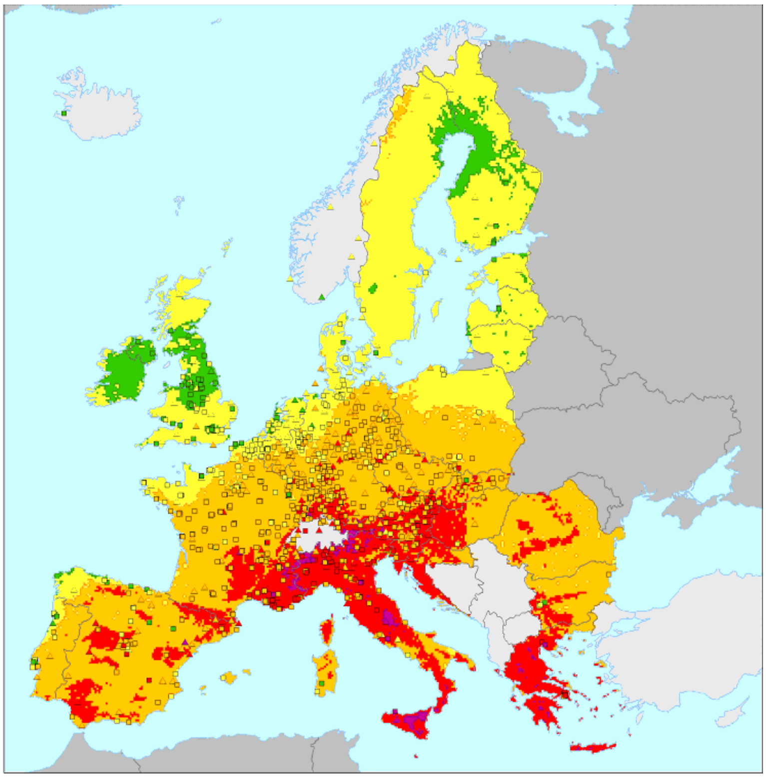
Interpolation and assimilation methods for European scale air quality assessment and mapping

Part II: Development and testing new methodologies



ETC/ACC Technical Paper 2005/8
December 2005

Final draft

Jan Horálek, Pavel Kurfürst, Bruce Denby, Peter de Smet, Frank de Leeuw, Marek Brabec and Jaroslav Fiala

European Topic Centre
on Air and Climate Change

The European Topic Centre on Air and Climate Change (ETC/ACC)
is a consortium of European institutes under contract of the European Environmental Agency
RIVM UBA-B UBA-V IIASA NILU AEAT AUTh CHMI DNMI NTUA ÖKO IEP TNO UEA

Front page picture

Combined rural and urban concentration map of ozone – SOMO35 for year 2003 based on spatial interpolated concentrations field and measured values at the indicated measuring points (units in $\mu g.m^{-3}.days$).

(The figure is taken from this report, Annex II, Figure A2. It represents the combined map as created by merging the interpolated rural map (combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals) and the interpolated urban map (using interpolation of urban increment Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, RO. Countries with missing population density information and therefore excluded from the mapping calculations: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR.)

DISCLAIMER

Interpolation and assimilation methods for European scale air quality assessment and mapping

Part II: Development and testing new methodologies

Contributing authors
Jan Horálek and Pavel Kurfürst, Czech Hydrometeorological Institute (CHMI), Praha
Bruce Denby, Norwegian Institute of Air Research (NILU), Oslo
Peter de Smet and Frank de Leeuw, Netherlands Environmental Assessment Agency (MNP)
Marek Brabec, National Institute of Public Health, Praha

EEA project manager: Jaroslav Fiala

Final Draft

TABLE OF CONTENTS

1 Introduction

This document is the second part of a report that reviews and recommends interpolation methods for air quality applications. In Part I a literature review is carried out of methodologies that can be applied for the interpolation or assimilation of observational data to produce maps of air quality on the European scale. In this Part II a number of these methodologies are tested and compared to established preferred methods of interpolation. In addition to the description and testing carried out here, a summary of the results obtained in this part are presented in Chapter 8 of Part I.

Maps of air quality are important for any spatially derived assessment of air quality effects, particularly in regard to human and ecosystem exposure. As a result air quality maps, even on the European scale, will require high resolution to cover the large gradients in population and land use. In the current study emphasis is put on the development of interpolation methodologies for ozone and PM_{10} with the subsequent preparation of high resolution maps that can resolve urban agglomerations (in support to the Structural Indicator work of EEA to DG ENV). Ozone and PM_{10} as air pollutants are chosen due to their high actuality and the need for input of spatial related information on these pollutants in support to recent policy development processes (e.g. CAFE).

A number of methods described in Part I of this paper are tested and intercompared in this Part II. Monitoring data from AirBase over a four-year period (2000-2003), for the pollutants ozone and PM_{10} and their relevant indicators, are applied in the interpolation tests. Both pure interpolation methods and methods that utilize supplementary information (such as dispersion model data, altitude or climatologic parameters) are applied. A short overview of the methods used is given in Chapter 2, more background information on these methods can be found in Part I.

The input data for the interpolation tests are described in Chapter 3. This chapter also deals with the different character of rural and urban air quality. Due to this different character the interpolations on rural and urban areas are handled separately.

In order to apply methods involving the use of supplementary information on top of measurement data it is necessary to determine the relevance of these additional data. The relationships between the measured concentrations and the additional data are examined. Only parameters considered relevant to concentrations will be treated. This examination is described separately for rural and urban areas and for ozone and PM_{10} in Chapter 4.

The comparison of different interpolation methods for rural areas is described in Chapter 5, separately for both ozone and PM_{10}. The map creation in the urban areas, especially in cities with no measurement, is the theme of Chapter 6.

Rural and urban air quality concentration maps must be merged together. This is carried out with the help of the population density grid. The concentration maps produced in this way are subsequently combined with the land cover grid and the population density grid in order to construct the maps of the crops and vegetation at risk and the maps overlaid by population density, in Chapter 7.

The final result of this work, i.e. high resolution maps are included as an Annex II.

2 Methodology

Three classes of methods are examined in the tests carried out here. The first class consists of methods that use monitoring data only. The second class of methods use model calculations to aid the interpolations and finally, the third class contains regression type of methods that combine all available supplementary data, including model fields, to improve the interpolation. The actual interpolation methods applied within these classes are selected as being representative of that class. Comparison of these methods is performed by cross-validation, using the root mean-square-error (RMSE) as described in Chapter 7 of Part I. This type of comparison provides an objective measure of interpolation quality; it is applicable for all examined methods and Chapters 5 and 6 describe the comparison results in more detail.

2.1 Interpolation methods using monitoring data only

This class of methods will be further referred to as **pure interpolation**. Several methods are investigated:

Inverse distance weighting method (IDW), see Section 2.2 in Part I. The required parameters are set as follows: the number of used neighbouring stations is n=15, the weighting power β is always optimized computationally in order to minimize RMSE.

Ordinary kriging (OK), see Section 2.5.1 in Part I. The parameters of the variogram (range, nugget and sill) are chosen manually in order to minimize RMSE in cross-validation. The anisotropy or the trend removal is not considered.

Ordinary cokriging (OC), see Section 2.5.1 in Part I. The secondary variable is especially altitude, but also climatological parameters are used. The variogram parameters are chosen in the same way as in the case of ordinary kriging.

Lognormal kriging (LK) and **lognormal cokriging (LC)**. This is the same as ordinary kriging and cokriging but performed after logarithmic transformation. This interpolated field is back-transformed by exponentiation $exp(Z+\sigma^2/2)$, where Z is the interpolated field and σ^2 is the kriging error (Cressie, 1993).

2.2 Interpolation methods using monitoring and modelling data

These methods are described in Chapter 4 of Part I. They will be further referred to here as **combination with model**. Two approaches will be examined:

Combination using plain subtraction of modelled values from measured data, similar to the methodology described in Section 4.1 of Part I. For every measuring point $s_1, \ldots, s_N$ differences (or residuals) are calculated

$$diff(\mathrm{s_i}) = Z(\mathrm{s_i}) - M(\mathrm{s_i}). \tag{2.1}$$

where $Z(\mathrm{s}_i)$ is the measured value in point s_i,

$M(\mathrm{s}_i)$ is the value of the model's result in the point of grid s_i.

These differences are then interpolated (using IDW, kriging and cokriging) into the field of differences D. This interpolation field is added to the model M:

$$\hat{Z}(\mathrm{s}_0) = M(\mathrm{s}_0) + D(\mathrm{s}_0) \tag{2.2}$$

where $\hat{Z}(\mathrm{s}_0)$ is the estimated value of the concentration in the point s_0,

Combination using fitted model, as described in Section 4.2 of Part I. It is similar to the plain combination method described above, but the results of the model are first fitted, through linear regression, to the observations. The differences are again interpolated using IDW, kriging and cokriging. In the case of interpolation by kriging, this approach is sometimes called "kriging with local varying mean" (e.g. Lloyd and Atkinson, 2004).

2.3 Interpolation methods using monitoring, modelled and supplementary data

These methods join together the approaches as described in both Chapters 3 and 4 of Part I, i.e. they use monitoring as well as supplementary and modelling data (these two types of data are subsequently sometimes for simplicity described together as "additional", because of their similar character in relation with the measured data). The aim is to use all available relevant data to improve the quality of the resulting estimation. These data should bring other additional information, which is not fully included in the model results (e.g. sunshine duration reflects the contribution of secondary organic particles into the whole PM_{10} concentration field, while these particles are not modelled by the EMEP PM_{10} model).

This approach is similar to the method of the fitted model, but based on a linear regression analysis on all the available and relevant data by

$$M'(s) = c + a_1.D_1(s) + ... + a_m.D_m(s) \qquad (2.3)$$

where $D_1(s), \ldots, D_m(s)$ are additional data in point *s*,

$c, a_1, \ldots, a_m$ are parameters of the linear regression model, computed at air quality monitoring stations.

The differences (or residuals) are calculated according to

$$diff(\mathrm{s_i}) = Z(\mathrm{s_i}) - M'(\mathrm{s_i}). \qquad (2.4)$$

These differences are interpolated, using ordinary kriging and cokriging, and the interpolated field of differences *D* is added to the field *M'* (the meaning of the symbols is the same as in the previous equations):

$$\hat{Z}(\mathrm{s_0}) = M'(\mathrm{s_0}) + D(\mathrm{s_0}) \qquad (2.5).$$

All these methodologies are subsequently applied on real data.

2.4 Comparison of the interpolation methods

Comparison of different interpolation methods is performed by cross-validation as described in chapter 7 of Part I. The cross-validation method computes the spatial interpolation for each measured point using all the available information except from that one point (i.e. it withholds one data point and then makes a prediction at the spatial location of that point). The predicted and measured values are then compared and the procedure is repeated for all points. In this way the behavior of the investigated methods, excluding the influence of the measurement, can be simulated. The used indicator is RMSE, i.e. root mean squared (cross-validated) error:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Z(s_i) - \hat{Z}(s_i))^2}\ , \qquad (2.6)$$

where $Z(s_i)$ is the measured value in the i-th point ,

$\hat{Z}(s_i)$ is the estimated value in the *i*-th point using other points.

A smaller RMSE generally means a better estimation.

3 Input data

3.1 Introduction

Input data requirements are dependent on the methodology used. The minimum input data required for the interpolation are the air quality measurements, with their geographical coordinates (and elevation). Furthermore, a number of meteorological parameters considered relevant are also included. Finally concentration fields calculated with the EMEP unified model are required for the model based interpolation methodologies. All input data used are described in this chapter.

3.2 Air quality parameters

The analysis is executed for the two pollutants ozone and suspended particles PM_{10} and their indicators, for the years 2000-2003. The parameters chosen for ozone are SOMO0 (annual sum of all maximum daily 8-hour concentrations, in $\mu g.m^{-3}$.days), SOMO35 (annual sum of maximum daily 8-hour concentrations above 70 $\mu g.m^{-3}$ (i.e. 35 ppb), in $\mu g.m^{-3}$.days), and AOT40 (sum of the differences between hourly concentrations greater than 80 $\mu g.m^{-3}$ (i.e. 40 ppb) and 80 $\mu g.m^{-3}$ over the three months from May to July, using only values measured between 7:00 and 19:00 UTC, in $\mu g.m^{-3}$.hours). For PM_{10} the annual average and 36^{th} maximum daily average value (both in $\mu g.m^{-3}$) are chosen.

The two pollutants are selected because of their major importance in the area of air quality. The selection of parameters is motivated by the existing legislation (AOT40, both PM_{10} parameters) and by the developing Structural Indicators of the Commission and Eurostat (SOMO35, SOMO0 for comparison purposes; annual average for PM_{10}). These parameters are also seen as the best descriptors for health, resp. vegetation impact.

3.3 Data sources

3.3.1 Measured air quality concentrations ("hard data")

Only background stations[1] (not industrial and traffic) are taken into account. For ozone 440 rural background stations and 830 urban/suburban background stations are used. For PM_{10} 205 rural background stations and 724 urban/suburban background stations are used. The air quality data have been extracted from AirBase, with addition of a few rural EMEP stations (there are 8 PM_{10} stations, helpful for spatial coverage). For every year only the stations that have temporal data coverage of at least 75 percent are used. Together with the measured data information concerning the measurement stations - geographic coordinates determining station's geographic location, altitude of the station and the type of area (i.e. rural-suburban-urban-unknown), have been extracted.

3.3.2 Additional data ("soft data")

The purpose of these data is to bring additional information to the regions not covered by measurement data. These data include:

- Results of photochemistry version of the EMEP Unified Eulerian model (revision rv2_1_2) – grid 50x50 km; for all the relevant years, subsequent parameters:
    - SOMO0 [$\mu g.m^{-3}$.day]
    - SOMO35 [$\mu g.m^{-3}$.day]

[1] Background type of station, as defined in AQ-DEM manual: "Station located such that its pollution level is not influenced significantly by any single source or street, but rather by the integrated contribution from all sources upwind of the station. These stations can be located both inside (urban background) as well as outside (regional/background) cities."

- AOT40 [µg.m^{-3}.hour] - as the sum of all the days from a year (using values from daylight hours, when solar zenith angle is equal or less than 89 degrees) and as such calculated in different way than the measured AOT40 values, which are composed from EC and UNECE legislation defined three months measurement period
- PM_{10}, annual average [µg.m^{-3}] – computed as the sum of primary $PM_{2.5}$, primary $PM_{2.5-10}$ and secondary inorganic particles
- PM_{10}, 36th maximum daily average value [µg.m^{-3}]

The model is described by Simpson et al. (2003) and Fagerli et al. (2004). The model results for ozone parameters are based on different emissions for each year (see Vestreng et al., 2004 and 2005). In case of PM_{10} the same dataset of PM emissions from the year 2000 has been used for all the years (namely CAFE Baseline scenarios, see Amann et al., 2004, filled with MSC-W expert estimates for countries not included in this dataset, Vestreng et al., 2004), because time series of these emissions have not been available. The model uses actual meteorological data.

- Altitude [m] – grid 30'' x 30'', source: GTOPO30 (Global Digital Elevation Model), ESRI (Redlands, California, USA, 2005) ;
- Climatologic parameters – grid 10'x10' (the means for the period 1960-1990), source: CRU CL 2.0, New et al. (2002), www.cru.uea.ac.uk/cru/data/
  - Temperature [°C]
  - Precipitation [mm.year^{-1}]
  - Sunshine duration [%]
  - Wind speed [m.s^{-1}]
  - Relative humidity [%]
- CORINE Land cover database – grid 250 x 250 m (CLC2000, version 7, summer 2005) (Source EEA – ETC/TE, lceugrid250v7), using parameters "LEVEL1" and "DRECL1";
- Population density [inhbs.km^{-2}] - grid 100 x 100 m. Source: JRC, draft version (pop01c00v3f, aug/sep 2005); it includes population data 2001 (EUROSTAT statistics) mapped on EEA's CLC2000 land cover data. As the cells have an area of 1 hectare (100x100m), dividing by 100 will give absolute population in each cell.

When considering which additional datasets to use their usefulness for air quality mapping, as well as their availability, played a key role. (E.g. it would be desirable to use actual meteorological data instead of the climatological ones and to use solar radiation instead of sunshine duration, if available.)

3.4 Rural and (sub)urban area differentiation

The different nature of urban/suburban and rural air quality is examined. Tables 3.1 and 3.2 show the average values from all AirBase background stations at all individual years by their types, rural background and urban/suburban background stations, for different pollutants and parameters. (Urban and suburban stations are handled together.)

Table 3.1. O_3, SOMO0 and SOMO35. Averages of all rural and urban/suburban background stations for the years 2000 – 2003, in ug/m^3.days.

SOMO0	2000	2001	2002	2003
rural	27394	27472	28160	30877
urb. + suburb.	23508	23888	24504	27587

SOMO35	2000	2001	2002	2003
rural	5219	5381	5777	8206
urban + suburban	3707	4249	4301	6802

Table 3.2. PM_{10}, annual average and 36th maximum daily mean values. Averages of all rural and urban/suburban background stations for the years 2000 – 2003, in ug/m^3.days.

annual average	2000	2001	2002	2003
rural	21.4	20.1	22.0	22.5
urban + suburban	29.6	25.5	26.7	29.4

maximum 36th d. v.	2000	2001	2002	2003
rural	37.3	35.0	38.5	40.1
urban + suburban	47.3	41.2	44.2	49.6

The difference in nature between urban/suburban and rural air quality, as shown in the table values, is reconfirmed in the examination of relations between measured and additional data (see Chapter 4). It can be stated that for ozone the rural concentration field is higher than the urban field and for PM_{10} the urban concentration field is higher than the rural field. These known empirical findings have obvious physico-chemical interpretation. During daylight hours ozone concentrations in the troposphere are controlled by the competition of NOx and VOC for OH radicals and hence ozone concentrations are dependent on the VOC to NO_x concentration ratio. In urban areas characterised with NO_x concentrations higher than is the VOC/NO_x optimal ratio for maximum ozone formation, the ozone concentrations are there generally lower than in rural areas because of terminating reactions of NO_x with OH radicals which prevails in NO_x rich regions of city centres. During the night ozone concentrations reduce more in polluted urban areas by direct chemical reactions of ozone with nitrogen oxides emitted by traffic, industry and heating, then in the rural areas. In the case of PM_{10} it is the higher emissions in the populated regions, through traffic, land use and industrial emissions that play a role.

Due to the different nature of rural and urban air quality it is not suitable to use the interpolation methods simply for the whole area together. If, for example, PM_{10} urban concentrations were assessed by interpolation of the surrounding rural stations, the concentrations are likely to be underestimated. Based on this hypothesis it is necessary to create rural and urban maps separately and to produce the final European maps by merging them. This approach is further elaborated in the following chapters.

4 Relationships between measurement, model and supplementary data

4.1 Introduction

In this chapter we examine the relationship between the air quality parameters derived from the measurement data ("hard data") and the different model and other additional input data as listed in Section 3.2 ("soft data"), in order to determine which additional data are suitable for usage in mapping. The relationship is examined by means of linear regression, separately for rural and urban/suburban areas. The most important outputs of that treatment are the parameters of linear regression and the coefficient of determination (or square multiple correlation, R^2). The parameters of linear regression are assessed by the standard ordinary least squares (OLS) method, see Rao (1973). The basic type of linear regression considered is

$$Z(s_i) = c + a*X(s_i) + \varepsilon(s_i) \tag{4.1}$$

where $Z(s_i)$ are the measured concentration in the points s_i ,
$X(s_i)$ are some additional data in the points s_i,
a, c are parameters of linear regression (intercept c and slope a), estimated by OLS,
$\varepsilon(s_i)$ are residuals in the points s_i .

For simplicity this type of regression is further termed *meas = c+ a*x.*

The significance of linear regression is tested using the F-test. In cases where regression is significant, the parameter c is tested to see whether it is significantly different from zero, using the t-test. If not, the simpler type of linear regression *meas = a*x* is used - the constant c is dropped. (These tests are performed, although the distribution of residuals $\varepsilon(s)$ in many cases is not normal; the high number of cases, i.e. measuring stations enables the usage of these tests.)

The closeness of the regression relation is given by the coefficient of determination, R^2:

$$R^2 = \frac{\sum_{i=1}^{n}(Z(s_i)-\overline{Z})(P(s_i)-\overline{P})}{\sqrt{(Z(s_i)-\overline{Z})^2}.\sqrt{(P(s_i)-\overline{P})^2}} \quad \left(= \frac{\sum_{i=1}^{n}\left(Z(s_i)-\overline{Z}\right)\left(X(s_i)-\overline{X}\right)}{\sqrt{\left(Z(s_i)-\overline{Z}\right)^2}.\sqrt{\left(X(s_i)-\overline{X}\right)^2}} \right) \tag{4.2}$$

where $Z(s_i)$ are measured concentrations in the points s_i , $i=1,\dots, n$,
$P(s_i)$ are calculated predictions of linear regression in the points s_i, e.g. $c + a*X(s_i)$.
$\overline{Z}$ is the arithmetic average of $Z(s_1), \dots, Z(s_n)$,
$\overline{P}$ is the arithmetic average of $P(s_1), \dots, P(s_n)$,
n is the number of measuring stations.

The closer R^2 is to 1 then the closer the relation. The value of R^2, multiplied by 100, expresses the percentage of the variability, which is explained by the regression.

Additionally, for different linear regression relationships in Chapter 4 also root-mean-square-error are computed (given in the same units as $Z(s)$), see Annex I, Section 5:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Z(s_i)-P(s_i)\right)^2}{n}} \tag{4.3}$$

The smaller *RMSE* means the more precious estimation by linear regression.

Rural and urban/suburban areas are dealt with separately, for reasons given in Section 3.4. For each of these area types the ozone and PM_{10} parameter regression relation with the model and supplementary data are examined. This results in four test groups:

- Rural areas - Ozone parameters (SOMO0, SOMO35 and AOT40) (Section 4.2)

- Rural areas - PM_{10} (Annual average, 36th maximum daily mean value) (Section 4.3)
- Urban/suburban areas - Ozone parameters (SOMO0, SOMO35 and AOT40) (Section 4.4)
- Urban/suburban areas - PM_{10} (Annual average, 36th maximum daily mean value) (Section 4.5).

At each of these 4 groups the model and additional input data examined by the linear regression are:

- EMEP model
- Altitude (taken from AirBase)
- Climatic parameters
  - sunshine duration
  - relative humidity
  - temperature
  - precipitation
  - wind speed

For the urban/suburban regions additional tests are performed to determine the regression relations between the geographical location and the air quality parameter values (Section 4.4.4 and 4.5.4). The reason for this is to utilize the spatial location in cases where spatial interpolation is not performed (see the discussion in Chapter 6).

For both the rural and urban/suburban area groups a final linear regression model is formulated given by

$$\textit{air quality parameter} = c + a1 * \textit{addit. parameter1} + a2 * \textit{addit. parameter2} + \ldots \qquad (4.4)$$

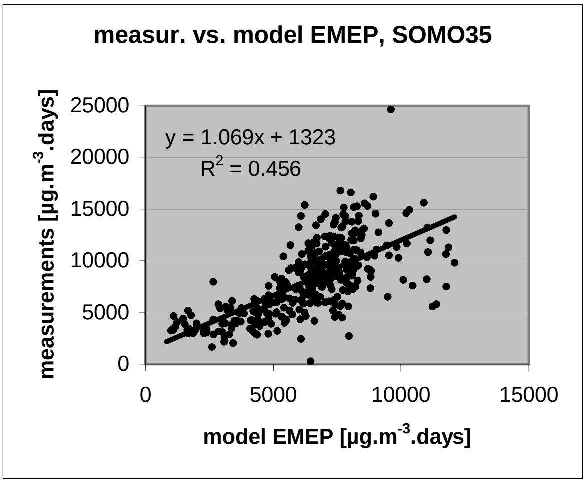
which combines the parameters with the most significance in the regression relation with the measurement parameters (Sections 4.2.4, 4.3.4, 4.4.5 and 4.5.5). Their selection is carried out by the so-called stepwise of backward type (for an example, see Annex I, Section 1) and it keeps in balance the variance of parameters and the predictive capability of the linear regression model. With an increasing number of parameters this capability is enhanced, but the variance of parameters increases, particularly in cases of high mutual correlation (collinearity) among some parameters. If for any parameter the square multiple correlation R^2 between this parameter and all of the other parameters is greater than 0.7 (i.e. there is high collinearity), this parameter is excluded (its correlation is considered as being of such a height that adding this parameter has only small added value to improved the mapping results). However, the main aim of this linear regression model is not to estimate the regression parameters, but to estimate the concentrations which are subsequently used in mapping.

Linear regression models are further utilized in Chapters 5 (rural areas) and 6 (urban areas) in map creation.

4.2 Rural areas – ozone

4.2.1 EMEP model

The relation between the EMEP model and measurement values is examined at the location of station measurements for each year separately. Figure 4.1 shows two examples of the tested regression relations. Table 4.1 presents the results of all correlation tests.



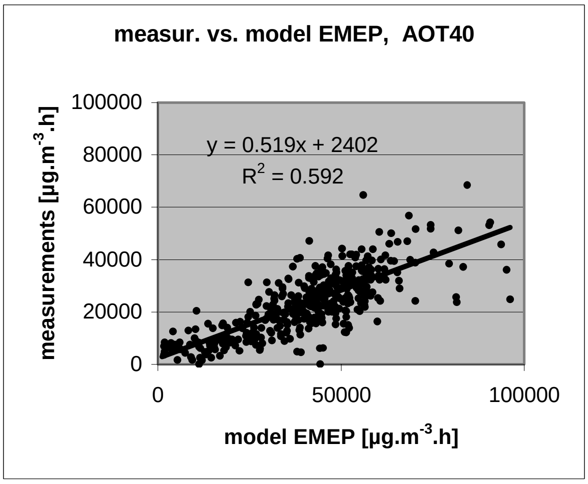


Figure 4.1. The graphs represent the linear regression results indicating the level of correlation between the respective ozone parameters SOMO35 (left) and AOT40 (right) of measurements of 2003 at the rural stations (y-axis) versus the same ozone parameters based on EMEP dispersion modelled values (x-axis). The AOT40 shows a higher R^2 than for the SOMO35.

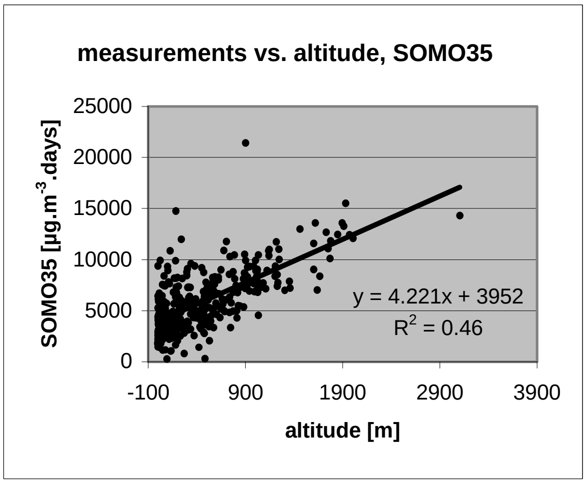
Table 4.1. Ozone parameters SOMO0, SOMO35 and AOT40 at rural background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against the EMEP dispersion model. The strongest relation shows AOT40 against EMEP model (R^2 is highest for all years) and the weakest is SOMO0 (lowest R^2 for all years). ("n. sign." = not significant)*

meas =	**SOMO0**				**SOMO35**				**AOT40**			
c + a * EMEP	**2000**	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**
c	n. sign.	n. sign.	n. sign.	n. sign	-751	n. sign.	910	1323	n. sign.	n. sign.	n. sign.	2402
a	0.95	0.98	0.94	1.04	1.04	1.03	0.77	1.07	0.44	0.55	0.43	0.52
R^2	**0.35**	**0.38**	**0.24**	**0.36**	**0.45**	**0.48**	**0.27**	**0.46**	**0.47**	**0.63**	**0.40**	**0.59**

Table 4.2 shows that the closest correlation, i.e. highest R^2 values at all years, exists between the modelled and the measured ozone values in the case of the ozone air quality parameter AOT40. This is surprising since AOT40 in the EMEP model is defined differently, than in the case of measurements (see Section 3.3.2). The weakest regression relation exits in the case of SOMO0 with the lowest R^2 values at all years. The absolute values of SOMO0 and SOMO35 are of the same order of magnitude for both the modelled and measured data as is indicated by *a* being close to 1, while in the case of AOT40 the model overestimates the measured values (i.e. station measured data) as indicated by the parameter *a* being much lower then 1. This overestimation is caused by the fact that AOT40 values are calculated for the whole year in the model, whilst measurements based AOT40 is calculated only from the 3 summer months, as defined in the legislations. A comparison of the years indicates that 2002 shows the weakest correlation for all the parameters, since R^2 is the lowest in each table.

4.2.2 Altitude

In the literature a considerable relation between altitude and ozone station measurements is detected (e.g. Brönnimann et al., 2000); e.g. ozone concentrations are higher in mountainous regions than in the other areas. Figure 4.2 shows two examples, and all the investigated comparison results are given in Tables 4.8 – 4.10. The strongest relation between ozone measurement data and altitude shows SOMO0 with the highest R^2 values in all years, while the weakest shows AOT40 with the lowest R^2. Better results for SOMO0 (and similar results for SOMO35) in comparing with EMEP model are probably given by better spatial resolution of altitude.



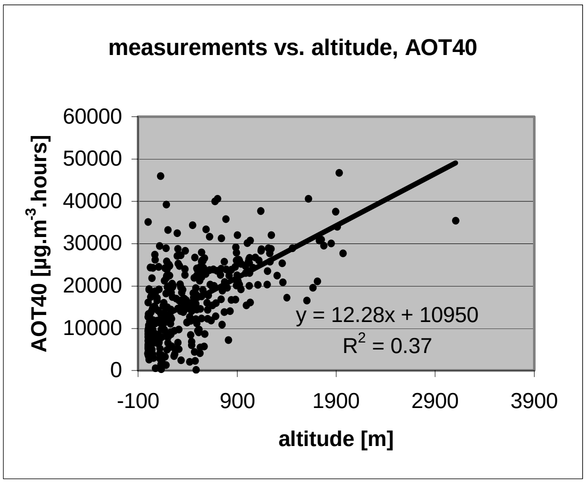


Figure 4.2. The graphs represent the linear regression results indicating the level of correlation between the respective ozone parameters SOMO35 (left) and AOT40 (right) of measurements of 2002 at the rural stations (y-axis) versus altitude (x-axis). The AOT40 shows a lower R^2 than the SOMO35, meaning the ozone parameter AOT40 provides a lower correlation of altitude and the station measurements of ozone than SOMO35.

Table 4.2. Ozone parameters SOMO0, SOMO35 and AOT40 at rural background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against the altitude of the area the station is situated. The strongest relation shows the SOMO0 against altitude (R^2 is highest for all years) and the weakest is AOT40 (lowest R^2 for all years).*

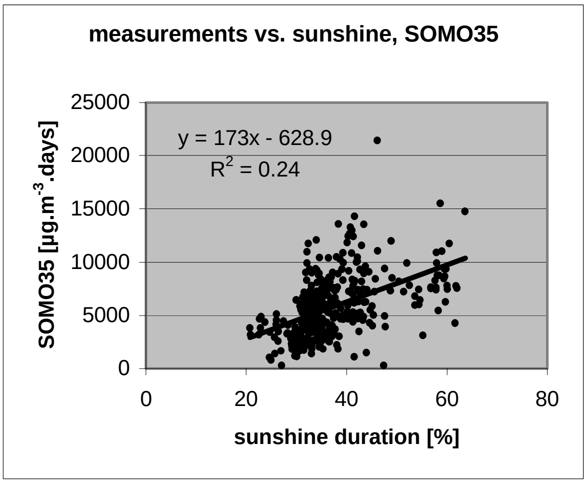
meas =	**SOMO0**				**SOMO35**				**AOT40**			
c + a * altitude	**2000**	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**
c	24279	24438	25288	27923	3236	3547	3952	6271	11081	12631	10950	18735
a	7.41	7.14	6.65	6.99	4.71	4.30	4.22	4.59	10.94	10.74	12.28	11.51
R^2	**0.52**	**0.55**	**0.49**	**0.47**	**0.47**	**0.47**	**0.46**	**0.37**	**0.28**	**0.32**	**0.37**	**0.19**

4.2.3 Climatic parameters

Sunshine duration

The relation with sunshine duration is well-known (e.g. Tidblad, 2004) and it has a clear physical background in the photo-chemical nature of ozone formation. It holds not only for (multi-)annual averaged concentration parameters, but also for daily concentrations at the individual measuring stations. Figure 4.3 gives, as example, the regression relation between sunshine duration and the ozone parameters SOMO35 and AOT40. The Table 4.3 provides the square multiple correlation results (R^2) for all three ozone parameters considered. The values show that the relation for sunshine duration is weaker than those for the EMEP model and altitude described above. Nevertheless, R^2 shows that sunshine duration explains about 20 percent of ozone's variability.

The table also shows SOMO0 has the weakest relation with sunshine. This is probably caused by the cumulative character of the other two parameters for concentrations above a certain threshold (70 and 80 $\mu g.m^{-3}$ for SOMO35 and AOT40 respectively), by which the influence of hot sunny days is extended.



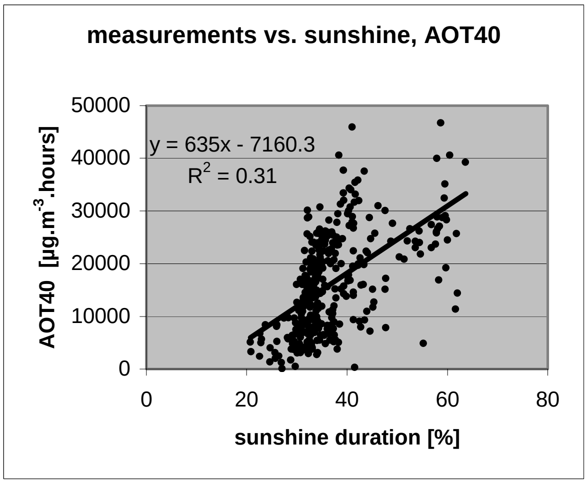


Figure 4.3 The graphs represent the linear regression results indicating the level of correlation between the respective ozone parameters SOMO35 (left) and AOT40 (right) of measurements of 2002 at the rural stations (y-axis) versus sunshine duration (x-axis). The AOT40 shows a higher R^2 than for the SOMO35, meaning the ozone parameter AOT40 provides a higher correlation of sunshine duration and the station measurements of ozone than SOMO35.

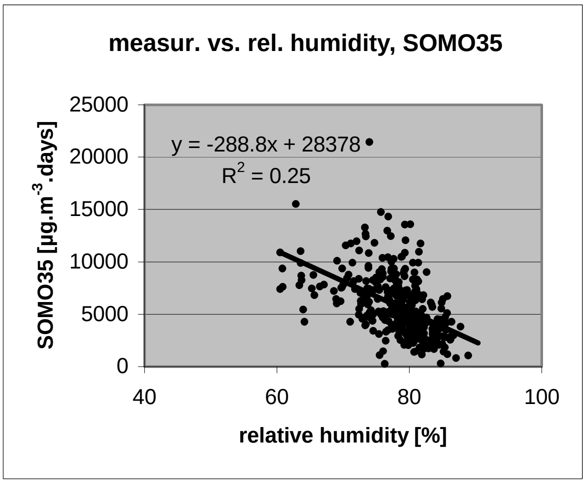
Table 4.3. Ozone parameters SOMO0, SOMO35 and AOT40 at rural background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against the sunshine duration. The strongest relation shows the SOMO35 against sunshine duration (R^2 is highest for all years). Both for SOMO35 and AOT40 the R^2 shows that sunshine duration explains about 20 percent of ozone variability. Weakest is SOMO0 with variability below 20 percent for all years.*

meas =	**SOMO0**				**SOMO35**				**AOT40**			
c + a * sunsh. d	**2000**	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**
c	16974	17872	19686	22889	-2928	-1995	-629	2084	-5994	-5412	-7160	-527
a	291.2	263.5	228.6	212.5	227.8	202.6	173.0	163.1	602.4	628.8	635.0	644.5
R^2	**0.18**	**0.19**	**0.18**	**0.15**	**0.24**	**0.27**	**0.24**	**0.17**	**0.19**	**0.27**	**0.31**	**0.20**

Relative humidity

Relative humidity has the opposite relation to ozone concentrations than does sunshine duration: the higher the relative humidity, the lower the ozone concentrations. Relative humidity and sunshine duration are mutually dependent and have a coefficient of determination R^2 of 0.69 (further details are not tabulated in this report).

The results of the regression calculation between relative humidity and measurement data are presented in Table 4.4. Figure 4.4 shows as illustration two cases from the tables. The graphs represent the scatter plot of the long-term average relative humidity versus the ozone parameters SOMO35 and AOT40 showing the inverse correlation between the climatic and air quality parameters. The closest relation with humidity is detected for AOT40.



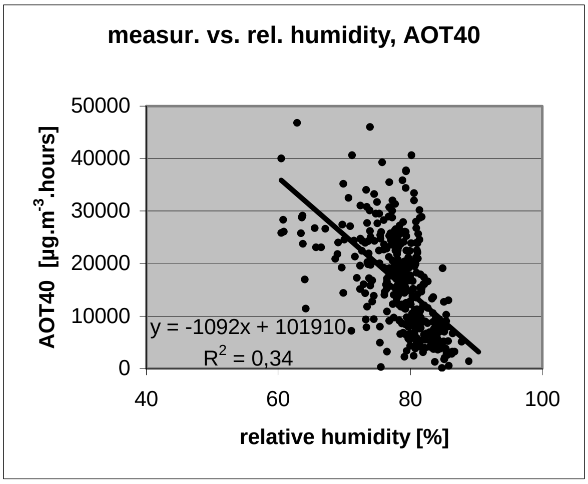


Figure 4.4 The graphs represent the linear regression results indicating the level of correlation between the respective ozone parameters SOMO35 (left) and AOT40 (right) of measurements of 2002 at the rural stations (y-axis) versus relative humidity (x-axis). The AOT40 shows a higher R^2 than for the SOMO35, meaning the ozone parameter AOT40 provides a higher correlation of relative humidity and the station measurements of ozone than SOMO35.

Table 4.4. Ozone parameters SOMO0, SOMO35 and AOT40 at rural background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against the relative humidity. The strongest relation with relative humidity shows the AOT40 (R^2 is highest for three out of the four all years). Weakest relation shows SOMO0 with variability up to 20 percent.*

meas =	SOMO0				SOMO35				AOT40			
*c + a * rel. hum.*	2000	2001	2002	2003	2000	2001	2002	2003	2000	2001	2002	2003
c	67945	62580	58090	62488	37867	32786	28378	32728	108364	109150	101910	118345
a	-513.5	-446.2	-382.5	-405.5	-413.4	-348.2	-288.8	-314.5	-1175.7	-1165.1	-1092.4	-1213.6
R^2	**0.19**	**0.20**	**0.19**	**0.20**	**0.28**	**0.29**	**0.25**	**0.22**	**0.27**	**0.30**	**0.34**	**0.26**

Temperature, precipitation and wind speed

The other three climatic parameters, temperature, precipitation and wind speed have no or only a weak regression relation with measured ozone concentrations as is demonstrated by the low correlations given in the summary Table 4.5 for SOMO0, SOMO35 and AOT40. If no significant linear regression (using F-test) is detected, "n. sign." is indicated in the tables, instead of the resulting R^2 value.

Table 4.5. Ozone parameters SOMO0, SOMO35 and AOT40 at rural background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against successively temperature, precipitation and wind speed. The summary of the R^2 values show little or no correlations of these climatic parameters with the ozone parameters. ("n. sign." = not significant)*

R^2	SOMO0				SOMO35				AOT40			
	2000	2001	2002	2003	2000	2001	2002	2003	2000	2001	2002	2003
temperature	0.03	0.01	0.02	n. sign	n. sign	n. sign	n. sign	n. sign	n. sign.	n. sign.	n. sign.	0.03
precipitation	0.08	0.08	0.02	0.04	0.07	0.07	0.02	0.04	0.02	0.04	n. sign.	0.01
wind speed	0.03	0.04	n. sign	0.05	0.10	0.14	0.05	0.12	0.09	0.19	0.12	0.20

The tables indicate no or very small correlation with temperature and precipitation. They do show some correlation with wind speed.

The absence of significant correlation between ozone and temperature is rather surprising. The reason is probably enclosed in the character of annual averaged parameters. (When considering hourly values, there is a known relation between ozone concentrations and the square of temperature, see Brönnimann et al., 2002.).

4.2.4 Linear regression model

Finally the parameters for a linear regression model are estimated, based on equation 4.4 and the above examined relations.

The model for SOMO0 and SOMO35 considered is:

$$O_3 = c + a1*EMEP + a2*altitude + a3*sunshine\ duration \tag{4.5}$$

And for AOT40 this is:

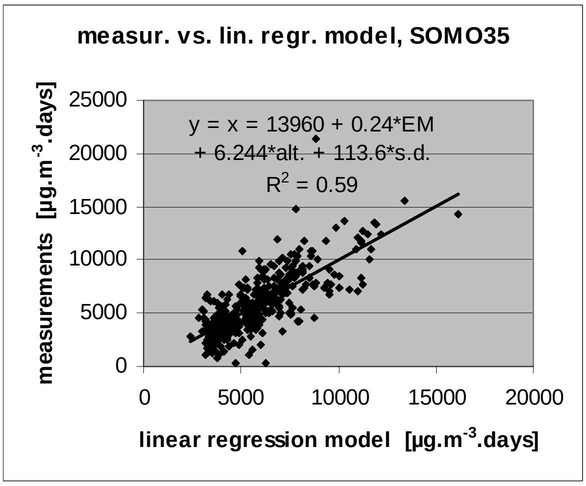
$$O_3 = c + a1*EMEP + a2*altitude + a3*sunshine\ duration + a4*rel.humidity \tag{4.6}$$

The estimated parameters *c*, *a1*, *a2*, *a3*, *a4* and the R^2 for each year are listed in the Table 4.6.

When determining the regression coefficients for each year, parameters not statistically significant (using t-test) are dropped and the whole regression is computed once again without those parameters. Subsequently the other parameters are re-estimated. This can lead to considerable differences in the resulting coefficients ultimately derived for each individual year.

As stated above, sunshine duration and relative humidity are closely related with each other. Therefore, these two parameters are not used together in the linear regression model (we consider they can be used interchangeably in the regression model). However, the usage of sunshine duration is generally more appropriate: R^2 mostly gives worse results, if relative humidity is used instead of sunshine duration in the linear regression model (see Annex I, Section 2).

Comparing the R^2 results in Table 4.6 on the linear regression model with those in Table 4.1, representing the R^2 values for the EMEP dispersion model only, one can conclude that the inclusion of other additional data next to the EMEP model improves the relation between the ozone parameter values from observed concentrations and those derived from the EMEP model calculations substantially. It holds especially for SOMO0 and SOMO35, where the R^2 are considerably higher for the linear regression model in Table 4.6 than in Table 4.1.



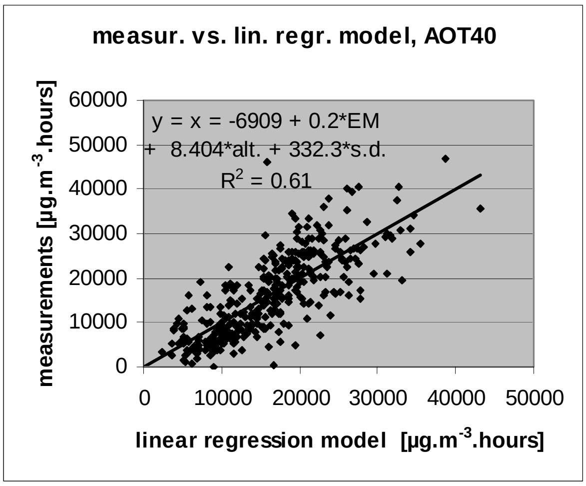


Figure 4.5 The graphs represent the linear regression model, eq. 4.5 (left) and 4.6 (right) results indicating the level of correlation between the respective ozone parameters SOMO35 (left) and AOT40 (right) of measurements of 2002 at the rural stations (y- axis) versus linear regression model (x-axis).

Table 4.6. Ozone measurement parameters SOMO0, SOMO35 and AOT40 at rural background stations tested at four years in a linear regression with an intercept constant c for the EMEP dispersion model combined with the altitude and sunshine duration, and additionally for AOT40 also the relative humidity. The summary of the R^2 values show – compared to the values in Table 4.1 - an increased relation between the concentrations derived by combining EMEP modelled values with the additional parameters and the measured concentrations, especially at SOMO0 and SOMO35. ("n . sign." = not significant)

linear regression model (eq. 4.4)	SOMO0				SOMO35			
	2000	2001	2002	2003	2000	2001	2002	2003
c (constant)	13960	11785	19760	9136	-1653	-871	n. sign.	1205
a1 (EMEP)	0.24	0.38	n. sign.	0.65	0.53	0.61	0.17	0.87
a2 (altitude)	6.244	5.816	6.089	5.383	3.377	3.123	3.592	3.308
a3 (sunsh. dur.)	113.6	70.5	155.8	n. sign	67.1	47.7	85.8	n. sign
R^2	**0.59**	**0.64**	**0.56**	**0.60**	**0.64**	**0.68**	**0.59**	**0.64**

linear regression model (eq. 4.5)	AOT40			
	2000	2001	2002	2003
c (constant)	23334	1512	-6909	24636
a1 (EMEP)	0.30	0.43	0.20	0.44
a2 (altitude)	5.432	5.493	8.404	5.060
a3 (sunsh. dur.)	n. sign.	n. sign.	332.3	n. sign
a4 (rel. humidity)	-262.0	n. sign.	n. sign.	-269.7
R^2	**0.54**	**0.69**	**0.61**	**0.64**

4.2.5 Summary for rural areas – ozone

Regression relations between parameters of ozone measurements and different additional parameters in the rural areas were investigated. Substantial relations with measured concentrations were found for concentrations from the EMEP model, altitude, sunshine duration and relative humidity, where sunshine duration and relative humidity are highly mutual correlated. For other parameters (temperature, wind speed and precipitation) only low or non-significant relations with measured values were found.

The final linear regression models developed and applied is composed of the EMEP dispersion model data, altitude, and sunshine duration and in the case of AOT40 relative humidity. This linear regression model is further utilized in the ozone rural map creation in Section 5.2, since it demonstrated that using additional parameters on top of the EMEP model contributes to an improved correlation between measured concentrations and other parameters.

4.3 Rural areas – PM_{10}

The structure of the air quality field is very complicated in case of PM_{10}. The usage of supplementary parameters could be really advantageous for map creation of PM_{10}. Therefore, we investigate here the relations of the PM_{10} parameters with different supplementary parameters.

4.3.1 EMEP model

The relation between the PM_{10} parameters from observations and those derived from the EMEP dispersion model calculations is not as clear as for ozone, as shows Figure 4.6 for the year 2003 as example.

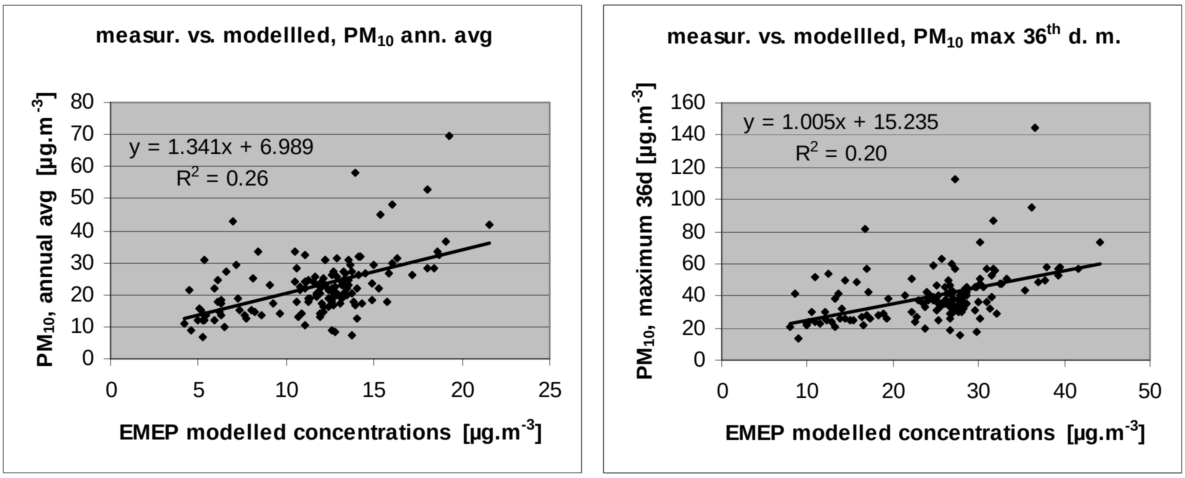


Figure 4.6 The graphs represent the linear regression results indicating the level of correlation between the respective PM_{10} parameters annual averages (left) and 36th maximum daily means (right)) of measurements of 2003 at the rural stations(y-axis) versus the EMEP dispersion modelled values (x-axis).

Table 4.7 presents the coefficients of determination, R^2, for the years 2000 - 2003. They are considerably lower then those given for a similar comparison but with the ozone parameters (Table 4.1). This indicates that a lower correlation exists between the EMEP modelled concentrations and the measured concentrations at the PM_{10} parameters than at the ozone parameters.

Another remarkable observation worth mentioning is that the EMEP dispersion model substantially underestimates observed concentrations. This can be concluded from the regression slope value *a* being considerably higher then 1 at both PM_{10} parameters and for most of the years. Figure 4.6 illustrates this well.

Furthermore, it is interesting to compare in Table 4.7 the results between the individual years. For example, in the case of the 36th maximum daily means it seems there is a different constant intercept coefficient *c* in the regression for the pairs of the years 2000-2001 and 2002-2003. So far, we could not find a plausible explanation for this striking difference.

Another interesting observation is that for the majority of years (i.e 2001–2003) there is a closer relation of the measured 36th maximum daily means with the EMEP modelled annual averages, than it has with the modelled 36th maximum daily means (see Annex I, Section 3).

Table 4.7. Measured PM_{10} parameters annual averages (left) and 36th maximum daily means (right) at rural background stations tested for the years 2000–2003 in a linear regression with an intercept constant c *against the EMEP dispersion model. The R^2 values show lower correlation between modelled and measured PM_{10} parameter values than was noticed for the rural background ozone parameters (Table 4.1). ("n. sign." = not significant)*

measurement =	**annual average**				**36th maximum daily mean**			
c + a * EMEP model	**2000**	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**
c	n. sign.	n. sign.	6.4	7.0	n. sign.	n. sign.	14.1	15.2
a	2.02	1.99	1.45	1.34	1.64	1.61	1.06	1.01
R^2	**0.53**	**0.44**	**0.29**	**0.26**	**0.58**	**0.33**	**0.28**	**0.20**

4.3.2 Altitude

The relation between altitude and PM_{10} is concluded as significant (although not strong). Table 4.8 shows the resulting R^2 values and Figure 4.7 presents as illustration the linear regressions graphs for the year 2002.

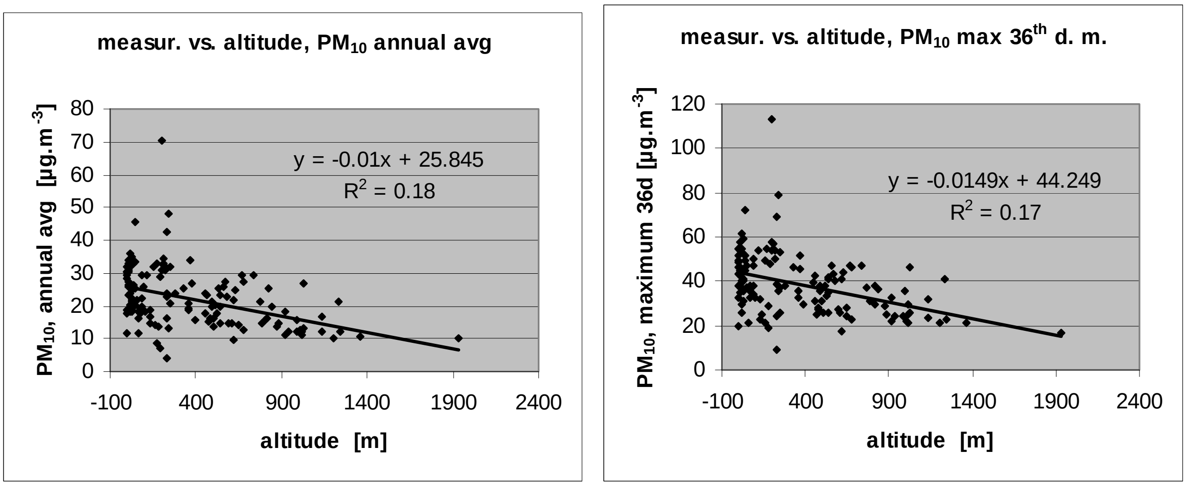


Figure 4.7 The graphs represent the linear regression results indicating the level of correlation between the respective PM_{10} parameters annual averages (left) and 36^{th} maximum daily means (right)) of measurements of 2002 at the rural stations(y-axis) versus the altitude (x-axis).

Table 4.8. Measured PM_{10} parameters annual averages (left) and 36^{th} maximum daily means (right) at rural background stations tested for years 2000 - 2003 in a linear regression with an intercept constant c *against the altitude. The R^2 values are considered to show significant (although not strong) relation between altitude and both PM_{10} parameters.*

measurement =	annual average				36^{th} maximum daily mean			
*c + a * altitude*	2000	2001	2002	2003	2000	2001	2002	2003
c	25.6	23.9	25.8	26.4	42.4	40.2	44.2	46.9
a	-0.0106	-0.0087	-0.0100	-0.0099	-0.0159	-0.0133	-0.0149	-0.0174
R^2	**0.20**	**0.13**	**0.18**	**0.14**	**0.17**	**0.13**	**0.17**	**0.13**

4.3.3 Climatic parameters

Sunshine duration, relative humidity, temperature, precipitation and wind speed

None of these examined five climatic parameters has any substantial relation to PM_{10} in case of rural pollution. In most cases there is no significant correlation as Table 4.9 shows.

Table 4.9. Measured PM_{10} parameters annual averages (left) and 36^{th} maximum daily means (right) at rural background stations tested for years 2000 - 2003 in a linear regression with an intercept constant c *against successively sunshine duration, relative humidity, temperature, precipitation and wind speed. The summary of the R^2 values shows hardly any significance with the measurement based parameters of PM_{10}. ("n. sign." = not significant)*

R^2	annual average				36^{th} maximum daily mean			
	2000	2001	2002	2003	2000	2001	2002	2003
sunshine duration	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	0.05	n. sign.
relative humidity	n. sign.	n. sign.	n. sign.	0.03	n. sign.	n. sign.	0.04	0.04
temperature	0.22	n. sign.	n. sign.	n. sign.	0.12	n. sign.	n. sign.	n. sign.
precipitation	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.
wind speed	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.

4.3.4 Linear regression model

Different supplementary parameters are joined together into one linear regression model with the help of stepwise of backward type, particularly the results of EMEP model and altitude. (The correlation with altitude detected in the EMEP model itself is weaker than the correlation between measurements and altitude). Surprisingly in this linear regression model sunshine duration appears to play a significant role and increases considerably the level of R^2, in comparison to the linear regression model without sunshine duration (for all the years – see Annex I, Section 4). Sunshine duration serves here as a surrogate parameter. It seems to improve the EMEP model in some respect. (Possibly it represents the missing component of secondary organic particles, which are not included in the EMEP model.)

The linear regression model considered is:

$$PM_{10} = c + a1*EMEP + a2*altitude + a3*sunshine\ duration \qquad (4.7)$$

Like at the calculations for rural ozone in Section 4.2.4, the parameters which are not significantly different from zero are dropped from the linear regression and the regression model is re-calculated. The resulting regression coefficients are given in Table 4.10.

In case of the 36th maximum daily means the results of the EMEP modelled 36th maximum daily means are used in the linear regression model for the year 2000, whilst for other years the EMEP model of annual average results are used (for motivation, see 4.3.1).

When comparing the R^2 results in Table 4.10 of the linear regression model that combines EMEP model with the supplementary parameters, with those for the EMEP dispersion model only (Table 4.7) one can conclude that inclusion of altitude and sunshine duration improves the correlation between measured parameter results and the modelled parameter results substantially.

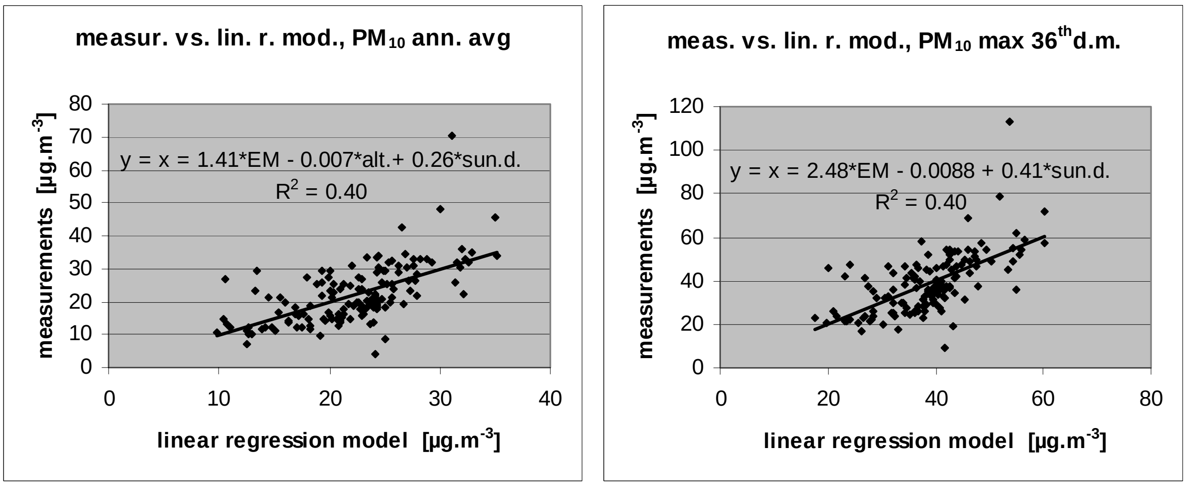


Figure 4.8 The graphs represent the linear regression model (eq. 4.7) results indicating the level of correlation between the respective PM_{10} parameters annual averages (left) and 36th maximum daily means (right)) of measurements of 2002 at the rural stations(y-axis) versus the linear regression model (x-axis).

Table 4.10. Measurement based PM_{10} parameters annual averages (left) and 36th maximum daily means (right) at rural background stations tested at four years in a linear regression with an intercept constant c *against the EMEP dispersion model, combined with the altitude and sunshine duration. ("n. sign." = not significant)*

linear regression model (eq. 4.7)	**annual average**				**36th maximum daily mean**			
	2000	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**
c (constant)	n. sign.	-13.6	n. sign.	n. sign.	-37.0	-22.8	n. sign.	n. sign.
a1 (EMEP model)	1.60	2.26	1.41	1.39	1.68	3.62	2.48	2.38
a2 (altitude)	-0.0077	-0.0054	-0.0070	-0.0064	-0.0075	-0.0079	-0.0088	-0.0109
a3 (sunshine duration)	0.22	0.37	0.26	0.24	1.15	0.66	0.41	0.45
R^2	**0.65**	**0.52**	**0.40**	**0.35**	**0.67**	**0.48**	**0.40**	**0.30**

4.3.5 Summary for rural areas – PM_{10}

The relationship between parameters of measured PM_{10} and a number of different additional parameters in the rural areas has been investigated. An obvious relation of measured values with EMEP model and altitude was detected. Besides, it was detected that sunshine duration improves the relation between measured values and EMEP model values, if it is taken together with the EMEP modelled values in the linear regression model. The sunshine duration tested separately against the measured values did not show a significant relation.

The final linear regression model we ultimately developed and apply is composed of the EMEP dispersion model values, the altitude and the sunshine duration. This linear regression model is subsequently utilized in the PM_{10} rural map creation in Section 5.3, since it demonstrated that using this combination of additional parameters results in an improved correlation between the modelled and measured concentrations.

4.4 Urban areas – ozone

For urban areas the same analysis is carried out as for the rural areas, but only for the health related SOMO0 and SOMO35 ozone parameters and not the AOT40, since that is vegetation and crops related ozone parameter and no subject to urban areas. In this section the urban and suburban stations are treated together, except in the case where the EMEP model is tested against the measurement values. There a more detailed analysis is carried out by testing the urban and suburban measurements separately against the EMEP model. This is to determine if possible distinctive characteristics between urban and suburban areas do play a role and will need separate treatment in further interpolation analyses. (Similar analysis has been performed at several other parameters, but these are not discussed here, since that goes beyond the main goals of this study.)

4.4.1 EMEP model

Urban and suburban measurements treated as one group of measurements

Despite the low resolution (50 x 50 km) of the EMEP long-range trans-boundary air pollution dispersion model, there is a considerable relation between measured urban/suburban concentrations and the model calculations. The reason is that ozone (as well as PM_{10}) has an important long-range contribution. The urban ozone parameter of determination R^2 in Table 4.11 is of about the similar magnitude as in the case of rural stations, while the slope (i.e. parameters *a* from the linear regression) is at the lower level than in the rural areas (Table 4.1). As illustration, Figure 4.9 shows for the year 2003 the relation between the measurement and the modelled SOMO0 and SOMO35 data.

Table 4.11. Ozone parameters SOMO0 and SOMO35 tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against the results of EMEP model.*

measurement =	**SOMO0**				**SOMO35**			
c + a * EMEP model	**2000**	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**
c	n. sign.	n. sign.	n. sign.	n. sign.	-928	-933	-504	295
a	0.81	0.85	0.82	0.92	0.77	0.92	0.74	0.96
R^2	**0.36**	**0.40**	**0.31**	**0.41**	**0.48**	**0.56**	**0.38**	**0.51**

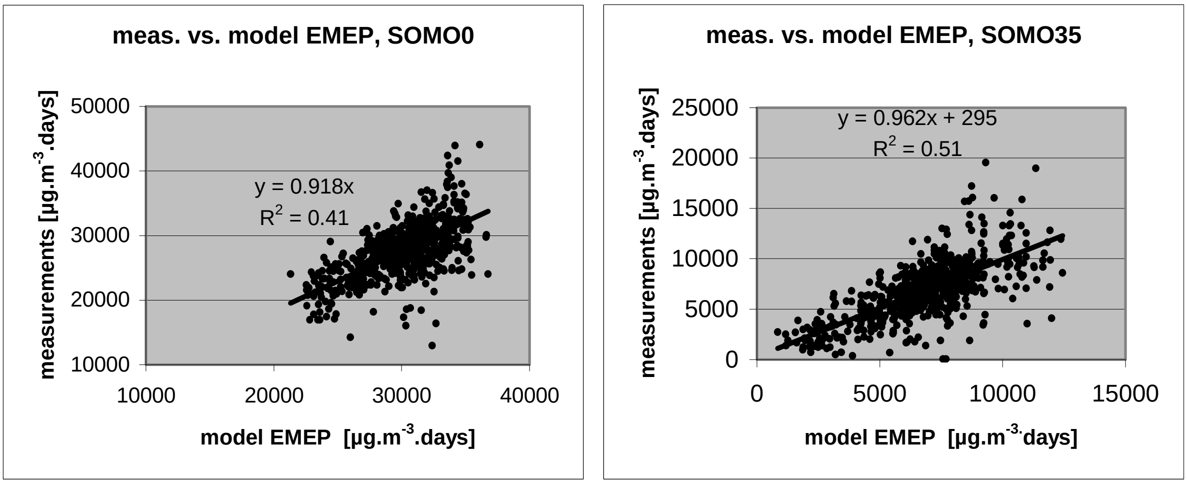


Figure 4.9. The graphs represent the linear regression results indicating the level of correlation between the respective ozone parameters SOMO0 (left) and SOMO35 (right) of measurements of 2003 at urban and suburban background stations taken together (y-axis) versus the same ozone parameters based on EMEP dispersion modelled values (x-axis).

Urban and suburban measurements treated as two separate groups of measurements

The results of an examination of the measurement-model relations for urban and suburban stations separately for SOMO0 are stated in the Table 4.12. (The results for SOMO0 are presented here only, since they are the best comparable due to the non-significant constant for all years.) The table shows that the level of correlation between modelled and measured concentrations is at the similar level for both urban and suburban stations (see R^2). Moreover, the differences between the slopes (e.g. *a* in the linear regression) for the urban and suburban areas were proved to be non-significant (tested by t-test for two linear regressions) for all the years.

Since the differences between the separate urban and suburban correlations in Table 4.12 are only small, no further separate urban and suburban measurement grouping is used in this report.

Table 4.12. Ozone parameter SOMO0 tested for the years 2000 – 2003 in a linear regression on separate grouping of urban background (left) and suburban background (right) stations demonstrate similar correlations at both stations classes. Therefore, further treatment of the two classes as separate groups the spatial interpolation calculations is assumed not to contribute with relevant.

measurement =	**SOMO0, urban**				**SOMO0, suburban**			
c + a * EMEP model	**2000**	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**
c	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.
a	0.79	0.83	0.80	0.90	0.83	0.87	0.84	0.94
R^2	**0.31**	**0.39**	**0.30**	**0.45**	**0.39**	**0.39**	**0.28**	**0.33**

4.4.2 Altitude

The relation between ozone measurements and altitude is much weaker at (sub)urban stations, than in the case of rural stations. See Table 4.13.

Table 4.13. Ozone parameters SOMO0 and SOMO35 at (sub)urban background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against the altitude of the area the station is situated.*

measurement = c + a * altitude	SOMO0				SOMO35			
	2000	2001	2002	2003	2000	2001	2002	2003
c	22131	22805	23429	26242	2868	3595	3534	5801
a	7.19	5.92	5.62	7.15	4.38	3.46	3.98	5.38
R^2	**0.13**	**0.07**	**0.08**	**0.11**	**0.15**	**0.08**	**0.11**	**0.13**

4.4.3 Climatic parameters

Sunshine duration

Sunshine is an important additional parameter for ozone urban/suburban air quality as the relative high R^2 values show in Table 4.13.

Table 4.14. Ozone parameters SOMO0 and SOMO35 at (sub)urban background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against the sunshine duration, with R^2 values showing a clear relation.*

measurement = c + a * sunshine dur.	SOMO0				SOMO35			
	2000	2001	2002	2003	2000	2001	2002	2003
c	14201	12367	14555	18432	-1217	-2177	-1462	1685
a	244.2	296.7	255.4	231.9	129.2	165.6	147.9	129.6
R^2	**0.27**	**0.39**	**0.33**	**0.27**	**0.24**	**0.40**	**0.30**	**0.17**

Relative humidity

The other important climatologic parameter for urban/suburban air quality is relative humidity, see Table 4.15. The relation with measured urban/suburban ozone concentration is even closer than in the case of rural air quality.

Table 4.15. Ozone parameters SOMO0 and SOMO35 at (sub)urban background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against the relative humidity, with R^2 values showing a clear relation.*

measurement = c + a * relative humidity	SOMO0				SOMO35			
	2000	2001	2002	2003	2000	2001	2002	2003
c	62108	63085	55714	60160	28794	30232	26938	29452
a	-495.8	-505.2	-402.8	-421.4	-322.3	-334.8	-292.2	-293.1
R^2	**0.30**	**0.32**	**0.26**	**0.27**	**0.39**	**0.47**	**0.37**	**0.26**

Temperature, precipitation and wind speed

Table 4.16 shows some relation of (sub)urban ozone measurements with temperature and with wind speed, although not as strong as for sunshine duration and relative humidity.

Table 4.16. Ozone parameters SOMO0 and SOMO35 at (sub)urban background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against successively temperature, precipitation and wind speed. The summary of the R^2 values shows rather low or no correlations of these climatic parameters with the ozone parameters. ("n. sign." = not significant)*

R^2	SOMO0				SOMO35			
	2000	2001	2002	2003	2000	2001	2002	2003
temperature	0.04	0.17	0.10	0.08	0.03	0.17	0.08	0.04
precipitation	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n. sign.	n.sign.
wind speed	0.03	0.05	n. sign.	0.11	0.16	0.22	0.14	0.28

4.4.4 Spatial coordinates

A rather high relation is found for latitude for (sub)urban ozone and a slight relation also for longitude, see Table 4.17. In Europe the ozone concentrations increase from North to South; this is caused especially by the high mutual correlation of latitude with sunshine duration (which is at the level R^2=0.79).

Table 4.17. Ozone parameters SOMO0 and SOMO35 at (sub)urban background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against successively the latitude and longitude coordinates of the stations. The summary of the R^2 values show a clear correlations of latitude position of the station with the ozone parameters.*

R^2	SOMO0				SOMO35			
	2000	2001	2002	2003	2000	2001	2002	2003
latitude	0.25	0.29	0.24	0.24	0.28	0.38	0.27	0.19
longitude	0.03	0.02	0.05	0.06	0.10	0.09	0.15	0.12

4.4.5 Linear regression model

On the basis of equation 4.4 and the relations of ozone with different additional parameters and with the help of stepwise of backward type the linear regression model is developed. The model considered is:

$$O_3 = c + a1*EMEP + a2*altitude + a3*sunshine\ duration + a4*longitude \qquad (4.8)$$

Relative humidity, latitude and temperature are not included due their high collinearity with other parameters.

Table 4.18. Ozone measurement parameters SOMO0 and SOMO35 at (sub)urban background stations tested at four years in a linear regression with an intercept constant c *and parameters for the EMEP dispersion model combined with the altitude, the sunshine duration and longitude. The summary of the R^2 values show – compared to the values in Table 4.11 - an increased relation between the concentrations derived by combining EMEP modelled values with the additional parameters and the measured concentrations, especially at SOMO0 and SOMO35.*

linear regression model (eq. 4.8)	SOMO0				SOMO35			
	2000	2001	2002	2003	2000	2001	2002	2003
c (constant)	6118	4690	8578	5507	-2908	-3180	-3896	-1181
a1 (EMEP)	0.296	0.340	0.165	0.051	0.443	0.511	0.213	0.714
a2 (altitude)	4.058	2.589	2.289	2.682	1.917	0.826	1.555	1.909
a3 (sunsh. duration)	190.9	220.3	242.1	138.5	82.7	101.0	141.9	56.7
a4 (longitude)	104.6	83.1	170.0	123.9	59.8	68.1	147.0	84.8
R^2	**0.45**	**0.51**	**0.47**	**0.49**	**0.54**	**0.63**	**0.60**	**0.58**

4.4.6 Summary urban - ozone

Regression relations between ozone measurement parameters and different additional parameters in the urban/suburban areas were examined. Treating the urban and suburban stations as separate groups in the interpolations appears not to be of relevance. At the final linear regression model the results of EMEP model, altitude, sunshine duration and longitude are utilized.

The results of this section are applied in Sections 6.3 and 6.7 as one of the alternatives for the creation of urban maps.

4.5 Urban areas - PM_{10}

The urban/suburban measured versus modelled air quality of PM_{10}, including additional parameters, is examined. Tables 4.19 – 4.21 show that the coefficient of determination R^2 is quite low for all the supplementary data studied. The best relationship is found between measurement PM_{10} parameter values and the EMEP modelled values.

4.5.1 EMEP model

There is some relation of PM_{10} measured concentrations with the EMEP model concentrations, see Table 4.19. However, these relations are much weaker than in the case of rural pollution described at Section 4.3.1.

Table 4.19. Measured PM_{10} parameters annual averages (left) and 36th maximum daily means (right) at (sub)urban background stations tested for years 2000 - 2003 in a linear regression with an intercept constant c *against the EMEP dispersion model. The R^2 values are lower here for the urban/suburban areas than those for the rural areas of Table 4.7.*

measurement = c + a * EMEP model	**annual average**				**36th maximum daily mean**			
	2000	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**
c	17.8	13.0	15.4	19.9	26.5	22.5	27.1	33.9
a	1.03	1.25	1.07	0.79	0.90	0.90	0.76	0.62
R^2	**0.16**	**0.19**	**0.17**	**0.09**	**0.17**	**0.11**	**0.12**	**0.07**

4.5.2 Altitude

As Table 4.20 shows there is almost no relationship between PM_{10} measurements and altitude in the case of urban/suburban air quality.

Table 4.20. Measured PM_{10} parameters annual averages (left) and 36th maximum daily means (right) at (sub)urban background stations tested for years 2000 - 2003 in a linear regression with an intercept constant c *against the altitude. The R^2 values are considered to show little or no relation between altitude and both PM_{10} parameters.("n.sign." = not significant)*

R^2	**annual average**				**36th maximum daily mean**			
	2000	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**
altitude	0.05	n. sign.	0.02	0.03	0.03	n. sign.	0.02	0.03

4.5.3 Climatic parameters

Sunshine duration, relative humidity, temperature, precipitation and wind speed

The highest relation, in the case of climatic parameters, is found for relative humidity (higher humidity giving lower concentrations). However this too has mostly very low correlation.

Table 4.21. Measured PM_{10} parameters annual averages (left) and 36th maximum daily means (right) tested for years 2000 - 2003 in a linear regression with an intercept constant c *against successively sunshine duration, relative humidity, temperature, precipitation and wind speed. The summary of the R^2 values shows hardly any considerable correlation with the measurement based parameters of PM_{10}. ("n. sign." = not significant)*

R^2	**annual average**				**36th maximum daily mean**			
	2000	**2001**	**2002**	**2003**	**2000**	**2001**	**2002**	**2003**
sunshine duration	0.17	0.02	0.01	0.02	0.17	0.02	n. sign.	n. sign.
relative humidity	0.18	0.02	0.04	0.04	0.16	0.02	0.03	0.03
temperature	0.09	0.02	n. sign.	n. sign.	0.06	n. sign.	n. sign.	n. sign.
precipitation	n. sign.	n. sign.	0.02	n. sign.	n. sign.	n. sign.	0.01	n. sign.
wind speed	0.05	n. sign.	0.02	0.02	0.04	n. sign.	0.03	0.01

4.5.4 Spatial coordinates

There relationship of measured PM_{10} concentrations with latitude and longitude is examined. There is some, but mostly low correlation.

Table 4.22. Measured PM_{10} parameters annual averages (left) and 36th maximum daily means (right) at (sub)urban background stations tested for the years 2000 – 2003 in a linear regression with an intercept constant c *against successively the latitude and longitude coordinates of the stations. The summary of the R^2 values show rather little relation between station location and both PM_{10} parameters.*

R^2	annual average				36th maximum daily mean			
	2000	2001	2002	2003	2000	2001	2002	2003
latitude	0.21	0.04	0.03	0.04	0.20	0.03	0.01	0.02
longitude	n. sign.	0.04	0.11	0.12	0.03	0.05	0.18	0.14

4.5.5 Linear regression model

The linear regression model considered here for urban/suburban PM_{10} air quality is:

$$PM_{10} = c + a1*EMEP + a2*altitude + a3*sunshine\ duration + a4*wind\_speed + a5*longitude \quad (4.9)$$

The individual variables are chosen on the basis of stepwise of backward type (see Section 3.1).

The estimated parameters of this linear model *c, a1, a2, a3, a4* and the R^2 for each year are presented in Table 4.23. (In case of 36th maximum daily means, the dispersion model for annual average is used for the years 2001–2003, while the modelled 36th max. daily means are used for the year 2000.) The results show that the PM_{10} variability explained by this regression model ranges between 30 and 40 percent (since R^2 shows values mostly between 0.3 – 0.4), which evaluates the quality of this model.

Table 4.23. Measurement based PM_{10} parameters annual averages (left) and 36th maximum daily means (right) at (sub)urban background stations tested at four years 2000 - 2003 in a linear regression with an intercept constant c *against the EMEP dispersion model combined with the altitude, sunshine duration, wind speed and longitude. ("n. sign." = not significant)*

linear regression model (eq. 4.9)	annual average				36th maximum daily mean			
	2000	2001	2002	2003	2000	2001	2002	2003
c (constant)	n. sign.	n. sign.	-4.9	-9.1	-18.0	-10.9	n. sign.	-16.9
a1 (EMEP model)	1.01	1.22	1.15	1.06	1.12	2.20	1.91	1.84
a2 (altitude)	n. sign.	n. sign.	0.0080	0.0089	n. sign.	n. sign.	0.0103	0.0158
a3 (sunshine duration)	0.53	0.25	0.33	0.44	1.16	0.52	0.46	0.69
a4 (wind speed)	n. sign.	0.80	0.94	1.14	n. sign.	2.29	n. sign.	2.46
a5 (longitude)	n. sign.	0.16	0.34	0.44	n. sign.	0.36	0.68	0.86
R^2	**0.36**	**0.28**	**0.33**	**0.29**	**0.42**	**0.27**	**0.36**	**0.29**

4.5.6 Summary urban - PM_{10}

Regression relations between PM_{10} measurement parameters and different parameters in the urban/suburban areas were investigated. Finally, a linear regression model was developed, using the results of EMEP model, altitude, sunshine duration, wind speed and longitude.

The results of this section are applied in Sections 6.3 and 6.7 as one of the alternatives for the creation of urban maps.

4.6 Overall conclusion

The relationships of observed ozone and PM_{10} with different spatial, climatic and model parameters have been investigated, for both rural and urban/suburban background stations. In both rural and urban areas a linear regression model is developed for every pollutant and indicator. In all cases the results of EMEP dispersion model, altitude and sunshine duration are used. In case of AOT40 in the rural areas, relative humidity is used in addition. In the urban areas longitude is used moreover, and for PM_{10} also wind speed.

In case of rural areas, the linear regression will be further applied in the creation of rural maps through interpolation, Chapter 5. In case of urban areas, it will be applied in Chapter 6 as one methodology for interpolating urban air quality.

5 Rural map interpolation

5.1 Introduction

The interpolation methods outlined in Chapter 2, along with the regression relationships established in Chapter 4, are now applied and the quality of the methods is tested through cross-validation (using root-mean-square error, RMSE). Tests are carried out separately for ozone and PM_{10}. The EMEP dispersion model (both unfitted and fitted by linear regression into the level of measured concentrations) is also tested, for comparison, using the similar methodology (RMSE, see Section 4.1). A summary of the using methods, stated in Chapter 2, is given below for reference. Not all methods are applied to all the tests and only relevant supplementary data is applied.

0. EMEP model, without interpolation
    a. Unfitted model
    b. Fitted model
1. Pure interpolation methods using monitoring data only
    a. IDW
    b. Ordinary kriging (OK)
    c. Ordinary cokriging with supplementary data (OC)
    d. Lognormal kriging (LK)
    e. Lognormal cokriging (LC)
2. Interpolation using monitoring and modelling data
    A. Plain subtraction of the (unfitted) EMEP model and measurements with interpolation of the residuals
        a. IDW
        b. Ordinary kriging (OK)
        c. Ordinary cokriging with supplementary data (OC)
    B. Combination using the fitted EMEP model with interpolation of the residuals
        a. IDW
        b. Ordinary kriging (OK)
        c. Ordinary cokriging with supplementary data (OC)
3. Interpolation using monitoring, modelled and supplementary data – combination by linear regression (EMEP model and supplementary data) with interpolation of the residuals
    a. IDW
    b. Ordinary kriging (OK)

5.2 Comparison of different interpolation methods, ozone

A comparison of the different interpolation methods is performed for the three ozone indicators (SOMO0, SOMO35 and AOT40) for the four-year period, 2000-2003. The three groups of methods, as outlined above, are applied and compared with the fitted and unfitted EMEP model values. (For further comparison with linear regression relationships without interpolation, see Annex I, Tables A5-A7). Due to time limitations the plain subtraction Methods 2A for all ozone indicators have not been carried out. Because of normal distribution of ozone data (against the area) there were not performed lognormal kriging and cokriging (Methods 1d and 1e).

In Tables 5.1 to 5.3 the comparison results over the four-year period are given. The values indicated in the tables are the RMSE (see Section 2.4 and Part I, Chapter 7) determined from cross-validation tests. The methods are ranked within groups and between groups with a ranking of 1 indicating the methodology that gives the lowest RMSE, i.e. the best interpolated estimates compared to the measurements.

Table 5.1. Comparison of the different interpolation methods showing the RMSE (in µg.m^{-3}.days) for SOMO0. The last two columns show the four-year average test results ranked with the best (lowest) RMSE results per method and per group marked as 1.

method		2000	2001	2002	2003	avg.	ranking	
0a	unfitted EMEP model	4060	3594	4195	3846	3924	2	4
0b	fitted EMEP model	3835	3561	3842	3692	3732	1	
1a	IDW	3523	3407	3476	3438	3461	4	2
1b	ord. kriging	3526	3379	3466	3421	3448	3	
1c	ord. cokriging (altitude)	3120	3032	2966	3116	3059	2	
1c	ord. cokriging (altitude, sunsh. dur.)	3038	2848	2784	2710	2845	1	
2Ba	comb. EMEP - fitted, IDW	3689	3365	3485	3492	3508	3	3
2Bb	comb. EMEP - fitted, OK	3539	3336	3465	3479	3455	2	
2Bc	comb. EMEP - fitted, OC (alt., s. d.)	3335	2997	3032	3194	3140	1	
3a	comb. EMEP+alt.+sunsh.d., IDW	2638	2539	2598	2760	2634	2	1
3b	comb. EMEP+alt.+sunsh.d., OK	2675	2520	2573	2747	2629	1	

Table 5.2. Comparison of the different interpolation methods showing the RMSE (in µg.m^{-3}.days) for SOMO35, including ranking of the four-year average from best (1) to weakest per method and per group.

method		2000	2001	2002	2003	avg.	ranking	
0a	unfitted EMEP model	2411	2123	2551	3074	2540	2	4
0b	fitted EMEP model	2357	2119	2417	2513	2351	1	
1a	IDW	2082	1998	2129	2305	2129	4	2
1b	ord. kriging	2055	1942	2079	2259	2084	3	
1c	ord. cokriging (altitude)	1900	1661	1836	2155	1888	2	
1c	ord. cokriging (altitude, sunshine dur.)	1704	1668	1873	2085	1833	1	
2Ba	comb. EMEP - fitted, IDW	2148	2028	2163	2411	2188	3	3
2Bb	comb. EMEP - fitted, OK	2113	1981	2096	2394	2146	2	
2Bc	comb. EMEP - fitted, OC (alt., s. d.)	1866	1704	1792	2199	1890	1	
3a	comb. EMEP+alt.+sunsh.dur., IDW	1629	1582	1649	2002	1716	1	1
3b	comb. EMEP+alt.+sunsh.dur., OK	1664	1564	1641	2001	1718	2	

Table 5.3. Comparison of the different interpolation methods showing RMSE (in µg.m^{-3}.hours) for AOT40, including ranking of the four-year average from best (1) to weakest per method and per group.

method		2000	2001	2002	2003	avg.	ranking	
0b	fitted EMEP model	6955	5376	7128	7744	6801	1	4
1a	IDW	6044	5103	5083	6953	5796	3	2
1b	ord. kriging	6028	4859	4577	6749	5553	2	
1c	ord. cokriging (altitude)	5843	4810	4571	6722	5487	1	
2Ba	comb. EMEP - fitted, IDW	6345	5144	5392	7090	5993	3	3
2Bb	comb. EMEP - fitted, OK	6324	5056	5199	7011	5898	2	
2Bc	comb. EMEP - fitted, OC (alt.)	5849	4672	5063	6946	5633	1	
3a	comb. EMEP+alt.+s.d.+r.h., IDW	5755	4595	4477	6634	5365	2	1
3b	comb. EMEP+alt.+s.d.+r.h., OK	5681	4535	4366	6647	5307	1	

The root-mean-square-error is given in the same units as the resulting map with interpolated ozone parameter values for the rural area. It expresses the uncertainty of the map.

When comparing the RMSE values with the individual ozone parameters for each year within a method, a similar tendency over all the years is observed. At all cases the best group of methods, i.e. the group which provides the lowest RMSE, is the combination of the EMEP model with atitude, sunshine duration (and in case of AOT40 relative humidity) using linear regression and interpolation of the residuals (group 3 of Section 5.1). They have as group in the above three tables a number 1 ranking as best result.

The geostatistical methods (i.e. OC and OK) give at all cases in general better results than IDW. This is most obvious at the two groups of Methods 1 and 2B. In the last group (3) the difference between IDW and OC is however minimal.

The inclusion of supplementary data, either through cokriging or through regression relationships, substantially reduces the RMSE in all cases.

The resulting maps for the three ozone indicators and for a number of the interpolation methods applied are presented in Figures 5.1 to 5.3. At each ozone parameter the mapping result of the method with the best RMSE in each group is presented as example as follows:

Top left: Pure interpolation of monitoring data – ordinary cokriging (OC), using altitude and sunshine duration (Method 1c)

Top right: Combination using the fitted EMEP model and ordinary cokriging of the residuals using altitude and sunshine duration (Method 2Bc).

Bottom left: Combination using linear regression of the EMEP model, altitude, sunshine duration and interpolation of the residuals using IDW (Method 3a)

Bottom right: Combination using linear regression of the EMEP model, altitude, sunshine duration and interpolation of the residuals using ordinary kriging (Method 3b)

The four maps in each figure are presented for comparison purposes. The results for the year 2002 are chosen, since it is commonly recognized as the most typical year out of the four, contrary to e.g. the climatological more extreme year 2003).

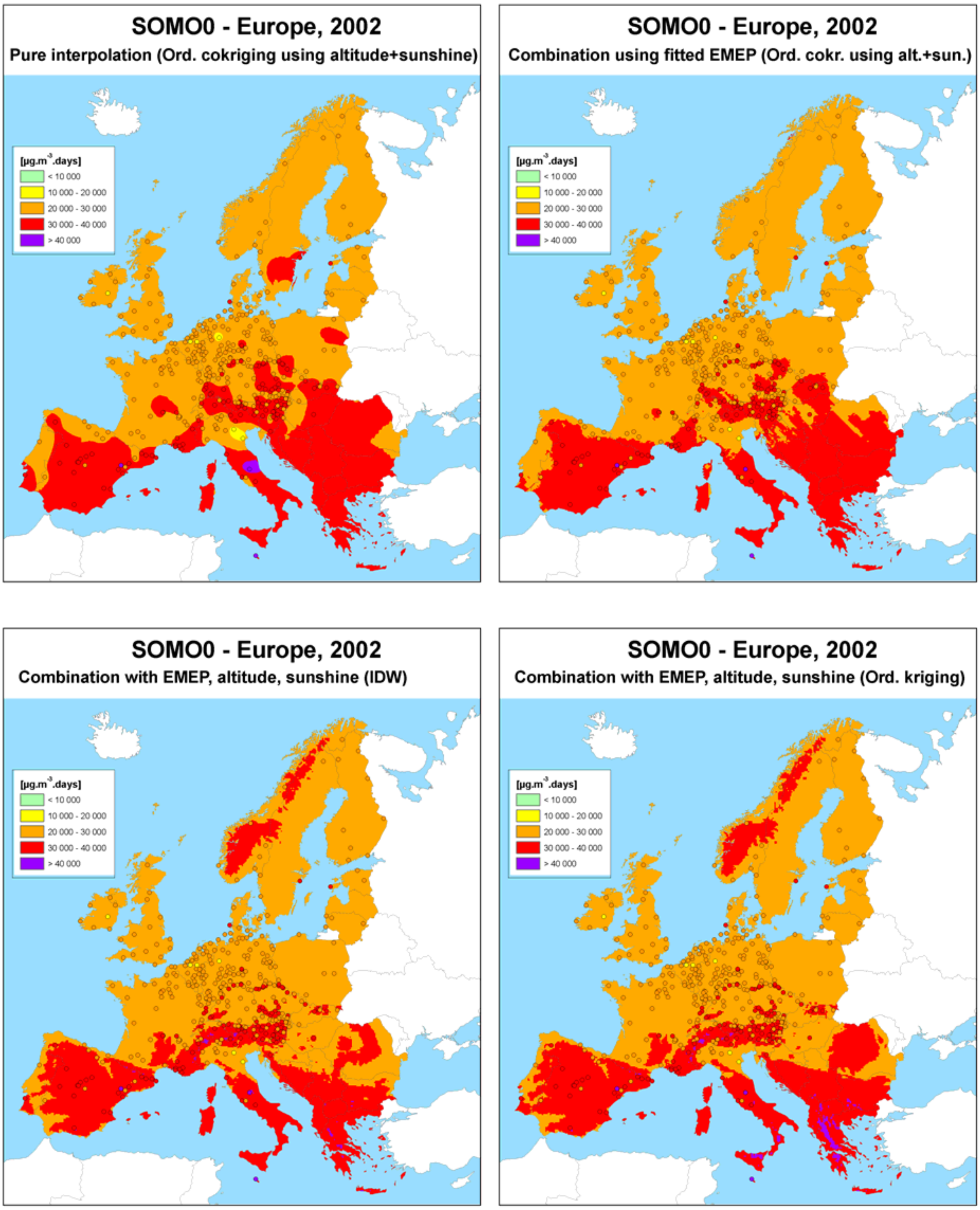


Figure 5.1. Maps showing the SOMO0 ozone parameter concentrations (µg.m^{-3}.days) on European scale for rural areas in a 10 km x 10 km grid resolution as a result of four different interpolation Methods. 1c (top left), 2Bc (top right), 3a (bottom left) and 3b (bottom right), using 2002 rural background monitoring data, whether or not combined with modelling and/or other supplementary parameters.

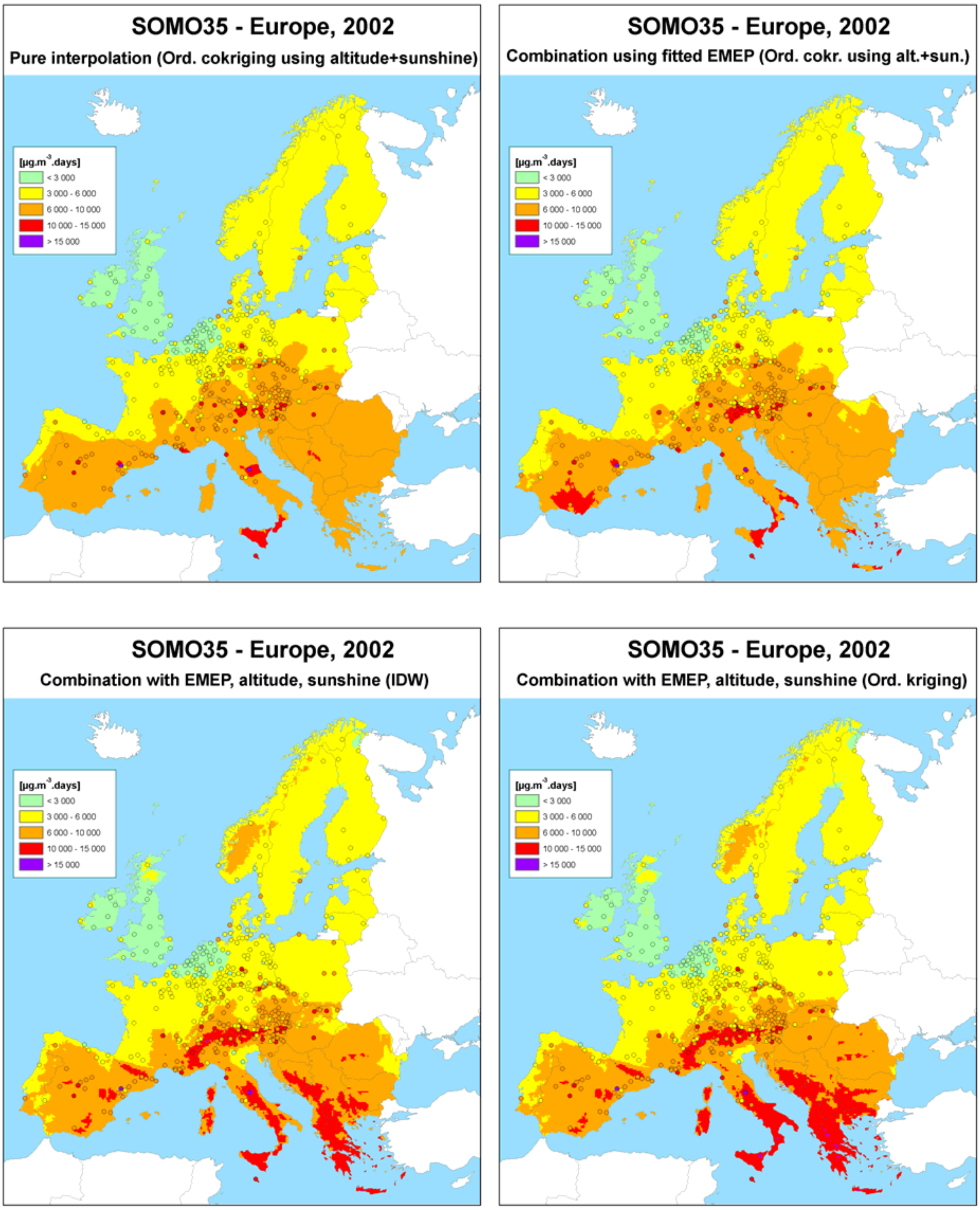


Figure 5.2. Maps showing the SOMO35 ozone parameter concentrations (µg.m^{-3}.days) on European scale for rural areas in a 10 km x 10 km grid resolution as a result of four different interpolation Methods 1c (top left), 2Bc (top right), 3a (bottom left) and 3b (bottom right), using 2002 rural background monitoring data, whether or not combined with modelling and/or other supplementary parameters.

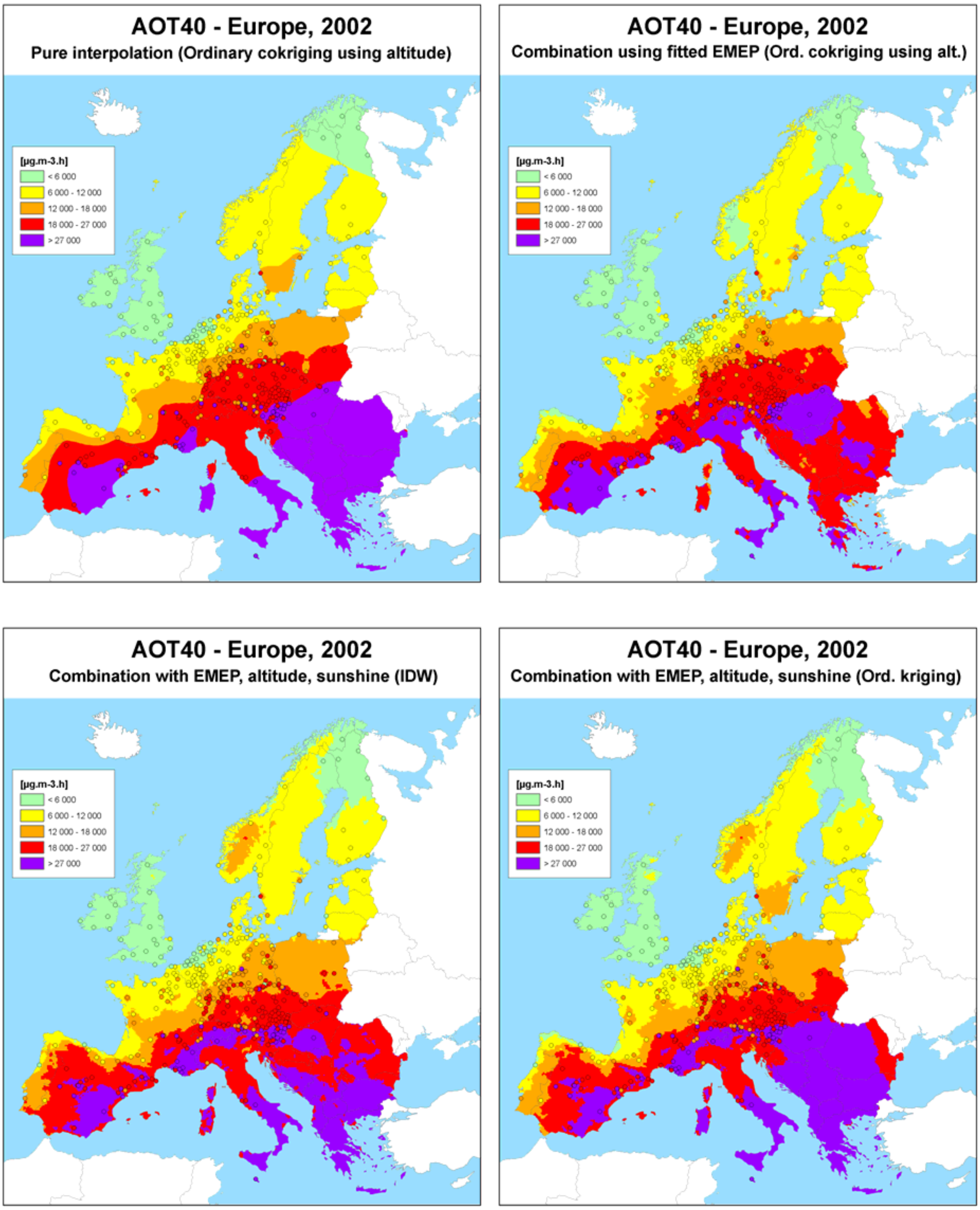


Figure 5.3. Maps showing the AOT40 ozone parameter concentrations (µg.m^{-3}.hours) on European scale for rural areas in a 2 km x 2 km grid resolution as a result of four different interpolation Methods 1c (top left), 2Bc (top right), 3a (bottom left) and 3b (bottom right), using 2002 rural background monitoring data, whether or not combined with modelling and/or other supplementary parameters.

5.3 Comparison of different interpolation methods, PM_{10}

The comparison of different interpolation methods is performed for the two PM_{10} indicators (annual averages and PM_{10} 36th maximum daily average values) for the four-year period, 2000–2003. The three groups of methods, outlined Section in 5.1, are applied and compared with the fitted and unfitted EMEP model values. (For further comparison with linear regression relationships without interpolation, see Annex I, Tables A8-A9). For PM_{10} lognormal kriging (LK) and lognormal cokriging (LC) are also tested, because of the lognormal distribution of PM_{10} concentrations (against area).

In Tables 5.4 and 5.5 the comparison over the four-year period is given. The values indicated in the tables are the RMSE determined from cross-validation tests. The methods are ranked within groups and between groups with a ranking of 1 indicating the methodology that gives the lowest RMSE in cross-validation, i.e the best interpolated estimates compared to the measurements (see Chapter 2 and Part I, Chapter 7).

Table 5.4. Comparison of the different interpolation methods showing RMSE (in µg.m^{-3}) for the annual average PM_{10} concentrations. The last two columns show the four-year average test results ranked with the best (lowest) RMSE results per method and per group marked as 1.

method		2000	2001	2002	2003	avg.	ranking	
0a	unfitted EMEP model	12.31	12.35	13.54	13.63	12.78	2	5
0b	fitted EMEP model	5.58	6.83	7.45	7.97	6.96	1	
1a	IDW	5.32	6.55	7.19	7.29	6.59	5	2
1b	ord. kriging	5.11	6.24	6.73	7.08	6.29	4	
1c	ord. cokriging (altitude)	3.67	5.96	6.23	5.86	5.43	2	
1d	lognorm. kriging	5.00	6.12	6.61	6.89	6.15	3	
1e	lognorm. cokriging (altitude)	3.81	5.72	5.79	5.44	5.19	1	
2Aa	comb. EMEP - plain, IDW	4.67	5.82	6.44	6.71	5.91	3	4
2Ab	comb. EMEP - plain, OK	4.98	5.67	6.15	6.66	5.86	2	
2Ac	comb. EMEP - plain, OC (alt.)	3.72	5.40	6.12	6.68	5.48	1	
2Ba	comb. EMEP - fitted, IDW	4.60	5.44	6.25	6.60	5.72	3	3
2Bb	comb. EMEP - fitted, OK	4.36	5.42	6.05	6.55	5.59	2	
2Bc	comb. EMEP - fitted, OC (alt.)	3.21	5.39	6.05	6.82	5.37	1	
3a	comb. EMEP+alt.+sunsh.d., IDW	3.26	5.01	5.59	5.98	4.96	2	1
3b	comb. EMEP+alt.+sunsh.d., OK	3.19	4.95	5.48	5.71	4.84	1	

Table 5.5. Comparison of the different interpolation methods showing RMSE (in µg.m^{-3}) for the 36th maximum daily average PM_{10} concentrations. The last two columns show the four-year average test results ranked with the best (lowest) RMSE results per method and per group marked as 1.

method		2000	2001	2002	2003	avg.	ranking	
0a	unfitted EMEP model	17.26	18.00	19.29	21.83	19.38	2	4
0b	fitted EMEP model	8.48	11.73	11.65	15.44	11.82	1	
1a	IDW	8.36	11.11	10.55	13.52	10.89	3	2
1b	ord. kriging	7.99	10.55	10.39	13.50	10.61	2	
1c	ord. cokriging (altitude)	5.49	9.96	9.36	12.05	9.22	1	
2Ba	comb. EMEP - fitted, IDW	6.87	9.04	9.59	12.81	9.58	3	3
2Bb	comb. EMEP - fitted, OK	6.18	9.26	9.33	12.70	9.37	2	
2Bc	comb. EMEP - fitted, OC	6.18	9.28	9.33	12.71	9.38	1	
3a	comb. EMEP+elev.+sun., IDW	5.83	8.37	8.44	11.88	8.63	2	1
3b	comb. EMEP+elev.+sun., OK	5.53	8.09	8.53	11.64	8.45	1	

The root-mean-square-error is given in $\mu g.m^{-3}$. It expresses the uncertainty of the map.

When comparing the RMSE values with the individual PM_{10} parameters for each year within a method, a similar tendency over all the years is observed. At all cases the best group of methods, i.e. the group which provides the lowest RMSE, is the combination of the EMEP model with altitude, and sunshine duration using linear regression and interpolation of the residuals (group 3 of Section 5.1).

The geostatistical methods (i.e. OC and OK) give at both PM_{10} parameters in general better results than IDW. This is most obvious at the groups of Methods 1, 2A and 2B. In the last group (3), the difference between IDW and OC is however minimal.

Lognormal interpolation is only carried out with the pure interpolation methods of group 1. In these cases it gives better results than kriging without logarithmic transformation. It is possible that this is also the case when applied to the other methods, but this has not been tested. For the 36th maximum daily average lognormal kriging is not carried out. However, it is recommended to investigate this option on its possible contributions to further improvement of the interpolation methodology at this parameter as well.

The inclusion of supplementary data, either through cokriging or through regression relationships, substantially reduces the RMSE in all cases, except for the inclusion of temperature.

The resulting maps for annual average PM_{10} and for a number of the interpolation methods applied are presented in Figure 5.4. The examples represent the following methods:

Top left: Pure interpolation of monitoring data – lognormal cokriging (LC), altitude (Method 1c variant)

Top right: Combination using the fitted EMEP model and ordinary cokriging of the residuals using altitude (Method 2Bc)

Bottom left: Combination using linear regression of the EMEP model, altitude and sunshine duration and interpolation of the residuals using IDW (Method 3a)

Bottom right: Combination using linear regression of the EMEP model, altitude and sunshine duration and interpolation of the residuals using ordinary kriging (Method 3b)

The four maps are presented for comparison purposes. The presented results are for the year 2002. The influence of the EMEP model can be seen especially in the areas without measurements, e.g. in Scandinavia. The influence of orography is visible in the two bottom maps.

5.4 Interpolation method selected for rural mapping

On the basis of the comparisons made in Sections 5.2 and 5.3 the best interpolation method for map construction has been selected. It is Method 3b, the combination of measured and supplementary data through linear regression with interpolation of the residuals using ordinary kriging. This method is consequently applied for all further rural and rural - urban merging interpolations.

The maps of AOT40 are created at a resolution of 2x2 km grid squares, the other maps (i.e. SOMO0, SOMO35, PM_{10}) are produced at a resolution of 10x10 km grid squares. These resolutions were considered as sufficient for mapping the rural areas on European scale.

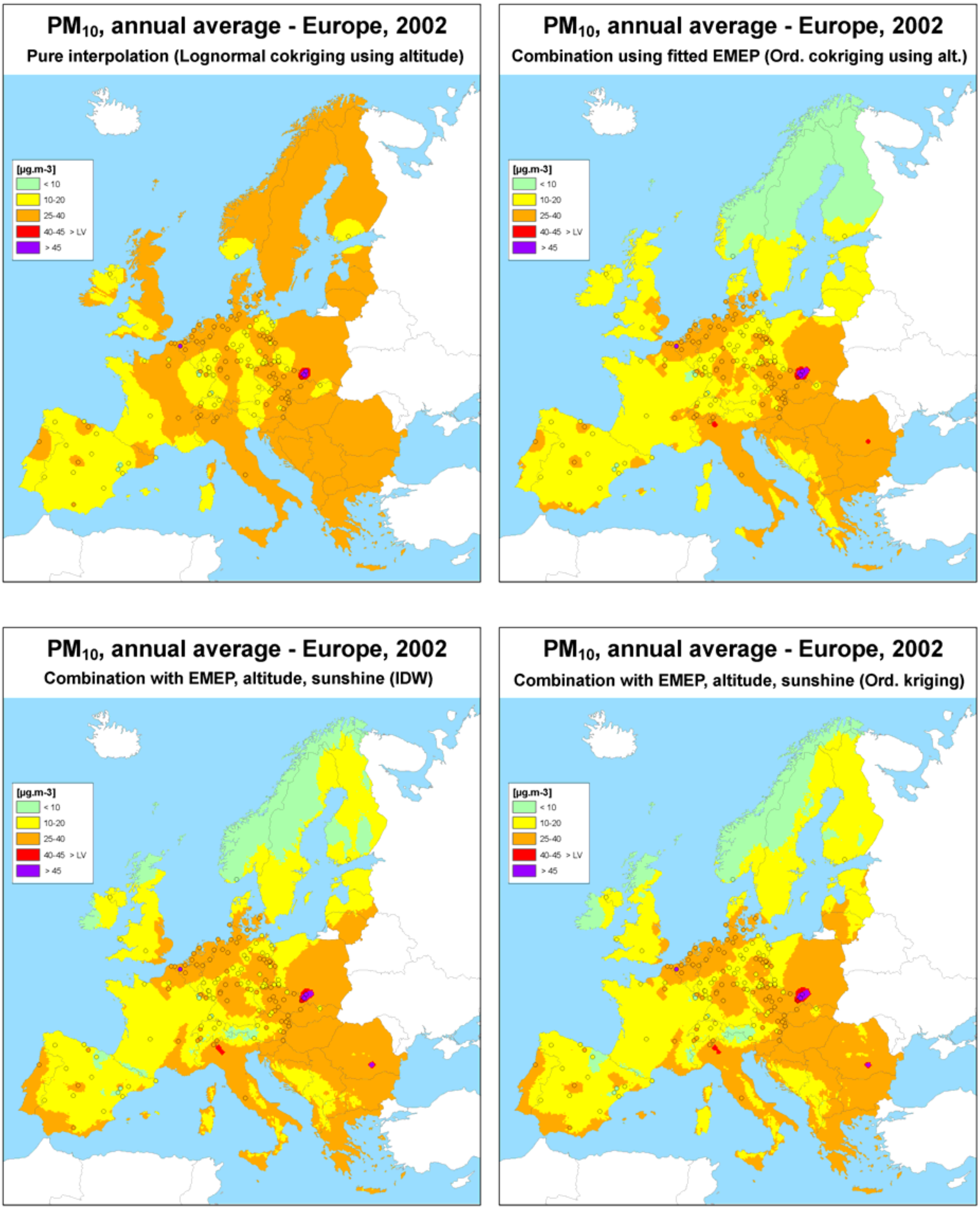


Figure 5.4. Maps showing the annual average PM_{10} concentrations ($\mu g.m^{-3}$) on European scale for rural areas in a 10 km x 10 km grid resolution as a result of four different interpolation Methods 1c (top left), 2Bc (top right), 3a (bottom left) and 3b (bottom right), using 2002 rural background monitoring data, whether or not combined with modelling and/or other supplementary parameters.

6 Urban map compilation

6.1 Introduction

The most important question to be solved in creating a concentration map for the urban areas on European scale is: How to map the cities with no measurements?

Because of the different nature of urban and rural air quality it is appropriate not to create air quality urban maps by interpolation based on neighbouring rural stations. Other possibilities are considered and tested in this chapter.

In some of the methods examined in this chapter the so-called Delta is applied. This is the difference between the urban and rural background concentrations or indicators. For all urban/suburban background stations Deltas are calculated for different pollutants and indicators, using the selected rural interpolation method (Section 5.4) as rural background concentration:

$$\Delta(s_i) = Z_{urb}(s_i) - \hat{Z}_{rur}(s_i) \tag{6.1}$$

where $Z_{urb}(s_i)$ is the measured value in the point s_i , being a urban/suburban station,

$\hat{Z}_{rur}(s_i)$ is the estimated value of the rural background field in the point s_i,

$\Delta(s_i)$ is Delta in the point s_i.

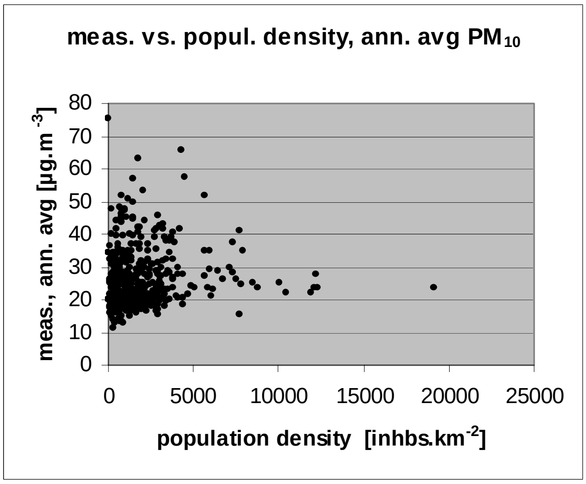
At first, in Sections 6.2 to 6.4 the approaches using linear regression are considered. Then, in Sections 6.5 and 6.6, the methods using interpolation outside the borders of the cities are discussed. In the Section 6.7 the introduced methods are mutually compared in order to examine their quality for the estimation of urban concentrations. On the basis of this comparison the method for consequent usage in the mapping of urban areas is selected.

6.2 Relation between urban concentrations (or Delta) and population density

The relation between urban/suburban measured concentrations (or Delta) and population density is investigated, to examine the possibility to simulate the urban concentrations in the cities with no measurements on the basis of this parameter. In the case of population density several grid resolutions are considered: 100 m, 1 km, 10 km and 20 km. Regression analysis indicates very poor correlation between population density and concentrations at all resolutions. The results for 10 km x 10 km grid are shown in Figure 6.1 and in Tables 6.1 and 6.2.

The 10 km x 10 km grid is derived from the original 100 m x 100 m raster in two steps. In the first step the grid 1 km x 1 km is created by merging and averaging 100 cells into one big cell. The second step is done similar by joining together 100 cells of the 1 km x 1 km grids created in the first step into one big cell of 10 km x 10 km. This second step is done for all 1 km x 1 km grids, in order to assure that all stations would be in the centre of the 10 km x 10 km grid cells. (The output is in fact the grid 1 km x 1 km, the values of each cell represent the average population density in surrounding area 10 km x 10 km).

The aggregation into a lower grid resolution is needed to keep the interpolation calculations within the limits of the calculation capacity of the hardware. Furthermore, the very high resolution of the 100 m raster is beyond the accuracy we can reach on the European scale air quality mapping, even if it concerns urban air quality.



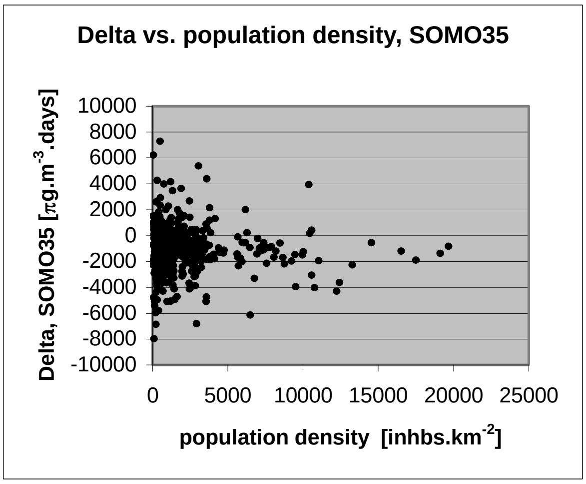


Figure 6.1 The graphs represent the linear regression results indicating the poor correlation between the annual average PM_{10} concentrations (left) in µg.m^{-3} and for ozone the SOMO35 Delta values (right) in µg.m^{-3} of measurements of 2002 at the urban/suburban background stations (y- axis) versus the population density data (inhbs.km^{-2}), that are step-wise aggregated from its original 100 m raster into a 10 km x 10 km grid resolution (x-axis).

Table 6.1. Ozone parameters SOMO0 and SOMO35 (left) and PM_{10} parameters annual averages and 36th maximum daily means (right) from measurements at urban/suburban background stations tested for the years 2000 – 2003 in a linear regression against the population density on the aggregated 10 km x 10 km grid resolution. The values are the coefficient of determinant R^2 and show little to no relation of the ozone or PM_{10} parameters with the 10 km x 10 km grids of population density.

ozone	2000	2001	2002	2003
SOMO0	0.09	0.01	0.03	0.04
SOMO35	0.05	n. sign.	0.01	0.02

PM_{10}	2000	2001	2002	2003
annual aerage	n. sign.	0.01	n. sign.	n. sign.
maximum 36th d. m.	n. sign.	0.01	n. sign.	n. sign.

Table 6.2. The Delta values for the ozone parameters SOMO0 and SOMO35 (left) and PM_{10} parameters annual averages and 36th maximum daily means (right) at urban/suburban background stations tested for the years 2000 – 2003 in a linear regression against the population density on the aggregated 10 km x 10 km grid resolution. The values are the coefficient of determinant R^2 and show little to no relation of the ozone or PM_{10} parameters with the 10 km x 10 km grids of population density.

ozone	2000	2001	2002	2003
SOMO0	0.05	n. sign.	0.05	n. sign.
SOMO35	n. sign	n. sign.	n. sign.	n. sign.

PM_{10}	2000	2001	2002	2003
annual aerage	0.03	0.04	0.03	0.01
maximum 36th d. m.	0.04	0.02	0.02	0.01

The situation changes slightly when the stations are binned into population density classes and then the air quality parameter values at the stations within individual classes are averaged out, instead assigning to each station's air quality parameter value the individual population density value detected at the location of the station. The results can be seen in Figure 6.2. For these air quality parameter averages of the stations assigned to the population density classes there is some relation with the classes of the population density. However, the variation within classes is large, see Figure 6.3.

The conclusion of this section is that there is very poor relation between the measured concentrations and population density in case of individual stations, without reference to grid resolution. Some relation shows up only when aggregation of population density into classes is done. The examined relations will be further utilized in the merging of urban and rural maps in Chapter 7.

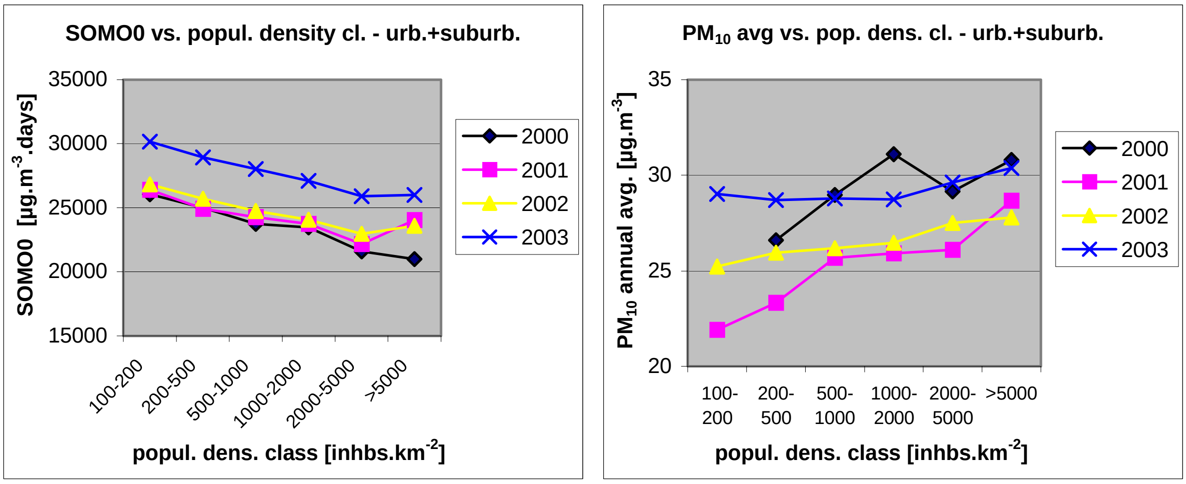


Figure 6.2. The graphs represent the relation of SOMO0 (left) and annual average of PM_{10} (right) with population density, if the averages of SOMO0 and PM_{10} of the urban/suburban stations in the population density classes are considered. Each line graph represents a separate year. The graphs with the average values of the population density classes show a decline for the SOMO0 and an increase for the PM_{10} annual average.

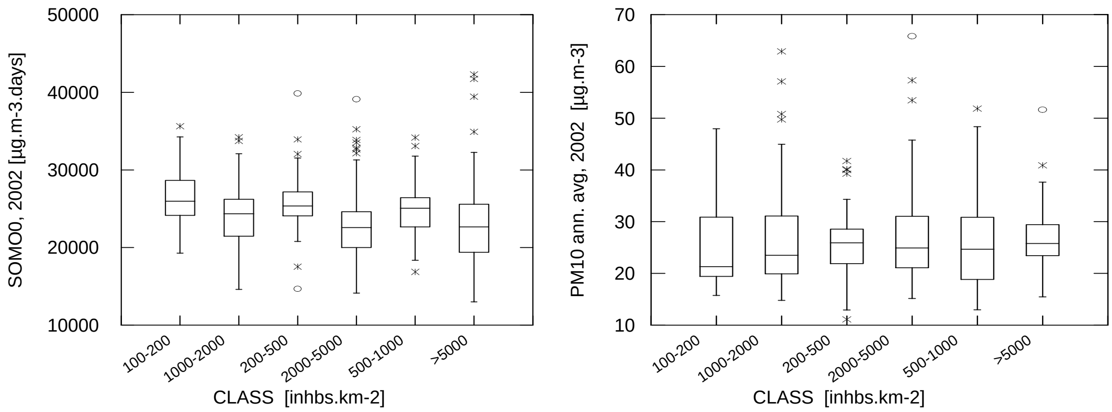


Figure 6.3 Ozone parameter SOMO0 (left) and PM_{10} annual averages (right) from measurements at urban/suburban background stations of the year 2002 (y-axis) versus the population density classes presented as box-and-whisker plot. For all population density classes the graphs show the median, 25th and 75th percentile ("box"), the fences ("whiskers") and the outliers. Big variability of the air quality parameter values within the individual population density classes is obvious.

6.3 Relation between urban concentrations and different supplementary data

The linear regression models based on available supplementary data, which were developed in Sections 4.4 and 4.5, can be utilized (see Tables 4.18 and 4.21). For comparison reasons, their multiple squared correlation coefficients R^2 are given again in Table 6.3. It can be seen that R^2 is quite high for ozone (0.5 – 0.6), but lower for PM_{10} (0.3-0.4). However, all these relations are more useful for urban mapping than the relations with population density (see Table 6.1).

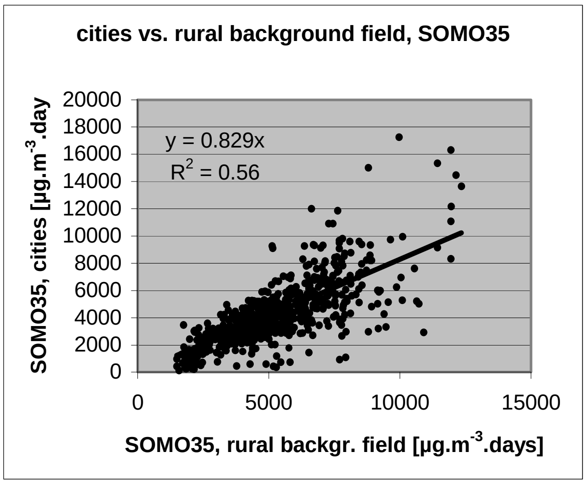
Table 6.3. Ozone parameters (left) and PM_{10} parameters (right) in linear regression models with supplemtary parameters. The values are the coefficient of determination R^2.

ozone	2000	2001	2002	2003
SOMO0	0.45	0.51	0.47	0.49
SOMO35	0.54	0.63	0.60	0.58

PM_{10}	2000	2001	2002	2003
annual aerage	0.36	0.28	0.33	0.29
maximum 36th d. m	0.42	0.27	0.36	0.29

6.4 Relation between urban concentrations and the rural background pollution field

The measured concentrations at the urban/suburban stations are compared to the underlying rural background field (in our case Method 3b, Chapter 5), for each pollutant and indicator. Figure 6.4 shows two examples of scatter plots and their associated linear regression relations.



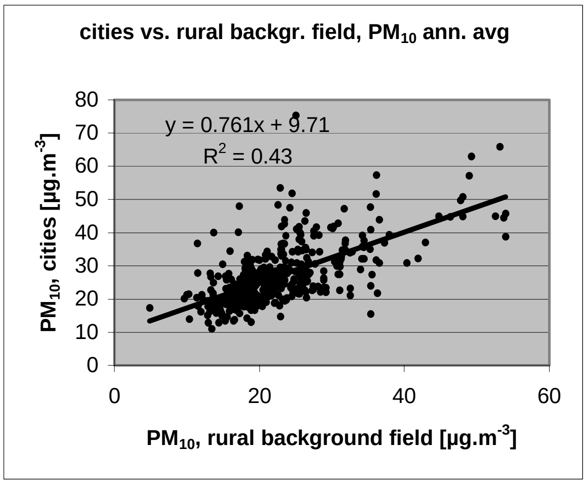


Figure 6.4 Regression relation between urban/suburban measured concentrations and underlying rural background concentration field for SOMO35 (left) and PM_{10} annual average (right) for the year 2002.

A summary of these two investigated relations for SOMO35 and PM_{10} annual averages are presented in Tables 6.4 and 6.5.

*Table 6.4 Regression parameters for SOMO35: urb = c + a * rural_map*

SOMO35	2000	2001	2002	2003
c	n. sign.	n. sign.	n. sign.	n. sign.
a	0.75	0.83	0.83	0.84
R^2	**0.54**	**0.51**	**0.56**	**0.41**

*Table 6.5 Regression parameters for PM_{10} (ann. avg): urb = c + a * rural_map*

PM_{10}, ann. avg	2000	2001	2002	2003
c	11.5	12.3	9.7	11.2
a	0.79	0.62	0.76	0.74
R^2	**0.29**	**0.33**	**0.43**	**0.38**

The correlation between the urban and rural background concentrations are as high as those for the supplementary data regressions described in Section 6.3. This clear correlation can be explained by two causes. Firstly, the urban concentrations for both ozone and PM_{10} are to large extend determined by the rural concentrations. Secondly, a number of the supplementary datasets used in the production of the rural maps are to the same extend relevant to the urban concentrations.

6.5 Interpolation of urban stations outside the city borders

A relationship between urban concentrations and latitude/longitude has been found at 4.4.4 and 4.5.4, particularly for ozone and latitude. This indicates that interpolation on air quality parameter values of urban/suburban stations outside of the borders of the cities might be useful leading to an improved interpolated urban mapping result. This approach is based on the idea that one can expect certain similarities in the air quality of neighbouring cities.

The interpolation is done the same way as for rural areas, using the methodology introduced in Chapter 2 (especially IDW and ordinary kriging). All the urban/suburban measured data are taken as the input data. The output is the map of European urban pollution, which is applicable only for the urban areas.

The advantage of the interpolation outside the border of the cities is that one and the same basic methodology for cities both with and without measurement stations can be used. However, more investigation is needed to confirm this as a realistic and appropriate method.

6.6 Interpolation of Deltas outside the city borders in addition with the rural background field

This approach is similar to the one in Section 6.5, except that it uses Deltas (see Section 6.1) instead of the measured urban concentrations in the interpolation, and that it adds this interpolation result into the rural background map.

The advantage of this approach is its simplicity: the already existing interpolation results of the rural background concentration field are utilized. This contains already interpolation improvements reached by utilizing the supplementary parameters, which are supposed to also apply to urban/suburban areas. As such, this one-time utilisation of additional parameters should guarantee the improved results and would affect the urban/suburban areas to a similar extend.

As example of this approach, Figure 6.5 shows the European map with the interpolated urban PM_{10} annual average concentrations for the year 2003. However, due to the definition of Delta this map shows in fact the air quality for the whole Europe as if it is one large urban area. In reality this is of course not the case. The map is applicable only in the urban areas and needs an additional merging with the map of ruaral air quality that is applicable for the rural areas. Chapter 7 deals with that merging.

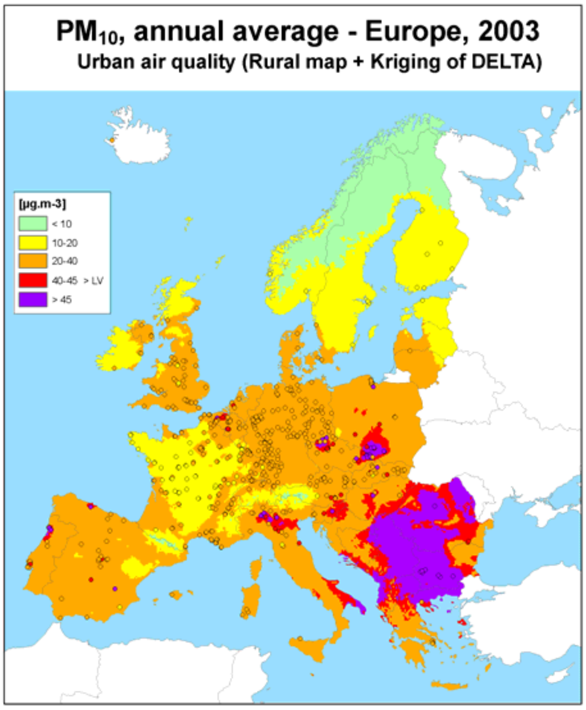


Figure 6.5. European map showing the interpolation result for the Delta values of PM_{10}, annual average for 2003. The use of Delta intrinsically deal with Europe as if it is one large urban area which is not according reality and needs refinement in its mapping methodology.

6.7 Comparison of different approaches

In this section a comparison of different interpolation methods for urban/suburban areas is performed to find the most suitable method to apply. The interpolations are done with a subset of 369 urban/suburban background stations for PM_{10} annual averages of 2003. The subset is composed such that it uses only one station per city (randomly selected). This is intended to simulate in cross-validations the cities with no measurement. (If there are more stations in one city, it would not be possible.) Different methods for the mapping of the urban air quality are compared by RMSE (Chapter 2 and Part I, Chapter 7) through cross-validation and the results are presented in Table 6.6. A summary of the methods applied is given below.

1. Regression relationships
    a. Linear regression between measured urban/suburban concentrations and population density (Section 6.2)
    b. Linear regression model between urban/suburban concentrations and other supplementary data (Section 6.3)
    c. Linear regression between measured urban/suburban concentrations and rural background concentrations field (Section 6.4)
2. Interpolation of measured urban concentrations outside city borders (Section 6.5)
    a. IDW
    b. Ordinary kriging (OK)
3. Interpolation of Deltas outside city borders in addition to the rural background field (Section 6.6)
    a. IDW
    b. Ordinary kriging

Table 6.6. Comparison of the different interpolation methods showing the RMSE (in µg.m^{-3}) for PM_{10} annual averages for the year 2003. The last column shows the test results ranked with the best (lowest) RMSE results per distinguished method or group of methods marked as 1.

	PM_{10} annual avg, 2003	**RMSE**	**ranking**
1a	**linear regression with population density**	10.20	5
1b	**linear regression model using supplementary data**	8.93	4
1c	**linear regression with rural background field**	8.33	3
2a	**interpolation of urban concentrations, IDW**	7.31	2
2b	**interpolation of urban concentrations, ord. kriging**	7.12	
3a	**rural backgr. field + interpolation of Deltas, IDW**	7.15	1
3b	**rural backgr. field + interpolation of Deltas, ord. kriging**	7.09	

The results show that interpolation outside the border of the cities for both urban concentrations (Methods 2a and 2b) and urban Deltas (Methods 3a and 3b) gives better results than the approaches based on linear regression relations only (Methods 1a, 1b, and 1c). The interpolation of Deltas shows a slightly lower RMSE value than the interpolation of just concentrations, to the intent that it appears to be the most precise method, although the differences of these two methods are not so big. However, these results will have to be checked in more detailed analyses.

The results of this section are applied in the final map construction of Chapter 7.

7 Final map construction

7.1 Introduction

Final maps are created by merging the rural and urban maps (with the exception of AOT40, where urban maps are not constructed because of the rural character of this indicator). For the rural map production the combination of the EMEP dispersion model and supplementary data through linear regression, with interpolation of the residuals using ordinary kriging, is used (Method 3b of Section 5.1).

Production of the final map of urban areas happens by using the method of interpolation of Deltas by ordinary kriging with the addition of the rural background concentrations field (Method 3b of Section 6.7). Using this methodology concentration maps of urban air pollution are constructed for ozone (SOMO35) and for PM_{10} (annual average and 36^{th} maximum daily average value). The maps are created on an aggregated grid resolution of 10x10 km (see Section 6.2).

The merging of rural and urban maps is the theme of the next section.

7.2 Merging the rural and urban maps

It is necessary to merge the rural and urban maps into one combined air pollution concentration map. Essential for this merging is a representative and accurate delimitation of rural and urban areas, as the rural maps represent the rural areas and the urban maps the urban/suburban areas. Three possibilities of this delimitation have been considered: 1) using the population density grid, 2) using administrative borders of the cities and 3) using representative circles around selected cities. Ultimately, the use of the population density grid is chosen as most favourable for several reasons: the spatial administrative boundary datasets are not accurately representing the (sub)urban - rural area delimitations; by definition the city related administrative areas consist of substantial rural areas belonging to the city administration; defining appropriate criteria on the circle sizes to be set around city centres is hardly possible and not representative for the real spatial distribution of urbanised areas; the high resolution population density dataset derived by JRC from EEA's (recognized as high quality) land cover dataset CLC2000 combined with recognised Eurostat population density statistics represents a better spatial distribution of populated areas, including its density; the population density dataset is fully supported by EEA and a typical European Union product; the population density grid provides clear compatibility in subsequent combination of air quality maps with population.

For delimitation of urban and rural areas the average population density of the 10 km x 10 km aggregated grid is applied. The results of examining the relationship between the ozone and PM_{10} measurements of the two station classes rural and urban/suburban with the population density classes are shown in Figure 6.6.

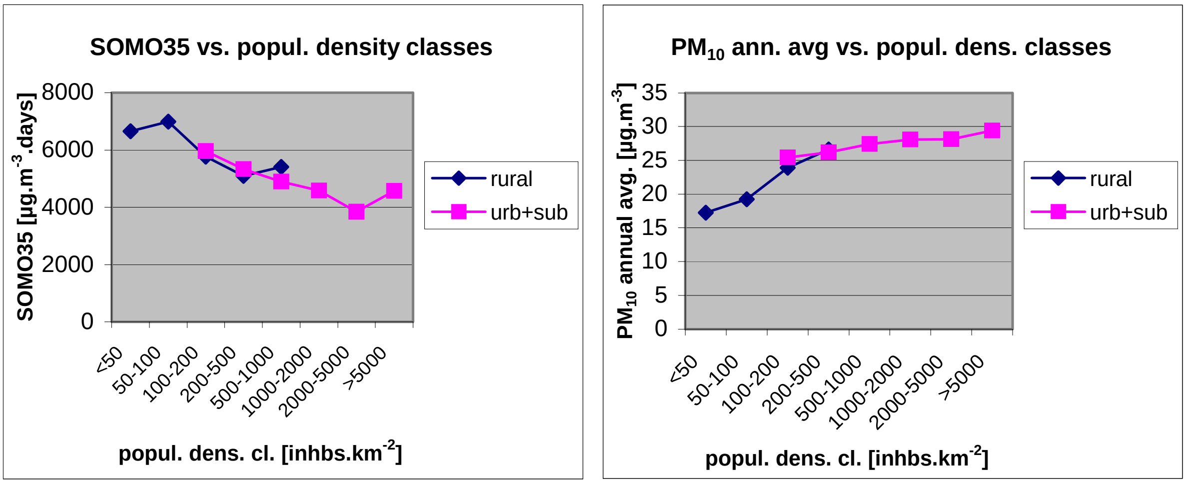


Figure 6.6. The SOMO35 (left) and PM_{10} annual average (right) averages versus the population density classes (x-axis), separately for rural and urban/suburban stations. For every population density class the average of all measured ozone and PM_{10} values stations is considered over the four-year period 2000-2003. The population density classes containing both rural and urban/suburban stations show at both air pollutions about the same level of concentrations and a continued tendency in the line graph characteristics going from the low to the high population density.

The figure shows that the rural and urban/suburban stations in the areas with the same overlapping population densities have approximately the same level of air pollution for both ozone and PM_{10}, independent of their type of area classification.

Therefore it is decided, for both ozone and PM_{10}, that for areas above 500 inhbs/km^2 the urban maps are used to indicate typical urban air pollution and for areas below 100 inhbs/km^2 the rural maps are used to indicate typical rural pollution. The areas between 100 and 500 inhbs/km^2 are mapped as the combination of rural and urban maps, weighted by the population.

In the routine computations the rural and urban maps are compared before the merging. At several locations, where the rural map for PM_{10} shows higher concentration levels than the urban map does, the resulting map is adapted towards the concentration levels of the urban map. The same procedure is followed at rural ozone map, where at locations the urban map shows higher the concentration levels. At these rural locations the resulting map will contain the concentration levels from the urban map. The motivation why we compensate for the assumed inconsistency with common behaviour of rural and urban air quality is explained in Section 3.4. These inconsistencies are mostly caused by the lack of rural stations in these regions. (However, this easy mechanism needs improvement in further development of the mapping methodology – one possibility is to apply this type of correction only at locations that are nearer to the urban stations than to the rural ones.)

7.3 Final air quality maps

The resulting maps are created using the EEA standard map projection ETRS89-LAEA5210, map extent 1c (http://www.eionet.eu.int/gis). The aggregated grid resolution of AOT40 maps is 2 x 2 km, the other maps (i.e. SOMO0, SOMO35, PM_{10}) are produced at a resolution of 10 x 10 km. The maps are included in the Annex II.

Combined urban and rural interpolated concentration maps are constructed for the ozone parameter SOMO35 for the years 2000 (Figure A1) and 2003 (Figure A2). For ozone parameter AOT40 only the rural interpolated maps for the years 2000 (Figure A3) and 2003 (Figure A4) are prepared with interpolation Method 3b of Section 5.1. No AOT40 are defined for urbanized areas, therefore no combined maps can be created.

Additionally, an AOT40 interpolated rural concentration map for 2003 based on ordinary cokriging with altitude (Method 1c of Section 5.1) is prepared, since it is needed as input to the EEA Core Set

Indicator describing the current state of ozone concentration levels and their excess over thresholds defined by the air quality framework directive. It is included in the Annex II as Figure A5 and illustrates when comparing to Figure A4 that using different interpolation methods certainly leads to the different spatial patterns of concentrations fields in the European map.

Combined urban and rural interpolated concentration maps for the PM_{10} parameters PM_{10} annual average and PM_{10} 36^{th} maximum daily means are created for the years 2001 and 2003. These maps are presented in the Figures A6 to A9 of the Annex II.

7.4 Combining the final maps with population and land cover

The final maps can be combined with population and land cover data in order to provide information related with these items. The outputs can be both maps and tables.

The AOT40 map is combined with the Corine land cover 250 m grid (CLC2000). The output of this combination is the map of the crops and vegetation at risk for the years 2000 and 2003 (Figures A10-A12). These maps present the values of AOT40 at those areas where the CLC class 2 on level 1 is represented, i.e. the subclasses “intensive agriculture” and “extensive agricultural” (described by parameter DRECL1). On basis of the AOT40 map and land cover grid the improved analysis for the EEA’s Core Set of Indicators (CSI) can be facilitated. For every country one can now calculate the total percentage of agricultural land falling belonging assigned for each individual concentration class. For every country one can now calculate percentages of the agricultural or natural land areas exposed to the different levels of air quality concentrations, leading to estimates of crops, vegetation and ecosystems at risk to damage or harvest reductions.

The map of SOMO35 and the maps of both indicators of PM_{10} are overlaid with the population density grid. The population density 10 km x 10 km grid map is illustrated by Figure A13. The subsequent results of the overlays of the air pollution parameter map (Figures A1, A2 and A6 to A9) with a transparent version of the population density map (Figure A13) are represented by the Figures A14 to A19. The overlays consist only of putting the population density gridded map as shown in Figure A13 literally as a transparent sheet on top of the map with the interpolated air quality concentration fields. From these overlays one can visually conclude that for ozone the concentration are as expected lower at several large cities (Figure A14 and to a less clear extend Figure A15). For PM_{10} one can visually conclude that large cities and more populated areas show higher concentration fields than the more rural areas (Figures A16 to A19). For PM_{10} this effect is stronger and possibly with higher and clearer peaks in urbanised areas when using the 36^{th} maximum daily mean value compared to annual average.

On the basis of these ozone and PM_{10} maps and the population density grid one the “population at risk” can be estimated. For every country one can now calculate the percentages of the population that lives in certain areas above certain concentration levels. The population weighted concentrations, i.e. the average concentration per inhabitant, can also be calculated. Now we become able to create population weighted concentration maps and population at risk maps (population percentages exposed to health damage or potential mortality) as a relative risk function (dose-response relation of population to concentration levels) of the spatial interpolated air quality parameter maps and population density maps.

8 Conclusions and recommendations

The different nature of urban and rural air quality has been confirmed. It is useful to create the urban and rural maps separately and to create the final maps by merging them.

8.1 Summary of rural interpolation studies

This section summarizes the results of tests carried out in this report to establish the preferred method of interpolating rural monitoring data with regard to ozone and PM_{10} at the European level. The interpolation methods tested, outlined in Section 2, are applied to the three ozone indicators (AOT40, SOMO0 and SOMO35) and to the two PM_{10} indicators (annual mean and 36^{th} maximum daily mean). Data for four years (2000-2003) have been selected from the AirBase dataset along with EMEP unified model calculations.

The additional supplementary data used in cokriging and in regression calculations includes altitude and 30 year climatological means of sunshine duration, relative humidity, temperature, precipitation and wind speed.

The results of these tests and analysis are described in detail, along with the resultant maps, in this document. Each of the interpolation methods is analysed in terms of the root mean square error (RMSE) derived by cross-validation analysis. RMSE is used to define the quality of the interpolation.

8.1.1 Modelled and supplementary data

In order to determine which additional datasets to use in the methods that combine monitoring with the modelled and supplementary data, the relations between the measured concentrations and these additional data have been examined. Regression analysis is carried out on the supplementary data. The following conclusions are drawn from the analysis.

- Results of EMEP model, altitude, sunshine duration and relative humidity give the most significant correlations for ozone indicators. Sunshine duration and relative humidity are highly mutually corellated.
- A weak correlation was found between climatological parameters – the means for 1969-1990 of temperature, precipitation and wind speed – and ozone indicators for recent years
- The final linear regression models developed and applied is composed of the EMEP dispersion model data, altitude, and sunshine duration
- Only results of EMEP model and altitude shows a significant correlation with PM_{10}. None of the other climatic variables show significant correlation in single regression. In the linear regression model sunshine duration in addition to the EMEP model and altitude plays role.
- Annual meteorological data, rather than 30 year climatological data, may be more appropriate as supplementary data for interpolation

8.1.2 Pure interpolation excluding model data

The following conclusions are drawn from the pure interpolation methods that do not include model data and that can include other supplementary data in the places of measurement only.

- Of the two pure interpolation methods without supplementary data tested, kriging is always found to be slightly superior to IDW
- For PM_{10}, interpolation using a logarithmic transformation is slightly superior to interpolation without this transformation
- Cokriging with supplementary data, such as altitude and sunshine duration, significantly improves the interpolation results

8.1.3 Interpolation including model data

The following conclusions can be drawn from interpolations carried out in combination with EMEP model data

- Cokriging with supplementary data significantly improves the results for both plain and fitted model combinations
- The combination of both chemical transport model and supplementary data using linear regression and ordinary kriging of the residuals gives the best result
- The exclusive use of EMEP model data, without any other supplementary data, does not lead to improvement of the interpolated field

8.1.4 Best interpolation methods

The following overall conclusions can be drawn from the interpolation tests

- The best interpolation method, for all the indicators examined, is the linear regression that includes both supplementary and EMEP model data in combination with ordinary kriging of the residual field (Method 3b)
- When model concentration fields are not available the best interpolation method tested is cokriging with the relevant supplementary data being altitude and, for ozone, sunshine duration (Method 1c)
- When no supplementary data is available for interpolation ordinary kriging (Method 1b) is always found to be slightly superior to IDW (Method 1a)

8.2 Summary of urban interpolation studies

A number of methodologies are tested for the interpolation of the urban/suburban concentration fields. Monitoring data from AIRBASE classified as urban or suburban background stations are used for the studies along with population density fields and a number of relevant supplementary data sets including geographical position and climatological parameters. In this case only the two ozone indicators SOMO0 and SOMO35 along with the annual mean and 36th highest daily mean PM_{10} concentrations are examined in the studies. The final interpolation methodology is only tested for annual mean PM_{10} concentrations.

Based on the results of the rural interpolation studies the focus of urban air quality estimation is to establish regression relationships between measured concentrations and other datasets, e.g. population density, climate etc. Besides that, measured urban concentrations (or delta, i.e. urban increments) can be used to create interpolated urban concentration fields where the interpolation is not limited to the city borders.

Testing of the methodologies is carried out in a quite similar fashion to the rural interpolation studies.

8.2.1 Regression analysis

The following conclusions can be made concerning the regression analysis between urban background monitoring stations and population density and supplementary datasets

- There is very almost no correlation between population density and measured concentrations for both ozone and PM_{10}
- When regions of similar population density are binned a trend for decreasing ozone, increasing PM_{10}, becomes apparent for increasing population density but with large variation

- Similar to the rural stations, there are significant regression relationships between environmental factors such as climate and geographical position and also the result of EMEP model
- There is strong correlation between the measured urban background concentrations and the interpolated rural field (Method 1c).

8.2.2 Best interpolation methods

Tests are carried out to establish the RMSE, based on cross-validation, of the different methodologies for annual average PM_{10} concentrations. The results of this can be summarized as:

- The best urban fields are created by interpolation of the urban increment Delta outside the city borders and by addition of the rural background concentrations field (Method 3b)
- The best method based purely on regression, i.e. no interpolation, is based on the linear regression of the urban background fields (Method 1c)
- The worst method based purely on regression, i.e. no interpolation, is based on the linear regression of the population density fields (Method 1a)

8.3 Summary of final map construction

Having established the urban (or delta) concentration field the urban and rural concentration fields are then combined using the methodology described in Section 7.2 that allows a smooth merging of rural and urban fields based on population density. This method is chosen based on an analysis of AirBase data, which demonstrates a convergence of rural and urban background station measurements at population densities between 100 and 500 inhabitants/km^2, and for being a practical methodology that allows merging of the two interpolation fields on a European wide scale.

High resolution air quality maps (resolving urban concentration) have been created for ozone and PM_{10} by merging the urban and rural maps. Consequently, the resulting concentration maps have been combined with the land cover and population density grids and the maps of the crops and vegetation at risk have been constructed, as well as the maps overlaid with the 10 km population density grid map.

8.4 Future recommendations

The mapping methods explored in this paper are intended for use in both the area of public information and policy assessment. There are thus a number of relevant application areas both now and in the future. Of specific relevance to the EEA is the inclusion of air quality in its 'In Your Neighbourhood' project. In addition, the European Union air pollution Directives and the EoI will, in the future, be fully integrated. The INSPIRE Directive will also have implications for spatial air quality reporting in the future.

It is clear that not all methodologies have been tested and there may be a number of untested supplementary datasets that can improve the interpolation results. A number of questions and possibilities have yet to be explored. The following list indicates areas that have not been completely addressed in this report and will require future attention

- The use of concurrent meteorological data rather than climatological data as a supplementary source of information for regression analysis
- The use of supplementary data in better spatial resolution
- Detailed analysis of uncertainty in the resulting interpolation fields
- Establishment of a robust system for the production of interpolated fields for a variety of compounds and indicators

- Further improvement of mapping methodology, especially of urban mapping and merging the rural and urban maps
- Assessment of interpolation methodologies that can be used on short time scales and that include temporal as well as spatial variability, in support of the EEA and its consultant to apply the mapping methods in the 'In Your Neighbourhood' project.
- The estimation of population weighted concentrations and the population in risk, including the comparing of results obtained by different methods;
- Improvement of the mapping methodology for the indicators, which are routinely computed by ETC/ACC (26th highest max8h value, maps of exceedances);
- Extension of these methods to other pollutants.

References

Amann, M., Bertok, I., Cofala, J., Gyarfas F., Heyes C., Klimont Z., Schöpp, W., Winiwarter, W. (2005). Baseline scenarios for the Clean Air for Europe (CAFE) Programme. IIASA, Laxenburg, Austria. www.iiasa.ac.at/rains/CAFE_files/CAFE-baseline-full.pdf

Brönnimann, S., Schuepbach, E., Zanis, P., Buchmann, B., Wanner, H. (2000). A climatology of regional background ozone at different elevations in Switzerland (1992-1998). Atmospheric Environment 34, pp. 5191-5198.

Brönnimann, S., Buchmann, B., Wanner, H. (2002). Near-surface concentrations in Switzerland: the 1990s. Atmospheric Environment 36, pp. 2841-2852.

Cressie, N. (1993). Statistics for spatial data. Wiley series, New York.

Fagerli, H., Simpson, D., Tsyro, S. (2004). Unified EMEP model: Updates. In: EMEP Status Report 1/2004. MNP, Oslo, Norway. www.emep.int/publ/reports/2004/Status_report_int_del1.pdf

Lloyd, C. D., Atkinson, P. M. (2004). Increased accuracy of geostatistical prediction of nitrogen dioxide in the United Kingdom with secondary data. Applied Earth Observation and Geoinformation 5, 293-305

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.

Rao, C. R. (1973): Linear statistical inference and its applications. Wiley. New York.

Simpson, D., Fagerli, H., Jonson, J. E., 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. www.emep.int/publ/reports/2003/emep_report_1_part1_2003.pdf

Tidblad, J., Mikhailov, A.A., Henriksen J., Kucera V. (2004). Improved prediction of ozone levels in urban and rural atmospheres. Protection of Metals, Vol. 40, pp. 67-76.

Vestreng, V., Adams, M., Goodwin, J. (2004). Inventory Review 2004 Emission Data reported to CLRTAP and under the NEC Directive. MSC-W Technical Report 1/04. EMEP/EEA Joint Review Report. www.emep.int/publ/reports/2004/emep_technical_1_2004.pdf

Vestreng, V., K. Breivik K., Adams, M., Wagner, A., Goodwin, J., Rozovskaya, O., Pacyna J. M. (2005). Emission Data reported to LRTAP Convention and NEC Directive. Initial review for HMs and POPs. MSC-W. Technical Report 1/05. MNP, Oslo, Norway. www.emep.int/publ/reports/2005/emep_technical_1_2005.pdf

Wackernagel, H. (2003). Multivariate geostatistics: An introduction with applications. 3rd ed, Springer, Berlin.

Annex I. Several examples and details of statistical computations

1. Stepwise strategy for identifying the useful subset of variables, backward type

Table A1. Example - O_3 rural, SOMO35, 2000. In each step there is selected and dropped the variable with the highest P, if greater 0.05. Moreover, if for any parameter the square multiple correlation R^2 between this parameter and all of the others parameters is greater than 0.7, the relevant variable is dropped. As the result, the linear regression model is developed, where P<0.05.

linear regression model (eq. 4.4)	step1		step2		step3		step4		step5	
	coeff.	P	coeff.	P	coeff.	P	coeff.	P	coeff.	P
c (constant)	-5782.7	0.199	-4799.7	0.272	-587.3	0.517	-1002.2	0.224	-1653.4	0.003
a1 (altitude)	3.41	0.000	3.60	0.000	3.50	0.000	3.39	0.000	3.38	0.000
a2 (temperature)	-54.2	**0.354**								
a3 (precipitation)	-0.74	0.131	-0.66	0.173	-0.50	**0.270**				
a4 (wind speed)	-198.9	0.147	-194.8	0.155	-137.1	0.268	-132.2	**0.286**		
a5 (sunsh. dur.)	98.8	0.001	84.5	0.001	67.5	0.001	68.5	0.001	67.1	0.001
a6 (rel. humidity)	60.8	0.236	48.9	**0.325**						
a7 (EMEP model)	0.54	0.000	0.51	0.000	0.50	0.000	0.50	0.000	0.53	0.000
R^2	**0.64**		**0.64**		**0.64**		**0.64**		**0.64**	

2. Comparison of the linear regression models with inclusion of sunshine duration or relative humidity, for ozone, rural areas

Table A2. In comparison of the linear regression models with including of sunshine duration and relative humidity is seen, that sunshine duration is more appropriate (square multiple correlation R^2 is higher).

meas = c + a1 *EMEP + + a2*altitude + a3 *sun.dur.	SOMO0				SOMO35			
	2000	2001	2002	2003	2000	2001	2002	2003
c (constant)	13960	11785	19760	9136	-1653	-871	n. sign.	1205
a1 (EMEP model)	0.24	0.38	n. sign.	0.65	0.53	0.61	0.17	0.87
a2 (altitude)	6.244	5.816	6.089	5.383	3.377	3.123	3.592	3.308
a3 (sunshine duration)	113.6	70.5	155.8	n. sign.	67.1	47.7	85.8	n. sign.
R^2	**0.59**	**0.64**	**0.56**	**0.60**	**0.64**	**0.68**	**0.59**	**0.64**

meas = c + a1 *EMEP + + a2*altitude + a3*rel.hum.	SOMO0				SOMO35			
	2000	2001	2002	2003	2000	2001	2002	2003
c (constant)	18019	9546	40846	9136	n. sign.	n. sign.	10784	1205
a1 (EMEP model)	0.41	0.56	n. sign.	0.65	3.32	3.05	3.39	0.87
a2 (altitude)	5.912	5.584	5.947	5.383	0.670	0.773	0.296	3.308
a3 (relative humidity)	-60.7	n. sign.	-194.9	n. sign.	n. sign.	n. sign.	-106.4	n. sign.
R^2	**0.58**	**0.63**	**0.53**	**0.60**	**0.63**	**0.67**	**0.56**	**0.64**

3. Comparison of the linear regression relationships of measured 36th maximum daily means and EMEP model results, if modelled 36th maximum daily means or modelled annual averages are used, for PM_{10}, rural areas

Table A3. In the table is visible that R^2 shows higher correlation of measured 36th maximum daily means with EMEP model for annual average in comparison with EMEP model for 36th maximum daily mean in the years 2001-2003.

meas. 36th max.d.m. = c + a * EMEP model	EMEP model of 36th max. daily mean				EMEP model of annual average			
	2000	2001	2002	2003	2000	2001	2002	2003
c	n. sign	n. sign.	14.1	15.2	n. sign.	n. sign.	12.9	14.2
a	1.64	1.61	1.06	1.01	3.34	3.34	2.35	2.22
R^2	**0.58**	**0.33**	**0.28**	**0.20**	**0.52**	**0.41**	**0.33**	**0.22**

4. Comparison of the linear regression models with and without the inclusion of sunshine duration, for PM_{10}, rural areas

Table A4. In comparison of the linear regression models with and without sunshine duration is visible, that the inclusion of sunshine duration brings the improvement of the model (R^2 is higher in all cases).

meas = c + a1 *EMEP + + a2*altitude + a3 *sun.dur.	annual average				36th maximum daily mean			
	2000	2001	2002	2003	2000	2001	2002	2003
c (constant)	n. sign.	-13.6	n. sign.	n. sign.	-37.0	-22.8	n. sign.	n. sign.
a1 (EMEP model)	1.60	2.26	1.41	1.39	1.68	3.62	2.48	2.38
a2 (altitude)	-0.0077	-0.0054	-0.0070	-0.0064	-0.0075	-0.0079	-0.0088	-0.0109
a3 (sunshine duration)	0.22	0.37	0.26	0.24	1.15	0.66	0.41	0.45
R^2	**0.65**	**0.52**	**0.40**	**0.35**	**0.67**	**0.48**	**0.40**	**0.30**

meas = c + a1 *EMEP + + a2 *altitude	annual average				36th maximum daily mean			
	2000	2001	2002	2003	2000	2001	2002	2003
c (constant)	6.5	n. sign.	11.5	11.7	n. sign.	n.sign.	20.0	22.8
a1 (EMEP model)	1.64	2.11	1.18	1.14	1.70	3.48	1.98	1.84
a2 (altitude)	-0.0071	-0.0039	-0.0059	-0.0059	-0.0049	-0.0047	-0.0077	-0.0106
R^2	**0.60**	**0.48**	**0.35**	**0.31**	**0.61**	**0.44**	**0.36**	**0.26**

5. Root-mean-square error (RMSE) for linear regression relations and linear regression models

Table A5. Comparison of the different methods showing the RMSE (in µg.m^{-3}.days) for ozone indicator SOMO0, rural areas

method	2000	2001	2002	2003	avg	ranking
linear regression with EMEP	3832	3556	3837	3687	3728	3
linear regression with altitude	3262	3030	3158	3348	3199	2
linear regression with sunshine	4309	4085	4020	4239	4163	4
linear regression with rel. hum.	4879	4461	4283	4495	4530	5
linear regr. mod. (EMEP,alt., s. dur.)	3027	2715	2916	2905	2891	1

Table A6. Comparison of the different methods showing the RMSE (in µg.m^{-3}.days) for ozone indicator SOMO35, rural areas

method	2000	2001	2002	2003	avg	ranking
linear regression with EMEP	2355	2117	2459	2510	2360	3
linear regression with altitude	2291	2111	2117	2697	2304	2
linear regression with sunshine	2765	2514	2522	3097	2724	4
linear regression with rel. hum.	3337	2934	2817	3367	3114	5
linear regr. mod. (EMEP,alt., s. dur.)	1892	1647	1843	2032	1853	1

Table A7. Comparison of the different methods showing the RMSE (in µg.m^{-3}.hours) for ozone indicator AOT40, rural areas

method	2000	2001	2002	2003	avg	ranking
linear regression with EMEP	6955	5376	7128	7744	6801	2
linear regression with altitude	8108	7331	7394	10929	8441	4
linear regression with sunshine	8578	7511	7668	10774	8633	5
linear regression with rel. hum.	8174	7353	7508	10442	8369	3
lin. regr. mod. (EMEP,alt., s. dur., r. h.)	6454	4859	5771	7285	6092	1

Table A8. Comparison of the different methods showing RMSE (in µg.m^{-3}), annual average of PM_{10}, rural areas

method	2000	2001	2002	2003	avg	ranking
linear regression with EMEP	5.6	6.8	7.5	8.0	7.0	3
linear regression with altitude	7.2	8.4	8.1	8.7	8.1	4
linear regr. mod. (EMEP,alt.)	5.0	6.6	7.2	7.8	6.7	2
linear regr. mod. (EMEP,alt., s. dur.)	4.9	6.3	6.9	7.5	6.4	1

Table A9. Comparison of the different methods showing RMSE (in µg.m^{-3}), 36th maximum daily PM_{10} value, rural areas

method	2000	2001	2002	2003	avg	ranking
linear regression with EMEP	8.5	11.7	11.6	15.4	11.8	3
linear regression with altitude	11.9	13.4	12.5	16.0	13.5	4
linear regr. mod. (EMEP,alt.)	8.3	10.8	10.9	14.7	11.2	2
linear regr. mod. (EMEP,alt., s. dur.)	7.6	10.3	10.7	14.4	10.7	1

Table A10. Comparison of the different methods showing the RMSE (in µg.m^{-3}.days) for ozone indicator SOMO0, urban/suburban areas

method	2000	2001	2002	2003	avg	ranking
linear regression with EMEP	3092	3350	3334	3204	3245	2
linear regression with altitude	3608	4091	3785	3893	3844	5
linear regression with sunshine	3294	3327	3232	3521	3344	3
linear regression with rel. hum.	3233	3979	3726	3917	3714	4
linear regr. mod. (EMEP,alt., s. dur.)	2879	2915	2859	2945	2899	1

Table A11. Comparison of the different methods showing the RMSE (in µg.m^{-3}.days) for ozone indicator SOMO35, urban/suburban areas

method	2000	2001	2002	2003	avg	ranking
linear regression with EMEP	1590	1552	1876	2022	1760	2
linear regression with altitude	2075	2241	2253	2706	2319	5
linear regression with sunshine	1920	1811	2000	2643	2094	3
linear regression with rel. hum.	1713	2087	2158	2728	2171	4
linear regr. mod. (EMEP,alt., s. dur.)	1486	1423	1509	1878	1574	1

Table A12. Comparison of the different methods showing RMSE (in µg.m^{-3}), annual average of PM_{10}, urban/suburban areas

method	2000	2001	2002	2003	avg	ranking
linear regression with EMEP	10.4	7.6	8.1	9.8	9.0	2
linear regr. mod. (EMEP,alt., s. dur.)	9.3	7.3	7.6	8.9	8.3	1

Table A13. Comparison of the different methods showing RMSE (in µg.m^{-3}), 36th maximum daily PM_{10} value, urban/suburban areas

method	2000	2001	2002	2003	avg	ranking
linear regression with EMEP	16.5	13.4	14.4	18.0	15.6	2
linear regr. mod. (EMEP,alt., s. dur.)	14.3	12.5	12.6	16.1	13.9	1

Annex II. Final maps

Contents:

(Maps of Figures A14-A19 are the maps of Figures A1-A2 and A6-A9 with on top of it as a transparency the map of Figure A13)

Codes of included countries

code	country	code	country	code	country	code	country	code	country
AD	Andorra	CY	Cyprus	GR	Greece	LV	Latvia	RO	Romania
AL	Albania	CZ	Czech Republic	HR	Croatia	LU	Luxembourg	SK	Slovakia
AT	Austria	DE	Germany	HU	Hungary	MK	FYR of Macedonia	SI	Slovenia
BA	Bosnia and Herzegovina	DK	Denmark	IE	Ireland	MT	Malta	SE	Sweden
BE	Belgium	EE	Estonia	IT	Italy	NL	Netherlands	TR	Turkey
BG	Bulgaria	ES	Spain	IS	Iceland	NO	Norway	UK	United Kingdom
CH	Switzerland	FI	Finland	LI	Liechtenstein	PL	Poland		
CS	Serbia and Montenegro	FR	France	LT	Lithuania	PT	Portugal		

Figure A1 Combined rural and urban concentration map of ozone – SOMO35, year 2000. Spatial interpolated concentration field and the measured values in the measuring points. Units: µg.m^{-3}.days. The combined map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals) and the urban map (Method 3b of Section 6.7 using interpolation of urban Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, RO. Countries with missing population density information and therefore excluded from the mapping: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR.

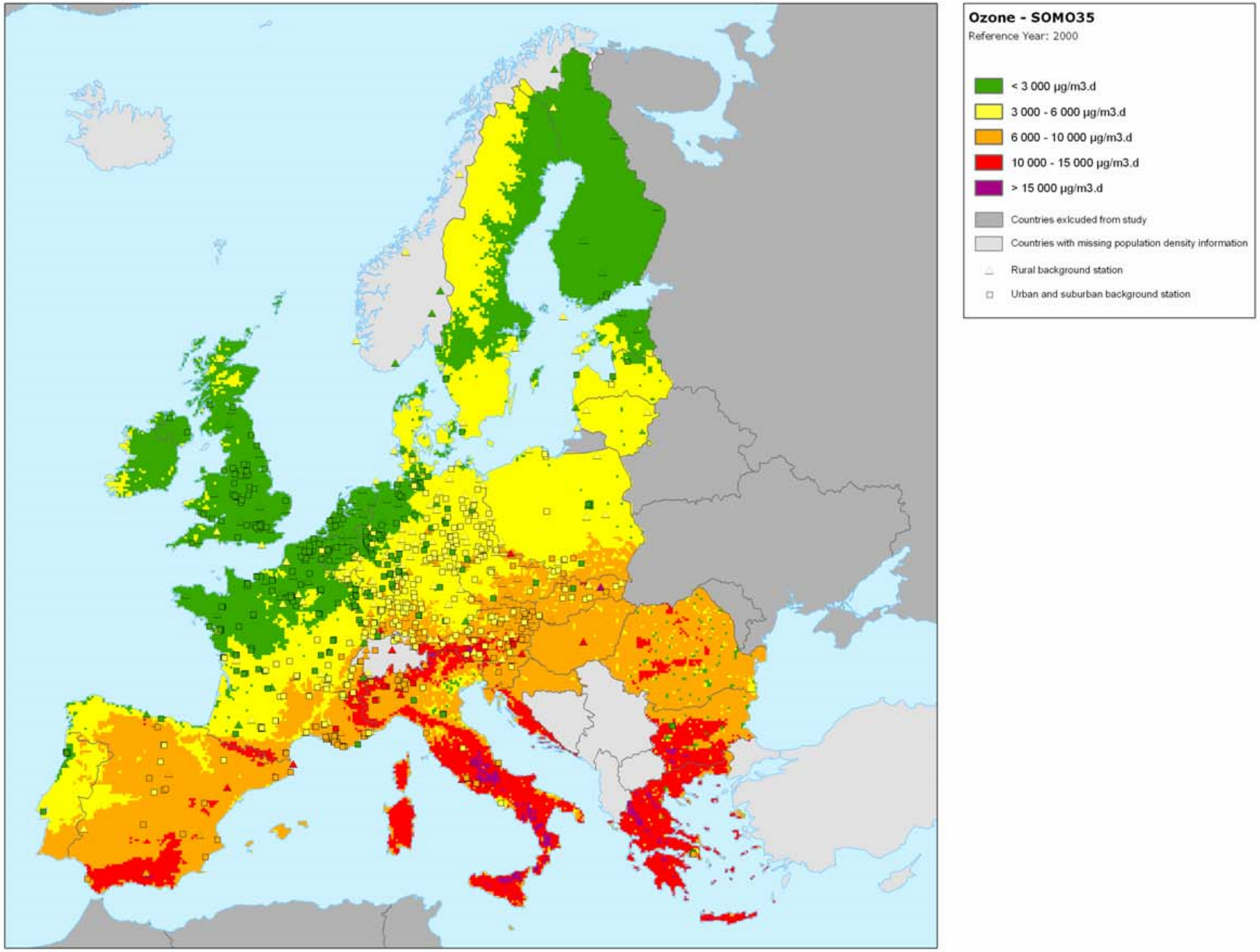


Figure A2 Combined rural and urban concentration map of ozone – SOMO35, year 2003. Spatial interpolated concentrations field and the measured values in the measuring points. Units: µg.m^{-3}.days. The combined map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, RO. Countries with missing population density information and therefore excluded: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR.

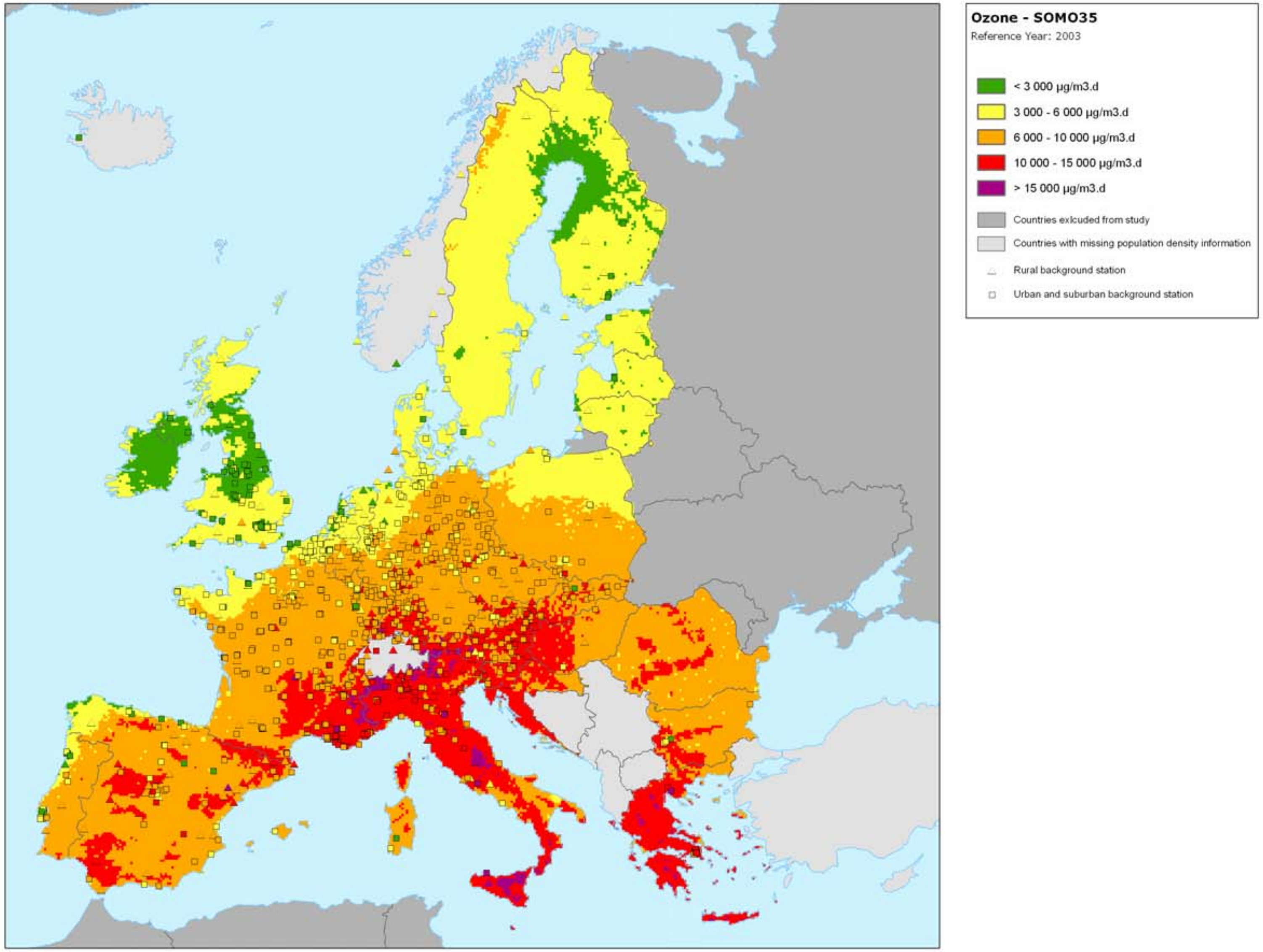


Figure A3 Rural concentration map of ozone – AOT40, year 2000. Spatial interpolated concentration field and the measured values in the measuring points. Units: μg.m^{-3}.hours. The final map is created by combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals. Countries with few ozone data and therefore excluded from mapping calculations: AL, BA, BG, CS, CY, GR, IS, MK, RO, TR.

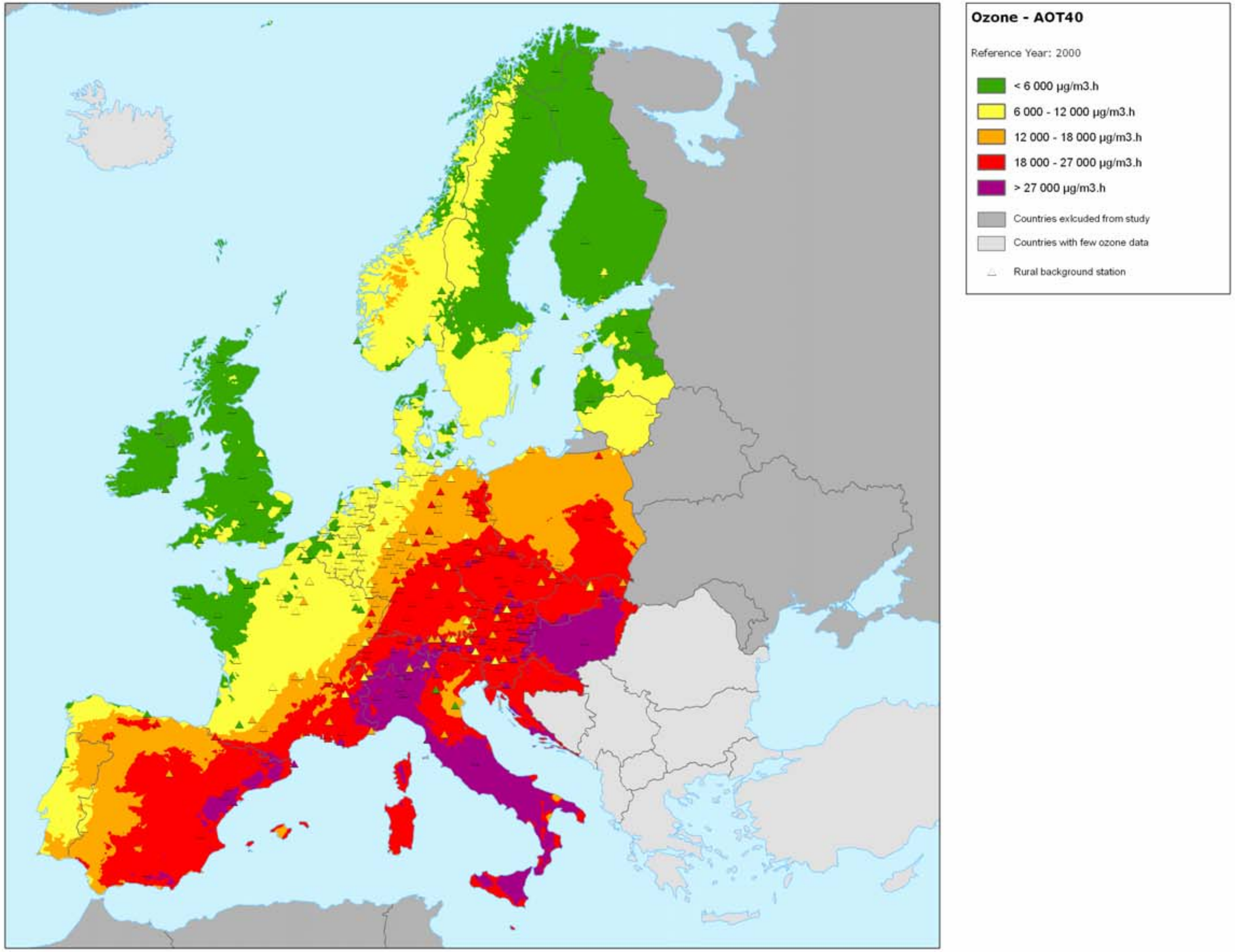


Figure A4 Rural concentration map of ozone – AOT40, year 2003. Spatial interpolated concentration field and the measured values in the measuring points.
Units: µg.m^{-3}.hours. The final map is created by combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals (Method 3b of Section 5.1). Countries with few ozone data and therefore excluded from the mapping calculations: AL, BA, BG, CS, CY, GR, IS, MK, RO, TR.

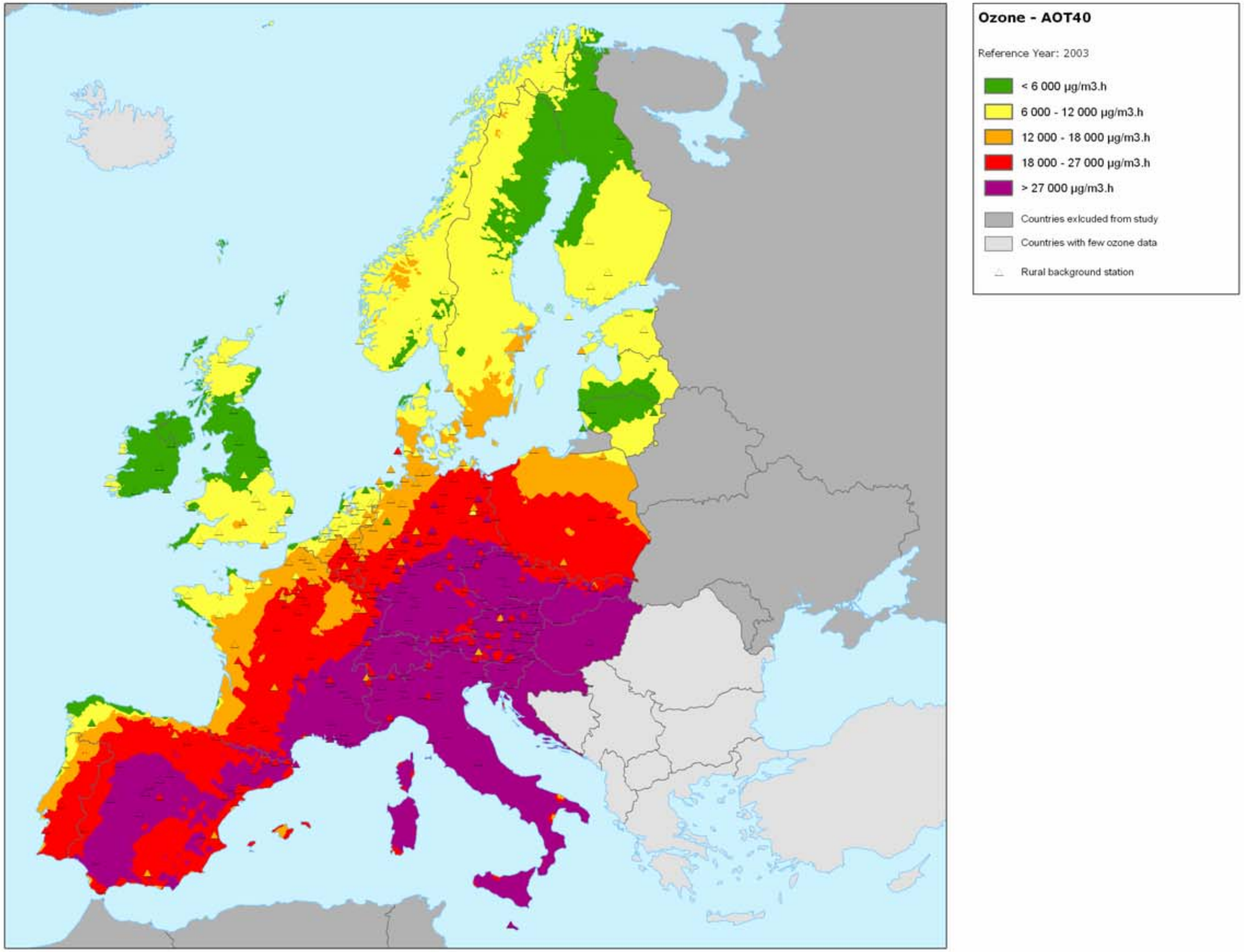


Figure A5 Rural concentration map of ozone – AOT40, year 2003. Spatial interpolated concentration field and the measured values in the measuring points.
Units: µg.m^{-3}.hours. The final map is created by ordinary cokriging, using altitude (Method 1c of Section 5.1). Countries with few ozone data and therefore excluded from the mapping calculations: AL, BA, BG, CY, CS, GR, IS, MK, RO, TR.

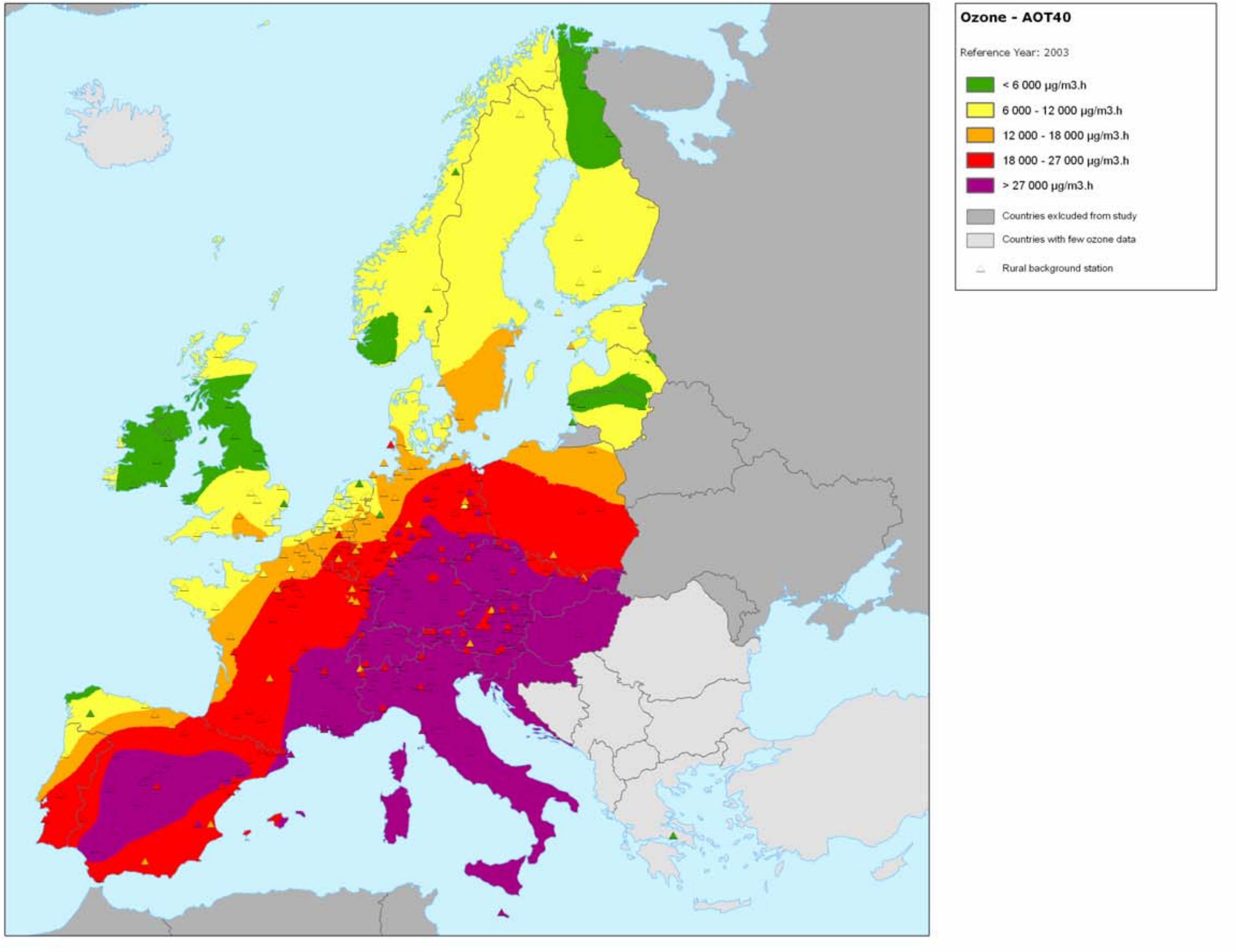


Figure A6 Combined rural and urban concentration map of PM_{10} – annual average, year 2001. Spatial interpolated concentration field and the measured values in the measuring points. Units: $\mu g.m^{-3}$.days. The combined map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ord. kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ord. kriging). Countries with interpolation based on additional data only: BG, GR, HR, HU, RO. Countries with missing population density information and therefore excluded: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR.

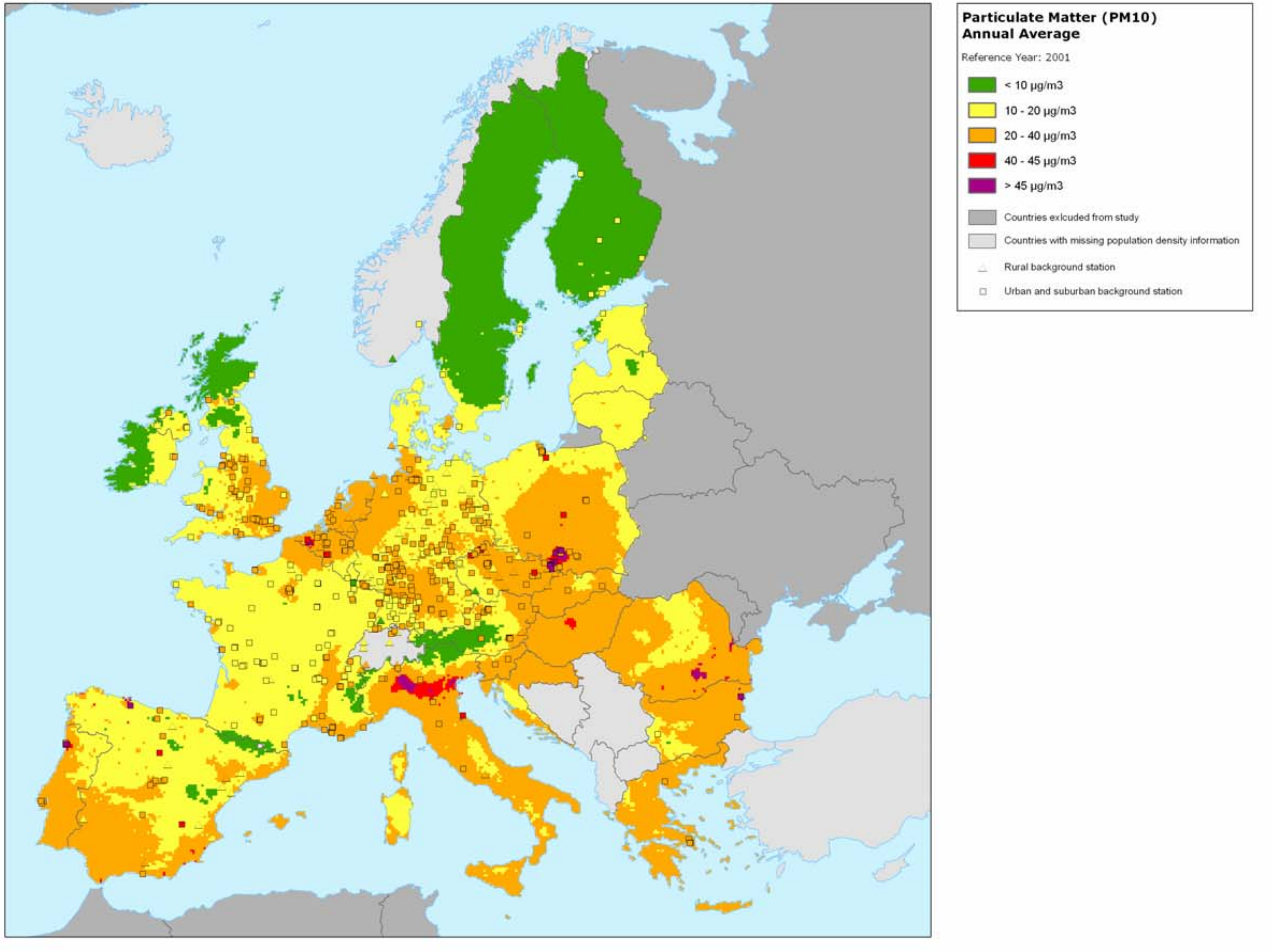


Figure A7 Combined rural and urban concentration map of PM_{10} – annual average, year 2003. Spatial interpolated concentration field and the measured values in the measuring points. Units: $\mu g.m^{-3}$. The combined map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, HR, HU, RO. Countries with missing population density information and therefore excluded: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR.

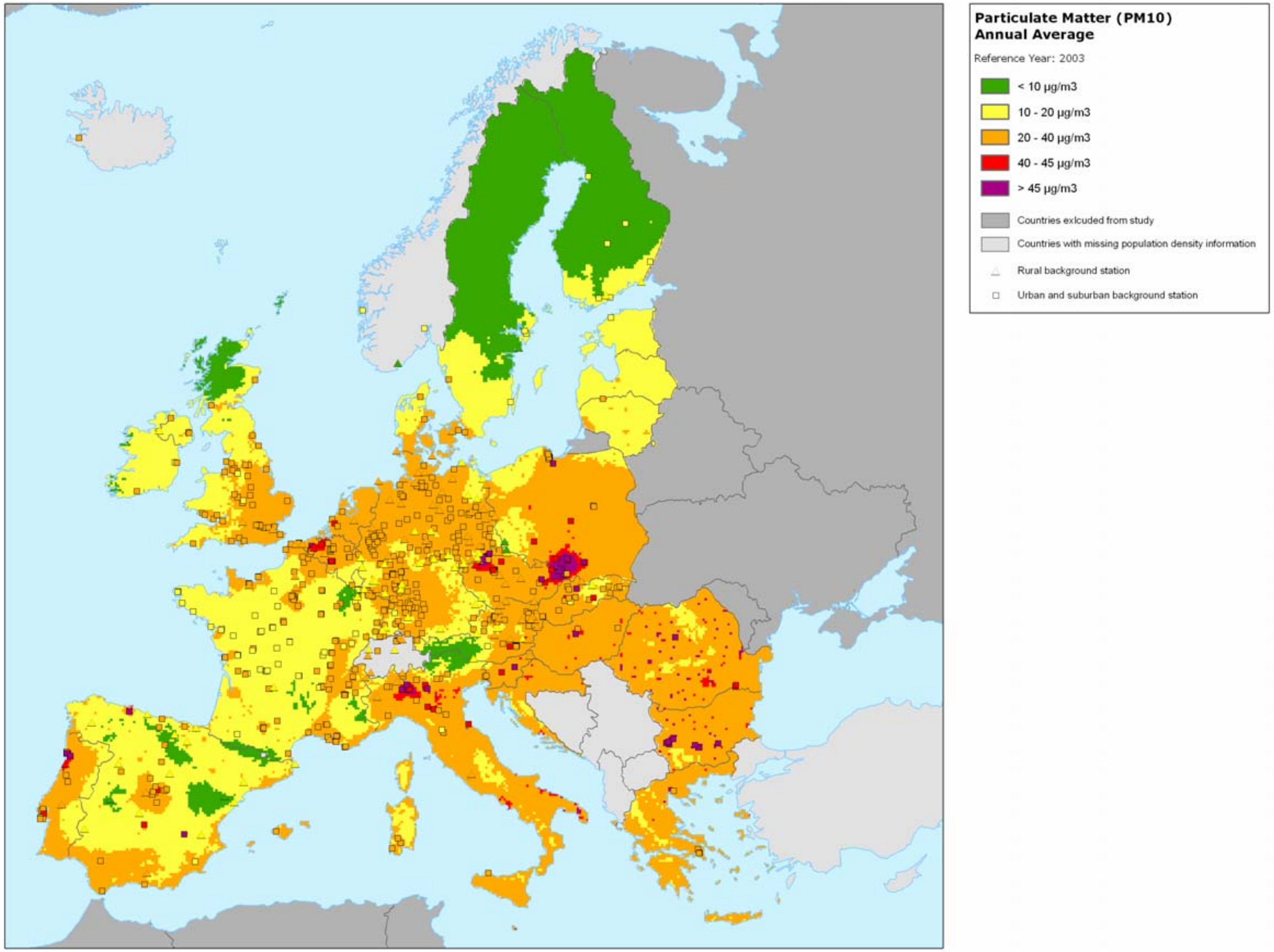


Figure A8 Combined rural and urban concentration map of PM_{10} – 36th maximum daily average value, year 2001. Spatial interpolated concentration field and the measured values in the measuring points. Units: $\mu g.m^{-3}$. The combined map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, HR, HU, RO. Countries with missing population density and therefore excluded: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR.

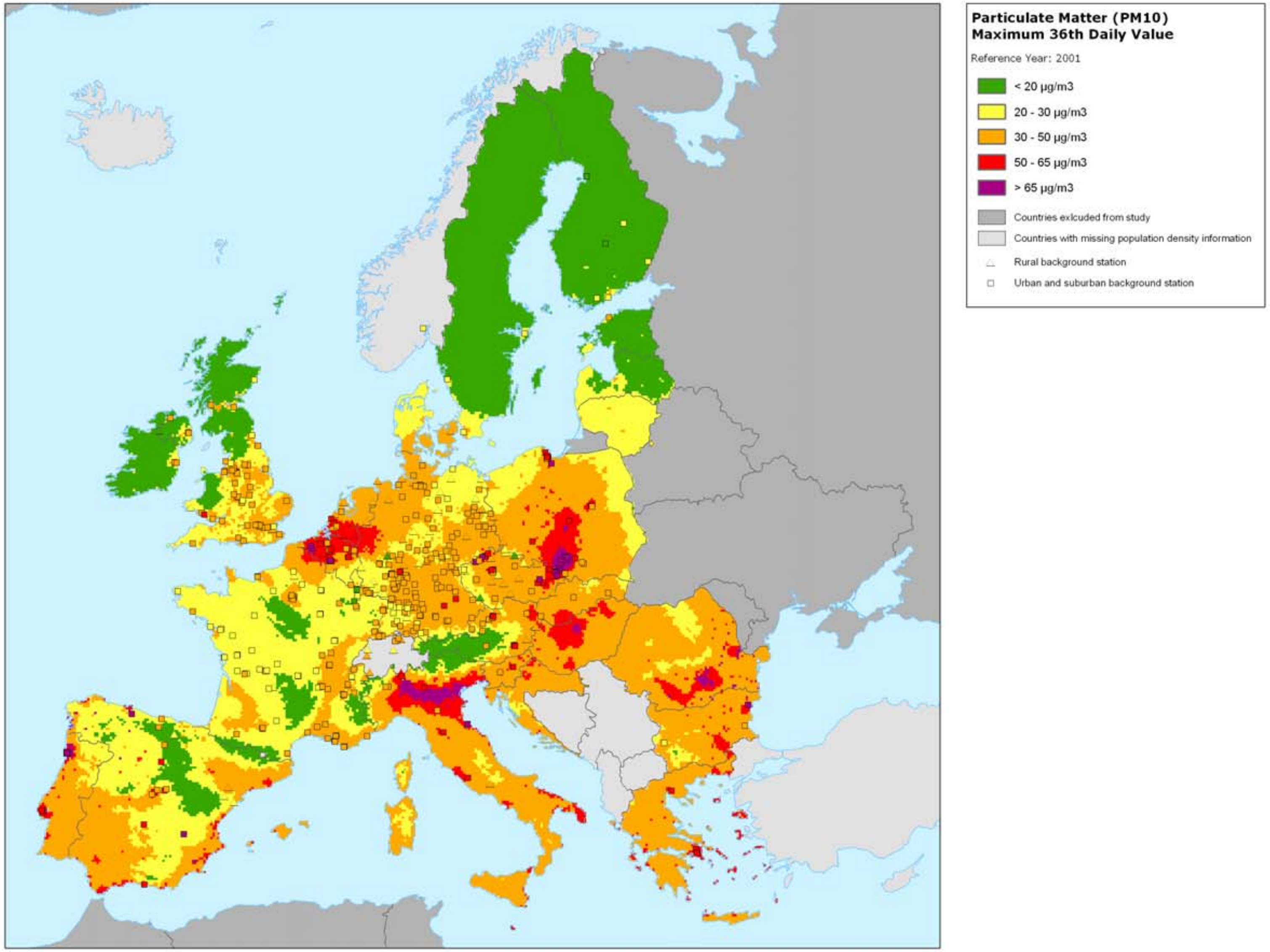


Figure A9 Combined rural and urban concentration map of PM_{10} – 36[th] maximum daily average value, year 2003. Spatial interpolated concentration field and the measured values in the measuring points. Units: $\mu g.m^{-3}$. The combined map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ord. kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ord. kriging). Countries with values based on additional data only: BG, GR, HR, HU, RO. Countries with missing population density information and therefore excluded: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR.

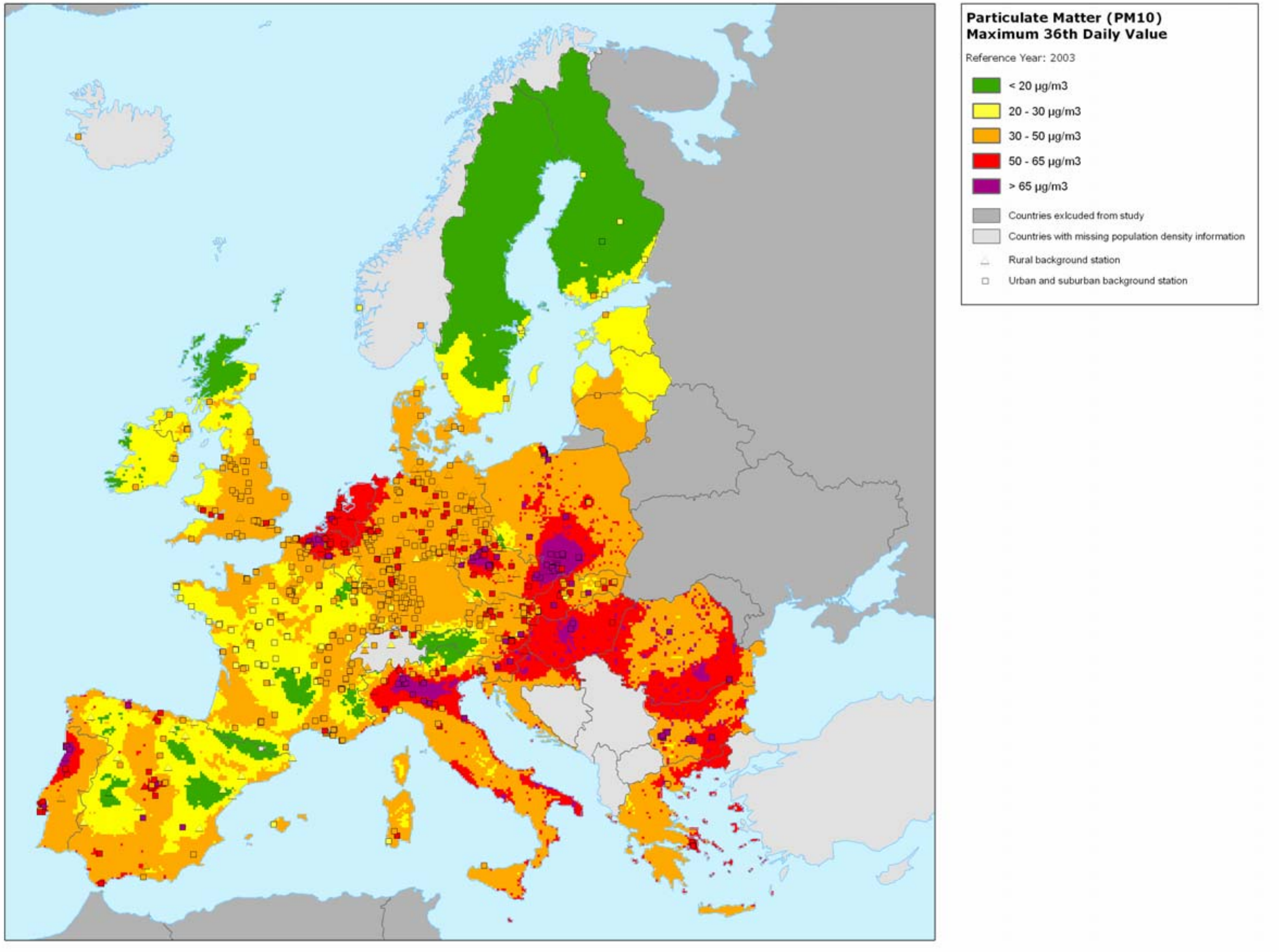


Figure A10 Crops and vegetation at risk / damage map – AOT40, year 2000. Spatial interpolated concentration field combined with land cover grid of CLC2000. Units: µg.m^{-3}.h. The spatial concentration field interpolation is done by combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals (Method 3b of Section 5.1). Countries with missing land cover information are excluded from the mapping: AD, CH, CS, IS, LI, NO, TR. Countries with few ozone data and therefore excluded from the mapping calculations: AL, BA, BG, CY, GR, IS, MK, RO, TR. (Compare with Figure A3).

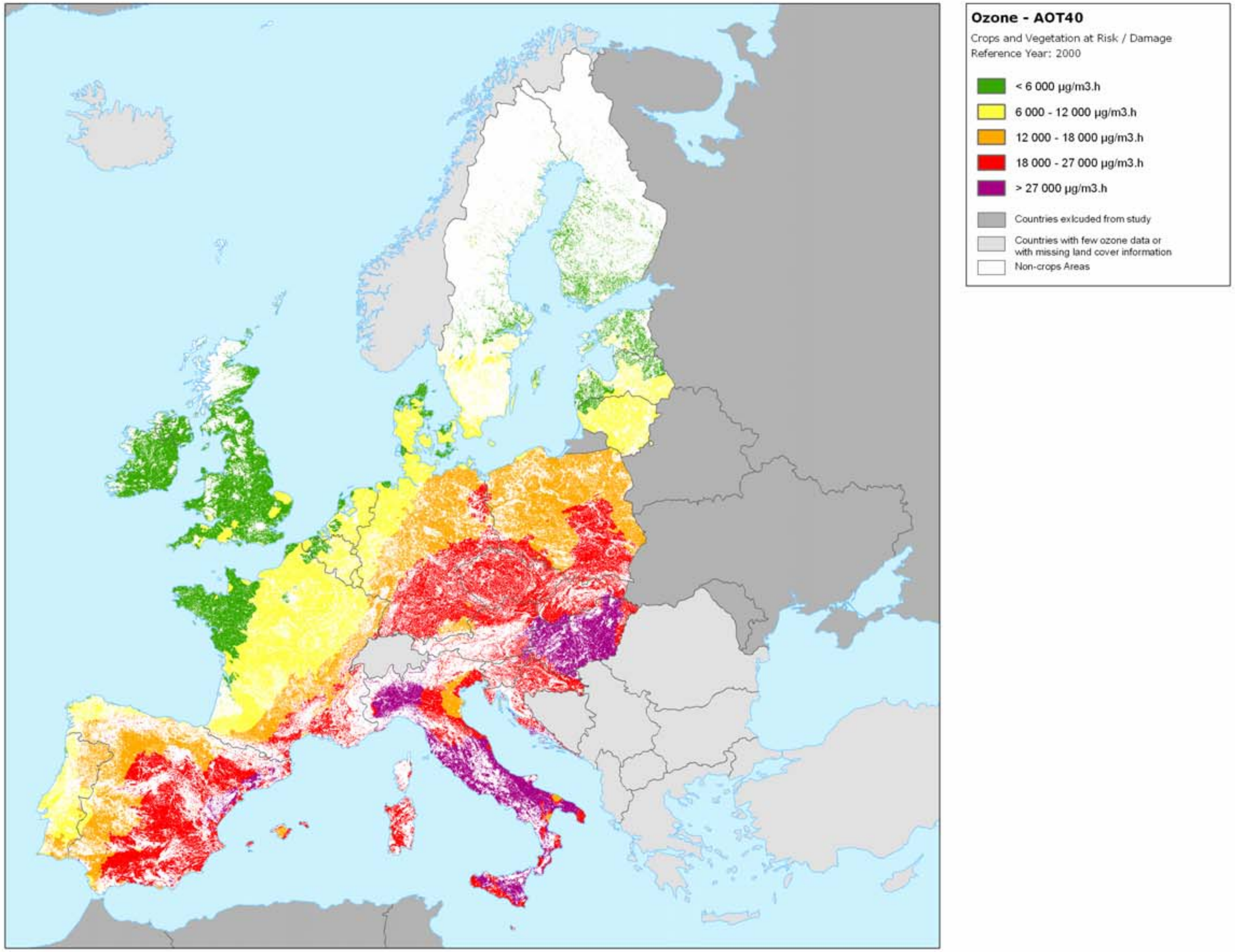


Figure A11 Crops and vegetation at risk / damage map – AOT40, year 2003. Spatial interpolated concentration field combined with land cover grid of CLC2000. Units: µg.m^{-3}.hours. The spatial concentration field interpolation is done by combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals (Method 3b of Section 5.1). Countries with missing land cover information are excluded from the mapping: AD, CH, CS, IS, LI, NO, TR. Countries with few ozone data and therefore excluded from the mapping calculations: AL, BA, BG, CY, GR, IS, MK, RO, TR. (Compare with Figure A4).

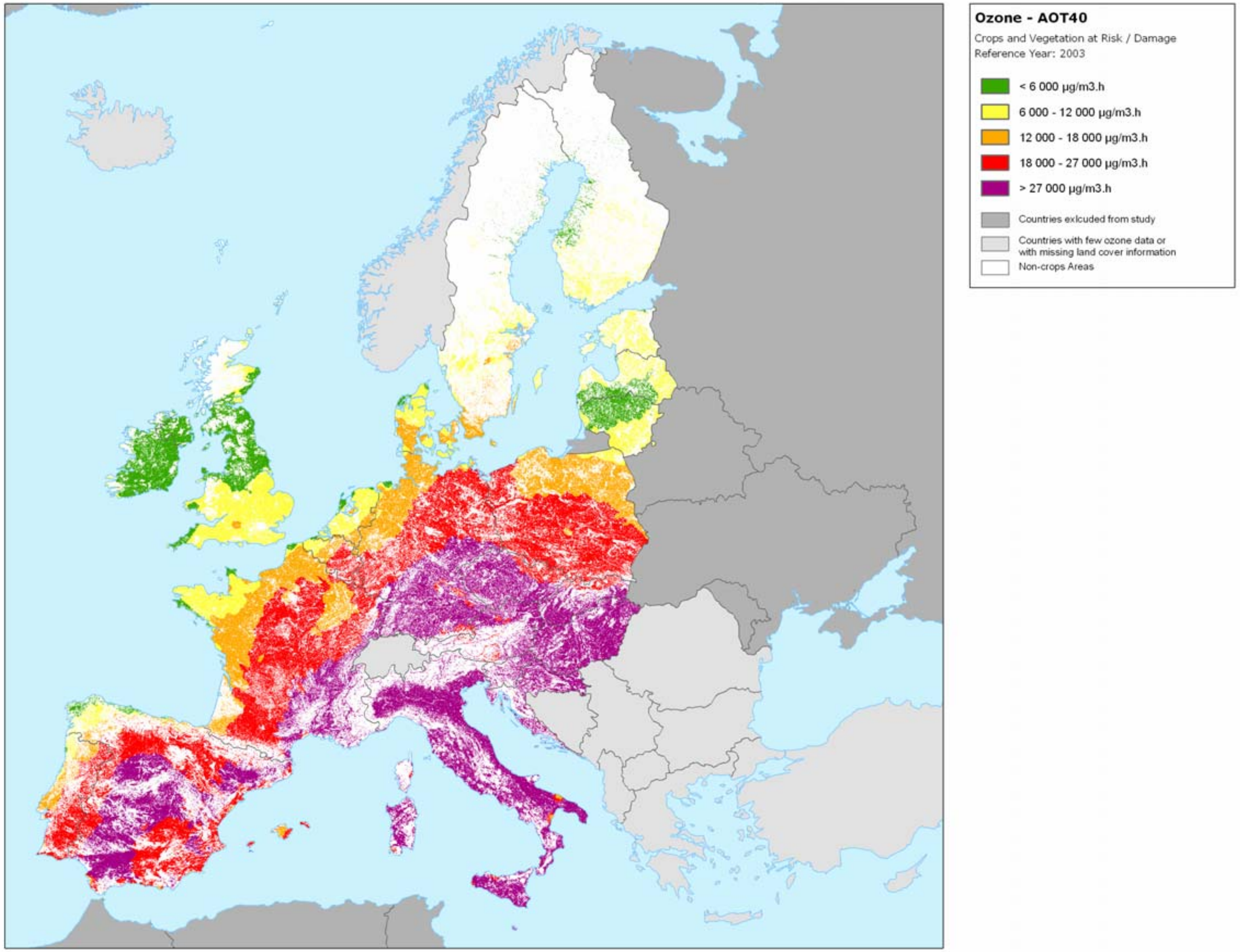


Figure A12 Crops and vegetation at risk / damage map – AOT40, year 2003. Spatial interpolated concentration field combined with land cover grid of CLC2000. Units: µg.m^{-3}.hours. The spatial concentration field interpolation is done by ordinary cokriging, using altitude (Method 1c of Section 5.1). Countries with missing land cover information are excluded from the mapping: AD, CH, CS, IS, LI, NO, TR. Countries with few ozone data and therefore excluded from the mapping calculations: AL, BA, BG, CY, GR, IS, MK, RO, TR. (Compare with Figure A5).

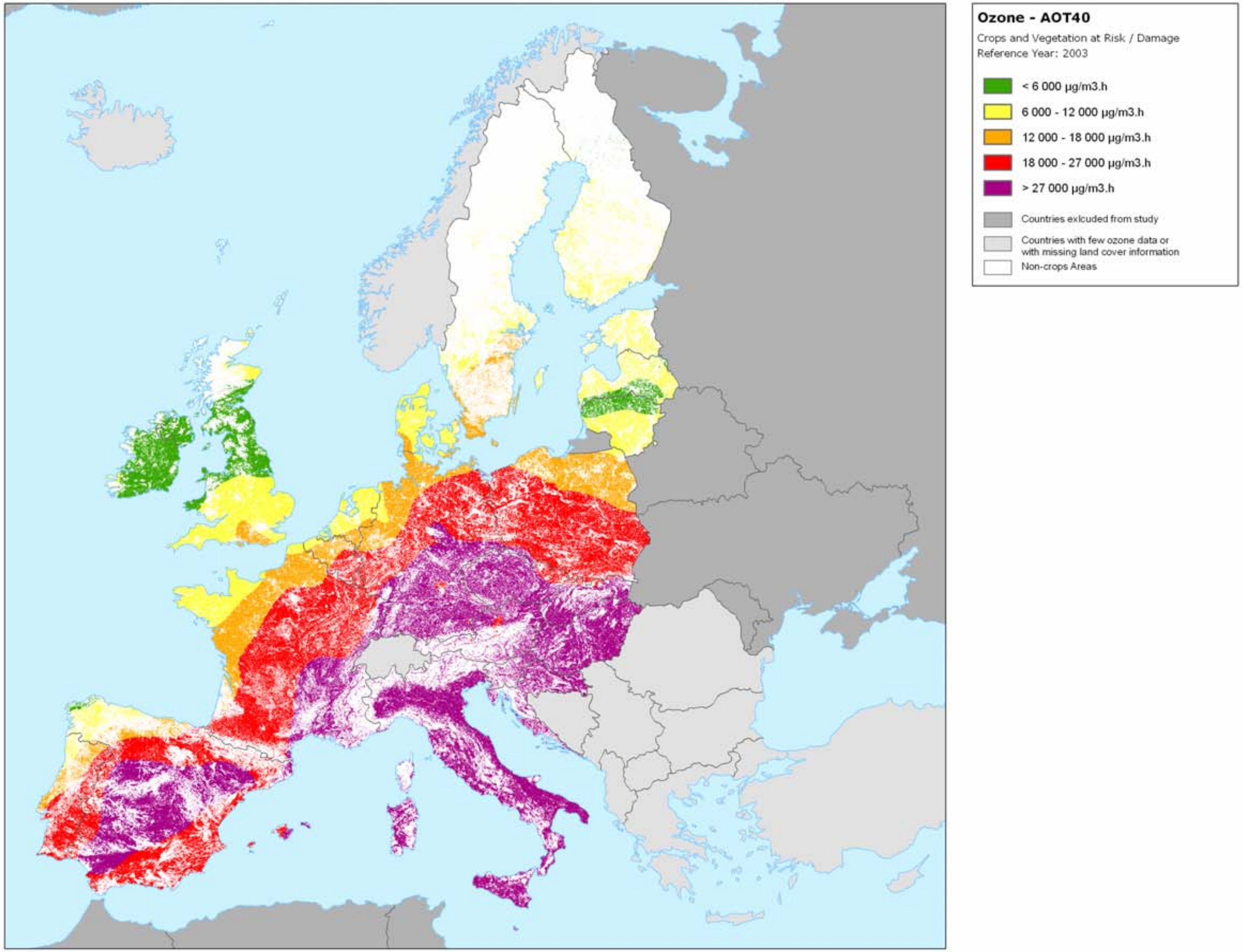


Figure A13 Population density map of Europe at a aggregated 10 km x 10 km grid resolution. Units: inhbs.km^{-2}.

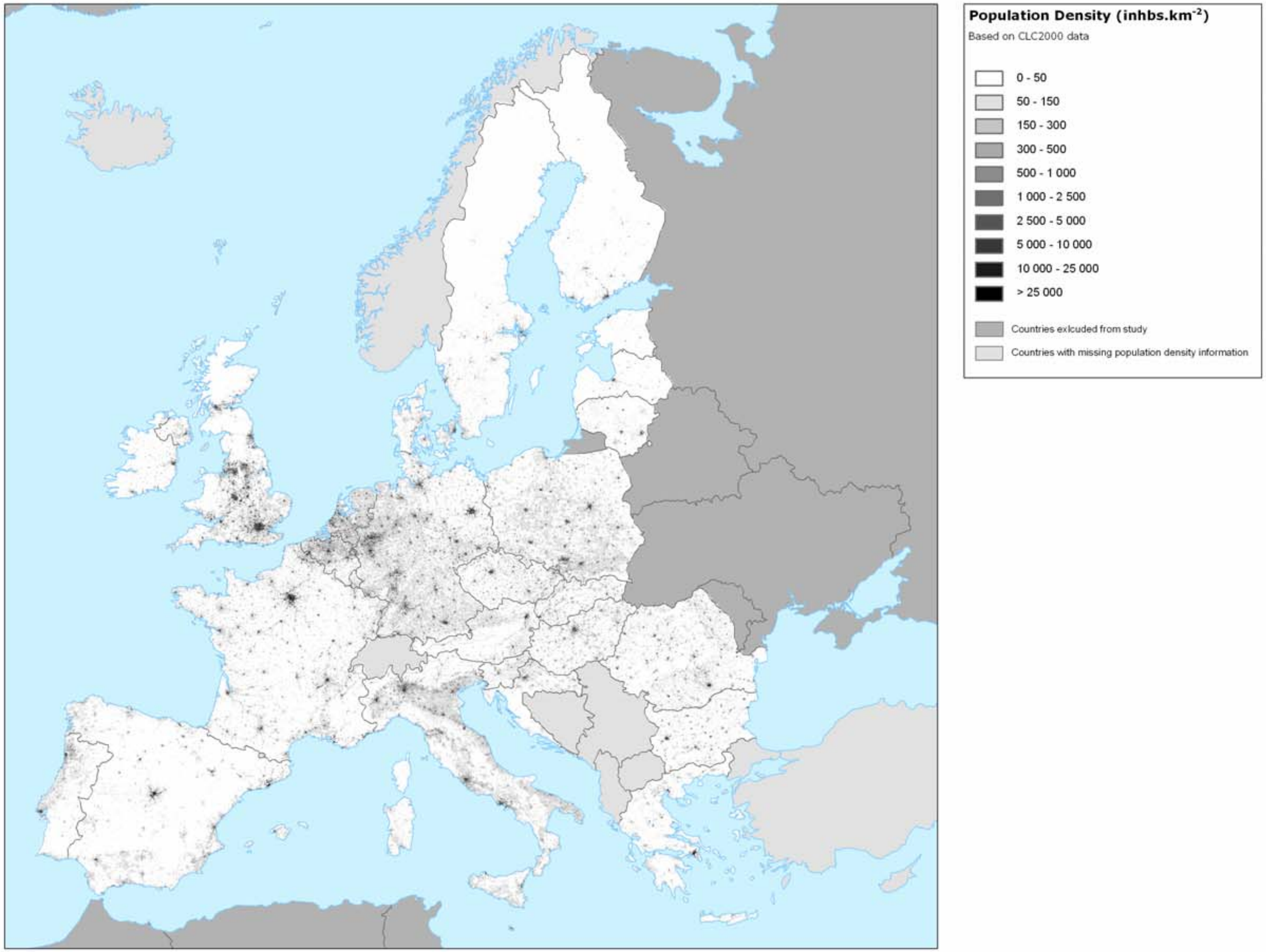


Figure A14 Combined rural and urban concentration map of ozone – SOMO35, year 2000. Spatial interpolated concentration field overlaid by population density grid. Units: µg.m^{-3}.days. This map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, RO. Countries with missing population density information and therefore excluded: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR.
(This map is the map of Figure A1 with on top of it as a transparency the map of Figure A13. The map shows the spatial match of relative low ozone concentrations at the more dense populated and urbanized areas even at those without measurements).

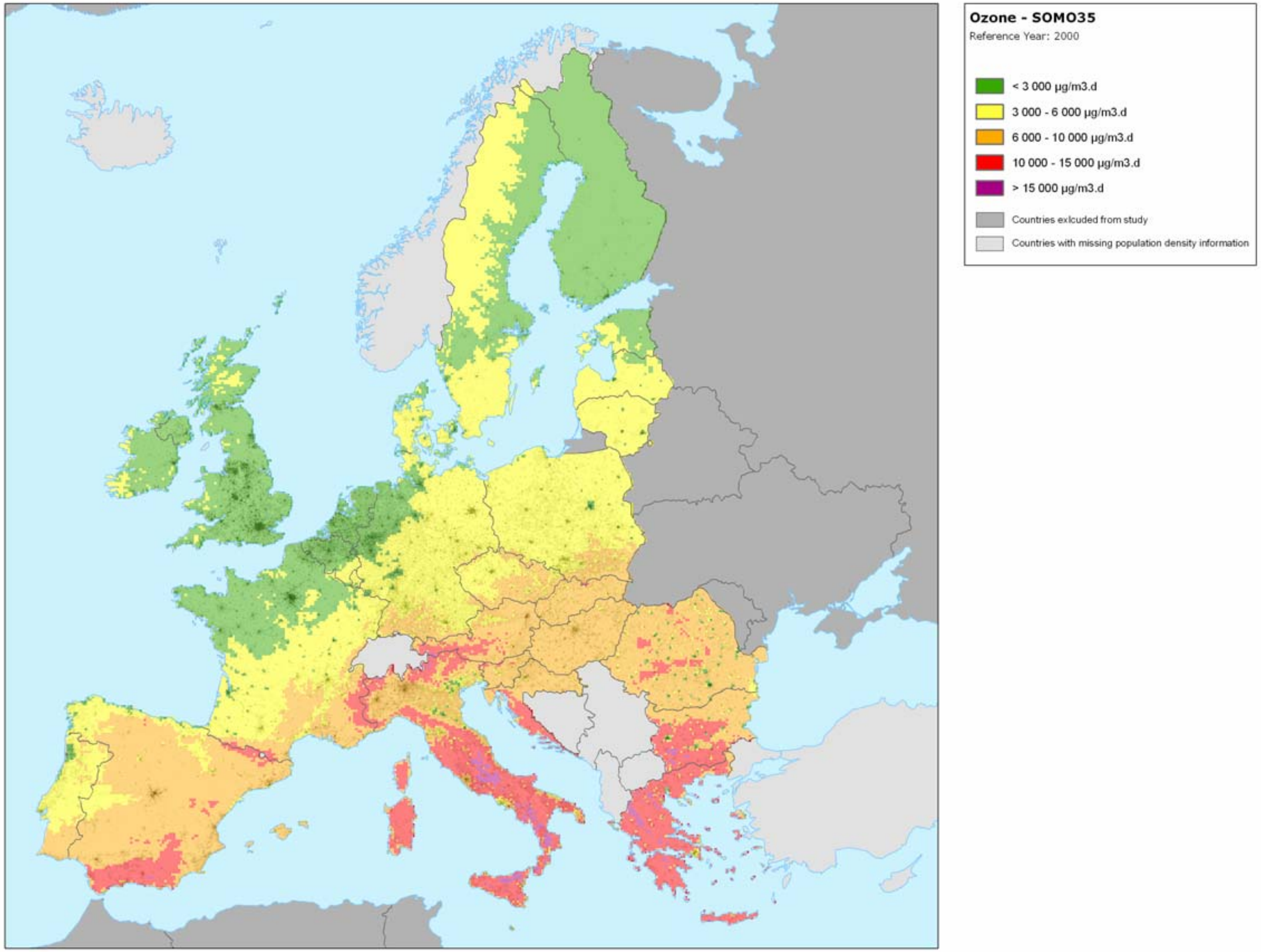


Figure A15 Combined rural and urban concentration map of ozone – SOMO35, year 2003. Spatial interpolated concentration field overlaid by population density grid. Units: µg.m^{-3}.days. This concentration map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, RO. Countries with missing population density information and therefore excluded from the mapping: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR. (This map is the map of Figure A2 with on top of it as a transparency the map of Figure A13. The map shows the spatial match of relative low ozone concentrations at the more dense populated and urbanized areas even at those without measurements).

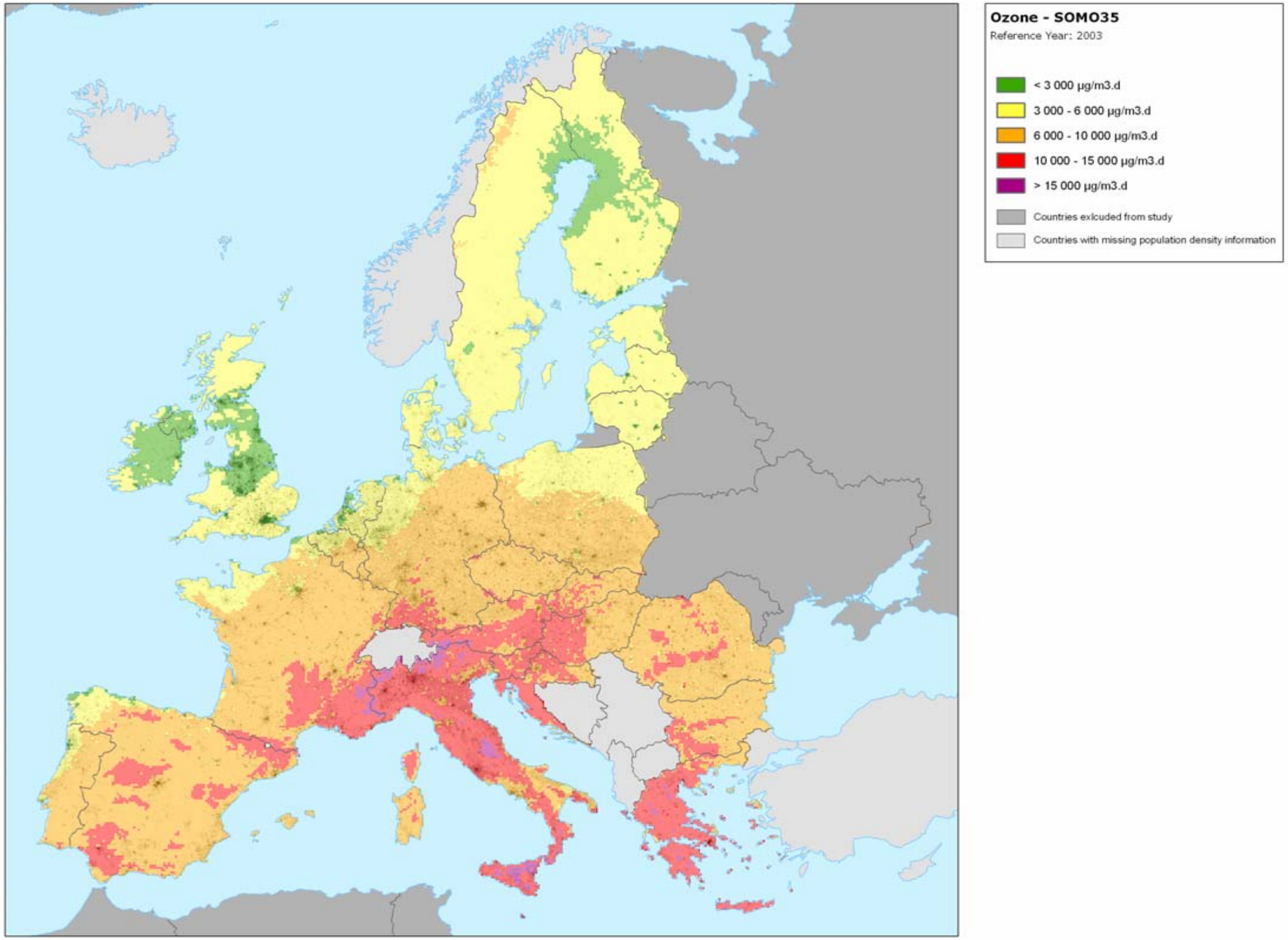


Figure A16 Combined rural and urban concentration map of PM_{10} – annual average, year 2001. Spatial interpolated concentration field overlaid by population density grid. Units: µg.m^{-3}.days. This concentration map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, HR, HU, RO. Countries with missing population density information and therefore excluded: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR.
(This map is the map of Figure A6 with on top of it as a transparency the map of Figure A13. The map shows the spatial match of relative elevated PM_{10} concentrations at the more dense populated and urbanized areas even at those without measurements).

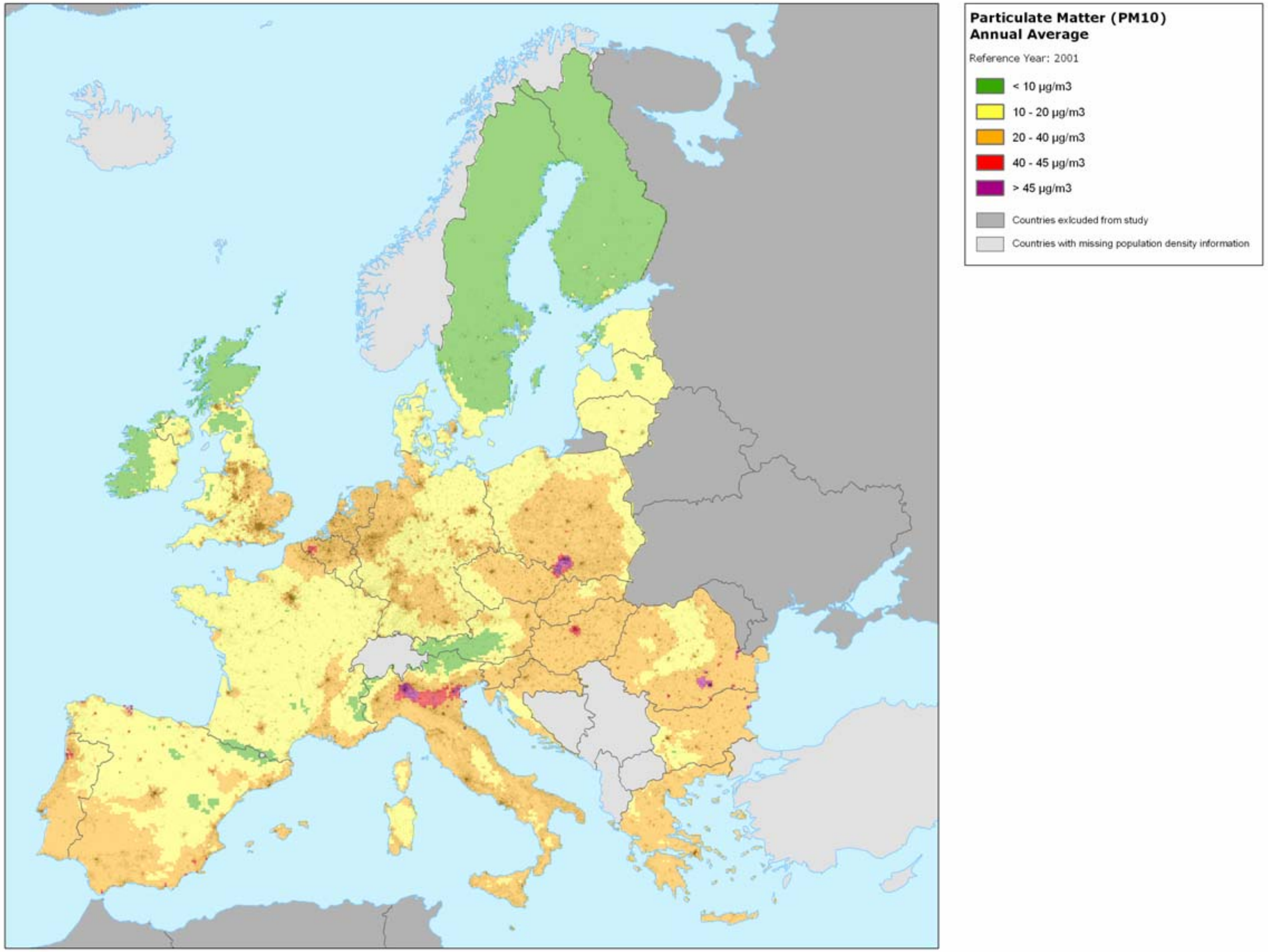


Figure A17 Combined rural and urban concentration map of PM_{10} – annual average, year 2003. Spatial interpolated concentration field overlaid by population density grid. Units: $\mu g.m^{-3}$. This concentration map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, HR, HU, RO. Countries with missing population density information and therefore excluded: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR.
(This map is the map of Figure A7 with on top of it as a transparency the map of Figure A13. The map shows the spatial match of relative elevated PM_{10} concentrations at the more dense populated and urbanized areas even at those without measurements).

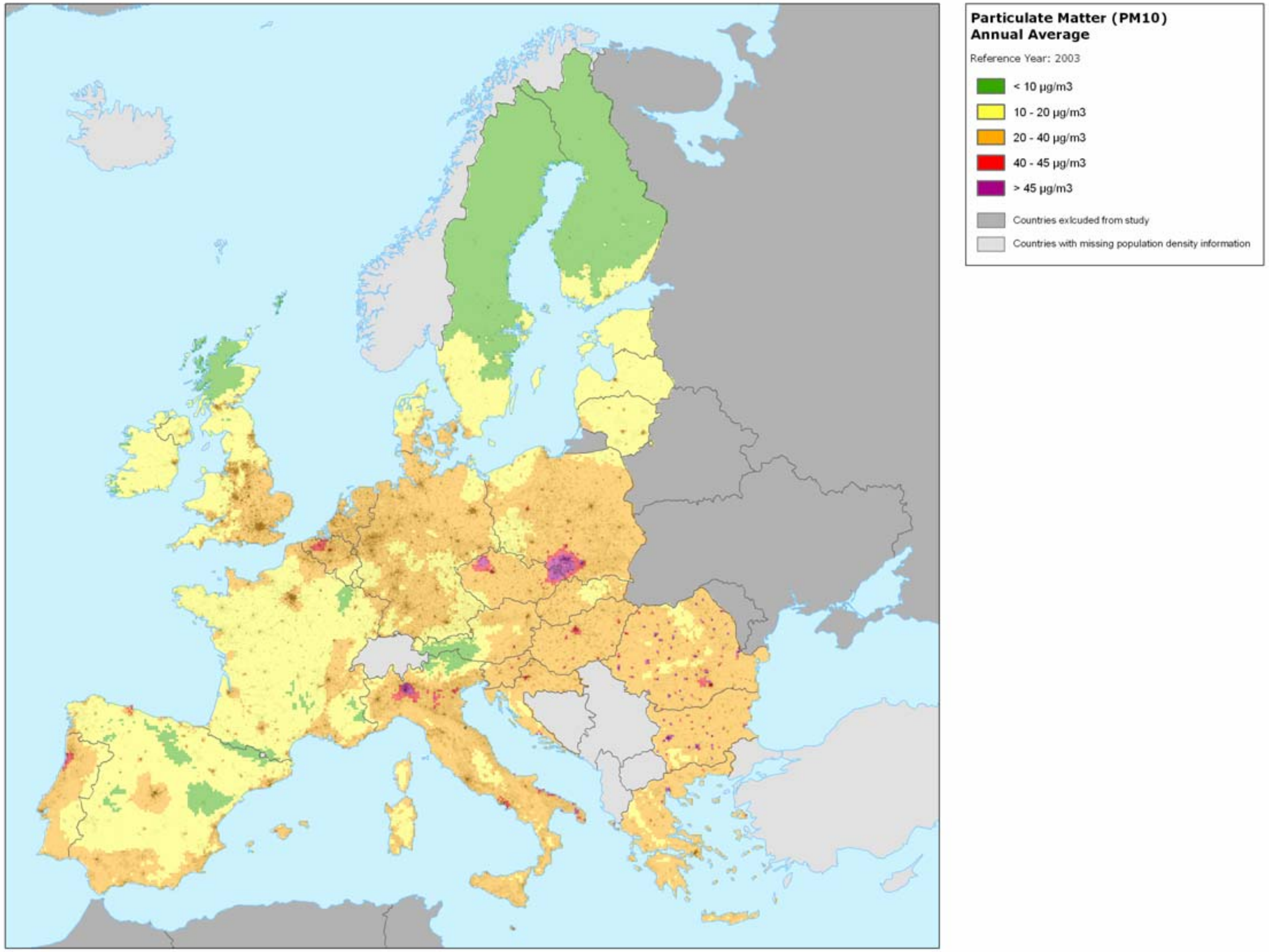


Figure A18 Combined rural and urban concentration map of PM_{10} – 36^{th} maximum daily average value, year 2001. Spatial interpolated concentration field overlaid by population density grid. Units: $\mu g.m^{-3}$. This map is created by merging the rural map (Method 3.b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration, using linear regression and ordinary kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, HR, HU, RO. Countries with missing population density information and therefore excluded: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR. (This map is the map of Figure A8 with on top of it as a transparency the map of Figure A13. The map shows the spatial match of relative elevated PM_{10} concentrations at the more dense populated and urbanized areas even at those without measurements).

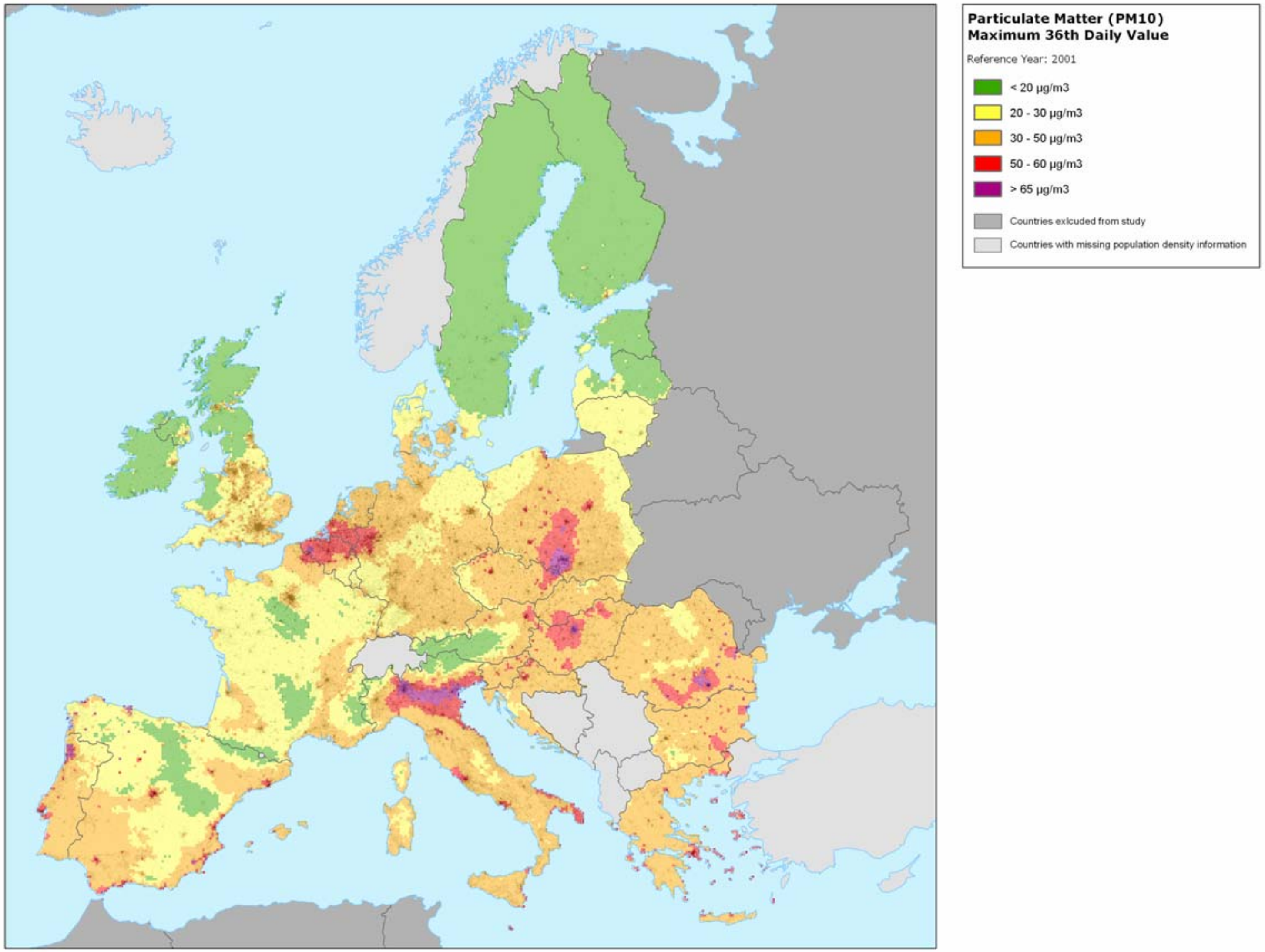


Figure A19 Combined rural and urban concentration map of PM_{10} – 36th maximum daily average value, year 2003. Spatial interpolated concentration field overlaid by population density grid. Units: $\mu g.m^{-3}$. This map is created by merging the rural map (Method 3b of Section 5.1: combination of measured values with EMEP model, altitude and sunshine duration , using linear regression and ordinary kriging of residuals) and the urban map (Method 3b of Section 6.7, using interpolation of urban Delta by ordinary kriging). Countries with interpolation based on additional data only: BG, GR, HR, HU, RO. Countries with missing population density information and therefore excluded: AD, AL, BA, CH, CS, CY, IS, LI, MK, NO, TR. (This map is the map of Figure A9 with on top of it as a transparency the map of Figure A13. The map shows the spatial match of relative elevated PM_{10} concentrations at the more dense populated and urbanized areas even at those without measurements).