this appendix the tables on which the exploration of the supplementary data sources for $PM_{2.5}$ interpolation (both log-normal and residual kriging) are provided. Best scoring statistical indicators are shown as red, the next as orange and the next as green. A total 'indicative' colour coding is given in the column 'Supplementary data' to indicate the total best scoring supplementary data set. These are to some extent subjective and are intended to be indicative to help the reader locate the appropriate values in the table.

A3.1 Rural background mapping

The following tables show the results of the exploratory assessment for the year 2007 and 2008.

Table A3.1. Results of the exploratory study using residual kriging for all the background stations in the year 2007.

RURAL BACKGROUND stations 2007: n = 253, observed mean = 12.73 μg/m^3						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
0	Log-normal kriging	0.003	3.385	0.447	0.751	0.874
1	EMEP	0.005	3.164	0.519	0.775	0.885
7	EMEP + sun(meteo) + wind(meteo) + alt + log(pop)	0.006	2.821	0.618	0.806	0.913
8	EMEP + wind(meteo) + alt + log(pop)	0.005	2.836	0.613	0.794	0.909
10	EMEP + sun(meteo) + wind(meteo) + log(pop)	0.004	2.964	0.576	0.798	0.905
11	EMEP + sun(meteo) + wind(meteo) + alt	0.007	2.907	0.596	0.791	0.905
12	EMEP + alt + log(pop)	0.002	2.952	0.58	0.806	0.897
13	EMEP + wind(meteo) + alt	0.006	2.912	0.594	0.794	0.909
14	EMEP + wind(meteo) + log(pop)	0.003	2.966	0.576	0.794	0.905

Table A3.2. Results of the exploratory study using residual kriging for all the background stations in the year 2008.

RURAL BACKGROUND stations 2008: n = 278, observed mean = 13.03 μg/m^3						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
0	Log-normal kriging	0.004	3.231	0.573	0.748	0.924
1	EMEP	0.008	3.136	0.605	0.741	0.921
7	EMEP + sun(meteo) + wind(meteo) + alt + log(pop)	0.005	2.84	0.673	0.766	0.917
8	EMEP + wind(meteo) + alt + log(pop)	0.006	2.833	0.676	0.763	0.924
10	EMEP + sun(meteo) + wind(meteo) + log(pop)	0.004	2.908	0.656	0.777	0.928
11	EMEP + sun(meteo) + wind(meteo) + alt	0.006	3.012	0.636	0.748	0.903
12	EMEP + alt + log(pop)	0.006	2.863	0.667	0.752	0.914
13	EMEP + wind(meteo) + alt	0.007	3.01	0.637	0.759	0.914
14	EMEP + wind(meteo) + log(pop)	0.006	2.901	0.658	0.766	0.921

A3.2 (Sub)urban background mapping

The following tables show the results of the exploratory assessment for the year 2007 and 2008.

Table A3.3. Results of the exploratory study using residual kriging for all the background stations in the year 2007.

URBAN BACKGROUND stations 2007: n = 928, observed mean =16.92 μg/m^3						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
0	Log-normal kriging	0.001	2.977	0.733	0.88	0.93
1	EMEP	0	3.019	0.725	0.874	0.923
7	EMEP + sun(meteo) + wind(meteo) + alt + log(pop)	-0.001	3.058	0.718	0.869	0.92
8	EMEP + wind(meteo) + alt + log(pop)	0	3.064	0.717	0.87	0.923
10	EMEP + sun(meteo) + wind(meteo) + log(pop)	0	3.055	0.719	0.863	0.918
11	EMEP + sun(meteo) + wind(meteo) + alt	-0.001	3.03	0.723	0.871	0.922

Table A3.4. Results of the exploratory study using residual kriging for all the background stations in the year 2008.

URBAN BACKGROUND stations 2008: n = 1025, observed mean = 17.65 μg/m^3						
Ref #	**Supplementary data**	**FB**	**RMSE**	**R^2**	**FAC25%**	**ECQO**
0	Log-normal kriging	-0.001	3.231	0.731	0.861	0.911
1	EMEP	0.002	3.35	0.71	0.857	0.902
7	EMEP + sun(meteo) + wind(meteo) + alt + log(pop)	0.001	3.385	0.704	0.85	0.897
8	EMEP + wind(meteo) + alt + log(pop)	0.001	3.394	0.703	0.848	0.895
10	EMEP + sun(meteo) + wind(meteo) + log(pop)	0	3.37	0.707	0.851	0.899
11	EMEP + sun(meteo) + wind(meteo) + alt	0.001	3.393	0.703	0.845	0.894

A3.3 Semi-variogram parameters

In the rural residual kriging and the (sub)urban log-normal kriging semi-variogram parameters were determined by fitting (using a least squares method) a spherical semi-variogram to the empirical semi-variogram data. The results for the two years are shown in table A3.5 below. The fitted parameters are consistent between the two years.

Table A3.5. Semi-variogram parameters used in the selected kriging procedure using the observed and pseudo $PM_{2.5}$ data. Note that the variance for lag distances greater than the range is the sum of the sill and the nugget.

Application	**Range (km)**	**Partial sill (log(μg/m^3)^2)**	**Nugget (log(μg/m^3)^2)**
Rural residual kriging 2007	655	0.048	0.022
Rural residual kriging 2008	588	0.046	0.025
Urban log-normal kriging 2007	890	0.068	0.022
Urban log-normal kriging 2008	856	0.065	0.025

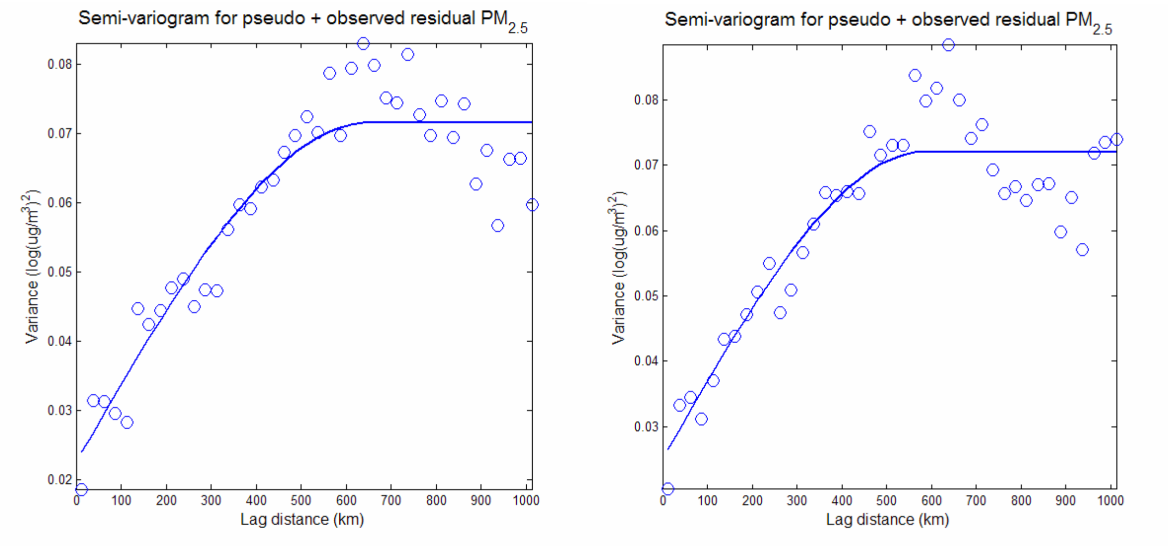


Figure A3.1 Semi-variograms derived for the residual kriging of rural stations. Observed values are shown as circles and the fitted spherical variogram is shown as a solid line. The fitted kriging parameters are given in Table A3.5. The two years, 2007 (left) and 2008 (right), are shown.

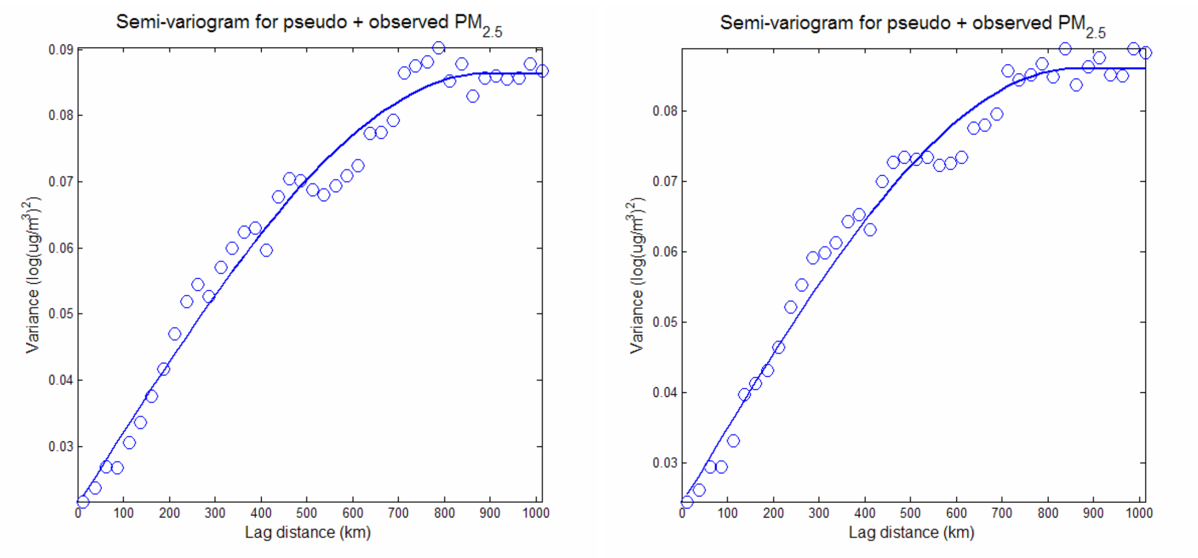


Figure A3.2 Semi-variograms derived for the log-normal kriging of (sub)urban stations. Observed values are shown as circles and the fitted spherical variogram is shown as a solid line. The fitted kriging parameters are given in Table A3.5. The two years, 2007 (left) and 2008 (right), are shown.

Appendix 4 Table of variogram parameters used for mapping

In this appendix the variogram parameters for the mapping (Section 7) are provided. These are the parameters used in the ArcGIS software for the interpolation.

Table A4.1. Variogram parameters used for the mapping shown in figures 7-8. Year 2007

Year 2007			
Application	**Range (km)**	**Partial sill ($(\log(\mu g/m^3)^2)$)**	**Nugget ($\log(\mu g/m^3)^2$)**
Residual kriging, rural, meas. + pseudo	588	0.046	0.025
Residual kriging, rural, measured only	588	0.058	0.025
Log-normal kriging, urban, meas. + pseudo	856	0.065	0.025
Log-normal kriging, urban, measured only	856	0.097	0.025
Residual kriging, joint, meas. + pseudo	588	0.043	0.025
Residual kriging, joint, measured only	588	0.043	0.025

Table A4.2. Variogram parameters used for the mapping shown in figures 12-13. Year 2008

Year 2008			
Application	**Range (km)**	**Partial sill ($(\log(\mu g/m^3)^2)$)**	**Nugget ($(\log(\mu g/m^3)^2)$)**
Residual kriging, rural, meas. + pseudo	588	0.046	0.025
Residual kriging, rural, measured only	588	0.127	0.025
Log-normal kriging, urban, meas. + pseudo	856	0.065	0.025
Log-normal kriging, urban, measured only	856	0.086	0.025
Residual kriging, joint, meas. + pseudo	588	0.043	0.025
Residual kriging, joint, measured only	588	0.043	0.025

Appendix 5 Country based exposure tables

The following tables provide estimates of population exposure based on the maps provided in Figures 7, 8, 12 and 13. These show:

- Table A5.1: Exposure table for 2007 including pseudo $PM_{2.5}$ stations (Figure 7)
- Table A5.2: Exposure table for 2007 using only observed $PM_{2.5}$ (Figure 8)
- Table A5.3: Exposure table for 2008 including pseudo $PM_{2.5}$ stations (Figure 12)
- Table A5.4: Exposure table for 2008 using only observed $PM_{2.5}$ (Figure 13)

Table A5.1. Country based exposure tables showing the population, the percentage of population exposed in particular concentration ranges and the population weighted average concentration based on the 10 x 10 km² maps. Tables are based on the combined pseudo and measured $PM_{2.5}$ concentrations (Figure 7) for the year 2007. The target value of 25 μgm^{-3} is to be met by 2010, which is set as a limit value to be met by 2015.

Country	Population UN 2007	PM2.5 – avg, exposed population, 2007[Percentage]						PM2.5 - avg07
		< 5 $\mu g/m^3$	5 – 10 $\mu g/m^3$	10 – 15 $\mu g/m^3$	15 – 25 $\mu g/m^3$	25 – 30 $\mu g/m^3$	> 30 $\mu g/m^3$	Pop. weighted
			< Target value			> Target value		
AD	61,237		30.6	69.4				11.8
AL	3,927,150			1.5	97.3	1.2		21.3
AT	8,270,625	0.0	4.8	47.7	47.6			14.6
BA	4,175,354			2.8	86.8	10.4		21.4
BE	10,579,323		1.7	36.1	62.2			15.2
BG	7,981,501		0.4	15.4	51.1	8.1	24.9	22.6
CH	7,237,927		5.6	90.0	4.4			13.1
CY	851,996			0.3	99.7			21.9
CZ	10,164,426		1.1	39.1	57.4	2.4		16.0
DE	82,110,675		5.3	88.7	6.0			12.8
DK	5,414,545	0.2	8.6	91.2				11.6
EE	1,335,153	0.0	55.3	44.7				9.7
ES	38,991,635	0.0	7.6	31.8	60.6			15.7
FI	5,129,224	1.5	85.1	13.4				8.2
FR	58,495,208		1.9	64.2	33.9			14.1
GR	10,966,809		0.1	11.1	87.0	1.8		19.7
HR	4,464,134			5.0	95.0			19.3
HU	10,127,814			0.2	99.8			18.6
IE	3,729,873	2.1	97.9					6.6
IS	178,476	42.2	57.8					5.3
IT	56,793,542	0.0	1.2	12.7	70.8	9.2	6.1	19.7
LI	66,973			100.0				13.8
LT	3,469,054		0.7	71.2	28.2			12.9
LU	425,239		8.7	91.3				12.3
LV	2,383,168		20.7	79.3				11.6
ME	713,492			6.3	93.7			20.3
MK	2,275,108			7.7	42.2	50.1		23.2
MT	394,641			13.7	86.3			16.9
NL	15,729,237		0.1	23.0	76.9			15.9
NO	3,187,004	9.6	57.8	32.5				8.4
PL	38,222,745		0.3	21.6	65.4	10.5	2.1	18.4
PT	9,906,444	0.1	16.9	39.3	43.7			13.2
RO	22,428,492		0.0	5.1	57.3	37.5		21.9
RS	10,735,972			0.6	46.8	47.3	5.3	24.6
SE	8,886,973	4.6	68.8	26.6				8.8
SI	2,030,370		0.0	11.1	88.9			17.2
SK	5,297,625			3.3	96.7			18.3
SM	20,427				100.0			16.8
UK	59,028,708	0.7	30.8	68.4	0.1			10.7
Total	**516,188,303**	**0.3**	**9.3**	**43.3**	**40.8**	**4.9**	**1.3**	**15.5**

Table A5.2. Country based exposure tables showing the population, the percentage of population exposed in particular concentration ranges and the population weighted average concentration based on the 10 x 10 km^2 maps. Tables are based on the measured $PM_{2.5}$ concentrations only (Figure 8) for the year 2007. The target value of 25 μgm^{-3} is to be met by 2010, which is set as a limit value to be met by 2015.

Country	Population UN 2007	PM2.5 – avg, exposed population, 2007[Percentage]						PM2.5 - avg07
		< 5 µg/m^3	5 – 10 µg/m^3	10 – 15 µg/m^3	15 – 25 µg/m^3	25 – 30 µg/m^3	> 30 µg/m^3	Pop. weighted
				< Target value		> Target value		
AD	61,237		30.6	69.4				10.9
AL	3,927,150		1.0	17.1	38.3	43.6		21.9
AT	8,270,625	0.6	12.2	32.7	54.6			15.4
BA	4,175,354		7.5	28.9	63.6			17.6
BE	10,579,323		2.1	23.1	74.8			15.7
BG	7,981,501		1.8	26.5	29.2	13.3	28.2	22.7
CH	7,237,927	1.3	10.3	36.5	51.9			14.3
CY	851,996			5.5	57.6	36.9		21.8
CZ	10,164,426		1.4	22.8	68.5	7.3		17.4
DE	82,110,675		4.7	56.0	39.2			14.0
DK	5,414,545	0.2	45.9	53.8				10.2
EE	1,335,153	0.0	57.7	42.2				9.3
ES	38,991,635	0.0	7.9	26.6	60.7	4.8		16.0
FI	5,129,224	4.1	81.3	14.5				7.9
FR	58,495,208	0.0	2.6	71.8	25.6			13.6
GR	10,966,809		1.3	18.9	30.3	40.2	9.3	21.5
HR	4,464,134		1.4	33.6	65.0			16.8
HU	10,127,814			12.6	87.4			17.7
IE	3,729,873	2.2	59.2	38.7				9.4
IS	178,476	42.6	23.9	33.5				6.9
IT	56,793,542	0.1	2.3	15.6	53.3	18.0	10.7	21.2
LI	66,973			42.9	57.1			14.9
LT	3,469,054		26.4	73.6				11.6
LU	425,239		6.8	93.2				13.2
LV	2,383,168		40.6	59.4				10.7
ME	713,492		16.8	26.6	30.3	26.3		17.3
MK	2,275,108		2.5	23.1	12.0	10.5	51.9	24.6
MT	394,641			7.9	92.1			22.3
NL	15,729,237		1.2	19.4	79.4			15.8
NO	3,187,004	17.4	54.6	28.0				8.1
PL	38,222,745		0.6	20.4	66.2	12.8		19.1
PT	9,906,444	0.1	9.8	39.6	50.4			14.9
RO	22,428,492		1.3	20.3	42.1	36.3		20.6
RS	10,735,972		0.9	9.9	49.8	29.3	10.2	22.1
SE	8,886,973	7.2	73.9	18.8				8.2
SI	2,030,370		1.2	44.3	54.5			16.6
SK	5,297,625		0.0	6.7	93.3			18.3
SM	20,427				100.0			16.7
UK	59,028,708	1.0	12.1	86.8	0.1			12.5
Total	**516,188,303**	**0.5**	**8.1**	**41.2**	**40.9**	**7.2**	**2.2**	**16.0**

Table A5.3. Country based exposure tables showing the population, the percentage of population exposed in particular concentration ranges and the population weighted average concentration based on the 10 x 10 km^2 maps. Tables are based on the combined pseudo and measured $PM_{2.5}$ concentrations (Figure 11) for the year 2008. The target value of 25 μgm^{-3} is to be met by 2010, which is set as a limit value to be met by 2015.

Country	Population Eurostat 2008	PM2.5 - avg, exposed population, 2008 [Percentage]						PM2.5 - avg08
		< 5 µg/m^3	5 – 10 µg/m^3	10 – 15 µg/m^3	15 – 25 µg/m^3	25 – 30 µg/m^3	> 30 µg/m^3	Pop. weighted
		< Target value				> Target value		
AD	75,000		30.6	69.4				10.1
AL	3,170,050		0.1	16.6	83.3			18.9
AT	8,318,592	0.0	4.4	40.4	55.2			15.5
BA	3,843,846		0.1	22.8	68.1	9.0		19.0
BE	10,666,866		0.0	6.2	93.8			16.9
BG	7,640,238		0.6	30.4	24.7	13.0	31.3	23.9
CH	7,593,494	0.0	7.6	64.6	27.8			13.9
CY	789,269			0.3	39.2	60.5		24.0
CZ	10,381,130		1.5	30.0	61.2	7.3		16.8
DE	82,217,837		4.6	70.6	24.8			13.5
DK	5,475,791	0.2	40.8	59.0				10.3
EE	1,340,935	0.0	77.8	22.2				8.8
ES	45,283,259	0.2	19.2	47.1	33.5			13.2
FI	5,300,484	4.9	95.1					7.1
FR	64,004,333	0.0	5.4	50.0	44.7			14.3
GR	11,213,785		0.1	19.4	69.6	9.4	1.4	20.2
HR	4,436,401		0.2	19.5	80.3			17.3
HU	10,045,401			0.1	99.9			18.9
IE	4,401,335		69.7	30.3				9.5
IS	315,459	39.8	60.2					6.2
IT	59,619,290	0.0	2.4	19.3	65.3	11.7	1.4	18.9
LI	35,356			100.0				14.3
LT	3,366,357		0.0	26.0	74.0			15.5
LU	483,799			68.9	31.1			14.3
LV	2,270,894		0.4	44.9	54.7			16.5
ME	627,508		1.6	44.0	54.4			17.0
MK	2,045,177		0.2	26.0	23.5	50.4		22.4
MT	410,290			22.6	77.4			14.8
NL	16,405,399		0.2	10.1	89.6			16.8
NO	4,737,171	16.7	62.7	20.7				7.8
PL	38,115,641		0.5	17.5	63.7	10.5	7.9	20.2
PT	10,617,575	1.5	38.6	59.8				10.3
RO	21,528,627		0.3	16.6	62.3	18.2	2.5	20.0
RS	9,518,646		0.1	9.9	42.3	36.7	11.0	23.3
SE	9,182,927	9.2	80.2	10.6				7.6
SI	2,010,269		0.1	29.4	70.6			16.7
SK	5,400,998			0.8	95.0	4.2		19.6
SM	31,000				100.0			16.8
UK	61,191,951	0.2	9.0	82.0	8.9			12.3
Total	**534,112,380**	**0.4**	**9.0**	**41.2**	**43.2**	**4.6**	**1.6**	**15.6**

Table A5.4. Country based exposure tables showing the population, the percentage of population exposed in particular concentration ranges and the population weighted average concentration based on the 10 x 10 km^2 maps. Tables are based on measured $PM_{2.5}$ concentrations only (Figure 12) for the year 2008. The target value of 25 μgm^{-3} is to be met by 2010, which is set as a limit value to be met by 2015.

Country	Population Eurostat 2008	PM2.5 - avg, exposed population, 2008 [Percentage] < 5 µg/m^3	5 – 10 µg/m^3	10 – 15 µg/m^3	15 – 25 µg/m^3	25 – 30 µg/m^3	> 30 µg/m^3	PM2.5 - avg08 Pop. weighted
			< Target value			> Target value		
AD	75,000	5.3	25.3	69.4				10.4
AL	3,170,050		1.1	15.8	39.5	43.6		21.7
AT	8,318,592	0.6	14.2	33.8	51.4			15.3
BA	3,843,846		4.7	26.7	68.6			17.8
BE	10,666,866		0.0	3.5	96.5			17.7
BG	7,640,238		1.1	17.2	37.7	19.8	24.2	23.2
CH	7,593,494	1.0	13.1	26.4	59.5			14.7
CY	789,269		0.1	9.9	53.2	36.9		21.9
CZ	10,381,130		7.0	26.8	56.4	8.9	0.8	16.9
DE	82,217,837		8.3	50.0	41.7			14.4
DK	5,475,791	0.1	47.4	52.4				10.1
EE	1,340,935		69.5	30.5				9.5
ES	45,283,259	0.1	13.2	57.6	29.0			13.6
FI	5,300,484	0.2	99.6	0.2				8.0
FR	64,004,333	0.0	4.6	44.0	51.3			15.0
GR	11,213,785		1.1	16.8	64.7	17.0	0.5	21.2
HR	4,436,401		1.9	25.9	72.2			17.1
HU	10,045,401			2.0	94.5	3.5		19.9
IE	4,401,335	0.3	55.4	6.8	37.5			10.8
IS	315,459	19.7	46.8	33.5				8.8
IT	59,619,290	0.0	2.0	13.5	57.9	18.8	7.8	21.2
LI	35,356			100.0				13.9
LT	3,366,357			3.7	96.3			18.9
LU	483,799			12.4	87.6			16.1
LV	2,270,894		0.0	23.7	54.0	22.3		18.3
ME	627,508		10.5	35.7	34.5	19.3		17.4
MK	2,045,177		1.9	18.2	17.5	56.3	6.1	24.4
MT	410,290		0.0	7.8	14.7	77.4		24.2
NL	16,405,399		0.4	8.4	91.2			17.0
NO	4,737,171	6.2	67.0	26.8				9.2
PL	38,115,641		0.0	4.1	60.3	18.9	16.6	23.4
PT	10,617,575	1.2	33.5	55.7	9.5			11.0
RO	21,528,627		0.3	6.7	48.2	38.9	5.9	23.3
RS	9,518,646		0.4	8.8	49.4	36.2	5.2	22.6
SE	9,182,927	1.1	75.7	23.2				8.9
SI	2,010,269		1.6	36.4	62.0			16.9
SK	5,400,998		0.0	1.4	84.8	13.8		20.9
SM	31,000				100.0			16.7
UK	61,191,951	0.5	10.2	28.4	60.8			13.8
Total	**534,112,380**	**0.2**	**9.4**	**28.6**	**51.1**	**7.8**	**2.9**	**16.9**

Appendix 6 Method for calculating aggregated exposure uncertainty

When assessing the average population exposure over a given area it is necessary to aggregate the available data. If we assume that we have gridded concentration and population data available (in this case at 10 km resolution for all of Europe) at each grid point then the population weighted concentration (C_{pw}) is a straight forward task carried out using the following equation:

$$C_{pw} = \frac{\sum_i^n C_i P_i}{\sum_i^n P_i} \tag{A6.1}$$

Where n is the number of grids in the aggregating area, C_i is the concentration at each grid and P_i is the population at each grid.

In addition to the gridded concentration data we also have an estimate of the uncertainty (variance) at each of the grid points, based on the kriging variance, and we wish to estimate the uncertainty of the aggregated data. However, there is a degree of spatial correlation involved that will affect the calculation of the aggregated uncertainty.

A6.1 Basic concept for determining the aggregated uncertainty

The variance (*var*) of the mean of a spatially varying parameter X is equivalent to the double sum of the covariance matrix (*cov*) as follows:

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_i^n \sum_{i=1,j=1}^{n} \mathrm{cov}\left(X_i, X_j\right) \tag{A6.2}$$

It is helpful to reformulate this in terms of the variance, the diagonal terms of the covariance matrix, and the covariance of the non-diagonal terms as follows:

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_i^n \mathrm{var}\left(X_i\right) + \frac{1}{n^2}\sum_i^n \sum_{i=1,j\neq i}^{n} \mathrm{cov}\left(X_i, X_j\right) \tag{A6.3}$$

For further brevity and clarity we replace the variance and covariance terms in Equation A6.3 with the following:

$$\sigma_i^2 = \mathrm{var}\left(X_i\right) \qquad \text{and} \qquad \rho_{ij}\sigma_i\sigma_j = \mathrm{cov}\left(X_i, X_j\right) \tag{A6.4}$$

Where ρ is the correlation coefficient. Equation A6.3 can then be rewritten as:

$$\sigma^2 = \frac{1}{n^2}\sum_i^n \sigma_i^2 + \frac{1}{n^2}\sum_i^n \sum_{i=1,j\neq i}^{n} \rho_{ij}\sigma_i\sigma_j \tag{A6.5}$$

Equation A6.5 then provides a description of how the variance σ^2 (uncertainty) of the mean of any aggregated parameter (e.g. mean concentration) can be written in terms of individual (gridded) variances and the correlation between these (gridded) variances.

Two special cases arise when the gridded data is totally uncorrelated or totally correlated. In the totally uncorrelated case the variance is written as:

$$\sigma_{uncor}^{2} = \frac{1}{n^{2}} \sum_{i}^{n} \sigma_{i}^{2} \qquad \text{(A6.6)}$$

For the case where the grid point concentrations are totally correlated then the uncertainty would be written as

$$\sigma_{cor}^{2} = \frac{1}{n} \sum_{i}^{n} \sigma_{i}^{2} \qquad \text{(A6.7)}$$

which is the same as the average variance. There is a large difference between these two uncertainty estimates, a factor of *n* where *n* is a large number (number of grid points).

A6.2 Determination of the spatial correlation

There will be a degree of spatial correlation that must be taken into account and so the correlation must be included in the calculation. This is one of the benefits of using kriging methods to determine the concentration at each grid point since not only is the variance at each grid point known but the correlation between each grid point can be determined directly from the variogram that was used for the kriging interpolation.

The spherical variogram, which has been employed in the residual interpolation to create the concentration fields, is given by:

$$\begin{aligned} \gamma(h) &= 0 && h = \delta h \\ &= c_0 + c_1\left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right) && \delta h < h < a \qquad \text{(A6.8)} \\ &= \mathrm{c} && h > a \end{aligned}$$

Where *h* is the lag distance, *a* the variogram range, c_0 the nugget value and c_1 the partial sill value and the total sill value is $c = c_0 + c_1$. Instead of using the variogram as a spatial parameter, the spatial correlation can also be used as follows:

$$\rho(h) = \frac{(c - \gamma(h))}{c} \qquad \text{(A6.9)}$$

By determining the distance (h_{ij}) between two grid points *i* and *j* and applying Equation A6.8 then the correlation ρ_{ij} between the two grid points is provided. Note that $\rho(h) = 0$ for $h > a$ (range) and that the minimum value for the correlation in the off-diagonal elements of the covariance matrix is $\rho(h) = c_0/c$.

A6.3 Addition of weighted uncertainties

In addition to the theory relating to spatial aggregation of data we also write the equation for adding weighted and correlated data:

$$X_w = \mathrm{var}\left(\sum_i^n X_i\, a_i\right) \tag{A6.10}$$

$$\sigma_w^2 = \mathrm{var}\left(\sum_i^n a_i^2 \sigma_i^2\right) = \sum_i^n a_i^2 \sigma_i^2 + \sum_i^n i \sum_{i=1,\, j\neq i}^n \rho_{ij} a_i\, \sigma_i\, a_j \sigma_j \tag{A6.11}$$

Where *a* is the weighting parameter. This equation can be used to add population weighted data from different regions that are population weighted. In the current application this involves combining the regions that are urban, mixed and rural. To carry out this calculation the correlation between the regions is required. This correlation, however, is generally unknown though the correlation is likely to be high. In this regard we assume the worst case, i.e. that the two regions are completely correlated, and Equation A6.11 reduces to:

$$\sigma_w^2 = \sum_i^n a_i\, \sigma_i^2 \tag{A6.12}$$