# Methodological improvements on interpolating European air quality maps

**ETC/ACC Technical Paper 2009/16**
**February 2010**

*Jan Horálek, Peter de Smet, Frank de Leeuw,*
*Markéta Coňková, Bruce Denby, Pavel Kurfürst*

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 Environment Agency
MNP UBA-D UBA-V NILU AEAT AUTh CHMI MET.NO ÖKO TNO REC

***Front page picture:***

*Anonymous image from the internet, symbolising the efforts we make to improve the representativeness of the air quality maps and exposure assessments through improved methodologies on data use, data treatment and particularly on merging rural and urban map layer data.*


***Author affiliation:***

*Peter de Smet, Frank de Leeuw: Netherlands Environmental Assessment Agency (PBL), Bilthoven, The Netherlands*

*Jan Horálek, Markéta Coňková, Pavel Kurfürst: Czech Hydrometeorological Institute (CHMI), Praha, Czech Republic*

*Bruce Denby: Norwegian Institute of Air Research (NILU), Kjeller, Norway*

***EEA project manager:***

*Anke Lükewille, EEA, Copenhagen, Denmark*


**DISCLAIMER**





*ETC/ACC Technical Paper 2009/16*


*European Topic Centre on Air and Climate Change*
*PO Box 303*
*3720 AH Bilthoven*
*The Netherlands*
*Phone +31 30 2743550*
*Fax +31 30 2744433*
*Email etcacc@pbl.nl*
*Website http://air-climate.eionet.europa.eu/*

# Contents

# 1 Introduction

In the past few years we developed and tested a methodology for spatially mapping the air quality on European scale (Horálek et al., 2005, 2007 and 2008; EEA, 2009a). Now, European maps of $PM_{10}$ and ozone indicators are routinely prepared using a standardised spatial interpolation methodology (De Smet et al., 2009 and 2010). However, partial improvements of the methodology are needed to reduce the observed uncertainties in the interpolations. In this paper three different categories for potential uncertainty reductions have our specific focus: alternative data sources or data treatment, rural versus urban area treatment in merging procedures into one final map, and changing grid resolutions. For each category one or more options for improvements are addressed and tested on their relevancy for implementation into the default standard mapping procedure. The uncertainty reductions, this paper focuses on, are selected from the recommendations and summary list of Denby et al. (2009).

***Log-normal concentration transposition, altitude transformation, introduction of land cover data***

In Chapter 4 of this paper, three options for potential improvements described in published literature are examined. We quantify their actual contribution in reaching reduced uncertainties at the maps obtained with our current method of using a linear regression model with spatial interpolation of the regression residuals. These potential improvements are:

(i) Instead using the straightforward observed pollutant concentrations, apply the *log-normal transposition* of the point observations for $PM_{10}$ and ozone, as most actual pollutants in the current policy arena (Denby et al. (2005, 2008a, 2008b), Brabec et al. (2005), Van de Kassteele (2006), Beelen et al. (2009))

(ii) Instead using the direct topographic altitude field as supplementary data source, apply a more sophisticated *altitude transformation*, (Beelen et al., 2009)

(iii) Using *land cover* as supplementary data source (Janssen et al., 2008, Hoek et al., 2008).

The purpose of this paper is to verify and conclude whether these provide significant enhancements of the current methodology and worth implementing, despite the calculations additionally involved.

***Rural and urban merger***

Another aspect we examine in more detail in Chapter 5 is the improvement of the *merging process based on population densities* of the separate rural and urban maps into one combined final European air quality map, based on the population density map for defining the rural and urban areas. The aim is to reduce uncertainties in the final map and its derivates, such as the population exposure numbers and health impact assessments.

Currently, the rural and urban maps are created separately through linear regression followed by spatial interpolation of the regression residuals based on the rural background stations and urban/suburban background stations, respectively. Subsequent merging takes place on basis of a European high resolution population density map and according pre-defined criteria on population densities and relative concentration levels for selecting grid values for the final map from either the rural map, or the urban map, or a weighted combination. On top of that the current methodology includes an instrumental *joint urban/rural* map, compensating for assumed inconsistencies in the relative concentration levels between the rural and urban maps at several areas in Europe. For example, ozone concentrations in urban areas are considered to be lower than in neighbouring rural areas due to the higher ozone depleting effect of the higher nitrogen oxide emissions in urbanised areas caused by its higher human activity, e.g. industry, domestic heating and traffic. For $PM_{10}$ the opposite condition is assumed: higher concentration levels are assumed in urbanised areas due to higher human activities leading to higher particulate matter emissions compared to the neighbouring rural areas. At areas where these conditions are not fulfilled the map values from a third map, this joint

urban/rural map, are selected to compensate for the inconsistency. The joint urban/rural map is derived by taking into account all stations disregard their classified type and the type of area.

So far, we had limited attention to further improve the merging process due to the limited knowledge on, and availability of, reliable and accurate methodologies for merging on basis of population density data sources. Meanwhile we experience imperfections in the merging methodology that would need improvement, or that we at least could give some quantitative uncertainty label. For example, we use the monitoring data from AirBase stations as classified and reported by the countries themselves: rural background stations, suburban background stations and urban background stations. When comparing the stations types with the underlying population density map and area type classification we apply on this map for defining the rural and urban areas, we conclude that there exists a level of inconsistency between the station types as reported by the countries and the corresponding area types as classified in the map. This inconsistency, which is not easy to eliminate or reduce, introduces a certain level of uncertainty in the final maps which we try to indicate and quantify. Additionally, based on the experiences and some better understanding of observed uncertainties at current merging process, we explore and compare in this paper three alternative options next to the current method in attempts to optimise the merging and reduce its uncertainty:

A. Apply the current merging method, but ignoring the application of the joint urban/rural map to compensate for spatially bound inconsistencies in the rural and (sub)urban concentration levels. This provides information on how much the uncertainty reduces in the combined final map when the joint urban/rural map is used to compensate for the assumed spatial concentration inconsistencies. If the use of such extra map layer shows no significant uncertainty reduction one could consider leaving it out, saving calculation efforts.

B. Apply all rural and (sub)urban background stations directly in the linear regression model and subsequent residual interpolation, disregarding simply the (sub)urban and rural characteristics of both the station types and the area types. In fact this is considering the joint urban/rural map being representative as the final European interpolated map. It is the simplest and thus most efficient 'merging' method we can think of when using rural and (sub)urban observational data in spatial interpolation mapping.

C. Using a type of merging through applying a weighting fraction on the rural and urban areas directly in the linear regression model itself, followed by a spatial interpolation of the common (i.e., joint rural and urban) regression residuals. In other words, using one linear regression model consisting of a weighted fraction representing the rural areas with its relevant supplementary data and background stations, and a weighted fraction consisting of the urban area and its representative supplementary data and relevant background stations. Subsequently the spatial interpolation takes place on the common regression residuals attributable to both the rural and the urban areas. This option is, from a calculation efficiency point-of-view, a statistical and methodological compromise between using the three separate rural, urban and joint urban/rural maps to derive a combined final European map and using just the joint urban/rural map as the representative European map.

Up to now we evaluated the uncertainties in the combined final map by estimating the levels for the rural and the urban map separately. In this paper we estimate now the uncertainty of the combined final map itself as obtained with the current method, and compare it with the uncertainties found at the final maps resulting from the three alternative approaches. Through this comparison we are able to estimate the quantitative contribution of each methodological refinement and whether we should implement it as part of our standard interpolation mapping methodology.

Next to the comparison with the alternate merging methods, we also suggest and test in Section 5.3 two improvements of just the current merging methodology. They consist of using a different population density class interval boundary between the rural and mixed population density grid cell types and a more detailed specification of the conditions for selecting the ultimate grid square from the separate interpolated rural, urban and instrumental *joint urban/rural* map.

### *Grid resolution*

The linear regressions and their residual interpolations are currently executed on an aggregated 10x10 km grid resolution, despite some supplementary data being available in a finer grid (Section 3). That is to keep the calculation demands within reasonable hardware and software limits. Nonetheless, it is known that spatial aggregations do contribute to increased uncertainties and sub-grid variability in the mapping results. There are two options of interest to try to tackle this spatial resolution bound uncertainties and sub-grid variability. The first is by increasing simply the grid resolution and thus reducing uncertainties and distinguishing sub-grid variability through calculations and presentation in smaller grids. However, the computational demand may become limitative due to the quadratic increases at augmented grid resolutions. The second is by trying to describe for the current grid resolution the sub-grid variability through applying sophisticated geo-statistical indicators. In Chapter 6 we focus on the first by quantifying the uncertainties obtained through executing the linear regression and its residual interpolations, as well as the merging of the separate rural and urban maps into the combined final map on the 1x1 km grids, and compare these with the uncertainties of the 10x10 km grid resolution calculations. The recommendation for implementing a higher grid resolution as improvement of the default mapping methodology depends on the level of uncertainty reduction obtained versus the considerably higher calculation capacity needed to create maps on a higher resolution.

### *$PM_{10}$ and ozone indicators of 2006 and 2007*

All the analyses have been done for $PM_{10}$ and ozone with data of 2006, and partially also 2007 allowing us to achieve at least some minimal insight on possible inter-annual variability that may mask results. Maps for 2006 and 2007 created according the current methodology are already presented in De Smet et al. (2009, 2010). Their results are used to compare and conclude on the results of the proposed alternatives of this paper.

Some of the methodological techniques used in this paper are highlighted in Chapter 2. Chapter 3 provides a brief overview of the data sources used as primary and supplementary input data. The three data treatments (log-normal concentration transposition, altitude transformation and land cover application) intended to provide improved mapping results, are described and discussed in Chapter 4. Chapter 5 focuses on the different options of applying the rural and urban stations and areas in alternative 'mergers' to obtain a combined final European concentration map. The contribution of the level of grid resolution to a reduced uncertainty in concentration maps and population exposure estimates is dealt with in Chapter 6. Chapter 7 provides a summarising overview on the contribution of the uncertainty reductions obtained with the tested alternatives. Based on that, and on the complexity of their applicability, we estimate which alternative calculation approaches are considered suitable and efficient enough to implement as part of the default spatial interpolation methodology for future routinely production of European wide air quality maps, and their derivates such as population exposures.

# 2 Methodology

In this chapter we first summarise in Section 2.1 the current mapping methodology used for the routinely production of interpolated air quality maps and derivates. Then methodological descriptions are given on the options we focus on for potential uncertainty reductions. It consists of a set of three categories: alternative data sources as supplementary data or specific data treatment, options of alternative rural and urban area treatment in merging procedures to attain one final map, and changing grid resolutions. Section 2.2 considers the log-normal concentration transposition and the altitude transformation as data treatments that could provide potential uncertainty reduction. As additional supplementary data source the inclusion of land cover is examined for its potential. Section 2.3 highlights three alternative methodological approaches for the rural and urban merging procedure and Section 2.4 describes a refined grid resolution on its merits. Finally, Section 2.5 explains which geo-statistical methods are used to compare between the existing and the alternative methodological approaches, necessary to come to an objective quantitative judgement of their contribution to the uncertainty reductions.

## *2.1 Current mapping method*

The basic mapping method currently used, and described in Horálek et al. (2007), is applying a linear regression model followed by kriging of its residuals (residual kriging). Interpolated maps for rural and urban areas are created separately. The rural map is based on the rural background stations and the urban map on the urban and suburban background stations. The interpolation is carried out according to the relation:

$$\hat{Z}_t(s_0) = c_t + a_{t1}X_{t1}(s_0) + a_{t2}X_{t2}(s_0) + ... + a_{tn}X_{tn}(s_0) + \eta_t(s_0) \qquad (2.1)$$

where $\hat{Z}_t(s_0)$ is the estimated value of the air pollution indicator for the type of the area $t$ at point $s_o$

$t$ is the type of area ($r$ - rural or $u$ - urban)

$X_{t1}(s_0), X_{t2}(s_0),\ldots, X_{tn}(s_0)$ are the $n$ number of individual supplementary variables participating in the linear regression for the type of area $t$ at point $s_o$

$c_t, a_{t1}, a_{t2},,\ldots, a_{tn}$ are the $n+1$ selected parameters of the linear regression model calculated for the type of area $t$, based on either the rural or the urban and suburban background stations depending on the type of area considered

$\eta_t(s_0)$ is the spatial interpolation of the residuals of the linear regression model for the type of the area $t$ at the point $s_o$.

The spatial interpolation of residuals is carried out for each of those maps using ordinary kriging (Cressie (1991), Wackernagel (2003)), based on variogram analysis using a spherical function with parameters *nugget, sill* and *range*. For different pollutants and area types, different supplementary data are used depending on their contribution to an improved fit of the regression. As Table 1.1 summarises, for $PM_{10}$ at rural areas the supplementary data used is EMEP model output, altitude, wind speed and solar radiation, and at urban areas it is EMEP model output only. For ozone at rural areas it is EMEP model output, altitude, and solar radiation, and at urban areas the EMEP model output, wind speed and solar radiation.

*Table 1.1 Supplementary data used in ETC/ACC routine mapping for individual pollutants and area types*

| Supplementary data sources | $PM_{10}$ rural | $PM_{10}$ urban | $O_3$ rural | $O_3$ urban |
|---|---|---|---|---|
| EMEP model | ✓ | ✓ | ✓ | ✓ |
| Altitude | ✓ | | ✓ | |
| Wind speed | ✓ | | | ✓ |
| Solar radiation | ✓ | | ✓ | ✓ |

The merging of the rural and urban map into one combined final map is done on basis of an aggregated 10x10 km grid resolution of the originally 100x100 m population density grid. For areas with a population density less than a defined value of $\alpha_1$ the rural map is applied, and for areas with a population density greater than a defined value $\alpha_2$ the urban map is applied. For areas with a population density within the interval $(\alpha_1,\alpha_2)$ a weighting function with $\alpha_1$ and $\alpha_2$ is applied. The merging of the separate map for rural areas and for urban areas is based on the assumption that in rural areas adjacent to urban areas the relative air pollution levels are lower for $PM_{10}$, and higher for ozone, due to more intensive human activities in urbanised areas, leading to higher emissions of $PM_{10}$ and ozone depleting substances (Horálek et al. 2005). However, this appears not to be the case at about 4.5% to 11.5% of the total area in the year 2006 and at about 5.4% to 8.8% of the total area in the year 2007, depending on the pollutant indicator. For these areas the *joint urban/rural map* is constructed by taking all background stations (both rural and urban/suburban) together and applying equation

$$\hat{Z}_j(s_0) = c_j + a_{j1}X_{j1}(s_0) + a_{j2}X_{j2}(s_0) + ... + a_{jn}X_{jn}(s_0) + \eta_j(s_0) \qquad (2.2)$$

where $\hat{Z}_j(s_0)$ is the estimated joint rural/urban value of the air pollution indicator at point $s_o$

$X_{j1}(s_0), X_{j2}(s_0),\ldots, X_{jn}(s_0)$ are the $n$ number of individual supplementary variables at point $s_o$

$c_j, a_{j1}, a_{j2},\ldots,a_{jn}$ are the $n+1$ selected parameters of the linear regression model calculated based on all background stations disregard the type

$\eta_j(s_0)$ is the spatial interpolations of the residuals of the linear regression model at the point $s_o$.

The supplementary data used to construct the joint urban/rural map is EMEP modelling output, altitude, wind speed and surface solar radiation, for both the $PM_{10}$ and the ozone indicators.

In the case of the $PM_{10}$ indicators the combined final European map is composed from the separate maps according the grid value selection criteria reflected in Equation 2.3,

$$\begin{aligned}
\hat{Z}(s_0) &= \hat{Z}_r(s_0) && \text{for } \alpha(s_0) \le \alpha_1 \text{ and } \hat{Z}_r(s_0) \le \hat{Z}_u(s_0) \\
&= \hat{Z}_u(s_0) && \text{for } \alpha(s_0) \ge \alpha_2 \text{ and } \hat{Z}_r(s_0) \le \hat{Z}_u(s_0) \\
&= \frac{\alpha_2 - \alpha(s_0)}{\alpha_2 - \alpha_1} \cdot \hat{Z}_r(s_0) + \frac{\alpha(s_0) - \alpha_1}{\alpha_2 - \alpha_1} \cdot \hat{Z}_u(s_0) && \text{for } \alpha_1 < \alpha(s_0) < \alpha_2 \text{ and } \hat{Z}_r(s_0) \le \hat{Z}_u(s_0) \\
&= \hat{Z}_j(s_0) && \text{for } \hat{Z}_r(s_0) > \hat{Z}_u(s_0)
\end{aligned} \qquad (2.3)$$

where $\hat{Z}(s_0)$ is the estimated value of concentration at point $s_o$

$\hat{Z}_r(s_0)$ is the concentration at point $s_o$ for the rural map, based on the rural background stations only

$\hat{Z}_u(s_0)$ is the concentration at point $s_o$ for the urban map, based on the urban and suburban background stations only

$\hat{Z}_j(s_0)$ is the concentration at point $s_o$ for the joint urban/rural map, based on all background stations

$\alpha(s_o)$ is the density of population at point $s_o$

$\alpha_1, \alpha_2$ are the population density classification interval parameters, set as $\alpha_1$ = 100 inhbs.km$^{-2}$ and $\alpha_2$ = 500 inhbs.km$^{-2}$.

The final map for ozone indicators is calculated similarly, but with *opposite* conditions for the relation between $\hat{Z}_r(s_0)$ and $\hat{Z}_u(s_0)$. Figure 2.1 illustrates for $PM_{10}$ the four selection criteria of Equation 2.3 used to determine the ultimate grid value of the combined final European map. The selection of the ultimate grid cell value to be inserted from the separate maps happens at each grid square of the combined final map according one of the four selection criteria and depends on the conditions in the population density and on the relative concentration levels at that specific grid cell.

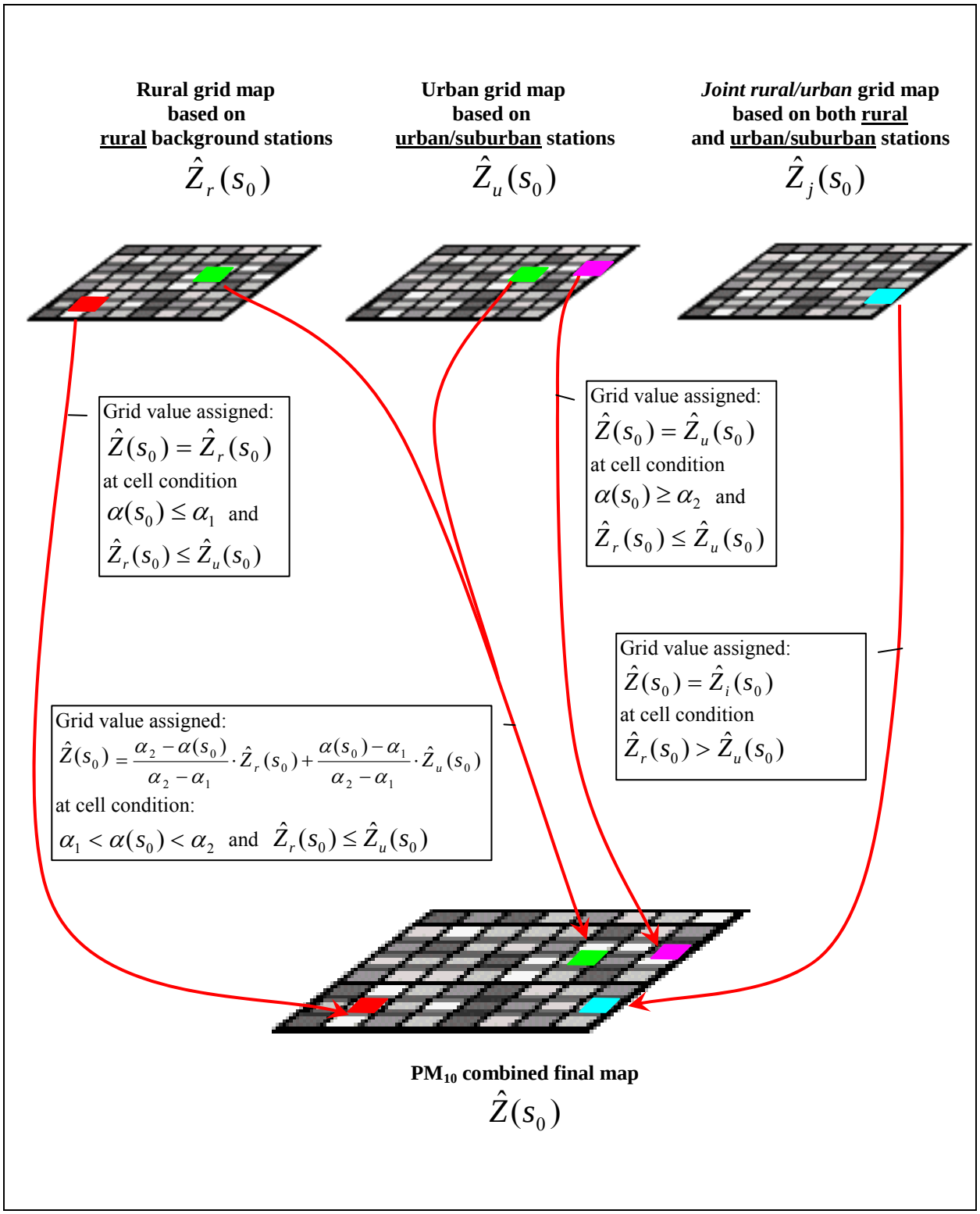


*Figure 2.1 The four selection criteria used to determine for $PM_{10}$ the ultimate grid value of the combined final European map. Each case depends on the conditions in the population density and on the relative concentration levels of the rural map and urban/suburban map.*

The values set for the parameters $\alpha_1$ and $\alpha_2$ in Equation 2.4 are based on the analysis presented in Chapter 7 of Horálek et al. (2005).

The resolution used for creating the maps is based on the EEA standard 10x10 km reference grid in the map projection ETRS89-LAEA5210, map extent 1c (http://www.eionet.eu.int/gis).

## 2.2 Data treatments and sources for methodological improvements

This section motivates and describes the log-normal concentration transposition and the altitude transformation as data treatments that could provide potential uncertainty reduction. As additional supplementary data source the inclusion of land cover is examined for its expected potential.

### 2.2.1 Log-normal concentration transposition

The measured air quality concentrations are in principle non-negative, thus their distribution is likely to be log-normal in character. If this is the case, then the interpolation could be carried out after logarithmic transformation with subsequent back-transformation of results. In Denby et al. (2008b) logarithmic transformations were also applied to daily mean $PM_{10}$ data. However, it is not necessarily the case that integrated indicators percentiles are log-normally distributed, but advantages of using logarithmic transformation are: (i) if the data is log-normally distributed then the transformation leads to a normal distribution for the interpolation; (ii) avoids unwanted negative values; (iii) leads to relative uncertainties, rather than absolute, in the interpolated fields, which is intuitively a more correct description of the uncertainty when dealing with a wide range of concentrations (Denby et al. 2009). In the work of Beelen et al. (2009) logarithmic distributions of integrated indicators were assumed.

In this paper we test whether the measured air quality data can be considered as to be normally or log-normally distributed. The goodness of fit is tested using the Kolmogorov-Smirnov test (Rao, 1973). This test is often used as a goodness of fit test, including the special case of testing for normality of the distribution. in that case the samples are standardized and compared with a standard normal distribution. In our case - the testing of log-normal distribution - the values are logarithmically transformed. The goodness of fit is represented by the level of probability that the statistics used in the test correspondents with the normality of the distribution, i.e. the *p-value* should be as close as possible to 1. In our test we apply a percentage level of 95%, i.e. the hypothesis of a normal distribution is rejected at a critical level of $p < 0.05$.

Next to this, we logarithmically transform concentrations from both the air quality measurements and the EMEP modelling output. Then we apply similar to Equation 2.1 the multiple linear regression, followed by residual kriging:

$$\hat{Y}(s_0) = c + a_1 \ln(X_1(s_0)) + a_2 X_2(s_0) + \ldots + a_n X_n(s_0) + \eta(s_0) \qquad (2.4)$$

where $\hat{Y}(s_0)$ is the estimated value of the logarithmical transformed concentration at point $s_o$

$X_1(s_0)$ is the output of dispersion model at point $s_o$

$X_2(s_0),\ldots, X_n(s_0)$ are the other supplementary variables at point $s_o$

$c, a_1, a_2,,\ldots, a_n$ are the $n+1$ selected parameters of the linear regression model calculated at the point of measurement $s_0$,

$\eta(s_0)$ is the spatial interpolation of the residuals of the linear regression model at the point of measurement.

The interpolated values are back-transformed by exponentiation with the kriging error (Cressie, 1993):

$$\hat{Z}(s_0) = \exp\left\{ Y(s_0) + \frac{\sigma^2(s_0)}{2} \right\} \quad (2.5)$$

where $\hat{Z}(s_0)$ is the estimated back-transformed concentration value at point $s_o$

$\sigma(s_0)$ is the kriging error value at point $s_o$.

This log-normal transposition of concentrations is applied separately for rural and urban areas and examined in Section 4.1 on its potential for reduced uncertainties in the combined final European map and its suitability for implementation as part of the standard interpolation methodology of specific pollutant indicators.

### 2.2.2 Altitude transformation

Altitude is an important supplementary data source since its represents the topography of an area, which affects air pollution patterns in various ways. It has its influence on the meteorological conditions, such as wind speed and rainfall, and associates with land use and emission intensity. Its relation with air quality concentrations is non-linear. In order to somewhat linearize this relation, the altitude values of the aggregated 10x10 km grid squares are transformed in line with Beelen et al. (2009) in the following way:

$$X'_{alt}(s_0) = \sqrt{\frac{X_{alt}(s_0) - \min(X_{alt})}{\max(X_{alt}) - \min(X_{alt})}} \quad (2.6)$$

where $X'_{alt}(s_0)$ is the transformed altitude at grid square $s_o$

$X_{alt}(s_0)$ is altitude at grid square $s_o$

$\min(X_{alt})$, $\max(X_{alt})$ are the minimal and maximal altitudes found across the complete mapping domain

This refined altitude approach is examined in Section 4.2 on its potential for improved interpolations at rural areas only, since at the current methodology altitude is applied as supplementary data source only at these areas. Thus the transformed altitude is used in the rural linear regression model (Eq.2.1, with $t = r$) as one of the supplementary variables $X_r$.

### 2.2.3 Land cover inclusion

Janssen et al. (2008) expected that next to population density the application of additional land use information would contribute to explain local characteristics of air pollution in Belgium. In order to reduce the high number of degrees of freedom the original 44 land cover classes are regrouped into 11 more general land cover classes (see Section 3.5 for details), similar to Janssen et al. To find out whether it could contribute also on a European mapping scale, we explore the inclusion of land cover as supplementary data source in the linear regression model (Eq.2.1). We follow two approaches.

First, each general land cover class is accounted for as a separate variable. The selection of the most suitable general classes for inclusion into the multiple linear regression model is done by using a backward stepwise eliminating regression, which was also earlier applied for the selection of the most 'contributable' supplementary data sources as given in Section 2.1. (Backward selection begins with all variables and removes the least useful as long as the regression fit is not substantially “worsened”. Stepwise regression allows “good” predictors to re-enter at any step into the model. All improvements should lie within defined statistical criteria.)

Parallel to that – as additional option – one joined land cover function is constructed from the 11 general classes, based on the multiple regression of air pollutant measurement data and 11 land cover classes:

$$X_{LC}(s) = \sum_{i=1}^{11} w_i X_{LCi}(s) \tag{2.7}$$

where $X_{LC}(s)$ is the joined land cover function at measurement point $s$

$X_{LCi}(s)$ are the individual general land cover classes distinguished at point $s$, $i$ = 1, ..., 11

$w_i$ are the weights of the distinguished general land cover classes, calculated on basis of the multiple linear regression (as its parameters).

Then this joined land cover function is inserted in the linear regression model (Eq.2.1) as a supplementary variable $X_{ti}$.

The contribution to improvements with land cover inclusion is examined in Section 4.3 for rural and urban areas separately.

## 2.3 Rural and urban merging options

As the introduction highlights, we explore and compare in this paper three alternative options next to the current method in an attempt to obtain a more optimal merging procedure of the rural and urban interpolated areas into one final European map, against reduced uncertainties. In Section 2.1, Equation 2.3 demonstrates the current merging methodology to obtain one combined final European map out of the separate urban, rural and joint rural/urban interpolated maps. The three alternative approaches to obtain a final European map are based on the equations given below:

**A.** The rural and urban interpolated maps are constructed according to Equation 2.1 and the combined final European map is merged according Equation 2.3, but ignoring the conditional case of the joint urban/rural map. This is expressed by Equation 2.8:

$$\begin{aligned} \hat{Z}(s_0) &= \hat{Z}_r(s_0) && \text{for } \alpha(s_0) \le \alpha_1 \\ &= \hat{Z}_u(s_0) && \text{for } \alpha(s_0) \ge \alpha_2 \\ &= \frac{\alpha_2 - \alpha(s_0)}{\alpha_2 - \alpha_1} \cdot \hat{Z}_r(s_0) + \frac{\alpha(s_0) - \alpha_1}{\alpha_2 - \alpha_1} \cdot \hat{Z}_u(s_0) && \text{for } \alpha_1 < \alpha(s_0) < \alpha_2 \end{aligned} \tag{2.8}$$

where $\hat{Z}(s_0)$ is the estimated value of concentration at point $s_o$

$\hat{Z}_r(s_0)$ is the concentration at point $s_o$ for the rural map

$\hat{Z}_u(s_0)$ is the concentration at point $s_o$ for the urban map

$\alpha(s_o)$ is the density of population at point $s_o$

$\alpha_1$, $\alpha_2$ are the population density classification interval parameters, set as $\alpha_1$ = 100 inhbs.km$^{-2}$ and $\alpha_2$ = 500 inhbs.km$^{-2}$.

The results of this approach are compared with the current method to see how much the uncertainty reduces in the combined final map when the joint urban/rural map is used to compensate for the assumed spatial concentration inconsistencies. If the use of such extra map layer shows no significant uncertainty reduction one could consider leaving it out, saving calculation efforts.

**B.** This alternative examines the simplest approach one can think of by not distinguishing between rural and urban areas and rural and (sub)urban station types in the linear regression model. One applies here Equation 2.2 for the whole mapping area, based on the whole set of background stations disregarding their rural or urban/suburban characteristics. In fact this is considering the joint urban/rural map being representative as the final European interpolated map.

**C.** A type of merging that uses a weighting factor for the rural and urban areas, applying the same population density criteria as before, but now based on the whole set of rural and (sub)urban background stations disregard their type directly in the regression model (Eq.2.1) itself, with subsequent spatial interpolation of the common (i.e., joint rural and urban) regression residuals. The linear regression model used is given by Equation 2.9:

$$\hat{Z}(s_0) = c + w_r.(a_{r1}.X_{r1}(s_0) + a_{r2}.X_{r2}(s_0) + ...) + w_u.(a_{u1}.X_{u1}(s_0) + a_{u2}.X_{u2}(s_0) + ...) + \eta(s_0) \quad (2.9)$$

with
$$w_r = 1 \quad \text{for } \alpha(s_0) \le \alpha_1$$
$$= \frac{\alpha_2 - \alpha(s_0)}{\alpha_2 - \alpha_1} \quad \text{for } \alpha_1 < \alpha(s_0) < \alpha_2$$
$$= 0 \quad \text{for } \alpha(s_0) \ge \alpha_2$$
$$w_u = 1 - w_r$$

where $\hat{Z}(s_0)$ is the estimated value of the air pollution indicator at point $s_o$

$w_r$ and $w_u$ are the weighting fractions for the rural and urban area respectively

$X_{r1}(s_0), X_{r2}(s_0), \ldots$ are the supplementary quantities at point $s_o$ for the rural area fraction

$X_{u1}(s_0), X_{u2}(s_0), \ldots$ are the supplementary quantities at point $s_o$ for the urban area fraction

$c, a_{r1}, a_{r2}, \ldots, a_{u1}, a_{u2}, \ldots$ are the rural and urban area parameters of the linear regression model, calculated at the points of measurement

$\eta(s_0)$ is the spatial interpolation of the residuals of the linear regression model at the point of measurement, disregard its station type

$\alpha(s_o)$ is the density of population at point $s_o$

$\alpha_1, \alpha_2$ are the population density classification interval parameters, set as $\alpha_1 = 100$ inhbs.km$^{-2}$ and $\alpha_2 = 500$ inhbs.km$^{-2}$.

This third option is, from a calculation efficiency point-of-view, a statistical and methodological compromise between the current merging methodology that uses three separate rural, urban and joint urban/rural map layers to derive a combined final European map and option B as simplest approach of just using the joint urban/rural map as the representative European map.

In Chapter 5 the three alternatives are examined and compared with the current merging methodology on their levels of uncertainty. Depending on the significance of the results we may change the merging procedure by implementing (parts of) an alternative into the standard interpolation mapping methodology. The complexity and efficiency of their applicability will play a role in the decision for implementation.

## *2.4 Grid resolution refinements*

The role of the grid resolution used in spatial interpolation, map merging and population exposure assessment is examined in Chapter 6 for both the two $PM_{10}$ and the two ozone indicators. The aim is to provide input for optimizing the grid resolution at each process step, accounting for both the suppression of uncertainties as well as keeping calculations within acceptable computational limits. The resolutions considered are those provided as the EEA standard resolutions for pan-European mapping goals: the 1x1 km grid and the 10x10 km grid. The examination focuses on three subsequent stages in the mapping and assessment process by comparing the cross-validation RMSE, MPE (i.e. bias) and scatter-plots of their (intermediate) products. The three stages relate to (1) the interpolation of the separate rural and urban maps, including at conditional cases the joint urban/rural map, (2) the merging into the combined final concentration map and (3) the ultimate accounting for population exposures on basis of that map.

At each of the three steps the following comparisons are executed:

(1) In Section 6.1 the separate rural and urban maps are constructed according Equation 2.1 on a 1x1 km and on a 10x10 km grid resolution for the health indicators of $PM_{10}$ and ozone. At conditional cases explained by equation 2.2 the joint urban/rural map is engaged as well. At each pollutant indicator the maps in the two different grid resolutions are mutually compared by cross-validation RMSE and scatter-plots (see Section 2.5).

(2) In Section 6.2 the influence of the grid resolution on the merging into the combined final concentration maps is examined. As input we take, in accordance with Equation 2.3, (i) the separate interpolated rural map, urban map and joint urban/rural map on a 1x1 km as well as on a 10x10 km resolution, and (ii) merging on basis of the population density grid applying Equation 2.3, in a 1x1 km, a 10x10 km, and additionally a 3x3 km floating grid resolution (see Section 3.6). Several combinations of (i) and (ii) are executed; not only the same resolution for both the interpolated maps and the merging process but also, for example, the 10x10 km separate maps merged on basis of the 1x1 km population density map into a 1x1 km combined final map. The comparisons are executed again according Section 2.5.

(3) The influence of the grid resolution on the population exposure estimates is investigated. The comparison is done with regard to (i) population weighted concentrations and (ii) exposure classes.

Figure 2.2 illustrates the examined combinations. From the above one can deduce that we examine a set of combinations of grid resolutions in the process steps of spatial interpolation and merging on basis of classes of the population density map into a combined final concentration map. Subsequently this concentration map is used to estimate the national and European population exposures by means of either the average population exposure or predefined exposure classes. The combinations are

A. *(10-10-10)* 10x10 km separate rural, urban and joint urban/rural maps - joint with 10x10 km grid merging - into 10x10 km combined final concentration map with population exposure estimates (i.e. current approach)

B. *(10-fl3-1)* 10x10 km separate rural, urban maps and joint urban/rural maps - joined with 3x3 km floating grid merger - into 1x1 km combined final concentration map with population exposure estimates

C. *(10-1-1)* 10x10 km separate rural, urban and joint urban/rural maps - joint with 1x1 km grid merging - into 1x1 km combined final concentration map with population exposure estimates

D. *(1-1-1)* 1x1 km separate rural, urban and joint urban/rural maps - joint with 1x1 km grid merging - into 1x1 km combined final concentration map with population exposure estimates

E. *(1-fl3-1)* 1x1 km separate rural, urban and joint urban/rural maps - joined with 3x3 km floating grid merger - into 1x1 km combined final concentration map with population exposure estimates

Details of and motivation for including a 3x3 km floating grid merger are given at Section 3.5 and 3.6.

Approach F is inserted for scrutinizing the effect spatial averaging has on the ultimate population exposure estimates expressed as national and European population weighted concentrations and exceedance classes.

F. *(aggrD-10)* approach D, followed by an additional step of spatially averaging the 1x1 km combined final concentration map into an aggregated final 10x10 km concentration map with population exposure estimates

The ultimate population exposure estimates obtained by the different combinations are compared in Section 6.3 for the individual countries and Europe as a whole on basis of their population exposure tables and scatter-plots.

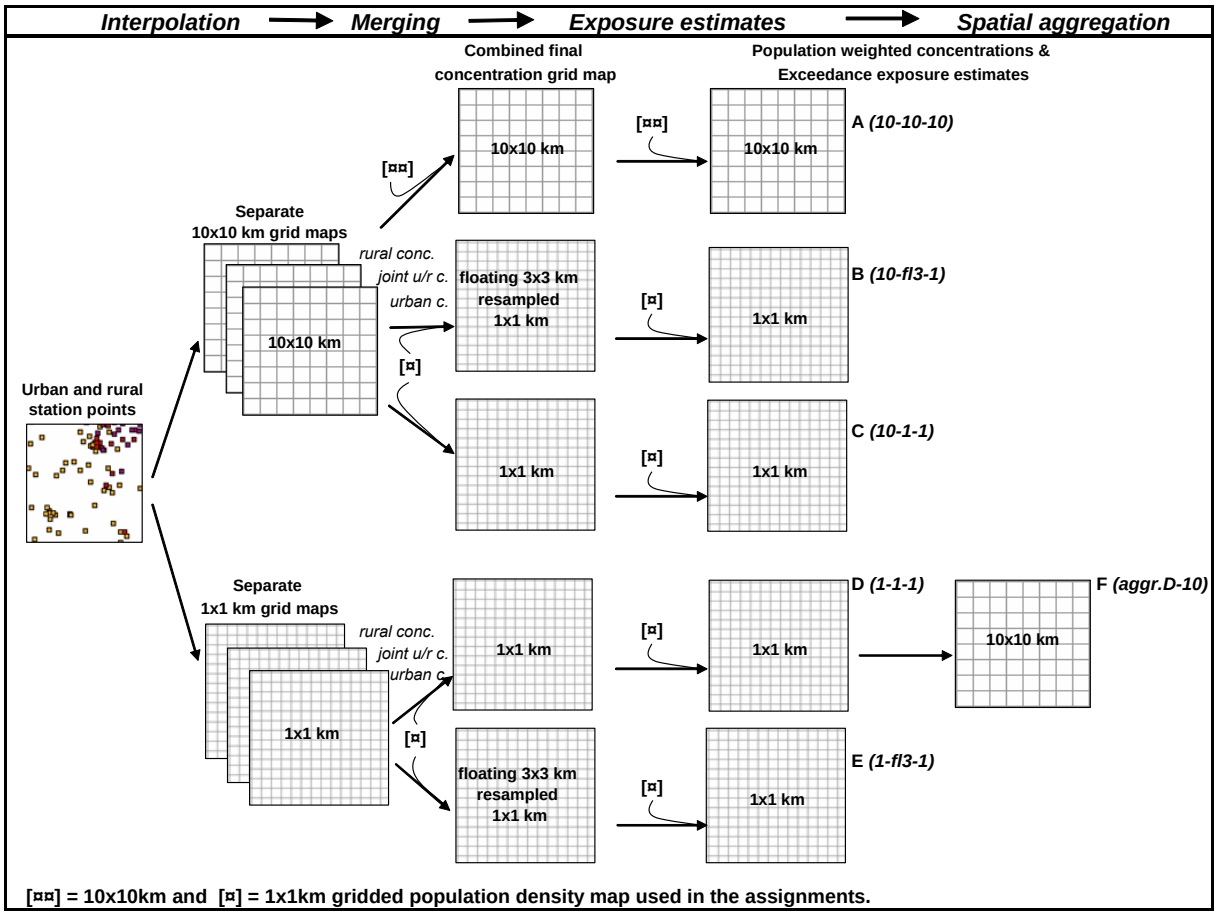


*Figure 2.2 Combinations of grid resolutions in the interpolation, merging and population exposure estimates as examined on their performance.*

## 2.5 Criteria for comparing mapping methods

The comparison of the uncertainty of mapping methods is done by cross-validation. The cross-validation method computes the quality of the spatial interpolation for each measurement point from all available information except from the point in question, i.e. it withholds one data point and then makes a prediction at the spatial location of that point. This procedure is repeated for all measurement points of the available set. The predicted and measured values at these points are compared by drawing their scatter plot. With help of statistical indicators, one demonstrates objectively the quality of the predictions. This method is quite exact – no suppositions have to be fulfilled. The advantage of this cross-validation technique lays in its nature: it enables to evaluate the quality of the predicted values at locations without measurements, as long as they are within the area covered by the measurements.

The main interpolation uncertainty indicators for comparison are the root mean squared error (RMSE) from cross-validation and the correlation following from the cross-validation between the predicted and the observed concentrations as presented in a scatter plot. In some cases, the mean prediction error (MPE), or the bias, is also used as additional indicator. RMSE and MPE are calculated according to equations

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

$$MPE = \frac{1}{n}\sum_{i=1}^{n}(Z(s_i) - \hat{Z}(s_i)) \tag{2.11}$$

where $Z(s_i)$ is the measured concentration at the $i^{th}$ point, $i = 1, \ldots, n$

$\hat{Z}(s_i)$ is the estimated concentration at the $i^{th}$ point using other information, without the measured concentration at that $i^{th}$ point

$n$ is the number of measuring points

A smaller RMSE generally means a better estimation, and MPE should be as close to zero as possible. In the cross-validation scatter plot, characterized by the linear regression equation $y=ax+c$ and by the coefficient of determination ($R^2$), $R^2$ should be as high as possible, the intercept $c$ should be as close to zero as possible and the slope $a$ should be as close to one as possible.

The uncertainty indicators RMSE and MPE from cross-validation are determined for the separate rural and urban maps. For the final European map, the most straightforward would be to calculate the overall uncertainty directly from all the stations together, disregard their types. However, that would lead to a rather unbalanced result due to the disproportionally large contribution of the relative high number of (sub)urban stations sited in a relative small urban area. The relative small total urban area contains a considerable higher number of urban and suburban background stations than the total number of rural background stations located in the significant larger total rural area. This has its nature in the primary interest for knowledge on negative human health impacts of air quality conditions. From this policy perspective monitoring priorities lay in urbanised areas with their higher population numbers potentially affected by air pollution. To obtain nevertheless an overall uncertainty estimate which is more representative to the entire final European map, one has to compensate for the unbalanced number of stations per type of area by introducing an alternative weighting options on basis of the fraction of area type or of the population representation per area type.

Thus, for expressing the interpolation uncertainty of the final European map, cross-validation indicators are determined separately for rural and urban/suburban stations to account for the total area of each type and determination of their subsequent contribution to the overall quality of the combined final map. The final cross-validation value of the final map is calculated from the two separate rural and urban values, using the following weighting equation

$$I_f = w_r I_r + w_u I_u \tag{2.12}$$

where $I_f$ is the final merged cross-validation indicator (RMSE or MPE) for the whole European area

$I_r$, $I_u$ are the cross-validation indicators for rural and urban areas, based on rural and urban/suburban background stations

$w_r$, $w_u$ are the weights, either based on size of area types or on percentage of population representation in each area type.

The criterion for the type of weighting can be two-fold. One can account for the fraction that the total rural area and the total urban area contribute to the total European mapping area. To facilitate this division between two area types, we consider as rural the area with a population density below $\alpha_1$ (Equation 2.3) and half of the area with population density between the values $\alpha_1$ and $\alpha_2$. Similarly, as urban is considered the area with population density above $\alpha_2$ and the other half of the area with population density between the values $\alpha_1$ and $\alpha_2$. In that case the weighting factor would become $w_r$ = 0.93 for the total rural area over $w_u$ = 0.07 for the total urban area. Alternatively, one can use as weighting factor the percentage of the total European population that lives in the rural areas and in the urban areas. In that case the weights would become $w_r$ = 0.41 for the rural area and $w_u$ = 0.59 for the urban area. These population weighting factors are based on the JRC population density data as described in Section 3.6.

When just spatial precision of the map is demanded, the area dependent weighting suffices. However, when the map should serve specifically health impact assessments, population density dependent weighting should be included as well.

# 3 Input data

## 3.1 Measured air quality data

The air quality data were extracted from the European monitoring database AirBase (Mol et al. 2009), supplemented by several rural EMEP stations which are not reported to AirBase. Only data from stations classified by AirBase and/or EMEP of the type *background* for the areas *rural*, *suburban* and *urban* are used. *Industrial* and *traffic* station types are not considered, since they represent local scale concentration levels not applicable at the mapping resolution employed. The following components and their indicators are considered:

$PM_{10}$ – annual average [$\mu g.m^{-3}$], years 2006 and 2007

– 36$^{th}$ maximum daily average value [$\mu g.m^{-3}$], years 2006 and 2007

Ozone – 26$^{th}$ highest daily maximum running 8-hour average value [$\mu g.m^{-3}$], years 2006 and 2007

– SOMO35 [$\mu g.m^{-3}$.day], years 2006 and 2007

– AOT40 for crops [$\mu g.m^{-3}$.hour], years 2006 and 2007

– AOT40 for forests [$\mu g.m^{-3}$.hour], years 2006 and 2007

For the indicators relevant to human health (i.e. both $PM_{10}$, and the ozone 26$^{th}$ highest daily maximum 8-hour average value and SOMO35) data from *rural*, *urban* and *suburban* background stations are considered. For the indicators relevant to vegetation (both AOT40 parameters for ozone) only *rural* background stations are considered, since no relevant vegetation is assumed to occur in (sub)urban areas.

In case of the annual indicators only stations with temporal data coverage of at least 75 percent are used. We excluded the stations from French overseas areas. Additionally, the ozone station IT1726A in Italy with highly questionable data has been excluded from the analysis of 2007.

## 3.2 Output from the Unified EMEP model

The established European chemical dispersion model used is the Unified EMEP model (revision rv2.7.10 for $PM_{10}$ for the year 2006, rv3.0.5 for ozone for 2006, and rv3.07 both for $PM_{10}$ and ozone for 2007), which is a Eulerian model with a resolution of 50x50 km. This model provides information at a 50x50 km scale. The disaggregation to the 10x10 km and 1x1 km grid cells is done as described in Section 4.4 of Horálek et al. (2007). Output from this model is used for the same parameter set as the set of measurement parameters in Section 3.1:

$PM_{10}$ – annual average [$\mu g.m^{-3}$], years 2006 and 2007

– 36$^{th}$ maximum daily average value [$\mu g.m^{-3}$], years 2006 and 2007

Ozone – 26$^{th}$ highest daily maximum 8-hour average value [$\mu g.m^{-3}$], years 2006 and 2007

– SOMO35 [$\mu g.m^{-3}$.day], years 2006 and 2007

– AOT40 for crops [$\mu g.m^{-3}$.hour], years 2006 and 2007

– AOT40 for forests [$\mu g.m^{-3}$.hour], years 2006 and 2007

The model is described by Simpson et al. (2003), Fagerli et al. (2004) and at the web site http://www.emep.int/OpenSource/index.html. The model results are based on the emissions for the relevant year (Vestreng et al., 2007; Mareckova et al. 2009) and actual meteorological data (from PARLAM-PS numerical weather prediction model for 2006, see Sandnes Lenschow and Tsyro (2002) and from HIRLAM numerical weather prediction model, version 7.1.3 for 2007). Details on the EMEP modelling are given for 2006 data for ozone by Tarrasón and Nyíri (2008) and for $PM_{10}$ by EMEP (2008) and for 2007 data by Benedictow et al. (2009).

## 3.3 Altitude

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

## 3.4 Meteorological parameters

Actual meteorological surface layer parameters are extracted from the Meteorological Archival and Retrieval System (MARS) of the ECMWF (European Centre for Medium-range Weather Forecasts) on a grid resolution of 0.25x0.25 arc-degrees (i.e. approx. 27x27 km). The derived parameters currently used and extracted from the ECMWF variables (details specified in Horálek et al. 2007, Section 4.5) are:

Wind speed – annual average [$m.s^{-1}$], years 2006 and 2007

Surface solar radiation – annual average [$MWs.m^{-2}$], years 2006 and 2007

The data were spatially disaggregated into the 10x10 km and 1x1 km EEA standard grid resolutions. For details, see Horálek et al. (2007).

## 3.5 Land cover

The input data from CORINE Land Cover 2000 (CLC2000) version 9/2007 (EEA, 2007) with a 100x100 m grid resolution is used (source and owner: EEA, lceugr100_00) for determining the most suitable land cover parameters for inclusion in the interpolation methodology comparison, which has been done on the aggregated 1x1 km resolution. The countries missing in this version are Iceland, Norway, Switzerland, and Turkey.

The ultimate exploration of the potential improvements obtained by introducing the earlier selected land cover parameters as most suitable additional supplementary data in the interpolation (Section 4.2) took place on a aggregated 10x10 km grid resolution, using a more recent 12/2009 version of CLC2000 (EEA, 2009), with Switzerland and Turkey missing only. The larger country coverage of this more extended version 12/2009, which was released during the execution of this study, plays an important role in getting the best estimate of its uncertainties. The more complete the total country coverage, the more precise we can determine the interpolation improvement potential of each land cover parameter and its involved uncertainty reductions.

In order to reduce the high number of degrees of freedom in the CORINE Land Cover description, the 44 CLC classes (http://etc-lusi.eionet.europa.eu/CLC2000/classes) were re-grouped in Table 3.1 into the 11 more general classes and as such in line with Janssen et al. (2008). The definition of the general land cover classes are taken from Janssen et al. (2008) just for conformity and to experience how results would work out on an European scale. However, in our case several general classes play only a minor role, as they represent air quality of more specific local or hotspot oriented areas instead providing an overall European picture. It concerns classes such as the general land cover class 3 (industrial) and classes 4, 5, and 6 (traffic) that are typically linked with industrial and traffic stations, which we do not consider in our European interpolation activities.

*Table 3.1 Definition of 11 general land cover classes, based on 44 Corine Land Cover (CLC) classes*

| General class | Description | CLC class |
|---|---|---|
| 1 | Coninuous urban fabric | 1 |
| 2 | Discontinuous urban fabric, green and sport | 2, 10, 11 |
| 3 | Industrial or commerial units | 3 |
| 4 | Road and rail networks and asociated land | 4 |
| 5 | Port areas | 5 |
| 6 | Airports | 6 |
| 7 | Mine, dump and construction sites | 7 - 9 |
| 8 | Arable land | 12 - 14 |
| 9 | Agricultural areas | 15 - 22 |
| 10 | Forests and semi-natural areas | 23 - 34 |
| 11 | Wetlands and water bodies | 35 - 44 |

For each general class we spatially aggregated the high land use resolution into the 1x1 km EEA standard grid resolution. The aggregated grid square value represents for each general class the total area of this class as percentage of the total 1x1 km square area.

A floating 3x3 km 'compiler' routine of ArcGis was applied for all the 1x1 km grid cells and for all general land cover classes in the cells. The resulting averaged 1x1 km grid square value for each general land use class represents the spatial average of the land use class at its surrounding 3x3 km area as is demonstrated by Figure 3.1. Section 3.6 motivates the application of such floating composite routine.

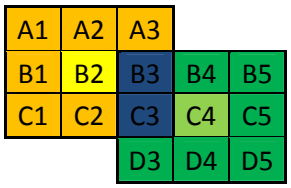


*Figure 3.1 Principle of the 3x3 floating grid cell routine. The 1x1 km grid cell B2 shows the percentage of chosen land cover class in its surrounding 3x3 km area (i.e. in the cells A1:A3, B1:B3, and C1:C3). Similarly, the 1x1 km grid cell C4 shows the percentage of chosen land cover class in its surrounding 3x3 km area (i.e. in the cells B3:B5, C3:C5, and D3:D5)*

For two reasons we examined the potential applying land cover classes in the spatial interpolations on the 1x1 km grid resolution: (i) to stay conform to the resolution used by Janssen et al, and (ii) to maintain the spatial relation between the land cover and air pollutant concentration as correlated as possible but without being confronted with computational limitations when applying calculations on a European mapping domain.

At the assessment of the potential contribution of land use data as supplementary data source in the interpolation, the most promising general classes are selected as significant contributors to the performance of the linear regression model (separately for rural and urban areas) at the resolution of 1x1 km. Subsequently, a spatial aggregation into the 10x10 km grid resolution has been done for these selected classes, to allow for verification of the performance at the specific case when applying land use data at this lower resolution, which is currently used in the routine mapping (see further Section 4.3.1).

## 3.6 Population density

Population density [inhbs.km$^{-2}$] is based on JRC data for the majority countries (Source EEA, popugrid01v4_1grid, official version Jan. 2008; Owner: JRC, 2008). For countries (Andorra, Albania, Bosnia-Herzegovina, Iceland, Liechtenstein, FYR of Macedonia, Montenegro, Norway, Serbia, Switzerland, and Turkey) and regions (Faroe Islands, Jersey, Guernsey, Man, Gibraltar, and northern

part of Cyprus) which are not included in this map we used population density data from an alternative source, namely the ORNL LandScan (2002) Global Population Dataset. An adjustment of the ORNL data into the level of JRC data was done according De Smet et al. (2010) to make the ORNL data better compatible with those of JRC.

The original 100x100 m grid resolution of the population density data is spatially aggregated into standard EEA grid resolutions of 1x1 km and the 10x10 km grids. These data are applied for the examination of interpolation uncertainties at the different grid resolutions in Chapter 6.

Next to this, a floating 3x3 km merging is executed the same way as at land cover, see Figure 3.1. The 3x3 floating neighbouring grid cell compiler routine is chosen as some kind of intermediate resolution merger positioned between the 1x1 km and 10x10 km grid resolution mergers. This specific technique is chosen simply because ArcGIS provides it as a standard routine, allowing for quick and easy compilations. Applying this feature on the 1x1 km grid resolution actually means using a 3x3 km floating window consisting of 9 grid cells for merging the separate maps into a new 1x1 km gridded combined final map. In the resulting map is accounted for possible nearby spatial correlation effects of the fluctuations in population densities classification from grid cell to grid cell. It could be an approach compensating for or tackle the problem that the merger function (Equation 2.3) makes use of the same population density class intervals at both the 10x10 km and 1x1 km resolution. These same intervals lead to the problem that exist on the one hand at merging with a 10x10 km resolution, which tends to 'overlook' small cities and towns located in more extended rural areas, and the problem that exists on the other hand at merging with a 1x1 km resolution where we may experience inaccurate or failing distinction in rural and urban areas when highly populated grid cells are neighbouring low or non populated grid cells. The answer to the question of how small a city or town should be to still be resolved by the mapping resolution, and above all by the merging resolution, is not straightforward. The 3x3 km floating merger could provide some answers to this issue and may appear to be an attractive alternative resolution for merging from the separate 1x1 km maps.

# 4 Evaluation of methodological improvements

This chapter evaluates the three options regarded in Section 2.2 as offering potential of improving the current methodology. They involve log-normal transposition of concentrations and altitude transformation as additional data treatments and inclusion of land cover as additional supplementary data source. Section 2.2 provides the motivation for their expected potential and highlights their methodological approach. The $PM_{10}$ and ozone indicators will be evaluated independently from each other and for the years 2006 and 2007. We test their performances on the rural and urban areas separately by using as main criteria the RMSE from the cross-validation (Eq. 2.10). The smaller the RMSE the better is the interpolation performance. Subsequently, the cross-validation predicted values at the station points are compared with the measurement values in a scatter plot and the coefficient of determination ($R^2$), slope and intercept of its linear regression equation provides information on the quality of the fit between station measurements and the interpolation predictions.

The cross-validation RMSE or scatter-plot coefficients of determination ($R^2$) should deliver an improved result of at least 3% to be considered as significant[1] for implementing the alternative data treatment or data source into the default interpolation methodology.

## *4.1 Log-normal transposition and altitude transformation*

### 4.1.1 $PM_{10}$

$PM_{10}$ aggregated data are sometimes assumed to be log-normally distributed (Beelen, 2009). We examined the distribution of $PM_{10}$ indicators (i.e. annual average and maximum 36$^{th}$ daily mean) for 2006 and 2007, separately for rural and urban areas. We tested whether the relevant data can be supposed to be normal or log-normal distributed. The goodness of fit was tested using the Kolmogorov-Smirnov test (Rao, 1973) with its results presented in Table 4.1. The goodness of fit is represented by the *p-value* which should be as close as possible to 1, see Section 2.2.1. In our test we apply a percentage level of 95%, i.e. the hypothesis of a normal (or log-normal) distribution is rejected at a critical level of $p < 0.05$.

*Table 4.1 The results of the Kolmogorov-Smirnov test of the goodness of fit of the normal and the log-normal distribution of $PM_{10}$ indicators annual average and 36$^{th}$ maximum daily mean for 2006 and 2007, separately for data from rural and (sub)urban background stations. The p-value should be as high as possible (at least 0.05).*

| $PM_{10}$ | Annual average | | | | 36$^{th}$ maximum daily mean | | | |
|---|---|---|---|---|---|---|---|---|
| | Rural | | Urban | | Rural | | Urban | |
| | 2006 | 2007 | 2006 | 2007 | 2006 | 2007 | 2006 | 2007 |
| normal distribution | 0.525 | 0.129 | 0.000 | 0.000 | 0.023 | 0.088 | 0.000 | 0.000 |
| log-normal distribution | 0.217 | 0.240 | 0.002 | 0.002 | 0.238 | 0.311 | 0.001 | 4.8E-05 |

Table 4.1 shows that at the rural data the $PM_{10}$ annual average at both years and 36$^{th}$ maximum daily mean for 2007 can be fitted by both log-normal and normal distributions. The rural data for 36$^{th}$ maximum daily mean for 2006 can be fitted by log-normal distribution only. The 36$^{th}$ maximum daily mean for 2006 does not meet with the critical level for accepting the hypothesis of being normally distributed. In the most cases log-normal distribution fits better than the normal one. At the urban data, neither normal nor log-normal distribution fits the data at the percentage level of 95%; however, the log-normal distribution fits the data just better than the normal one. It seems reasonable to assume that

[1] Horálek et al. (2005 and 2006) presented for 5 years consistently that log-normal kriging/cokriging gives better results for about 2 – 4 % in comparison with ordinary kriging/cokriging. From that we stated that the log-normal variant gives better results. Therefore, we define here a similar level of significance.

the log-normal transformation could improve the interpolated mapping results; this hypothesis is tested further.

As illustration of the better fitting log-normal distribution Figure 4.1 presents the red histograms of the $PM_{10}$ indicators in red fitted by the log-normal distribution in blue for 2007.

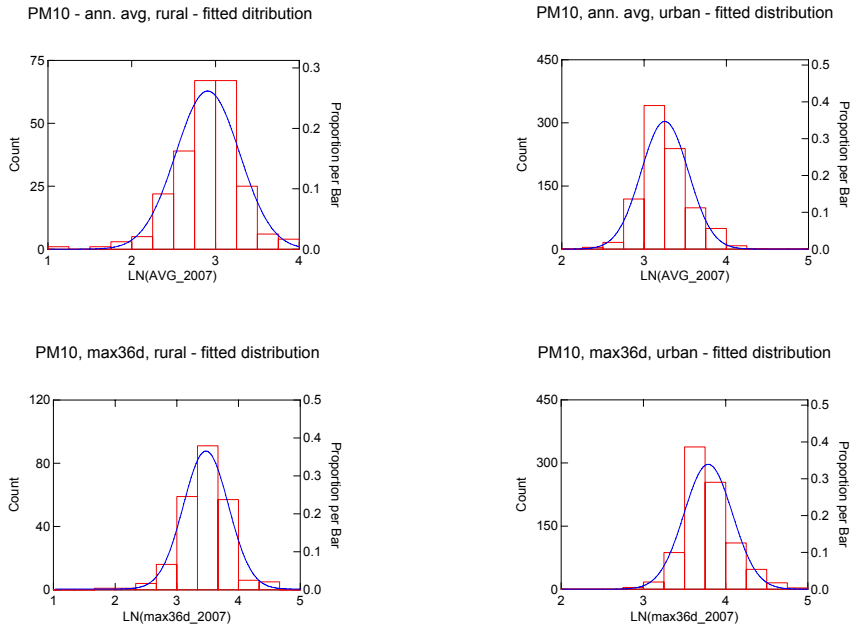


*Figure 4.1 Histograms (red) of $PM_{10}$ indicators annual average (top) and 36$^{th}$ maximum daily mean (bottom) for 2007, for rural (left) and (sub)urban (right) background stations, fitted by log-normal distribution (blue). X-axis is logarithmically transformed.*

At evaluation of the log-normal transposition of $PM_{10}$ concentration data in the mapping procedure, this transposition has been applied on both the station measurements and the EMEP model output and compared with the current methodology. Next to this, also altitude transformation according Equation 2.6 is compared. However, this transformation is executed at the rural areas only, since altitude has been selected as significant parameter in the linear regression model of the current methodology at the rural areas and not at the urban areas. For comparison of the interpolation improvements of the different interpolation variants Table 4.2 presents the RMSE resulting from the cross-validation for the $PM_{10}$ indicators annual average and the 36$^{th}$ maximum daily average.

By drawing scatter-plots of the cross-validation predicted values against the measurement values at the station points for $PM_{10}$ annual averages and 36$^{th}$ maximum daily averages we can conclude on the additional contribution of the additional data treatment in the interpolation to the fit with the observational values. Table 4.3 presents these results for 2006 and 2007 expressed by their linear regression equation $y=ax+c$ and coefficient of determination $R^2$. The better the interpolated predictions match with the measurements, the closer to 1 the $R^2$ and the slope $a$ will be, and the closer to zero its intercept $c$ will be.

*Table 4.2 The interpolation results of the current interpolation method, with additional log-normal transposition or altitude transformation, expressed as the RMSE (in µg.m$^{-3}$) for the $PM_{10}$ indicators annual average and 36$^{th}$ maximum daily average for 2006 and 2007 in rural and urban areas. The smaller the RMSE the better the interpolation performance is. Altitude is not applied at urban areas (gray shades)*

| $PM_{10}$ | RMSE (µg.m$^{-3}$) | | | | | |
|---|---|---|---|---|---|---|
| | Rural | | | Urban | | |
| | 2006 | 2007 | average | 2006 | 2007 | average |
| **Annual average** | | | | | | |
| lin.regr. + residual kriging (current method) | 5.79 | 4.60 | 5.19 | 6.09 | 4.96 | 5.53 |
| lin.regr. + residual kriging, with log-normal transposition | 5.65 | 4.59 | 5.12 | 6.06 | 4.96 | 5.51 |
| lin. regr.+ residual kriging, with altitude transformation | 5.77 | 4.52 | 5.14 | - | - | - |
| **36th maximum daily mean** | | | | | | |
| lin.regr. + residual kriging (current method) | 9.85 | 7.99 | 8.92 | 11.67 | 9.07 | 10.37 |
| lin.regr. + residual kriging, with log-normal transposition | 9.52 | 7.92 | 8.72 | 11.60 | 9.05 | 10.33 |
| lin. regr.+ residual kriging, with altitude transformation | 9.89 | 7.88 | 8.88 | - | - | - |

*Table 4.3 The linear regression equation and the coefficient of determination ($R^2$) of the scatter-plots do represent the correlation between the cross-validation predicted values and the measurements at station points for the $PM_{10}$ annual average and 36$^{th}$ maximum daily averages of 2006 and 2007 for the current interpolation method, and with additional log-normal transposition for rural and urban areas, or with altitude transformation for rural areas only.* $R^2$ *and the slope* a *(from the linear regression equation* $y = a \cdot x + c$*) should be as close to 1 as possible, the intercept* c *should be as close to 0 as possible.*

| $PM_{10}$ | 2006 | | | | 2007 | | | |
|---|---|---|---|---|---|---|---|---|
| | Rural | | Urban | | Rural | | Urban | |
| | lin.regr.eq. | $R^2$ | lin.regr.eq. | $R^2$ | lin.regr.eq. | $R^2$ | lin.regr.eq. | $R^2$ |
| **Annual average** | | | | | | | | |
| current interpolation method | y=0.570x+ 9.49 | 0.522 | y=0.705x+ 8.66 | 0.692 | y=0.619x+ 7.68 | 0.589 | y=0.673x+ 8.88 | 0.657 |
| interp. with log-normal transpos. | y=0.567x+ 9.68 | 0.543 | y=0.689x+ 9.04 | 0.695 | y=0.668x+ 6.74 | 0.598 | y=0.653x+ 9.38 | 0.657 |
| interp. with altitude transform. | y=0.565x+ 9.55 | 0.524 | - | - | y=0.631x+ 7.44 | 0.604 | - | - |
| **36$^{th}$ maximum daily mean** | | | | | | | | |
| current interpolation method | y=0.608x+15.00 | 0.556 | y=0.686x+15.67 | 0.681 | y=0.623x+13.11 | 0.598 | y=0.661x+15.80 | 0.648 |
| interp. with log-normal transpos. | y=0.576x+16.27 | 0.582 | y=0.674x+16.19 | 0.685 | y=0.654x+12.09 | 0.611 | y=0.635x+16.88 | 0.650 |
| interp. with altitude transform. | y=0.607x+14.97 | 0.553 | - | - | y=0.632x+12.79 | 0.609 | - | - |

Based on Tables 4.2 and 4.3 the currently used and alternative approaches are compared. The tables show that, compared to the current method, log-normal transposition provides slight improvements of about 0-3 % of reduced RMSE and 2-5 % of increased $R^2$ for the rural areas, and very slight improvement of about 0-0.5 % of reduced RMSE and of increased $R^2$ for the urban areas. The improvements obtained at rural areas enable us to consider implementing log-normal transposition in the routine mapping procedure. Such implementation is supported by the important argument that the Kolmogorov-Smirnov test confirms that log-normal distribution fits $PM_{10}$ data better than the normal distribution. For methodological uniformity and consistency reasons the transposition should be implemented then at both rural and urban areas, despite the very limited improvements obtained at urban areas.

The improved results of the maps constructed by using log-normal transposition agrees with findings of Horálek et al. (2005, 2007), where the log-normal variant of kriging and cokriging showed consistently better results for $PM_{10}$, over different years, at both annual averages and 36$^{th}$ highest daily means. (Note that in these reports interpolation results from insertion of the log-normal transposed values straight into the linear regression and residual kriging have not been investigated).

As Tables 4.2 and 4.3 demonstrate, the altitude transformation provides only very limited improvements of just about 0-2 % of reduced RMSE and increased $R^2$ and is considered as not significant at both years. The resulting improvements on the separate areas are that small that considering their common improvement is not relevant.

We come to the conclusion that the log-normal transposition should be implemented, and not the altitude transformation, into the default $PM_{10}$ interpolation methodology for the construction of the separate rural and urban maps, and the combined final European map.

As the log-normal transposition is recommended for future routinely use, we prepared the combined final maps for 2007 on both annual average and 36$^{th}$ highest daily mean and compared them with maps according the current methodology. Figures 4.2 and 4.3 show the map results. Note that merging of the rural and the urban maps is done using the current approach.

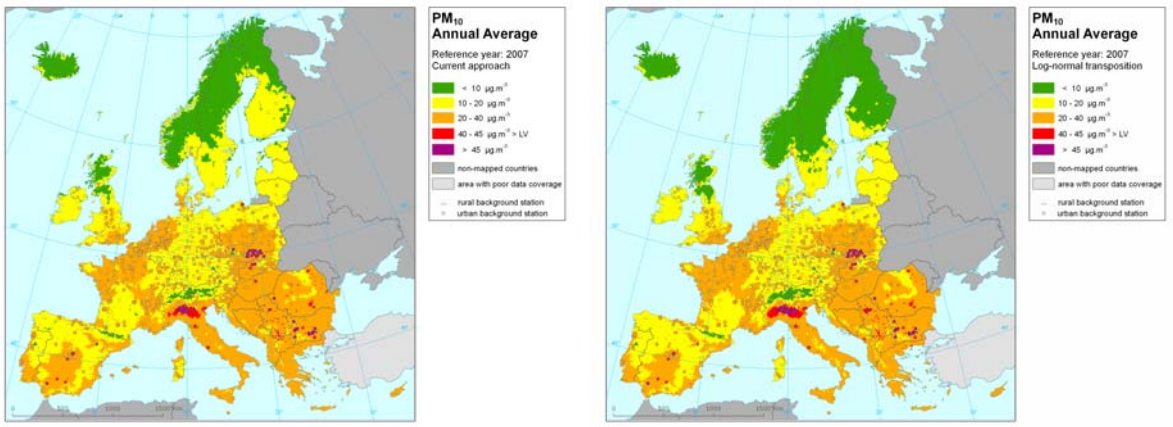


*Figure 4.2 Maps showing the annual average $PM_{10}$ values (in µg.m$^{-3}$) on the European scale in 2007, 10x10 km grid resolution, using current approach (left) and with log-normal transposition of concentrations (right).*

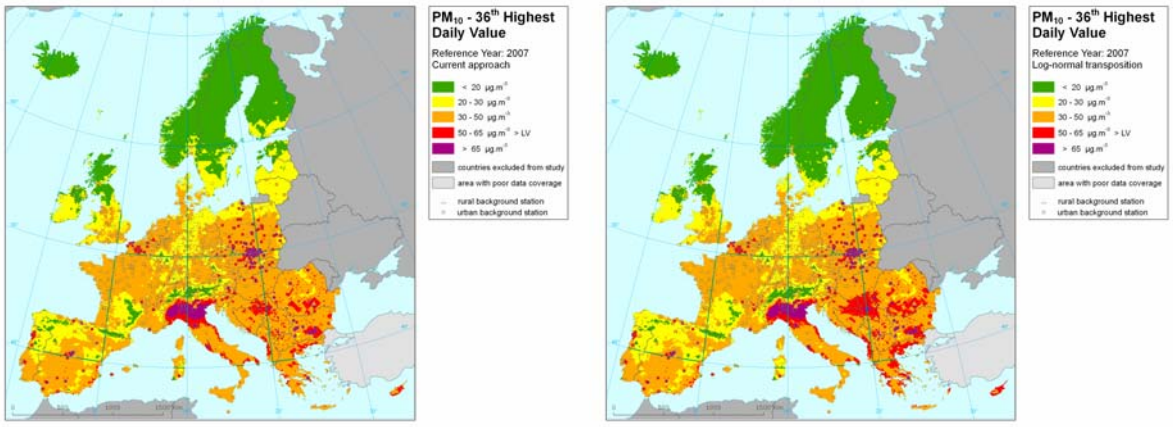


*Figure 4.3 Maps showing the 36$^{th}$ maximum daily average $PM_{10}$ values (in µg.m$^{-3}$) on the European scale in 2007, 10x10 km grid resolution, using current approach (left) and with log-normal transposition of concentrations (right).*

The maps illustrate that the main pattern remains unchanged when applying log-normal transposition, however, with a slight shifts towards lower classes or higher classes depending on the region. For example, in southern Sweden and Finland the yellow area shifts into a more extended green area and in the Balkan and Italian Po Valley there is an extension from the red areas.

## 4.1.2 Ozone

The same evaluation is done for ozone indicators, with again no altitude for urban areas as significant supplementary data source. Additionally the indicators AOT40 for crops and forests are evaluated. However, only rural areas are considered here, since crops and forests are assumed not to be present at the urban areas.

The results of the test of goodness of fit for normal and log-normal distributions are presented in Table 4.4.

*Table 4.4 The results of the Kolmogorov-Smirnov test of the goodness of fit of the normal and the log-normal distribution of $PM_{10}$ indicators annual average and 36$^{th}$ maximum daily mean for 2006 and 2007, separately for data from rural and (sub)urban background stations. The p-value should be as high as possible (at least 0.05).*

| Ozone | 26$^{th}$ highest max. daily 8h | | | | SOMO35 | | | | AOT40c | | AOT40f | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rural | | Urban | | Rural | | Urban | | Rural | | Rural | |
| | 2006 | 2007 | 2006 | 2007 | 2006 | 2007 | 2006 | 2007 | 2006 | 2007 | 2006 | 2007 |
| normal distribution | 0.083 | 0.211 | 0.001 | 0.037 | 0.010 | 0.0109 | 0.0001 | 1E-06 | 0.185 | 0.0075 | 0.009 | 0.0274 |
| log-normal distribution | 0.003 | 0.021 | 0.000 | 0.007 | 0.339 | 0.1532 | 0.000 | 1E-05 | 0.000 | 0.0002 | 0.0004 | 0.0001 |

Figure 4.4 illustrates the histograms fitted by the normal distribution for 2007, as the best fitting distribution for the different ozone indicators.

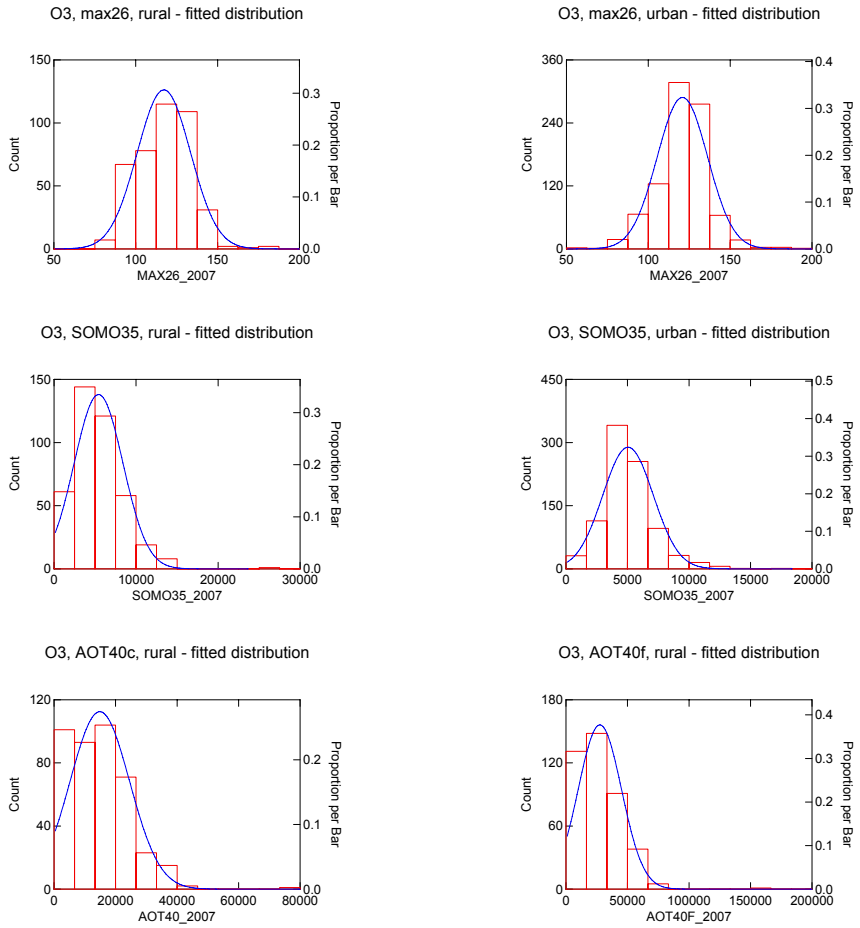


*Figure 4.4 Histograms (red) of ozone indicators 26$^{th}$ highest maximum daily running 8-hour mean value (top) and SOMO35 (middle) for rural (left) and (sub)urban (right) background stations, and of AOT40 for crops (bottom, left) and for forests (bottom, right) for rural background stations, fitted by normal distribution (blue), for the year 2007.*

Table 4.4 shows that ozone rural data for 26th highest maximum 8-hour daily mean at both years can be fitted by normal distribution. Contrary to that, rural data for SOMO35 can be fitted by log-normal distribution, at both years. AOT40 for crops can be fitted by normal distribution in 2006, but not satisfactory for 2007 data and neither for AOT40 for forests at both years (i.e. the p-value is below the 0.05 critical level); however, in all these cases the normal distribution gives better than the log-normal one. At the urban data, both normal and log-normal distributions give poor results at both health-related indicators and both years. In the case of 26th highest maximum 8-hour daily mean, the results for normal distribution are slightly better than the log-normal distribution.

Potential benefits of the log-normal transposition and altitude transformation for interpolated mapping are further examined in cross-validation analysis. The cross-validation RMSE results of all evaluated ozone indicators are presented in Table 4.6. Table 4.7 presents analysis of the scatter-plots with the cross-validation predictions against the station measurements.

*Table 4.5 Interpolation results of the current interpolation method, with additional log-normal transposition, or with altitude transformation, expressed as the RMSE (in µg.m$^{-3}$, µg.m$^{-3}$.d or µg.m$^{-3}$.h resp.) for ozone indicators 26th highest maximum 8-hour average, SOMO35 and AOT40 for crops and forests for 2006, 2007 and their average in rural and urban areas. The smaller the RMSE the better the interpolation performance is. At urban areas altitude is not applied as relevant data source and AOT40 indicators are not calculated (gray shades).*

| Ozone | RMSE | | | | | |
|---|---|---|---|---|---|---|
| | Rural | | | Urban | | |
| | 2006 | 2007 | average | 2006 | 2007 | average |
| **26th highest maximum daily 8-hour average** (µg.m$^{-3}$) | | | | | | |
| lin.regr. + residual kriging (current method) | 11.17 | 8.79 | 9.98 | 10.17 | 8.92 | 9.54 |
| lin.regr. + residual kriging, with log-normal transposition | 11.16 | 8.83 | 10.00 | 10.24 | 8.98 | 9.61 |
| lin. regr.+ residual kriging, with altitude transformation | 11.11 | 8.62 | 9.87 | - | - | - |
| **SOMO35** (µg.m$^{-3}$.d) | | | | | | |
| lin.regr. + residual kriging (current method) | 2075 | 1801 | 1939 | 1471 | 1260 | 1366 |
| lin.regr. + residual kriging, with log-normal transposition | 2123 | 1857 | 1982 | 1589 | 1316 | 1453 |
| lin. regr.+ residual kriging, with altitude transformation | 2099 | 1796 | 1948 | - | - | - |
| **AOT40 for crops** (µg.m$^{-3}$.h) | | | | | | |
| lin.regr. + residual kriging (current method) | 7531 | 5876 | 6704 | - | - | - |
| lin.regr. + residual kriging, with log-normal transposition | 7835 | 6465 | 7150 | - | - | - |
| lin. regr.+ residual kriging, with altitude transformation | 7485 | 5864 | 6675 | - | - | - |
| **AOT40 for forests** (µg.m$^{-3}$.h) | | | | | | |
| lin.regr. + residual kriging (current method) | 11850 | 10195 | 11023 | - | - | - |
| lin.regr. + residual kriging, with log-normal transposition | 12259 | 11067 | 11663 | - | - | - |
| lin. regr.+ residual kriging, with altitude transformation | 11840 | 10194 | 11017 | - | - | - |

*Table 4.6 The linear regression equation and the coefficient of determination ($R^2$) of the scatter-plots do represent the correlation between the cross-validation predicted values and the measurements at station points for the ozone indicators 26th highest daily maximum running 8-hour average, SOMO35, AOT40 for crops and forests (in µg.m$^{-3}$, µg.m$^{-3}$.d or µg.m$^{-3}$.h resp.) for 2006 and 2007 for the current interpolation method, and with additional log-normal transposition for rural and urban areas, or with altitude transformation for rural areas only.* $R^2$ *and the slope* a *(from the linear regression equation* $y = a \cdot x + c$*) should be as close to 1 as possible, the intercept* c *should be as close to 0 as possible.*

| Ozone | 2006 | | | | 2007 | | | |
|---|---|---|---|---|---|---|---|---|
| | Rural | | Urban | | Rural | | Urban | |
| | lin. regr. eq. | $R^2$ | lin. regr. eq. | $R^2$ | lin. regr. eq. | $R^2$ | lin. regr. eq. | $R^2$ |
| **26th highest max. daily 8hr avg.** | | | | | | | | |
| current interpolation method | y=0.539x+57.96 | 0.495 | y=0.602x+48.1 | 0.567 | y=0.732x+31.4 | 0.707 | y=0.685x+35.7 | 0.659 |
| interp. with add. LN transp. | y=0.536x+58.51 | 0.495 | y=0.601x+48.3 | 0.562 | y=0.726x+32.3 | 0.704 | y=0.680x+36.3 | 0.653 |
| interp. with altitude transf. | y=0.545x+57.23 | 0.501 | - | - | y=0.735x+31.1 | 0.718 | - | - |
| **SOMO35** | | | | | | | | |
| current interpolation method | y=0.495x+3325 | 0.468 | y=0.511x+2461 | 0.487 | y=0.642x+1921 | 0.635 | y=0.685x+1359 | 0.668 |
| interp. with add. LN transp. | y=0.493x+3359 | 0.455 | y=0.462x+2685 | 0.427 | y=0.662x+1871 | 0.621 | y=0.682x+1419 | 0.641 |
| interp. with altitude transf. | y=0.485x+3396 | 0.457 | - | - | y=0.635x+1955 | 0.637 | - | - |
| **AOT40 for crops** | | | | | | | | |
| current interpolation method | y=0.575x+11184 | 0.543 | - | - | y=0.677x+4777 | 0.634 | - | - |
| interp. with add. LN transp. | y=0.589x+10810 | 0.514 | - | - | y=0.700x+4827 | 0.584 | - | - |
| interp. with altitude transf. | y=0.579x+11127 | 0.548 | - | - | y=0.674x+4842 | 0.634 | - | - |
| **AOT40 for forests** | | | | | | | | |
| current interpolation method | y=0.525x+17055 | 0.505 | - | - | y=0.684x+8647 | 0.664 | - | - |
| interp. with add. LN transp. | y=0.545x+16200 | 0.478 | - | - | y=0.697x+8632 | 0.615 | - | - |
| interp. with altitude transf. | y=0.529x+16999 | 0.504 | - | - | y=0.676x+8898 | 0.664 | - | - |

The application of an additional log-normal transposition results in about all cases at both years and at both area types in a limited to moderately worse RMSE for the human health ozone indicators 26th highest maximum 8-hour means (ca. 0.1 – 0.7 %) and SOMO35 (ca. 2 – 6 %). Furthermore, the coefficients of determination ($R^2$) between the cross-validated predictions and the stations measurements show small decreases (ca. 0.5 – 12 %). We conclude that it would lead to even slightly worse interpolated human health ozone indicator maps when using a log-normal transposition of the concentrations. At the AOT40 for crops and forests the RMSE decrease by about 5.5 - 8 % and 5 % respectively, which we consider as a significant decline in the interpolation results compared to those of the current methodology. Therefore we do not even further consider the cross-validation and directly conclude that the log-normal transposition of concentrations is not suitable for the AOT40 vegetation indicators as well.

The altitude transformation, applied at rural areas only, leads at both years to either just very slightly improved or worse RMSE values that are considered as not significant (all between about 0.01 – 2%) compared to the current interpolation method. The cross-validation between the predicted and measurement values show at all four indicators that their coefficient of determination either slightly increases or declines (all between 0 – 2.5%) and do not indicate an added value to the current methodology. Concluding from this, we should not implement any additional altitude data treatment on top of current default methodological approach. The additional efforts do not provide any significant improvement in the interpolated mapping results.

## *4.2 Land cover inclusion*

The inclusion of land cover is considered two-fold. First, by inserting all eleven general land cover classes of Section 3.5 into Equation 2.1 and selecting the best and significant performing general land cover classes through a stepwise backward multiple linear regression (MLR) analysis In the stepwise backward MLR analysis all the variables with $p < 0.05$ are selected. And secondly, the joined land cover of Equation 2.7 is inserted into this Equation 2.1. Then, the best (significant) performing variants, i.e. either general land cover classes or the joined land cover, are selected. This is done for each area type (rural and urban) separately. Finally, for each area type the most promising land cover classes are used as additional supplementary data source in the interpolation method. Its results are compared with those of the current interpolation method where no land cover is being used.

The potential improvement of inclusion of land cover is examined on two resolution levels: one is the 1x1 km grid resolution to be consistent with the approach of Janssen et al. (2008). The other is the 10x10 km grid resolution, which is up to now used as the default resolution for the interpolated European mapping.

To be able to cope better with the quite computational demand of the calculations on the high grid resolutions, a two-step approach is chosen for the analysis. First, we determine on the 1x1 km resolution and for 2006 only which land cover classes are significant for each area type. Then we extend for those classes the analysis to the 10x10 km grid resolution for both years 2006 and 2007.

At the 1x1 km grid resolution the land cover version 9/2007 (EEA, 2007) is used for selecting the best land cover classes, and for 10x10 km grid resolution analysis we used the newer version 12/2009 (EEA, 2009b) which provides in a larger country coverage leading to a more complete and consistent spatial interpolation result and uncertainty estimate representative throughout the mapping domain.

### 4.2.1 $PM_{10}$

At both the $PM_{10}$ annual average and the 36$^{th}$ maximum daily averages the stepwise backward linear regression analysis on 1x1 km resolution selected as best the general land cover class number 10 "Forests and semi natural areas" (Table 3.1) for addition to the linear regression model (Equation 2.1) for the rural areas. Similarly, the joined land cover function (Equation 2.7) is selected as best for the urban areas. These selected variants were used in the regressions and their residual kriging on the 1x1 km grid resolution and their results are given in Table 4.7 in comparison with the current interpolation method on 1x1 km.

*Table 4.7 The interpolation results of the current interpolation method and with additional land cover inclusion, expressed as the RMSE (in µg.m$^{-3}$) for the $PM_{10}$ indicators annual average and 36$^{th}$ maximum daily mean for 2006 in rural and urban areas at the 1x1 km grid resolution.*

| $PM_{10}$ - 1x1 km resolution, 2006 | RMSE (µg.m$^{-3}$) | | | |
|---|---|---|---|---|
| | Annual average | | 36$^{th}$ max. daily average | |
| | Rural | Urban | Rural | Urban |
| lin.regr. + residual kriging (current method) | 5.62 | 6.15 | 6.10 | 11.74 |
| lin.regr. + residual kriging, with land cover | 5.50 | 6.79 | 6.06 | 12.96 |

Land cover seems to improve slightly the quality of the $PM_{10}$ indicator maps for the rural areas (RMSE reduces with about 0.7 – 2.5%) at the level of 1x1 km resolution for 2006. At the urban areas the RMSE increases significantly with about 11 to 20%, meaning a poorer quality of the urban $PM_{10}$ indicator maps when using land cover as additional data source. Therefore, we will further investigate possible improvements when of using additionally the land cover at the rural areas on the 10x10 km resolution for both 2006 and 2007 but not for the urban areas, where such addition is not recommended.

For the verification of the interpolation quality by additionally using land cover data on the 10x10 km resolution at rural areas we applied the more actual version 12/2009 of CLC2000, where only Switzerland and Turkey are missing (see Section3.5). Table 4.8 gives the results of this comparison with the currently used interpolation method and Figure 4.5 shows for the 10x10 km resolution the scatter-plots with the predicted and measured values for both $PM_{10}$ indicators at the station points in the rural areas.

*Table 4.8 The interpolation results of the current interpolation method and with additional land cover inclusion expressed as the RMSE (in µg.m$^{-3}$) for the $PM_{10}$ indicators annual average and 36$^{th}$ maximum daily mean for 2006 and 2007, and their averages, in rural areas at the 10x10 km resolution.*

| $PM_{10}$ - 10x10 km resolution, Rural | RMSE (µg.m$^{-3}$) | | | | | |
|---|---|---|---|---|---|---|
| | Annual average | | | 36$^{th}$ max. daily average | | |
| | 2006 | 2007 | average | 2006 | 2007 | average |
| lin.regr. + residual kriging (current method) | 5.81 | 4.59 | 5.20 | 9.84 | 7.92 | 8.88 |
| lin.regr. + residual kriging, with land cover | 5.64 | 4.57 | 5.11 | 9.59 | 7.85 | 8.72 |

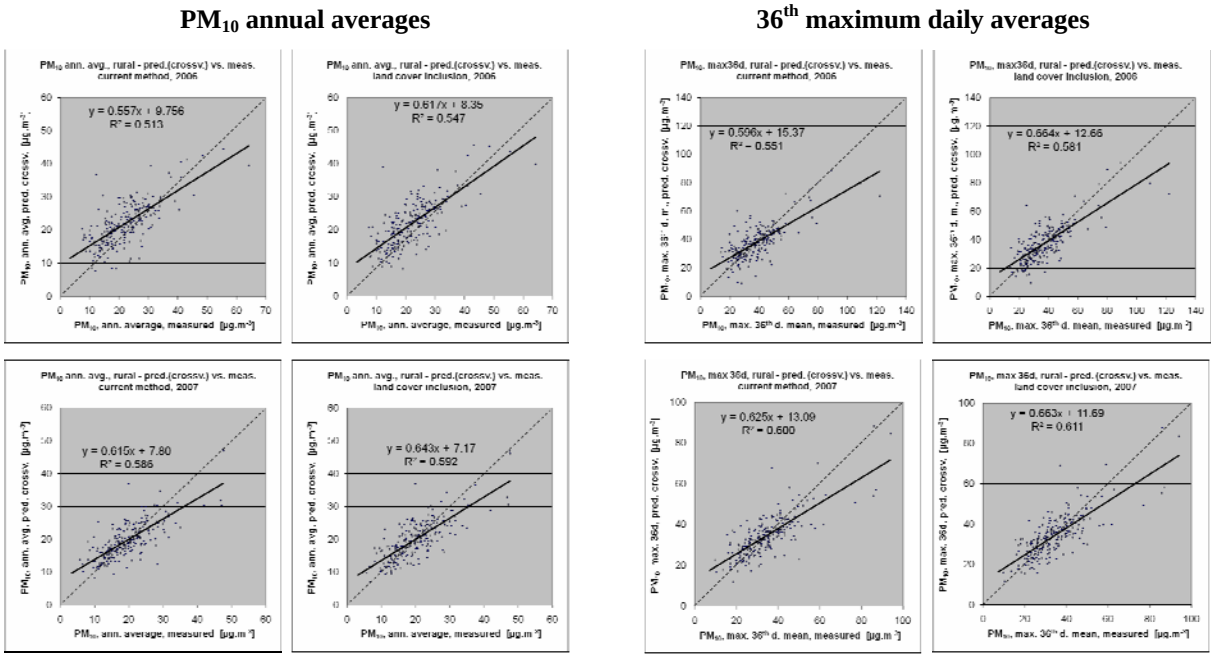


*Figure 4.5 Scatter-plots showing the correlation between cross-validation predicted values (y-axis) and measurements (x-axis) at station points for the $PM_{10}$ annual average and 36$^{th}$ maximum daily averages of 2006 (top) and 2007 (bottom) for the current interpolation method (left) and for the interpolation with additional land cover data (right) at the rural areas. Swiss stations are excluded due to lack of land cover data for this country.*

The RMSE in Table 4.8 reduces on average by about 2% when at the rural areas the general land cover class number 10 "Forests and semi natural areas" is introduced additionally to the interpolation method. The coefficients of the scatter-plots in Figure 4.5 improve with about 5 - 7% in 2006 and 1 - 2% in 2007 meaning that the interpolation provides better predictions when using additional land cover data. The improvement of $R^2$ is of such significance that it is worth considering the inclusion of land cover data into the default methodology. However, as the coverage of land cover data is not complete (CH and TR are missing), the implementation of this data into the routine procedure would complicate the mapping of areas lacking land cover data. Particarly when considering the additional efforts involved in including the land cover data into the interpolation methodology, this leads to the conclusion that land cover is not worth for inclusion in the $PM_{10}$ indicator mapping on the 10x10 km grid resolution as long as the mapping domain is not covered completely by land cover data. When in the future the land cover data will be completed for the whole of Europe, inclusion in the routine $PM_{10}$ mapping methodology should be reconsidered.

## 4.2.2 Ozone

At ozone 26$^{th}$ highest maximum daily running 8-hour average the stepwise backward linear regression analysis for ozone on 1x1 km resolution selected two classes as the best land cover data for addition to the linear regression model (Equation 2.1) for the rural areas: the general land cover classed number 6 "Airports" and number 10 "Forests and semi natural areas" (Table 3.1). For SOMO35 at the rural areas

three classes are selected: general land cover class numbers 8 "Arable land", 10 "Forests and semi natural areas" and 11 "Wetlands and water bodies". At both ozone indicators for the urban areas the general land cover class number 1 "Continuous urban fabric" is selected as most promising. For AOT40 for crops the joined land cover function (Equation 2.7) and for AOT40 for forests the general land cover class number 8 "Arable land" is selected. The classes are used in the regressions and their residual kriging on the 1x1 km grid resolution and their results are given in Table 4.9 for comparison with the current interpolation method on 1x1 km.

*Table 4.9 The interpolation results of the current interpolation method and with additional land cover inclusion, expressed as the RMSE (in µg.m$^{-3}$, resp. in µg.m$^{-3}$.d or in µg.m$^{-3}$.h) for the ozone indicators 26$^{th}$ highest maximum daily 8-hour mean and SOMO35 in rural and urban areas and AOT40 for crops and AOT40 for forests in rural areas for 2006 at the 1x1 km resolution.*

| Ozone - 1x1 km resolution, 2006 | RMSE | | | | | |
|---|---|---|---|---|---|---|
| | 26$^{th}$ hmd8-h avg (µg.m$^{-3}$) | | SOMO35 (µg.m$^{-3}$.d) | | AOT40c (µg.m$^{-3}$.h) | AOT40f (µg.m$^{-3}$.h) |
| | Rural | Urban | Rural | Urban | Rural | Rural |
| lin.regr. + residual kriging (current method) | 11.40 | 10.25 | 1969 | 1473 | 7499 | 11730 |
| lin.regr. + residual kriging, with land cover | 11.37 | 10.01 | 1964 | 1440 | 7504 | 11700 |

Inclusion of land cover appears to improve the quality of the interpolated ozone indicator maps at both area types on the 1x1 km grid resolution, with the exception of AOT40 for crops. However, the improvements are at both area types considered not to be significant enough to include them in the default interpolation methodology: rural areas show less than 0.5% and urban areas about 2.5% RMSE reduction.

We verified for the urban areas what the potential improvement of the interpolation quality is at the inclusion of land cover classes "Continuous urban fabric" on the 10x10 km grid resolution for 2006 and 2007 using the CLC2000 version 12/2009. Table 4.10 presents its results compared to the current interpolation method.

*Table 4.10 The interpolation results of the current interpolation method and with additional land cover inclusion expressed as the RMSE (in µg.m$^{-3}$, resp. in µg.m$^{-3}$.d) for the ozone indicators 26$^{th}$ highest maximum daily 8-hour mean and SOMO35 for 2006 and 2007, and their averages, in urban areas at the 10x10 km resolution.*

| Ozone - 10x10 km resolution, Urban | RMSE | | | | | |
|---|---|---|---|---|---|---|
| | 26th h.m.d.8-h avg (µg.m$^{-3}$) | | | SOMO35 (µg.m$^{-3}$.d) | | |
| | 2006 | 2007 | average | 2006 | 2007 | average |
| lin.regr. + residual kriging (current method) | 10.13 | 8.91 | 9.52 | 1467 | 1264 | 1366 |
| lin.regr. + residual kriging, with land cover | 10.12 | 8.85 | 9.49 | 1462 | 1260 | 1361 |

**26th highest max. daily running 8-hour mean** **SOMO35**

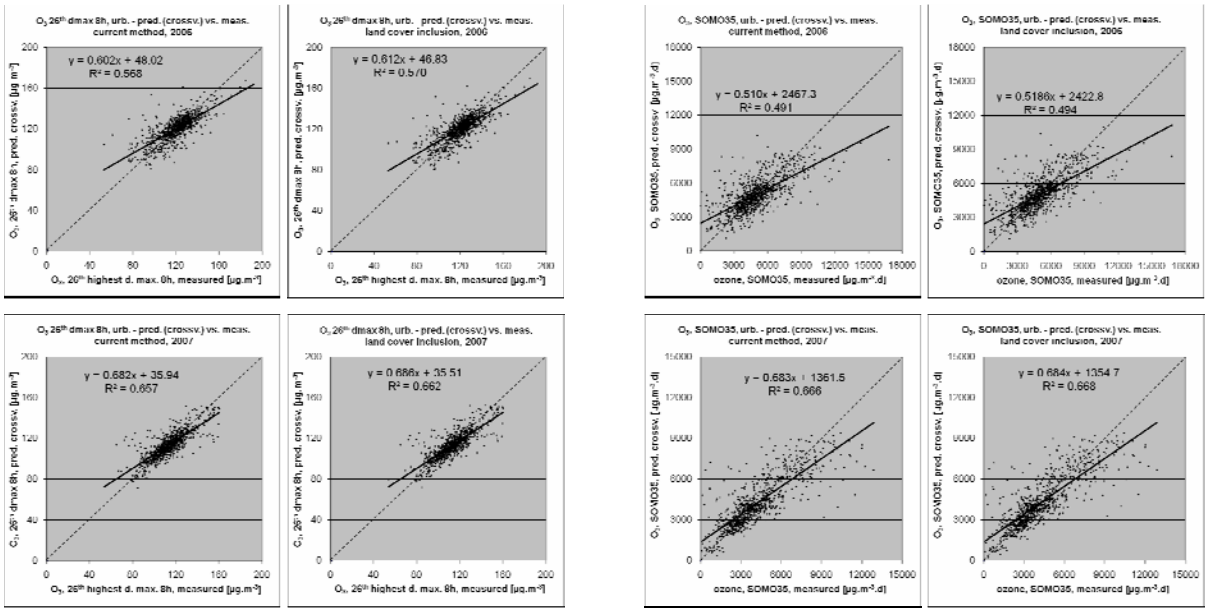


*Figure 4.6 Scatter-plots showing the correlation between cross-validation predicted values (y-axis) and measurements (x-axis) at station points for the ozone indicators 26th highest maximum daily 8-hour mean and SOMO35 of 2006 (top) and 2007 (bottom) for the current interpolation method (left) and for the interpolation with additional land cover data (right) at the urban areas.*

Table 4.10 shows that the reduced RMSE at the inclusion of land cover in the interpolations are less than 0.5% and hence not significant for urban areas at the 10x10 km resolution as well. Furthermore, Figure 4.6 with the cross-validation scatter-plots of the predicted against the measured values at the station points in the urban areas show that at both ozone indicators the improvement with land cover inclusion is very little, with an increased $R^2$ of less than 1%.

This leads to an overall conclusion that for the ozone indicators the land cover inclusion does not contribute significantly to an improved quality of the spatial interpolation at both the 1x1 km and the 10x10 km resolution. Therefore, it is decided not to implement the land cover data into the default spatial interpolation methodology of ozone.

## *4.3 Summary of conclusions and recommendations*

***Log-normal transposition and altitude transformation***

**$PM_{10}$**

*Log-normal transposition*

The log-normal transposition of concentrations provides slight and consistent improvements of cross-validation results in comparison with the current method. Next to this, the log-normal distribution fits $PM_{10}$ data better than the normal distribution does. Therefore, the log-normal transposition on both measurement concentrations and EMEP dispersion modelled concentrations will be implemented into the default interpolation methodology for the construction of the $PM_{10}$ rural, urban and combined final European map (Table 4.11).

*Altitude transposition*

For the altitude transformation at both $PM_{10}$ annual average and 36th maximum daily average the resulting improvements on the rural area types are that small, that considering their implementation into the combined final European map is not of any relevance. (Altitude has been selected as significant supplementary data source at rural areas only. Therefore no altitude transformation is applied at urban areas). Thereof, we conclude that the altitude transformation will not be implemented

as a part of the default $PM_{10}$ spatial interpolation methodology for the construction of the separate rural and urban concentration maps or the combined final European concentration map (Table 4.11).

**Ozone**

*Log-normal transposition*

Introducing a log-normal transposition on the ozone indicators would lead to even slightly worse spatially interpolated human health indicator maps and should not be implemented. At both vegetation indicators AOT40 for crops and forests the decline in interpolation quality is significant compared to the quality of the current methodology. Next to this, with the exception of SOMO35, the log-normal distribution does not fit ozone data better than the normal distribution does. Therefore, we conclude that an additional log-normal transposition of ozone concentrations should not be implemented at any of the two human health and two vegetation indicators (Table 4.11).

*Altitude transposition*

The altitude transformation has been compared with the current methodology only at the rural areas, because only at these areas the altitude parameter was originally selected as significant supplementary data source for inclusion in the current interpolation methodology. From the statistical analysis the altitude transformation appeared not to improve the spatial interpolation quality significantly. We conclude that no implementation of additional altitude data treatment should be done for any of the ozone indicators (Table 4.11).

***Land cover inclusion***

**$PM_{10}$**

The inclusion of land cover at urban areas on 1x1 km resolution declines the interpolation quality. Therefore, no further analysis on the 10x10 km resolution is executed for urban areas, since also there a decline is to be expected. When applying the general class 10 "Forests and semi natural areas" i.e. CLC classes 23-34 at rural areas, observed improvements indicate that it is worth considering the inclusion of land cover data into the default methodology. However, as the coverage of land cover data is lacking some countries, the implementation of this data into the routine procedure would complicate the mapping of areas lacking the land cover data. Particularly when considered the additional efforts involved in including the land cover data into the interpolation methodology. This leads to the conclusion that land cover is not worth for inclusion in the $PM_{10}$ indicator mapping on the 10x10 km grid resolution as long as the mapping domain is not covered completely by land cover data. When the land cover data will cover the Entire mapping domain one should reconsider its inclusion in the routine $PM_{10}$ mapping of the rural areas (Table 4.11).

**Ozone**

Inclusion of land cover appears to improve the quality of the spatially interpolated ozone indicator maps at both area types on the 1x1 km grid resolution. However, the improvements are considered not to be significant enough to include them in the default interpolation methodology. The observed best improvements obtained at urban areas on the 1x1 km resolution have been evaluated on the 10x10 km resolution as well. At both resolutions the interpolation improvements were not considered to be significant. For the ozone indicators no implementation of land cover data into the default spatial interpolation methodology should take place (Table 4.11).

*Table 4.11 Recommended methodological improvements to be implemented as part of the default interpolation methodologies as listed in Table 7.1 of Horálek et al. (2008).*

| Options methodological improvement | PM10 | Ozone |
|---|---|---|
| Additional log-normal transposition | Implement | -- [(1)] |
| Altitude transformation | -- [(1)] | -- [(1)] |
| Land cover inclusion | (Reconsider in future) [(2)] | -- [(1)] |

[(1)] *No significant improvement found; do not implement.*

[(2)] *Land cover inclusion should be reconsidered when it covers the entire European mapping domain.*

# 5 Rural and urban merger

## *5.1 Analysis of the current merger method*

### 5.1.1 Consistency between station types and population density classes

We compared the type of the stations used for the map interpolations with the type of area as assigned to the corresponding grid cells in the population density field. The station types are used as reported for 2006 to *AirBase*. The area types in the gridded population density field are assigned by applying of defined population density classification criteria for the determination if a grid square has to be considered as rural or urban. The aim is to get an impression of how good the ultimate 'direct' correspondence is between the *AirBase* defined monitoring station type at a measurement point and the defined population density type at the corresponding grid cell at that point. From the comparison we can conclude to what extend the separate rural and urban maps, as created in the current methodology, do match with the actual grid cell types as assigned by our classification criteria used on the population density field. In the ideal case the rural stations should be all sited in population density grid cells classified as rural and all urban/suburban stations should be sited in urban grid cells.

The analysis verifies how well the station types as reported for 2006 attain to this ideal match with a corresponding population density grid cell type. Simultaneously, we verify whether the currently defined class intervals of the population density types, expressed as the parameters $\alpha_1$, $\alpha_2$ of Equation 2.3, are in accordance with reality, i.e. whether the urban and the rural areas are distinguished in accordance with the relevant stations types.

The comparison has been done on the currently used 10x10 km grid resolution, and additionally on the potentially available 1x1 km grid resolution, to demonstrate the possible influence of the grid resolution on the matching uncertainty. Table 5.1 and 5.2 present the comparison results for $PM_{10}$ and ozone.

*Table 5.1 Comparison of the $PM_{10}$ background monitoring station types as reported for 2006 to AirBase with the area types rural and urban as derived from applying our classification criteria on the population density field, for both the 10x10 km and the 1x1 km grid resolution. The percentages represent the number of stations per station type located in the population density type. Green indicates consistent matches between the station type and the population density area type, and red the mismatches. Not considered is the marking of the more complex 'mixed' area type.*

| population density (p.d.) | | 10x10 km grid | | | | | | 1x1 km grid | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | rural stations | | | (sub)urban stations | | | rural stations | | | (sub)urban stations | | |
| | | number | | % | number | | % | number | | % | number | | % |
| Class [inhbs/ km2] | Type | p.d. class | p.d. type | p.d. type | p.d. class | p.d. type | p.d. type | p.d. class | p.d. type | p.d. type | p.d. class | p.d. type | p.d. type |
| 0-50 | rural | 78 | 124 | 54.4% | 13 | 39 | 4.4% | 143 | 161 | 70.6% | 29 | 43 | 4.9% |
| 50-100 | | 46 | | | 26 | | | 18 | | | 14 | | |
| 100-200 | mixed | 54 | 88 | 38.6% | 64 | 278 | 31.5% | 18 | 41 | 18.0% | 10 | 45 | 5.1% |
| 200-400 | | 12 | | | 148 | | | 22 | | | 24 | | |
| 400-500 | | 22 | | | 66 | | | 2 | | | 11 | | |
| 500-750 | urban | 3 | 16 | 7.0% | 122 | 566 | 64.1% | 9 | 25 | 11.0% | 28 | 795 | 90.0% |
| 750-1000 | | 3 | | | 90 | | | 4 | | | 35 | | |
| >1000 | | 10 | | | 354 | | | 12 | | | 732 | | |
| total | all | 228 | | 100% | 883 | | 100% | 228 | | 100% | 883 | | 100% |

*Table 5.2 Comparison of the ozone background monitoring station types as reported for 2006 to AirBase with the area types rural and urban as derived from applying our classification criteria on the population density field, for both the 10x10 km and the 1x1 km grid resolution. The percentages are the number of stations per station type located in the population density type. Green indicates consistent matches between the station type and the population density area type, and red the mismatches. Not considered is the marking of the more complex defined 'mixed' area type.*

| population density (p.d.) | | 10x10 km grid | | | | | | 1x1 km grid | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | rural stations | | | (sub)urban stations | | | rural stations | | | (sub)urban stations | | |
| | | number | | % | number | | % | number | | % | number | | % |
| Class [inhbs/ km2] | Type | p.d. class | p.d. type | p.d. type | p.d. class | p.d. type | p.d. type | p.d. class | p.d. type | p.d. type | p.d. class | p.d. type | p.d. type |
| 0-50 | rural | 190 | 273 | 61.6% | 31 | 64 | 6.7% | 298 | 321 | 72.5% | 51 | 72 | 7.5% |
| 50-100 | | 83 | | | 33 | | | 23 | | | 21 | | |
| 100-200 | mixed | 71 | 137 | 30.9% | 79 | 297 | 30.9% | 27 | 63 | 14.2% | 19 | 66 | 6.9% |
| 200-400 | | 56 | | | 149 | | | 30 | | | 30 | | |
| 400-500 | | 10 | | | 69 | | | 6 | | | 17 | | |
| 500-750 | urban | 9 | 33 | 7.4% | 128 | 599 | 62.4% | 19 | 59 | 13.3% | 39 | 822 | 85.6% |
| 750-1000 | | 5 | | | 98 | | | 11 | | | 39 | | |
| >1000 | | 19 | | | 373 | | | 29 | | | 744 | | |
| total | all | 443 | | 100% | 960 | | 100% | 443 | | 100% | 960 | | 100% |

The red marked fields in Tables 5.1 and 5.2 show that a reasonable number of stations lie in a different population density type than would be expected on basis of their station type as reported by the countries. These stations are used currently for creating the separate maps of another type of area than that they are sited in, according the population density field classification. This 'inconsistent' station siting per population density type is at both grid resolutions for both the rural and the urban population density areas slightly more frequent for the ozone stations than for the $PM_{10}$ ones. Furthermore, at both pollutants it is (proportionally) more frequent at the rural population density type than it is at the urban type.

The defined class intervals of the population density types have been checked as well by comparing the distributions of the number of rural and urban/suburban stations over the individual population density classes, compared to the total number of stations per class. In case of $PM_{10}$ at the 10x10 km grid resolution, the class boundary $\alpha_1$ between the rural and mixed grid type set at a lower level of 50 inhbs.km$^{-2}$, instead of 100 inhbs.km$^{-2}$, seems to be better. In population density class 50-100 inhbs.km$^{-2}$ the distribution of the number of rural stations is relatively limited compared to the number of urban and suburban stations and the total number of stations for that class. This indicates a rather poor match between the station types and the population density class type: this class seems to represent rather a mixed area type than a pure rural. This suggested change of class boundary value is further tested in Section 5.3.1.

An increased grid resolution - smaller grid cells - of the population density map contributes in two different points of view to a better match between station type and grid cell type.

First, we consider the increase (i.e. improvement) in the number of stations of a specific station type located in its corresponding area type in the 1x1 km gridded population density map, compared to the 10x10 km gridded map. It concerns the number of stations and percentages in the given tables for each population density type. Conclusion is that considerably more rural stations were found to be located in corresponding rural grid cells and more urban and suburban stations in urban grid cells. Furthermore, one notices that about the same number of stations are still 'wrongly' located. However, as positive side effect there are fewer stations of both types located in the grid cells of the mixed type. They are clearly 'redistributed' from the rather coarse 10x10 km grid cells into their better corresponding 1x1 km grid cell area types. The urban population density shows a greater improvement in matching stations types then at the rural types.

A second point of view is to consider as improvement the increase in the number of stations per type in relation to all stations located in each corresponding area type in the 1x1 km gridded population density map, compared to its situation in the 10x10 km gridded map. Conclusion here is that in the total number of stations per population density type of the type rural and urban does increase considerably and that they reduce at the mixed type; a matter of 'redistribution' as described above. Additionally, the percentage – for reasons of difficult interpretation not presented in the tables - of 'wrongly' assigned station types compared to all stations per population density type changes in most cases hardly: the highest change of just less than 3 % is attributable to the rural grid cells for $PM_{10}$. Overall is this not or hardly significant. It demonstrates that the redistribution of the stations from the mixed population density type over the types rural and urban takes place according about the same ratio of 'correctly' and 'wrongly' located stations types per population density grid cell type.

The level of 'inconsistent' location match of the type of station with the population density grid cell type can have several causes:

- as indicated above, the grid resolution plays a role (a lower resolution leads in general to a poorer match)
- the quality of the population density 10x10 km and 1x1 km gridded map (derived from census data on administrative units converted to grids by incorporating land cover data, each containing its limitations and inaccuracies; filling national gaps in the JRC data with ORNL data on yet another grid resolution)
- the criteria followed to determine the boundaries of the population density class intervals (used to define the type of area of the grid cell)
- correctness of the station type as reported by the country (according the EoI classification 'background', 'traffic' or 'industrial')
- correctness of the station area as reported by the country (according the EoI classification 'urban', 'suburban' or 'rural', based on different criteria then used at the population density)
- correctness of the station's geographical coordinates reported by the country.

As mentioned above, it is possible or even likely that a number of stations in AirBase is wrongly classified according the EoI definitions. It is however, too simple to evaluate this on population density alone. The influence of nearby sources should be included as well. Nevertheless, it might be worth to evaluate the classification of the stations lying in areas with non-corresponding population density. Furthermore relating the EoI classification, it should be noted that the population density approach makes no distinction between background and hotspot (traffic, industrial). A number of urban stations might be classified wrongly as background and not a as hotspot (or vice versa). This is not examined here.

### 5.1.2 Cross-validation of the separate rural, urban and combined final maps

***$PM_{10}$***

We did a cross-validation analysis for the separate rural and urban maps and for the combined final map as a whole. The calculations have been carried out for the different combinations of station types (rural and urban/suburban) and population density area types as distinguished in Equation 2.3 separately, because of the computational purposes. Figure 5.1 presents the cross-validation scatter plots of the combined final map constructed according to the current methodology for the 2006 annual average and 36$^{th}$ maximum daily value of $PM_{10}$.

If we would express the overall absolute mean uncertainty of the final map as RMSE from all background stations disregarding their types, it would be 7.1 $\mu g.m^{-3}$ for annual average and 12.8 $\mu g.m^{-3}$ for 36$^{th}$ maximum daily value (the 'no weight' column in Table 5.3). Next to the absolute values, the uncertainty can be expressed also in a relative way as the absolute RMSE uncertainty being a percentage of the mean air pollution indicator value for all stations. This relative mean uncertainty –

still calculated on basis all stations disregard their types – would be 25.6 % for annual average and 27.2 % for 36th maximum daily value. However, as mentioned in Section 2.5, that would be a rather unbalanced result due to the disproportionally large contribution of the urban stations: the relative small area classified in total as urban contains a considerable higher number of urban and suburban background stations than the total number of rural background stations in the significant larger total rural area.

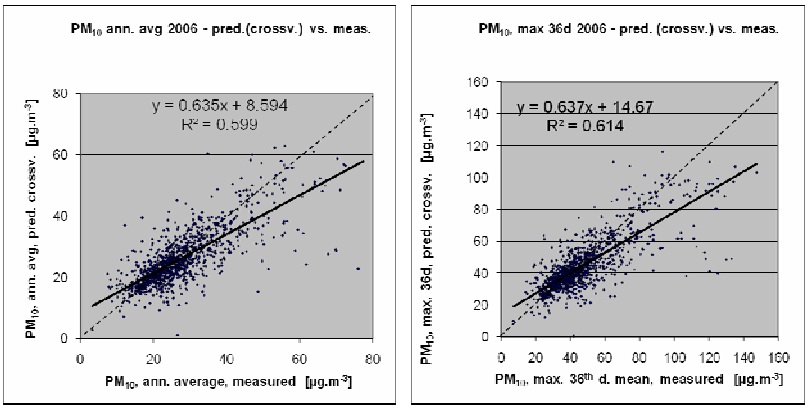


*Figure 5.1 Correlation between cross-validation predicted station values (y-axis) and station measurements (x-axis) for the $PM_{10}$ indicators annual average (left) and 36th maximum daily mean (right) in 2006, for the current methodology.*

To obtain an overall uncertainty estimate more representative to the entire final European map, we used the alternative weighting options according Equation 2.12 on basis of the fraction of area type and of population representation per area type. Table 5.3, column 'area' presents the RMSE of their absolute mean uncertainties. Based on the area fraction weighting, which appears to be more suitable for evaluating the map's quality than population weighting (column 'population'), the overall uncertainty of the final map is 5.6 µg.m$^{-3}$ for annual average and 9.5 µg.m$^{-3}$ for 36th maximum daily value. If we express this weighted RMSE as a percentage of the mean concentration value, the relative mean uncertainty is 25.2 % for annual average and 24.4 % for 36th maximum daily value.

*Table 5.3 Cross-validation parameter RMSE (in µg.m$^{-3}$), for the rural, the urban and the combined final map at the rural and the urban station points, and for the combined final map at all station points based on three different weighting criteria of the rural and urban contributions, for $PM_{10}$ indicators annual average and 36th maximum daily mean for 2006.*

| | annual average | | | | | 36th max. daily mean | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | rural stat. | urban stat. | all stat. - RMSE weighted by | | | rural stat. | urban stat. | all stat. - RMSE weighted by | | |
| | | | no weight | area | population | | | no weight | area | population |
| rural map | 5.79 | | | | | 10.12 | | | | |
| urban map | | 5.72 | | | | | 10.94 | | | |
| combined final map | 5.45 | 7.41 | 7.05 | 5.59 | 6.60 | 9.16 | 13.53 | 12.76 | 9.47 | 11.73 |

In addition, the absolute mean uncertainty has also been calculated separately for the rural background stations in relation to both the rural map according Equation 2.1 and the final map. The same has been done for the urban/suburban background stations in relation to both the urban map according Equation 2.1 and the final map. Table 5.3 shows the resulting RMSE values and Table 5.4 shows the resulting MPE values. Comparing the uncertainty RMSE values per station type related to each map type one can conclude that at both station types the final map shows higher uncertainties (higher values) than the separate urban map and lower values than the separate rural map.

*Table 5.4 Cross-validation parameter MPE (in µg.m$^{-3}$), for the rural, the urban and the combined final map at the rural and the urban station points, and for the combined final map at all station points based on three different weighting criteria of the rural and urban contributions, for $PM_{10}$ indicators annual average and 36$^{th}$ maximum daily mean for 2006.*

| | annual average | | | | | 36$^{th}$ max. daily mean | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | rural stat. | urban stat. | all stat. - MPE weighted by | | | rural stat. | urban stat. | all stat. - MPE weighted by | | |
| | | | no weight | area | population | | | no weight | area | population |
| rural map | 0.08 | | | | | 0.22 | | | | |
| urban map | | 0.11 | | | | | 0.14 | | | |
| combined final map | 0.99 | -2.06 | -1.44 | 0.78 | -0.81 | 1.56 | -3.37 | -2.36 | 1.21 | -1.35 |

As one can see in Table 5.4, the MPE is positive at the rural stations in case of the corresponding grid squares of the combined final map and close to zero in case of grid squares of the rural map. This means that the predicted (grid square) values in the combined final map are somewhat overestimated compared to the corresponding station point measurements. This occurs at some rural stations sited in moderately to highly populated areas.

On the other hand, the MPE is negative at the urban/suburban stations for the combined final map and close to zero in case of the rural map, meaning the predicted grid square values in the combined final map are on average lower than the measurement values at the corresponding stations. This occurs in lower or moderately populated areas with urban/suburban stations. Exactly in these areas the combined final map differs from the urban map (with its MPE close to zero) which is attributable to the influence of measurements originating from specifically small but highly polluted cities located in quite clean rural areas (ignoring the possibility of wrong station coordinates or population density grid values). Some underestimation in the combined final map could be expected, but that would be smaller than MPE currently shows. That has most likely its cause in the fact that we compare quantities representative for a different types of resolution: station observation *point* values against the corresponding 10x10 km spatial mean *grid square* values of the combined final map. Thus, a 'not-so-big' negative MPE value for the combined final grid map at the corresponding set of urban/suburban stations may be considered as a good result.

### *Ozone*

Cross-validation analysis is done by the same approach as in the case of $PM_{10}$.

The cross-validation scatter-plots of the complete final map constructed according to the current methodology, for ozone indicators 26$^{th}$ highest daily maximum 8-hour value and SOMO35 for 2006 are presented in Figure 5.2.

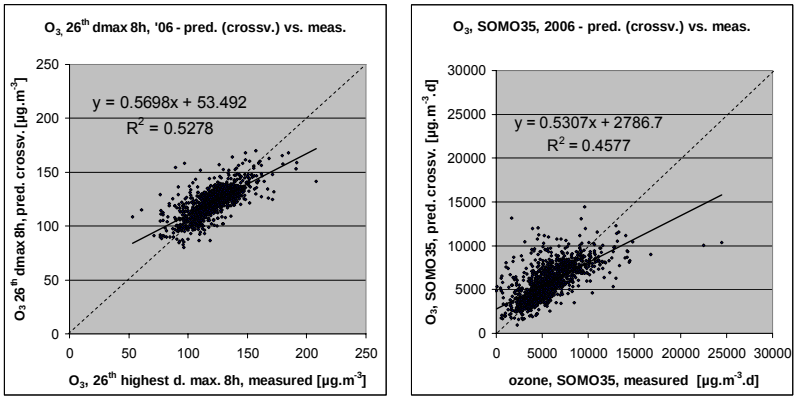


*Figure 5.2 Correlation between cross-validation predicted station values (y-axis) and station measurements (x-axis) for the ozone indicators 26$^{th}$ highest daily maximum 8-hour value (left) and SOMO35 (right) in 2006, for the current methodology.*

The absolute mean uncertainty of the final map expressed by RMSE from all stations disregarding their types is 10.8 µg.m$^{-3}$ for 26$^{th}$ highest daily maximum 8-hour value and 1840 µg.m$^{-3}$ for SOMO35

(the 'no weight' column in Table 5.5). Alternatively, this uncertainty can be expressed as the absolute RMSE uncertainty being a percentage of the mean air pollution indicator value for all stations. This relative mean uncertainty is 8.8% for 26$^{th}$ highest daily maximum 8-hour value and 33.1% for SOMO35.

If we express the uncertainty by the RMSE weighted by area according to Equation 2.12, the overall uncertainty of the combined final map is 11.0 µg.m$^{-3}$ for 26$^{th}$ highest daily maximum 8-hourly mean and 2037 µg.m$^{-3}$.d for SOMO35. Expressed as relative mean uncertainties this is 8.8% for the 26$^{th}$ highest daily maximum 8-hourly mean and 31.6% for SOMO35.

The absolute mean uncertainty, expressed as RMSE, has also been calculated again separately for the rural background stations in relation to both the rural map according Equation 2.1 and for the combined final map. The same has been done for the urban/suburban background stations in relation to both the urban map according Equation 2.1 and the combined final map. Next to this, the bias, expressed by MPE has been calculated. These RMSE and MPE values are given in Tables 5.5 and 5.6. Comparing the RMSE uncertainty values, one can conclude that the combined final map shows higher uncertainty values than the separate urban map and slightly lower values than the separate rural map.

Similar as to the case of $PM_{10}$, but in the opposite way due to the different ('opposite') behaviour of ozone concentration fields in rural and urban areas, the MPE is positive at the urban/suburban stations for the corresponding grid squares of the combined final map and close to zero in case of corresponding grid squares of the rural map. This means that the predicted grid square values of the combined final map show on average higher values than the corresponding urban/suburban point measurements at the stations. This occurs in lower or moderately populated areas in which urban/suburban stations are located. It is again attributable to the influence of measurements originating from small cities being located in predominantly rural areas. Some overestimation in the combined final map could be expected, but that would probably be lower than MPE currently shows. It has most likely the cause in the fact that we compare different quantities representing different resolution types: station *point* observations versus corresponding 10x10 km spatial mean *grid square* predictions of the combined final map. Thus, a 'not-so-big' positive MPE value for the grid squares of the combined final map at the corresponding set of urban/suburban stations may be considered as a good result. Nevertheless, this effect is smaller than in the case of $PM_{10}$.

*Table 5.5 Cross-validation parameter RMSE (in µg.m$^{-3}$), for the rural, the urban and the combined final map at the rural and the urban station points, and for the combined final map at all station points based on three different weighting criteria of the rural and urban contributions, for different types of the stations, for ozone indicators 26$^{th}$ highest daily maximum 8-hour value (left) and SOMO35 (right), for 2006.*

| | 26th high. daily max. 8hr | | | | | SOMO35 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | rural stat. | urban stat. | all stat. - RMSE weighted by | | | rural stat. | urban stat. | all stat. - RMSE weighted by | | |
| | | | no weight | area | population | | | no weight | area | population |
| rural map | 11.17 | | | | | 2077 | | | | |
| urban map | | 10.16 | | | | | 1471 | | | |
| combined final map | 11.02 | 10.74 | 10.84 | 11.00 | 10.86 | 2062 | 1717 | 1840 | 2037 | 1859 |

*Table 5.6 Cross-validation parameter MPE (in µg.m$^{-3}$), for the rural, the urban and the combined final map at the rural and the urban station points, and for the combined final map at all station points based on three different weighting criteria of the rural and urban contributions, for different types of the stations, for ozone indicators 26$^{th}$ highest daily maximum 8-hour value (left) and SOMO35 (right), for 2006.*

| | 26th high. daily max. 8hr | | | | | SOMO35 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | rural stat. | urban stat. | all stat. - MPE weighted by | | | rural stat. | urban stat. | all stat. - MPE weighted by | | |
| | | | no weight | area | population | | | no weight | area | population |
| rural map | 0.23 | | | | | 8 | | | | |
| urban map | | 0.02 | | | | | -3 | | | |
| combined final map | -0.19 | 1.40 | 0.86 | -0.08 | 0.74 | -149 | 342 | 176 | -115 | 140 |

### 5.1.3 Comparison of the point measured values with the predicted grid value

***$PM_{10}$***

A simple (not cross-validated) comparison between the station measurement values and interpolated 10x10 km grid values has been made. This comparison shows to what extend the predicted value of the corresponding grid cell represents the point measurement values covered by that cell. This comparison has been done on the one hand for all the stations and the combined final map, and on the other hand separately for the different combinations of station types and the grid types according the population density classification, in order to analyse possible differences in the results.

Figure 5.3 presents the regression results for the combined final map and shows that a better correlated relation (i.e. higher $R^2$, smaller intercept and the slope closer to 1) between station measurements and the interpolated values of the corresponding grid cells for both rural and urban areas exists, rather than it does at the point cross-validation predictions of Figure 5.1. This has its cause in the fact that the simple comparison between point measurements and gridded interpolated values shows the uncertainty at the actual station locations (points) itself, while the cross-validation simulates the behaviour of the interpolation at positions without actual measurements within the area covered by measurements.

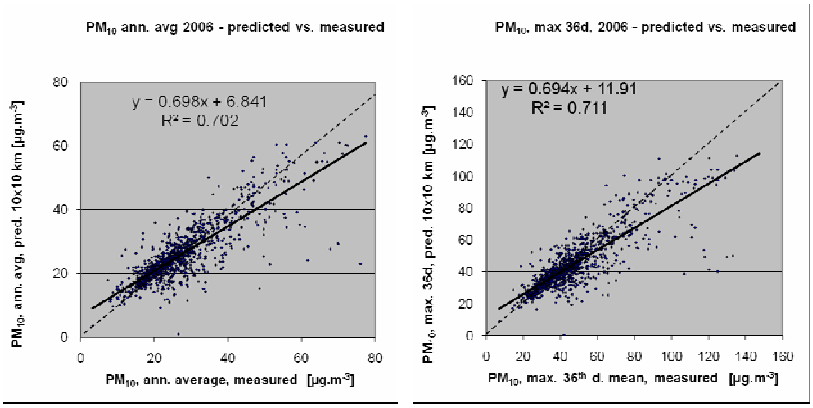


*Figure 5.3 Correlation between the predicted 10x10 km grid values (y-axis) and all the station measurements (x-axis) for the $PM_{10}$ indicators annual average (left) and 36$^{th}$ maximum daily mean (right) in 2006 for the current methodology.*

A number of outliers are visible in the right lower corner at both the annual average and 36$^{th}$ maximum daily mean. To examine this closer, linear regressions are applied on different combinations of the station types (rural and urban/suburban) and population density grid types (as distinguished in Equation 2.3).

In general, the $PM_{10}$-rural interpolated concentration field should be lower than the $PM_{10}$-urban concentration field at each grid cell. In case the rural concentration is higher than the urban, we included all (i.e. both rural and urban) stations in the interpolation leading to a 'joint urban/rural' interpolated field as an effort to compensate for this 'inconsistent' situation in the grid cell (see Section 2.1). The results are summarised in Table 5.7.

*Table 5.7 The matching number of stations, the linear regression equation and coefficient of determination $R^2$ from the scatter plots of the predicted values aggregated into the 10x10 km grid field versus the station measurement values for the different station type - area type combinations for $PM_{10}$ indicators annual average and 36th maximum daily mean in 2006 as a result of the current methodology, in relation to the different possible occurrences of combinations of station types and area types.*

| relation between urban and rural maps | station type | popul. density | | annual average | | | 36th max. daily mean | | |
|---|---|---|---|---|---|---|---|---|---|
| | | class values | class type | N of stat. | equation | $R^2$ | N of stat. | equation | $R^2$ |
| rural > urban | all | all | all | 212 | y = 0.931x + 2.705 | 0.74 | 232 | y = 0.819x + 8.093 | 0.85 |
| rural ≤ urban | rural | < 100 | rural | 121 | y = 0.797x + 4.303 | 0.86 | 121 | y = 0.869x + 4.858 | 0.92 |
| | rural | 100 - 500 | mixed | 77 | y = 0.724x + 7.836 | 0.83 | 74 | y = 0.791x + 10.75 | 0.87 |
| | rural | > 500 | urban | 14 | y = 0.794x + 9.262 | 0.66 | 12 | y = 0.818x + 14.99 | 0.77 |
| | urban/suburb. | < 100 | rural | 35 | y = 0.548x + 3.881 | 0.30 | 35 | y = 0.412x + 14.29 | 0.15 |
| | urban/suburb. | 100 - 500 | mixed | 223 | y = 0.430x + 10.72 | 0.47 | 220 | y = 0.404x + 20.44 | 0.4 |
| | urban/suburb. | > 500 | urban | 434 | y = 0.805x + 5.664 | 0.87 | 422 | y = 0.777x + 11.11 | 0.84 |
| all | all | all | all | 1116 | y = 0.698x + 6.841 | 0.70 | 1116 | y = 0.694x + 11.91 | 0.71 |

Table 5.7 shows clearly that the worst regression results (lowest $R^2$ and regression slope, highest intercept) are obtained at the (sub)urban stations in the areas represented by rural or mixed population density grid cells. This indicates a very poor relation between the measured and predicted grid values. Among just these stations, a group of 9 stations has been detected as 'outliers' under the regression analysis, i.e. stations with the worst agreement between the measured and predicted values. Five of these stations are Bulgarian, and Poland, Slovakia, Spain and Norway have each one of these stations sited. At all of these urban/suburban stations much higher values are measured than predicted in the underlying rural grid. Different reasons may play role here: small polluted cities in a relative clean rural area, wrongly assigned coordinates of the relevant stations causing allocation in the wrong area type, but also underestimation of the surrounding rural area which can be the case when there is a lack of rural measurement data representing the rural area.

Problematic is also the estimation of the urban/suburban stations in the areas with "mixed" grid cells. At annual average 212 stations and at 36th highest maximum daily mean 232 stations, which is around 20 % of the total number of stations, the population density grid cells show predictions with higher rural than urban values. As stated in Section 2.1, it concerns about 4.5 – 11.5 % of the total grid cells in the 2006 map. The major role in this phenomenon plays the non-applying of the correction factor at French stations in 2006: 163 (resp. 160) from the above mentioned stations are French and predominantly urban and suburban. Relating the other stations, evidently there is in some parts of Europe some clustering of stations in areas where the rural concentration exceeds the urban concentration. It could indicate that the station and/or population density classification has not been applied completely optimal or consistent, or that the chosen spatial distribution of the monitoring stations is not optimal compared to the population density grid cells types or vice versa. Also the spatial pattern of the areas types and the layout of monitoring networks covering these areas may play a role and might need improvement. A more detailed analysis to come to a plausible explanation of this phenomenon, including options to reduce inconsistencies we recommend as follow-up, since it was outside the scope of this paper.

Figure 5.4 shows the same regression results as Figure 5.3, but now cleared from the group of outliers (detected under regression analysis), which consists just of several urban/suburban stations in the rural or mixed population density areas: excluding these stations removes the outliers from the graphs, providing increased correlation ($R^2$) and regression slopes at both indicators.

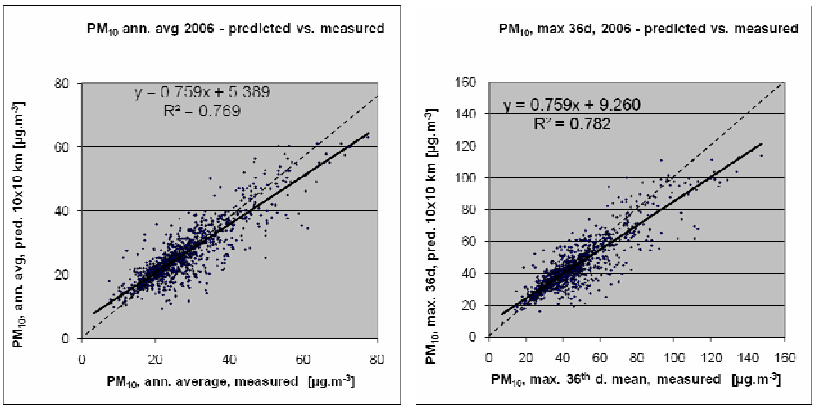


*Figure 5.4 Correlation between the predicted 10x10 km grid values (y-axis) and the station measurements (x-axis) for the $PM_{10}$ indicators annual average (left) and 36$^{th}$ maximum daily mean (right) in 2006 for the current methodology, without the group of 9 urban/suburban stations in the rural or mixed population density areas detected as outliers in the regression analysis, for at least one $PM_{10}$ indicator.*

The depicted discrepancy in such cases (urban stations in the rural areas) can be explained – at least in some cases – as following: imagine a small highly polluted town in a quite clean rural area. For the predicted value of the 10x10 grid cell the rural area will be predominant and under the current methodology the high measurement value at the urban station in the highly polluted town will have no influence on the rural predominance. The resulting mapped concentration of the corresponding grid cell will by itself reflect the rural value only.

One solution to reduce this discrepancy would be to use a finer population density grid resolution, either for the whole approach or at least for the rural/urban merging process. This is examined closer in Chapter 6.

***Ozone***

Like at $PM_{10}$, a simple comparison between the station measurements and interpolated values in a 10x10 km grid values has been made for ozone indicators. It shows to what extend the predicted value of the corresponding grid cell represents the point measurement values covered by that cell. This comparison has been done on the one hand for all the stations and the combined final map, and on the other hand separately for the different combinations of station types and the grid types according the population density classification, in order to analyse possible differences in the results.

Figure 5.6 presents the regression results of this comparison for the combined final map and demonstrates a better correlated relation between station measurements and the interpolated values of the corresponding grid cells for both rural and urban areas than it does at the point cross-validation predictions of Figure 5.2. This has its cause in the fact that the simple comparison between point measurements and gridded interpolated values shows the uncertainty at the actual station locations (points) itself, while the cross-validation simulates the behaviour of the interpolation at positions without actual measurements within the area covered by measurements.

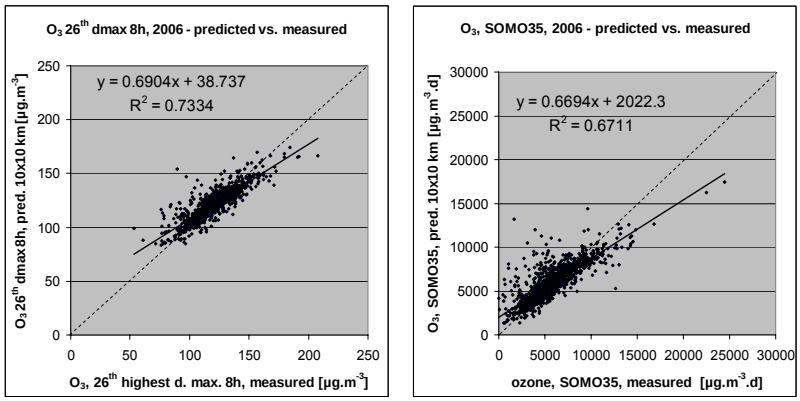


*Figure 5.6 Correlation between predicted values for 10x10 km grid (y-axis) and point measurements (x-axis) for the ozone indicators 26th highest daily maximum 8-hour value (left) and SOMO35 (right) in 2006, for presently used methodology.*

In Table 5.6 we indicate at each station type and its spatially corresponding population density grid cell types the current methodology provides the weakest correlations. The table shows clearly that the worst results (lowest $R^2$ and regression slope, highest intercept) are obtained - like at $PM_{10}$ - at the urban stations in the areas represented by the rural population density grid cells. This indicates a very poor relation between the measured (points) and predicted grid values. Again problematic is also the estimation of the urban/suburban stations in the areas with "mixed" grid cells, especially in the case of SOMO35.

The urban stations in the rural areas are visible in Figure 5.6, specifically at SOMO35, as the cloud above and left of the dashed line $y = x$, with low measurement values against clearly higher estimated grid values.

*Table 5.6 The matching number of stations, the linear regression equation and coefficient of determination $R^2$ from the scatter plots of the predicted values aggregated into 10x10 km grid field versus the station measurement values, for the different station type – area type combinations for ozone indicators 26th highest daily maximum 8-hour value (left) and SOMO35 (right) in 2006 as a result of the current methodology, in relation to the different possible occurrences of combinations of station types and area types.*

| relation between urban and rural maps | station type | population density | | 26th high. daily max. 8h | | | SOMO35 | | |
|---|---|---|---|---|---|---|---|---|---|
| | | class values | class type | N of stat. | equation | $R^2$ | N of stat. | equation | $R^2$ |
| rural > urban | all | all | all | 301 | y = 0.68x + 39.7 | 0.77 | 182 | y = 0.649x + 1835 | 0.76 |
| rural ≤ urban | rural | < 100 | rural | 237 | y = 0.79x + 26.2 | 0.87 | 251 | y = 0.704x + 1992 | 0.83 |
| | rural | 100 - 500 | mixed | 113 | y = 0.66x + 41.5 | 0.79 | 124 | y = 0.675x + 1772 | 0.77 |
| | rural | > 500 | urban | 22 | y = 0.80x + 18.5 | 0.70 | 26 | y = 0.509x + 1807 | 0.61 |
| | urban/suburb. | < 100 | rural | 41 | y = 0.29x + 91.0 | 0.13 | 42 | y = 0.254x + 6050 | 0.07 |
| | urban/suburb. | 100 - 500 | mixed | 225 | y = 0.58x + 54.9 | 0.61 | 255 | y = 0.510x + 3423 | 0.35 |
| | urban/suburb. | > 500 | urban | 408 | y = 0.77x + 28.7 | 0.79 | 466 | y = 0.700x + 1530 | 0.76 |
| all | all | all | all | 1347 | y = 0.69x + 38.7 | 0.73 | 1346 | y = 0.669x + 2022 | 0.67 |

Like at $PM_{10}$, under the current methodology such urban station measurement value has no influence on the rural predominance. The resulting mapped concentration of the corresponding grid cell will reflect the rural values only. The phenomena observed at the ozone indicators do correspond with those of the $PM_{10}$ indicators discussed earlier in this section and are not discussed here again.

## 5.2 Alternative merging methods

As is shown in Section 5.1, particular limitations exist in the current method of merging the rural and urban concentration fields into one combined final concentration map. An important issue is the relation between the station type classification and population density grid. Section 5.1 examined this relation and showed that in many cases there exists a discrepancy between the two attributes.

One can choose different alternative approaches to reduce this discrepancy. For example, the EU funded project APMoSPHERE (Briggs et al., 2005; Beelen et al., 2009) applied a reclassification of the type assigned to the air quality stations. Contrary to that project, we consciously do not make such changes. We insist to respect the station characteristics as formally reported under EU legislation by the countries to the European air quality database *AirBase*. Dealing with this limitation we examine in this section, next to the currently used merging approach (Equation 2.3), the three alternative 'merging' methods as described in the introductory chapter and in Section 2.3:

A. current method, but ignoring the use of a joint urban/rural map to compensate for assumed inconsistencies between rural and urban concentrations (Equation 2.8);

B. the simplest approach by not making a distinction between rural and urban areas and between rural and (sub)urban background stations (Equation 2.2);

C. merging, by applying a weighting fraction on the rural and urban areas directly in the linear regression model itself, using the same population density classification criteria as in current merging method and using the stations disregard their type in the regression model, followed by a spatial interpolation of the common regression residuals (Equation 2.9).

In this section current merging and its alternatives are mutually compared for $PM_{10}$ and ozone.

### 5.2.1 $PM_{10}$

Table 5.7 through 5.10 represent for the four (i.e. three alternative and the current) spatial interpolation methods the statistical cross-validation indicators RMSE and the mean prediction error (MPE) or bias. They express the uncertainty estimates for the whole interpolated mapping domain of the $PM_{10}$ annual average and the 36$^{th}$ maximum daily average respectively. The tables present the interpolation uncertainties estimates at the rural and urban station points, and at the combined final map. The uncertainty of the final map is calculated in three different ways. As Section 2.5 explains, most straightforward would be to calculate the overall uncertainty for the final European map, expressed as the RMSE (or MPE), directly from all the stations irrespective their type, i.e. no weighting. However, that would lead to rather unbalanced results due to the disproportionally large contribution of the urban stations: the relative small area classified in total as urban contains a considerable higher number of urban and suburban background stations than the total number of rural background stations in the significant larger total rural area. To reach an overall uncertainty estimate more representative to the entire final European map, one compensates for the unbalanced numbers of stations per type of area by introducing weighting options into the uncertainty calculation, on basis of either the fraction of area type or the population representation per area type (Equation 2.12).

*Table 5.7 Uncertainty estimates, expressed as RMSE (in µg.m$^{-3}$), of the interpolation results for the $PM_{10}$ indicator annual average for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of RMSE, using four different 'merging' methods. RMSE should be as small as possible.*

| | rural stations | urban stations | final map - RMSE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| current approach (with joint urban/rural map) | 5.45 | 7.40 | 7.05 | 5.59 | 6.60 |
| method A (current, without joint urban/rural map) | 5.50 | 7.37 | 7.03 | 5.63 | 6.60 |
| method B (all stations together) | 8.08 | 6.26 | 6.67 | 7.95 | 7.01 |
| method C (merging function, all stations) | 7.75 | 6.34 | 6.65 | 7.65 | 6.92 |

*Table 5.8 Uncertainty estimates, expressed as MPE (in µg.m$^{-3}$), of the interpolation results for the $PM_{10}$ indicator annual average for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of MPE, using four different 'merging' methods. MPE should be as close to zero as possible.*

| | rural stations | urban stations | final map - MPE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| current approach (with joint urban/rural map) | 0.99 | -2.06 | -1.44 | 0.78 | -0.81 |
| method A (current, without joint urban/rural map) | 0.99 | -2.12 | -1.14 | 0.77 | -0.84 |
| method B (all stations together) | 4.65 | -1.14 | 0.05 | 4.25 | 1.24 |
| method C (merging function, all stations) | 3.20 | -0.67 | 0.12 | 2.92 | 0.92 |

*Table 5.9 Uncertainty estimates, expressed as RMSE (in µg.m$^{-3}$), of the interpolation results for the $PM_{10}$ indicator 36$^{th}$ maximum daily average for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of RMSE, using four different 'merging' methods.*

| | rural stations | urban stations | final map - RMSE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| current approach (with joint urban/rural map) | 9.16 | 13.53 | 12.76 | 9.47 | 11.73 |
| method A (current, without joint urban/rural map) | 9.41 | 13.48 | 12.75 | 9.70 | 11.81 |
| method B (all stations together) | 13.62 | 11.70 | 12.11 | 13.48 | 12.49 |
| method C (merging function, all stations) | 13.06 | 11.48 | 11.82 | 12.95 | 12.13 |

*Table 5.10 Uncertainty estimates, expressed as MPE (in µg.m$^{-3}$), of the interpolation results for the $PM_{10}$ indicator 36$^{th}$ maximum daily average for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of MPE, using four different 'merging' methods.*

| | rural stations | urban stations | final map - MPE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| current approach (with joint urban/rural map) | 1.56 | -3.37 | -2.36 | 1.21 | -1.35 |
| method A (current, without joint urban/rural map) | 1.53 | -3.43 | -2.41 | 1.18 | -1.39 |
| method B (all stations together) | 7.38 | -1.77 | 0.10 | 6.74 | 1.99 |
| method C (merging function, all stations) | 4.98 | -1.20 | 0.06 | 4.55 | 1.34 |

Most striking observation from the tables is that in all cases (at both indicators, for both RMSE and MPE, and all four merging approaches) the uncertainty levels in the final map are mainly determined by the uncertainty levels of the urban stations, if no weighting of area or population contribution is applied, whereas these stations do represent a relative small area of the total mapping domain. This confirms our hypothesis that a weighting with a representative variable such as the fraction of area or population representation per area is needed to obtain a more balanced overall uncertainty statement for the final map, irrespective the type of merging method is chosen. Subsequently, one notices that the level of uncertainty in the combined final map is clearly under influence of the parameter type that is applied in the weighting and if the approach does include the station type as discriminator in the linear regressions model used at the interpolation method (current and alternative A). At these cases area type fraction related weighting provides the lowest overall uncertainty, i.e. lowest RMSE and MPE nearest to zero. It means that the current method and alternative A give better results than alternatives B and C.

At both $PM_{10}$ indicators the current approach and alternative A provide the best interpolation results at the rural station points, but the worst at urban station points when comparing among merging methods.

The two proposed alternatives B and C show the most reduced uncertainty (lowest RMSE values) at the urban stations in Tables 5.7 and 5.9. However, with a co-existing enhanced uncertainty at the rural areas, which may have its cause in the effect the urban/suburban stations with high concentrations

located in the low populated areas have on the rural predicted values. Another cause may be that in the interpolation of the residuals $\eta(s_0)$ in Equations 2.2 and 2.9 respectively no weighting function is applied, i.e. one and the same variogram is applied for both the rural and the urban/suburban background stations. Contrary to that, the interpolations at the current method and the alternative A take place separately on the rural and on the urban/suburban background stations.

As can be seen in Tables 5.8 and 5.10, the MPE is positive for all methods at the rural background stations, meaning in all cases the predicted corresponding grid values of the final map will be somewhat overestimated. Besides, the alternative methods B and C show a higher positive bias than the current approach and alternative A. This high overestimation at B and C is caused by influence of urban stations with high measured concentrations being located in rural areas. At method B this involves all urban stations, whereas at method C only those located in low populated areas. Therefore, the overestimation at rural stations is highest at method B. Somewhat reduced overestimation of grid squares in the final map occurs at current method and alternative A. (The limited overestimation of grid squares at some rural stations in lower and moderately populated areas is already discussed at Section 5.1.2.) One can conclude that alternative approaches B and C show some increased uncertainties at the rural areas.

Compared to the current approach and the alternative A, the smaller absolute values of MPE for the (sub)urban background stations at the alternative methods B and C indicate a smaller underestimation of the predicted corresponding grid values of the final map in relation to the (sub)urban point measurements. However – as shown in Section 5.1.2 – in case of the current approach the predicted grid squares of the combined final map are on average lower than the corresponding (sub)urban station measurements. This occurs in particular in lower or moderately populated and relative clean rural areas with urban/suburban stations representing small highly polluted cities. Bearing in mind that the final grid represents 10x10 km spatial means, the lower predicted value of the 10x10 km grid cell in such cases is a more realistic result. Thus, the alternatives B and C show better agreement of the final grid with the urban/suburban station measurements than at the current approach and alternative A. However, we cannot conclude that B and C show better agreement of the final grid with reality considered as 10x10 spatial mean.

In conclusion, the alternative methods B and C show some increased uncertainties at the rural areas. The role of rural station density variability seems to play a role here.

In the Figure 5.7 the final maps of the 36$^{th}$ maximum daily mean for 2006, constructed according the four merging approaches are presented.

In comparison with the current merging approach and the approach A, the alternative merging approaches B and C result, in general, in higher predicted grid values at large areas of Europe. This is caused specifically by the fact that the urban stations are used for mapping of rural areas in these two approaches. Most obvious it is in method B, where no distinction is made between (sub)urban and rural stations. The major difference between the maps based on the current approach and alternative B is observed in the rural areas with a lack of rural stations and/or high values of urban stations, particularly in the Balkan region, but also in Poland, Scandinavia and in smaller areas of southern Spain. The major differences of alternative C compared to the current approach and alternate A, exist in areas with a quite high distinction between the rural and the urban station measurements and simultaneously a high density of urban stations. This is caused probably by the joint interpolation of the common regression residuals from both the rural and urban stations (Equation 2.9); this joint interpolation of both urban and rural residuals results in a larger deviation from the current method than the application of a changed differentiation between the rural and urban stations (i.e. based on population density instead of EoI classification). We see this effect in Balkan area with its high concentrations at many urban stations, but also in France with its lower reported concentrations at the urban stations than at the rural ones. So far, we can conclude that the current method and alternative A give better results than alternatives B and C.

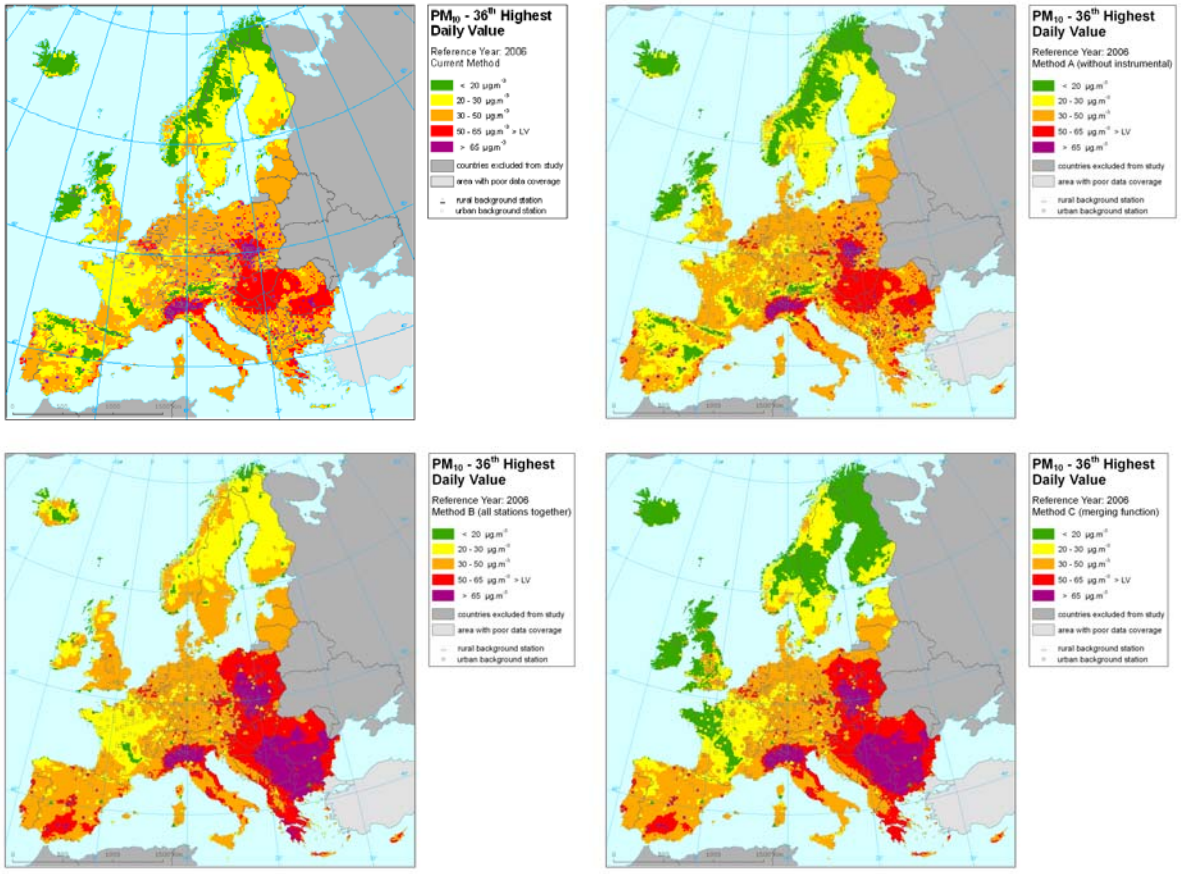


*Figure 5.7 Maps showing the 36th maximum daily average $PM_{10}$ values (in $\mu g.m^{-3}$) on the European scale in 2006, 10x10 km grid resolution, as a result of the current merging approach (top left), alternative A (top right), alternative B (bottom left) and alternative C (bottom right).*

When subsequently comparing the maps constructed by the current method and alternative A, only limited differences are visible. The major difference is in France, where at the current merging method the rural part of the combined final map is constructed from the rural stations and adjusted with the joint urban/rural map (see Section 2.1) encompassing the lower reported values of the urban stations. Another example of such difference is given in Figure 5.8, with the maps of the 36th maximum daily average for 2007 for the current merging method and alternative A. It demonstrates that in central Italy the overestimated rural grid at method A, is adjusted at the currently used merging approach by application of the joint urban/rural grid, which encompasses the lower urban measurement values. This overestimation at alternative A of the rural grid in central Italy is caused by a lack of rural stations in this area and by high $PM_{10}$ values in northern Italy. The map sections of the figure show that the use of a joint urban/rural map option in the merging procedure brings improved results in the combined final map.

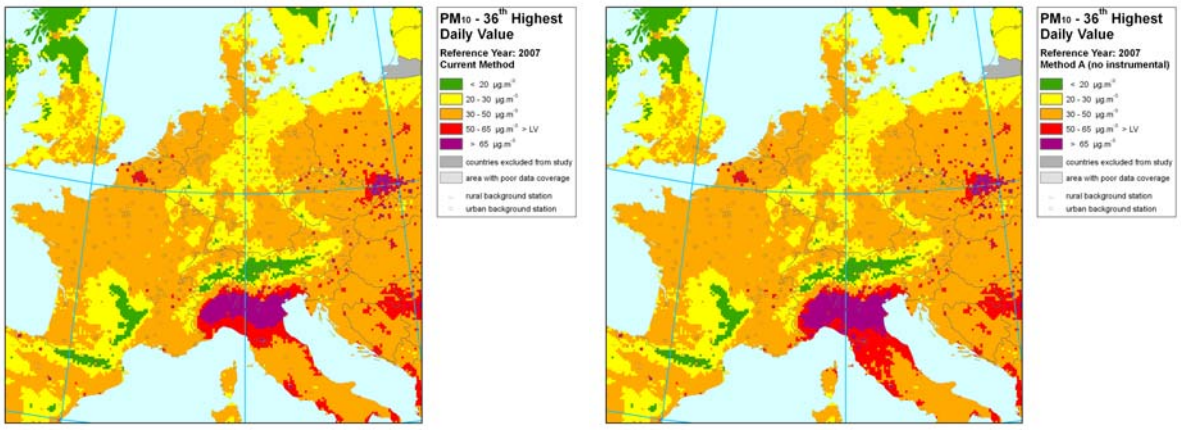


*Figure 5.8 Maps showing the 36th maximum daily average $PM_{10}$ values (in µg.m$^{-3}$) for the central part of Europe in 2007, 10x10 km grid resolution, as a result of the current merging method (left) and alternative A (right). In the current approach, joint urban/rural map is used in merging process, while under alternative A it is not.*

From all these findings we can conclude that the proposed alternative merging approaches (applying Equations 2.8, 2.2 and 2.9) do not improve the $PM_{10}$ mapping results at the end. In other words, the current merging method used so far gives in general already the best results.

## 5.2.2 Ozone

As for $PM_{10}$ also for ozone the comparison of the current and three alternative merging approaches is executed. Tables 5.11 and 5.12 present the uncertainty of the interpolation results expressed as the statistical indicators RMSE and MPE of the cross-validations for the rural and (sub)urban station types and for the final map results obtained through the current merging method and its three alternatives for 26th highest daily maximum 8-hour average of 2006. In the Tables 5.13 and 5.14 the same indicators are shown for SOMO35.

*Table 5.11 Uncertainty estimates, expressed as RMSE (in µg.m$^{-3}$), of the interpolation results for the ozone indicator 26th highest daily maximum 8-hour average for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of RMSE, using four different 'merging' methods. RMSE should be as small as possible.*

| | rural stations | urban stations | final map - RMSE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| current approach (with joint urban/rural map) | 11.02 | 10.74 | 10.84 | 11.00 | 10.86 |
| method A (current, without joint urban/rural map) | 11.06 | 10.80 | 10.89 | 11.04 | 10.90 |
| method B (all stations together) | 11.45 | 10.32 | 10.72 | 11.37 | 10.79 |
| method C (merging function, all stations) | 11.28 | 10.24 | 10.60 | 11.21 | 10.67 |

*Table 5.12 Uncertainty estimates, expressed as MPE (in µg.m$^{-3}$), of the interpolation results for the ozone indicator 26th highest daily maximum 8-hour average for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of MPE, using four different 'merging' methods. MPE should be as close to zero as possible.*

| | rural stations | urban stations | final map - MPE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| current approach (with joint urban/rural map) | -0.19 | 1.40 | 0.86 | -0.08 | 0.74 |
| method A (current, without joint urban/rural map) | -0.31 | 1.46 | 0.86 | -0.19 | 0.73 |
| method B (all stations together) | -3.04 | 1.52 | -0.02 | -2.72 | -0.35 |
| method C (merging function, all stations) | -1.34 | 0.82 | 0.09 | -1.19 | -0.07 |

*Table 5.13 Uncertainty estimates, expressed as RMSE (in µg.m$^{-3}$), of the interpolation results for the ozone indicator SOMO35 for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of RMSE, using four different 'merging' methods.*

| | rural stations | urban stations | final map - RMSE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| current approach (with joint urban/rural map) | 2062 | 1717 | 1840 | 2037 | 1859 |
| method A (current, without joint urban/rural map) | 2081 | 1724 | 1852 | 2056 | 1871 |
| method B (all stations together) | 2247 | 1575 | 1830 | 2200 | 1851 |
| method C (merging function, all stations) | 2181 | 1556 | 1791 | 2137 | 1813 |

*Table 5.14 Uncertainty estimates, expressed as MPE (in µg.m$^{-3}$.d), of the interpolation results for the ozone indicator SOMO35 for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of MPE, using four different 'merging' methods.*

| | rural stations | urban stations | final map - MPE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| current approach (with joint urban/rural map) | -149 | 342 | 176 | -115 | 140 |
| method A (current, without joint urban/rural map) | -141 | 353 | 186 | -107 | 150 |
| method B (all stations together) | -755 | 369 | -10 | -676 | -93 |
| method C (merging function, all stations) | -396 | 195 | -5 | -355 | -48 |

When the area weighting is applied to obtain a more balanced overall uncertainty indication of the combined final European map again the current merging method and alternative A provide the lowest RMSE and an MPE nearest to zero for the same reasons as given at $PM_{10}$.

Again the interpolation uncertainties are at the rural stations the smallest at current merging approach and also at alternative approach A. But the extreme high uncertainties at the urban station points are less explicit and less systematic in ranking among the four approaches.

Like at $PM_{10}$, also at ozone one can observe the best improvement of the two proposed alternatives B and C for the urban station points, and on the other hand the uncertainty in the rural areas is higher but of a much smaller magnitude than at $PM_{10}$. The MPE values for the alternatives B and C compared to the current merging approach and the alternative A show at ozone a similar pattern as at $PM_{10}$ and for the same reasons.

The MPE is negative at all merging methods at the rural background stations, i.e. an underestimation of the predictions at the corresponding grid squares at the rural station points. The large underestimation of the grid square values at methods B and C is caused by the influence of low measurements at urban stations located in rural areas. At method B it involves all urban stations, whereas at method C only those located in low populated areas. Therefore, the underestimation at rural stations is larger at method B. Somewhat reduced underestimation of grid squares in the final map occurs at current method and alternative A. (The limited underestimation of grid squares at some rural stations in lower and moderately populated areas is already discussed in Section 5.1.2.) One can conclude that alternative approaches B and C show some increased uncertainties at the rural areas.

The positive MPE at the (sub)urban background stations indicate higher predicted values at the corresponding grid squares in comparison with some (sub)urban station points, especially in case of small cities in rural areas (see Section 5.1.2). This may have its cause for similar reasons as given at $PM_{10}$ but then with ozone concentrations in rural areas being high compared to urban concentrations. Like at $PM_{10}$, also at ozone the smaller MPE at (sub)urban stations at approach B and C indicates that these alternatives show a slightly better agreement of the 10x10 km grid with the station point data than the current approach, but no better agreement of the 10x10 km grid with the reality considered as 10x10 spatial mean.

We conclude that the alternative approaches (applying Equations 2.8, 2.2 and 2.9) do not improve the mapping results for ozone.

## *5.3 Improvements of the current merging methodology*

### 5.3.1 Alternative population density class intervals

As concluded in the Section 5.2, the examined alternative approaches do not improve the mapping results, thus we stick to the current merging method. However, as suggested in Section 5.1.1., some improvements of this current method might be realised if we would change the population density classification interval boundary $\alpha_1$ (Equation 2.8), being the class boundary between the rural and mixed population density grid cell types, from 100 inhbs.km$^{-2}$ into a lower 50 inhbs.km$^{-2}$. This would in particular contribute to improve the mapping of the small but highly polluted cities located in rather sparsely populated rural areas.

**$PM_{10}$**

The comparison of the combined final $PM_{10}$ maps constructed according to the currently used and suggested class boundary between the rural and mixed grid types, $\alpha_1$, is executed by the same way as has been done for the different merging methods in Section 5.2. Tables 5.15 and 5.16 present the results of the comparisons, for cross-validation indicators RMSE (in µg.m$^{-3}$) and MPE or bias (in µg.m$^{-3}$).

*Table 5.15 Uncertainty estimates, expressed as RMSE (in µg.m$^{-3}$), of the interpolation results for the $PM_{10}$ indicators annual average and 36$^{th}$ maximum daily average for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of RMSE, using current merging method with currently used 100 inhbs.km$^{-2}$ and alternative 50 inhbs.km$^{-2}$ class boundary value between the rural and mixed grid types. RMSE should be as small as possible.*

| $PM_{10}$ | rural stations | urban stations | final map - RMSE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| **annual average** | | | | | |
| current approach, current setting $\alpha_1$ = 100 | 5.45 | 7.40 | 7.05 | 5.59 | 6.60 |
| current approach, alternative setting $\alpha_1$ = 50 | 5.45 | 6.53 | 6.32 | 5.53 | 6.09 |
| **36th maximum daily mean** | | | | | |
| current approach, current setting $\alpha_1$ = 100 | 9.16 | 13.53 | 12.76 | 9.47 | 11.73 |
| current approach, alternative setting $\alpha_1$ = 50 | 9.19 | 12.15 | 11.61 | 9.40 | 10.93 |

*Table 5.16 Uncertainty estimates, expressed as MPE (in µg.m$^{-3}$), of the interpolation results for the $PM_{10}$ indicators annual average and 36$^{th}$ maximum daily average for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of RMSE, using current merging method with currently used 100 inhbs.km$^{-2}$ and alternative 50 inhbs.km$^{-2}$ class boundary value between the rural and mixed grid types. MPE should be as close to zero as possible.*

| $PM_{10}$ | rural stations | urban stations | final map - MPE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| **annual average** | | | | | |
| current approach, current setting $\alpha_1$ = 100 | 0.99 | -2.06 | -1.44 | 0.78 | -0.81 |
| current approach, alternative setting $\alpha_1$ = 50 | 1.24 | -1.08 | -0.60 | 1.08 | -0.12 |
| **36th maximum daily mean** | | | | | |
| current approach, current setting $\alpha_1$ = 100 | 1.56 | -3.37 | -2.36 | 1.21 | -1.35 |
| current approach, alternative setting $\alpha_1$ = 50 | 1.93 | -1.76 | -1.01 | 1.67 | -0.24 |

Table 5.15 shows that the RMSE values, expressing the mean uncertainty for the grid cells in the final map, are improved in almost all cases when applying the alternative class boundary of 50 inhbs.km$^{-2}$. The major improvement occurs – as expected – at the urban stations, where one observes a better agreement between the grid values of the final map and corresponding (sub)urban station measurements in moderately to low populated areas. RMSE weighted by area gives only slight (ca.1%), while RMSE weighted by population density stronger (7-8%) improvement. Applying the alternative class boundary provides MPE with some improvement in urban areas and a slight worsening, i.e. increased overestimation, in rural areas. Nevertheless, the suggested boundary change results specifically in an improvement of the mapping of small cities in larger more sparsely populated rural areas, which had been detected as most likely one of the weakest points in the current mapping methodology. All in all, we suppose that the alternative class boundary of $\alpha_1$ = 50 inhbs.km$^{-2}$ would provide slight improvement of the map and will provide subsequently greater improvement at the impact assessments. As such, we could recommend to use this alternative class boundary of $\alpha_1$ = 50 inhbs.km$^{-2}$ from now on as part of the routine merging methodology of $PM_{10}$. Although in Section 6.2 we examine the use of a finer grid resolution as alternate option to improve the merger, which appears to be significantly more effective. See Chapter 7 for final recommendations.

### Ozone

As for $PM_{10}$ also for ozone the comparison of the combined final maps constructed with the use of the current and suggested boundary between the rural and mixed grid types is executed. Tables 5.16 and 5.17 present the uncertainty of the interpolation results expressed as the cross-validations indicators RMSE and MPE for the rural and (sub)urban station types and for the final map results obtained through the current merging method and its three alternatives for both ozone indicators for 2006.

*Table 5.16 Uncertainty estimates, expressed as RMSE (in µg.m$^{-3}$ and µg.m$^{-3}$.d resp.), of the interpolation results for the ozone indicators 26$^{th}$ highest daily maximum 8-hour average and SOMO35 for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of RMSE, using current merging method with currently used 100 inhbs.km$^{-2}$ and alternative 50 inhbs.km$^{-2}$ class boundary value between the rural and mixed grid types.*

| Ozone | rural stations | urban stations | final map - RMSE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| **26th highest daily maximum 8-hourly mean** | | | | | |
| current approach, current setting $\alpha_1$ = 100 | 11.02 | 10.74 | 10.84 | 11.00 | 10.86 |
| current approach, alternative setting $\alpha_1$ = 50 | 10.99 | 10.64 | 10.76 | 10.97 | 10.78 |
| **SOMO35** | | | | | |
| current approach, current setting $\alpha_1$ = 100 | 2062 | 1717 | 1840 | 2037 | 1859 |
| current approach, alternative setting $\alpha_1$ = 50 | 2062 | 1680 | 1818 | 2035 | 1837 |

*Table 5.17 Uncertainty estimates, expressed as MPE (in µg.m$^{-3}$ and µg.m$^{-3}$.d resp.), of the interpolation results for the ozone indicators 26$^{th}$ highest daily maximum 8-hour average and SOMO35 for 2006 at the rural and urban station points, and for the final European map based on three different weighting criteria of the rural and urban contributions of RMSE, using current merging method with currently used 100 inhbs.km$^{-2}$ and alternative 50 inhbs.km$^{-2}$ class boundary value between the rural and mixed grid types.*

| Ozone | rural stations | urban stations | final map - MPE weighted by | | |
|---|---|---|---|---|---|
| | | | no weight | area | population |
| **26th highest daily maximum 8-hourly mean** | | | | | |
| current approach, current setting $\alpha_1$ = 100 | -0.21 | 1.41 | 0.86 | -0.10 | 0.74 |
| current approach, alternative setting $\alpha_1$ = 50 | -0.41 | 1.28 | 0.71 | -0.29 | 0.59 |
| **SOMO35** | | | | | |
| current approach, current setting $\alpha_1$ = 100 | -149 | 342 | 176 | -115 | 140 |
| current approach, alternative setting $\alpha_1$ = 50 | -194 | 311 | 141 | -158 | 104 |

The results presented in Tables 5.16 and 5.17 reflect the same tendency as discussed for $PM_{10}$, which leads to the same conclusions as obtained at $PM_{10}$, i.e. the alternative class boundary of $\alpha_1$ = 50 inhbs.km$^{-2}$ could be used for improving of the routine mapping methodology of ozone. However, Section 6.2 suggests that applying a higher grid resolution as merger is likely a better alternative. See Chapter 7 for the final recommendations.

## 5.3.2 Specification of the merging

As mentioned in Section 2.1, the final map is constructed by combining the rural map, the urban map, and – in limited areas where the rural map shows higher $PM_{10}$ values, resp. lower ozone values than in the urban map – the joint urban/rural map. The reason for using this joint urban/rural map in these areas is following:

Though it is not impossible for a rural $PM_{10}$ station to be higher than an urban $PM_{10}$ station (and vice versa for ozone stations) it is expected that such a situation should be limited in its spatial extent, e.g. that we do not expect islands of low urban $PM_{10}$ concentrations in a sea of high rural $PM_{10}$ concentrations. Though we agree that it is possible for a rural station to show higher $PM_{10}$ concentrations than an urban one, the original reason for splitting the rural and urban stations into two maps is based on an implicit assumption that the rural and urban concentration fields are different, i.e. that urban $PM_{10}$ tends to be higher than the rural $PM_{10}$ (and vice versa for ozone). However, in the areas with actual lower urban grid $PM_{10}$ concentrations than rural grid values (vice versa for ozone)

this assumption seems not to be valid, and thus in these areas the grid values are derived from the map that is composed from all rural and urban/suburban background stations taken together in the interpolation, the so-called joint urban/rural grid map.

In general, one can assume that the joint urban/rural grid value lies between the urban and the rural corresponding grid value. However, in some cases this in not true. Reasons for such deviations are, for example, that different supplementary variables or different kriging parameters are used to create the individual urban, rural and joint urban/rural maps. Nevertheless, grid cell values of this joint rural/urban map should fall within the interval of the corresponding grid cell values from the separate rural and urban maps. Thus, Equation 2.3 should be specified for $PM_{10}$ more precisely as

$$\begin{aligned}
\hat{Z}(s_0) &= \hat{Z}_r(s_0) && \text{for } \alpha(s_0) \le \alpha_1 \text{ and } (\hat{Z}_r(s_0) \le \hat{Z}_u(s_0) \text{ or } \hat{Z}_j(s_0) > \hat{Z}_r(s_0) > \hat{Z}_u(s_0) \text{ or} \\
&&& \quad \hat{Z}_r(s_0) > \hat{Z}_u(s_0) > \hat{Z}_j(s_0)) \\
&= \hat{Z}_u(s_0) && \text{for } \alpha(s_0) \ge \alpha_2 \text{ and } (\hat{Z}_r(s_0) \le \hat{Z}_u(s_0) \text{ or } \hat{Z}_j(s_0) > \hat{Z}_r(s_0) > \hat{Z}_u(s_0) \text{ or} \\
&&& \quad \hat{Z}_r(s_0) > \hat{Z}_u(s_0) > \hat{Z}_j(s_0)) \\
&= \frac{\alpha_2 - \alpha(s_0)}{\alpha_2 - \alpha_1} \cdot \hat{Z}_r(s_0) + \frac{\alpha(s_0) - \alpha_1}{\alpha_2 - \alpha_1} \cdot \hat{Z}_u(s_0) && \text{for } \alpha_1 < \alpha(s_0) < \alpha_2 \text{ and } (\hat{Z}_r(s_0) \le \hat{Z}_u(s_0) \text{ or} \\
&&& \quad \hat{Z}_j(s_0) > \hat{Z}_r(s_0) > \hat{Z}_u(s_0) \text{ or } \hat{Z}_r(s_0) > \hat{Z}_u(s_0) > \hat{Z}_j(s_0)) \\
&= \hat{Z}_j(s_0) && \text{for } \hat{Z}_r(s_0) > \hat{Z}_u(s_0) \text{ and } \hat{Z}_r(s_0) \ge \hat{Z}_j(s_0) \ge \hat{Z}_u(s_0) \qquad (5.1)
\end{aligned}$$

For ozone the similar equation applies with *opposite* conditions for the relation between $\hat{Z}_r(s_0)$, $\hat{Z}_u(s_0)$ and $\hat{Z}_j(s_0)$.

## 5.4 Discussions

The current methodology of merging of the separately interpolated urban and rural maps into the rural and urban areas of the combined final European interpolation map for $PM_{10}$ and ozone indicators has been reconsidered with the aim to reduce the current uncertainties further. Different options for improvements in the merging technique that could lead to reducing the uncertainties in the final map have been proposed and tested against the current merging method.

***Current method***

Potential weak elements of the current merging methodology have been examined and evaluated in Section 5.1. Elements we have considered are:

- the level of correlation and consistency between the station type (as reported by the countries according the EoI classification) and the corresponding grid cell type according the population density classification criteria used to define the rural and urban area types in the final maps;
- the level of correlation between the station measurement values itself and the corresponding predicted grid cell values in the final map, including the involved level of uncertainties;
- the subsequent assignment of the predicted values into the final combined map from either the separate rural or separate urban map, based on the population density classification criterion of rural area (<100 inhbs.km$^{-1}$), urban areas (>500 inhbs.km$^{-1}$), and a weighting function for the 'mixed' areas in between, including the involved uncertainties of the predicted grid values in the final map related to the corresponding station measurements.

The classification of the measurement stations was compared with the classification of the corresponding grid cells of the population density field. In some cases we found classification inconsistencies between the station types used in the spatial interpolation and the corresponding population density grid cells used for the assignment of either a value of the separately interpolated

urban/suburban map or rural map, or even a weighted value into the final combined map. For those stations that have a classification not corresponding with the population density type of the area in which they are sited. It is recommended to verify the correctness of their classification and their coordinates (Table 5.18).

Concluded was that for the urban areas the match between the station type and the corresponding population density grid cell type is better when higher resolution of 1x1 km grid is used instead of the currently applied 10x10 km grid (Table 5.18). Particularly in the smaller 1x1 km grid cells considerably more rural stations were found to be located in corresponding rural grid cells, more urban and suburban stations in urban grid cells and fewer stations were located in the grid cells of the type 'mixed', which certainly contributes to lower interpolation uncertainties due to reduced inconsistencies in the match of the station type and corresponding grid cell type. The urban population density shows a greater improvement in matching stations types then at the rural types. In Chapter 6 this resolution related interpolation improvement and uncertainty reduction is further evaluated when changing interpolation resolution from a 10x10 km into a 1x1 km grid for the whole mapping.

We found that the variability in the regional distribution does play a role in the consistency of the match between the station types and corresponding grid cell types. Evidently there is some clustering of stations in areas where the rural concentration exceeds the urban concentration in some parts of Europe. It could indicate that the station and/or population density classification has not been applied completely optimal or consistent, or that the chosen spatial distribution of the monitoring stations is not optimal compared to the population density grid cell types or vice versa. Using more optimal population density class intervals has been investigated as improvement of the current interpolation methodology, see further. Also the spatial pattern of the area types and the layout of monitoring networks covering these areas may play a role and might need improvement (Table 5.18). We did not further deal with it, since it was outside the scope of this paper.

We observed an overestimation of $PM_{10}$, and an underestimation of ozone, in grid cells of the final map at some rural stations sited in moderately or highly populated areas. However, the greatest inconsistency of the current approach was found for urban/suburban stations sited in low populated areas. Those $PM_{10}$ stations appeared to represent specifically small but highly polluted cities in quite clean rural areas. In the current methodology the measured values from such urban/suburban stations have no influence on the corresponding rural grid cells of the final map. The final map is composed by assignment of grid cell values from either the separate rural or urban map on basis of the population density grid classification: to the rural grid cells the values from separate rural map are assigned. The separate rural map is based on a spatial interpolation of rural stations only. As such these highly $PM_{10}$ polluted small cities in low polluted rural areas are 'overlooked' this way, causing uncertainties in the map that might be tackled. Similar effect exists at ozone mapping. In an attempt to reduce the uncertainties related to this methodological effects, different potential methodological improvement have been investigated: three alternative merging approaches, different population density class interval boundaries (in Equation 2.3) but with the currently used methodology, and finally executing the spatial interpolation at a higher grid resolution (see Chapter 6).

***Alternative merging methods***

Three alternative merging approaches that could aim at reaching reduced uncertainties in the final European maps were evaluated against the current merging method in Section 5.2. Uncertainty indicators RMSE and MPE have been used in a non-weighted context as well as in a context of weighting by area or by population density. We introduced the weighting to compensate the overall estimated uncertainty for the disproportionally large number of the urban and suburban background stations representing the relative small total urban area compared to the limited number of rural stations representing the significant larger total rural area. For all $PM_{10}$ and ozone indicators the best uncertainty results, for both RMSE and MPE, were obtained when weighting by area at the currently used merging method and alternative A (i.e. current method, but ignoring the use of the *joint urban/rura*l map).

Next to this, MPE is positive at $PM_{10}$ (resp. negative at ozone) for all methods at the rural background stations, meaning that in all cases the predicted corresponding final grid values are somewhat

overestimated (resp. underestimation). The alternatives B and C show greater overestimation (resp. underestimation at ozone) than the current merging method and alternative A. This high overestimation (resp. underestimation) at B and C is caused by the influence of the high (resp. low) measured concentrations at (sub)urban stations sited in rural areas. At method B this involves all urban stations, whereas at method C only those located in low populated areas. Therefore, the overestimation (resp. underestimation) at rural stations is highest at method B. Somewhat more reduced overestimation (resp. underestimation) of grid squares in the final map occurs at current method and alternative A. The limited overestimation of grid squares takes place at some rural stations in lower and moderately populated areas (see Section 5.1.2).

These findings are clearly visible, when comparing the maps constructed based on different merging approaches. The alternatives B and C show for $PM_{10}$ in general higher (resp. lower for ozone) predicted grid values at large areas of Europe caused by the overestimation (resp. underestimation) of the rural areas, due to the fact that these two approaches take into account (sub)urban stations with their high values in the spatial interpolations of rural areas, which is most obvious in alternative B where no distinction is made between urban and rural stations. One can conclude that alternative approaches B and C show some increased uncertainties at the rural areas.

When comparing current merging method with alternative A, only limited differences have been found in the uncertainty analysis, due to the similar character of these approaches. The only difference is that at the current merging method the grid squares of the combined final map are assigned from the joint rural/urban map at areas where the rural map shows higher $PM_{10}$ concentrations (resp. lower at ozone) than the urban map. Concluding, the application of such joint rural/urban map does contribute to a better combined final map at the $PM_{10}$ and ozone indicators.

From all the findings we can conclude that the proposed alternative merging approaches do not improve the mapping results at the end. In other words, the current merging method used so far gives in general already the best results (Table 5.18).

***Improvements of the current merging methodology***

Sticking to the current method, we observed that the some improvement could be reached when selecting different class interval boundaries for population density classes ($\alpha_1$ and possibly $\alpha_2$ of Equation 2.3). Section 5.3.1 concludes that an alternative lower value of $\alpha_1$ = 50 inhbs.km$^{-2}$, instead the current 100 inhbs.km$^{-2}$, would provide slight improvement of the interpolated map and will provide subsequently improved impact assessments. Specifically improved interpolation results will be obtained at the mapping of small cities in larger more sparsely populated rural areas. This 'overlooking' of such small cities had been detected as most likely one of the weakest points in the current mapping methodology. Therefore, the alternative class boundary of $\alpha_1$ = 50 inhbs.km$^{-2}$ could be used as part of the routine mapping methodology of $PM_{10}$ and ozone (Table 5.18). However, it is recommended to verify for $PM_{10}$ whether this alternative class boundary provides also improvements at the log-normal transpositions. Next to this, an alternate option on how to improve the merger is on basis of the grid resolution; see Section 6.2 where it is suggested that applying a higher grid resolution as merger is likely a better alternative. Chapter 7 comes with final recommendations.

On top of that, Section 5.3.2 specifies more precisely the conditions for selecting the ultimate grid cell value for the combined final map from the separate interpolated maps at those areas where higher rural grid cell values in the rural $PM_{10}$ map, (resp. lower at ozone) are observed than their corresponding urban grid cell values of the urban map. The use of the more precisely specified merging function, Equation 5.1 (as refinement of Equation 2.3) is recommended for implementation in the default methodology for the mapping of the $PM_{10}$ and ozone health indicators (Table 5.18).

*Table 5.18 Recommended methodological improvements to be implemented as part of the default mapping methodologies as listed in Table 7.1 of Horálek et al. (2008).*

| Options methodological improvement | $PM_{10}$ | Ozone |
|---|---|---|
| **At current methodology** | | |
| Improve quality reported stations classification & cooridnates | Implement [1] | Implement [1] |
| Improve quality spatial layout monitoring networks | Implement [1] | Implement [1] |
| Apply pop.dens. class boundary $\alpha_1$ = 50 inhbs.km$^{-2}$ (instead of 100) | Implement occasionally [2][3] | Implement occasionally [2] |
| Refine merging criteria (apply Eq. 5.1 instead of Eq. 2.3) | Implement | Implement |
| **Introduce alternative merging methods** | | |
| Current method (with joint urban/rural map) | Keep | Keep |
| Method A (current method, without joint urban/rural map, Eq. 2.8) | -- [4] | -- [4] |
| Method B (all stations together in one interpolation, Eq. 2.2) | -- [4] | -- [4] |
| Method C (merging function, all stations, Eq. 2.9) | -- [4] | -- [4] |

[1] *Implementation is primarily at control and responsibility of the countries and DG ENV.*

[2] *To apply occasionally in mapping processes in case application of a finer 1x1 km grid merger would be too computational and/or time demanding. Nevertheless, the use of a finer 1x1 km grid as merging resolution is a better alternative to this, see Section 6.2.*

[3] *Verify at application of a log-normal transposition at $PM_{10}$ (Chapter 4) if indeed improvements are obtained. (Not tested in this paper).*

[4] *No significant improvement found; do not implement.*

# 6 Grid resolution

The role of the grid resolution used in spatial interpolation, map merging and population exposure assessment is examined in this chapter for both the two $PM_{10}$ and the two ozone indicators. The linear regressions and their residuals interpolations of separate rural and urban maps, as well as the subsequent merging into the combined final concentration map are currently executed on an aggregated 10x10 km grid resolution. That is to keep the calculation demands within reasonable hardware and software limits. Nonetheless, it is known and confirmed in Section 5.1 that spatial aggregations in the different production steps do contribute to increased uncertainties and sub-grid variability in the mapping results and successive exposure estimates.

The aim of this chapter is to provide input for optimizing the grid resolution at each process step, accounting for both the suppression of uncertainties as well as keeping calculations within acceptable computational limits. Basically, we compared the uncertainties of the 10x10 km grid resolution calculations currently applied with those of a 1x1 km grid resolution and a series of combinations at each process step. Additionally, a 3x3 floating grid resolution is considered as kind of intermediate merging resolution.

Section 2.4 describes in detail the combinations of resolutions used in the three different processing steps going from monitoring data to national and European population exposure estimates. Summarized they are: interpolation of separate maps – merging into a final concentration map – the final concentration map and its subsequent population exposure estimates. In this Chapter 6 each examined combination is encoded according section 2.4: A (10-10-10), B (10-fl3-1), C (10-1-1), D (1-1-1), E (1-fl3-1) and F (aggrD-10).

Section 6.1 compares the uncertainties of the separate rural and urban maps as obtained at different interpolation resolutions. Section 6.2 extends on the uncertainties of the combined final concentrations maps attained from merging at different resolutions. In Section 6.3 the ultimate population exposure estimates obtained with the different resolution combinations A – F are compared for both the population weighted concentrations and exposure exceedance classes for individual countries and Europe as a whole.

## *6.1 Separate rural and urban concentration maps*

As first step the analysis of separately created rural and urban maps according Equation 2.1 on a 10x10 km and a 1x1 km grid resolution has been executed for both the two $PM_{10}$ and the two ozone indicators. The interpolation uncertainty for the health indicators is presented in Table 6.1 for $PM_{10}$ and in Table 6.2 for ozone as the RMSE from cross-validation at the two different resolutions and for both the rural and urban maps. Figure 6.1 shows the cross-validation scatter-plots for $PM_{10}$ and Figure 6.2 for ozone. Due to the large similarities in results at both pollutants, we discuss them here simultaneously.

The tables indicate that refining the grid resolution of the interpolation brings slight improvement of 1-2 % for $PM_{10}$ and 2-6% for ozone at the rural areas, whereas no improvement is observed at both pollutants at the urban areas. The observed improvement is caused by the fact that altitude, which is used at rural areas only, is the only supplementary variable that is available in a smaller resolution than 10x10 km and therefore shows a better fit with interpolations at a reduced grid resolution of 1x1 km than it does at 10x10 km. The considerable improvement of about 6% at SOMO35 is probably caused by its more pronounced relationship of ozone with the finer resolution of the altitude field.

The scatter-plots in Figures 6.1 and 6.2 give similar results. The improvement in the correlation of the cross-validation predicted concentrations of the interpolation on 1x1 km with station measurements in the rural areas is clearly visible: the increase of $R^2$ and slope is between 2-4 % at the $PM_{10}$ indicators and almost 5% at 26$^{th}$ highest daily maximum 8-hour ozone mean; at SOMO35 the $R^2$ increases almost 14% and the slope about 11%. Contrary to that, slight worsening occurs at the urban areas: $R^2$ and slope reduce less than 1% for $PM_{10}$ and about 1-3 % for ozone.

## $PM_{10}$

*Table 6.1 The interpolation results of the current interpolation method, executed at 10x10km and 1x1 km grid resolution, expressed as the RMSE (in µg.m-3) for the $PM_{10}$ indicators annual average and 36th maximum daily mean in rural and urban areas. The smaller the RMSE the better the interpolation performance is.*

| $PM_{10}$ | RMSE (µg.m$^{-3}$) | | | |
|---|---|---|---|---|
| | annual average | | 36th max. daily mean | |
| | rural | urban | rural | urban |
| 10 x10 km grid resolution (currently used) | 5.79 | 6.09 | 9.85 | 11.67 |
| 1x1 km grid resolution | 5.66 | 6.12 | 9.79 | 11.70 |

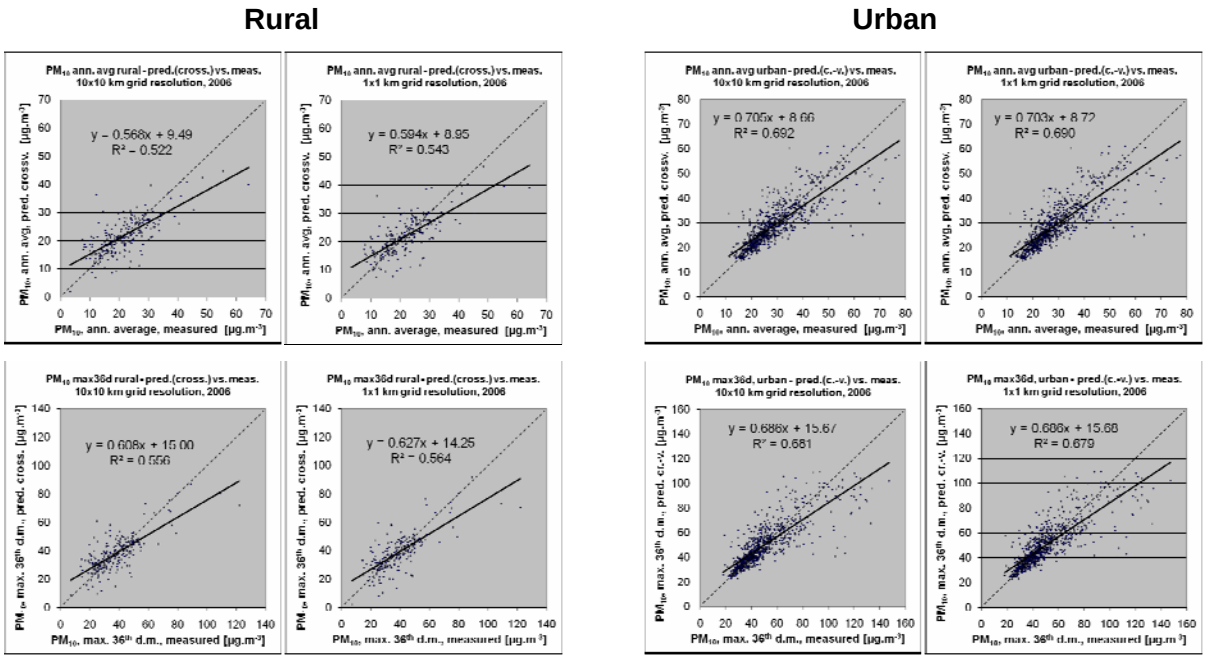


*Figure 6.1 Scatter-plots showing the correlation between cross-validation predicted values (y-axis) and measurements (x-axis) at station points for the $PM_{10}$ annual average (top) and 36th maximum daily averages (bottom) for 2006 at rural (left) and urban (right) areas on the 10x10 km (left columns) and 1x1 km (right columns) resolution.*

## Ozone

*Table 6.2 The interpolation results of the current interpolation method, executed on 10x10km and 1x1 km expressed as the RMSE (in µg.m$^{-3}$, resp. µg.m$^{-3}$.d) for the ozone indicators SOMO35 and 26th highest daily maximum 8-hourly mean in rural and urban areas.*

| ozone | RMSE | | | |
|---|---|---|---|---|
| | 26th high. m. d. 8h | | SOMO35 | |
| | rural | urban | rural | urban |
| 10 x 10 km grid resolution (currently used) | 11.17 | 10.17 | 2077 | 1471 |
| 1 x 1 km grid resolution | 10.91 | 10.37 | 1946 | 1488 |

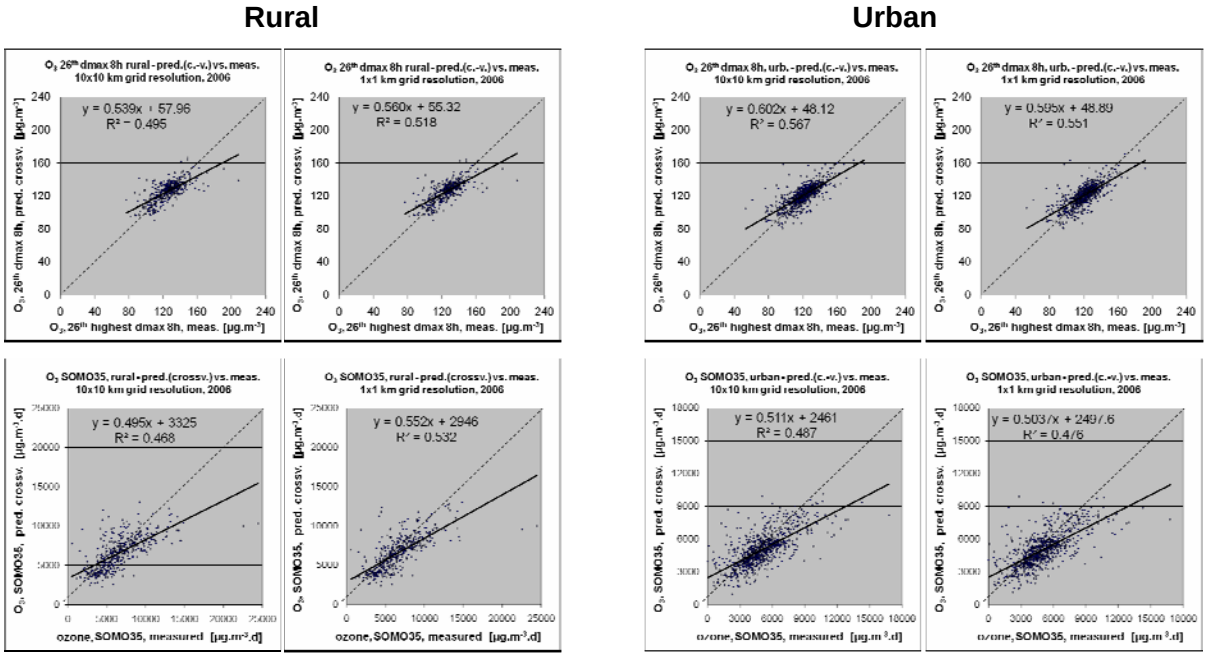


*Figure 6.2 Scatter-plots showing the correlation between cross-validation predicted values (y-axis) and measurements (x-axis) at station points for the ozone indicators 26th highest daily maximum 8-hourly mean (top) and SOMO35 (bottom) for 2006 at the rural (left) and urban (right) areas on the10x10 km (left columns) and 1x1 km (right columns) resolution .*

## 6.2 Combined final concentration maps

The comparison of different grid resolutions is executed also for combined final concentration maps that result from merging the separate urban and rural map. The currently used merger Equation 2.3 with currently used population density interval parameters of 100 and 500 inhbs.km$^{-2}$ are applied.

The results of Section 6.1 show that the differences between 10x10 km and 1x1 km maps are only small, except for rural SOMO35. Yet, Chapter 5 reveals the high influence of the merging method used to compose the combined final maps. Particularly at those areas where within short distances large inter-grid variations of urban and rural areas do occur. The influence of the merging resolution itself is also of influence. That is why we do not only compare the 10x10 km combined final map composed from separate urban and rural 10x10 km grid maps with a 10x10 km grid merger, or the 1x1 km combined final map composed from separate urban and rural 1x1 km grid maps with a 1x1 km grid merger. We also examine the variant with the separate 10x10 km rural and urban maps combined by means of a 1x1 km merger. Additionally examined is the performance of applying the 3x3 km floating merger (Section 3.6) on the separate rural and urban maps on both 10x10 km and 1x1 km resolution. The arguments for applying this floating aggregation are given at Section 3.6.

Contrary to Chapter 5, the comparison of the uncertainty calculation is executed at the points of the rural and urban stations and at all station points in the combined final maps without weighting only. The reason for not applying any area type or population density weighting in RMSE and MPE is that the percentages of the total rural and urban areas, as well as the percentages of the total population living in the rural and urban areas are different at the two resolutions of 1x1 km and 10x10 km. This would result in a different weighting at each resolution and leading to non-comparable uncertainties results. Nevertheless, as no weighting is included in the RMSE and MPE for the final map, the uncertainty values at the rural and urban stations have to be interpreted with care due to the unbalance between urban and rural areas and their number of station representations, as explained in detail in Section 2.5.

The comparison is executed separately for $PM_{10}$ and ozone.

**$PM_{10}$**

In Tables 6.3 and 6.4 the uncertainty indicators RMSE and MPE for combined final maps based on the separate rural and urban maps and merger in different grid resolutions.

*Table 6.3 Uncertainty estimates, expressed as RMSE (in $\mu g.m^{-3}$), of the interpolation results for the $PM_{10}$ indicators annual average and 36$^{th}$ maximum daily value for 2006 at rural and urban station points, and at all station points for the final European map (with no weighting criteria) and constructed from the separate rural and urban maps and merger at different grid resolutions. RMSE should be as small as possible.*

| $PM_{10}$ | | annual average | | | 36$^{th}$ maximum daily value | | |
|---|---|---|---|---|---|---|---|
| | | rural stations | urban stations | final map no weight | rural stations | urban stations | final map no weight |
| A (10-10-10) | 10x10 sep. maps, 10x10 merger | 5.43 | 7.54 | 7.14 | 9.02 | 14.05 | 13.12 |
| B (10-fl3-1) | 10x10 sep. m., 3x3 float. merger | 5.78 | 7.17 | 6.90 | 9.35 | 13.47 | 12.69 |
| C (10-1-1) | 10x10 separate, 1x1 merger | 5.65 | 7.24 | 6.93 | 9.11 | 13.58 | 12.75 |
| D (1-1-1) | 1x1 separate maps, 1x1 merger | 5.62 | 7.21 | 6.90 | 9.57 | 13.45 | 12.71 |
| E (1-fl3-1) | 1x1 sep. m., 3x3 floating merger | 5.71 | 7.16 | 6.87 | 9.73 | 13.38 | 12.68 |

*Table 6.4 Uncertainty estimates, expressed as MPE (in $\mu g.m^{-3}$), of the interpolation results for the $PM_{10}$ indicators annual average and 36$^{th}$ maximum daily value for 2006 at rural and urban station points, and at all stations points for the final European map (with no weighting criteria) and constructed from the separate rural and urban maps and merger at different grid resolutions. MPE should be as close to zero as possible.*

| $PM_{10}$ | | annual average | | | 36$^{th}$ maximum daily value | | |
|---|---|---|---|---|---|---|---|
| | | rural stations | urban stations | final map no weight | rural stations | urban stations | final map no weight |
| A (10-10-10) | 10x10 sep. maps, 10x10 merger | 0.96 | -1.47 | -0.95 | 1.51 | -2.58 | -1.70 |
| B (10-fl3-1) | 10x10 sep. m., 3x3 float. merger | 1.26 | -0.50 | -0.12 | 2.08 | -0.97 | -0.31 |
| C (10-1-1) | 10x10 separate, 1x1 merger | 1.00 | -0.64 | -0.29 | 1.70 | -1.23 | -0.60 |
| D (1-1-1) | 1x1 separate maps, 1x1 merger | 0.85 | -0.54 | -0.24 | 1.44 | -1.08 | -0.53 |
| E (1-fl3-1) | 1x1 sep. m., 3x3 floating merger | 1.09 | -0.43 | -0.10 | 1.74 | -0.86 | -0.30 |

Table 6.3 demonstrates that applying a finer grid resolution for interpolation of the separate maps and to a larger extent for the merging provides on the one hand a reduced uncertainty (i.e. reduced RMSE) at the urban stations and at all the stations for the combined final maps (without weighting criteria considered). On the other hand an increased uncertainty occurs in the final combined maps at the rural stations, despite the slightly lower uncertainty found at Table 6.1 for the rural interpolation map itself at the finer 1x1 km resolution than at the interpolation on 10x10 km. The higher uncertainty obtained at merging on the higher 1x1 km resolution can be explained by the fact that at this finer resolution more rural stations are being sited in areas with a high or moderate population density. It appears that the rural stations sited in areas with population density classified as mixed at the 10x10 km resolution, but classified as urban at the 1x1 km resolution (see Table 5.1) play the major role in this slightly larger uncertainty.

Table 6.4 shows MPE results of a comparable order of magnitude at the rural stations at each of the different resolutions (with the exception of 3x3 km floating merger executed at the 10x10 km separate maps). At the urban stations the finer resolution provides considerable improvements: the underestimation of urban stations reduces by about or more than a half against relative limited worsening of the MPE at the rural stations. This improvement is also considerable compared to those in Table 5.16 with its alternative class boundary setting, where in all cases the improvement of the MPE at urban stations brings considerable more worsening of the MPE at rural stations then at the changed merging resolutions of Table 6.4. We can conclude that the use of a finer 1x1 km grid resolution for merging is more suitable for solving the issue of the small cities in the lower or moderately populated areas than using an alternative population density class boundary (Section 5.3.1).

Furthermore, does it seem that the resolution of the merger plays the major role in the efforts of reducing the uncertainty: the results of applying the 1x1 km merger (and also 3x3 km floating merger) are almost the same at both the 10x10 km and 1x1 km separate maps, especially concerning the RMSE. Therefore, and due to the computational demands of the interpolation calculations on the 1x1 km grid, it seems worth to apply the finer resolution for the merger only, not for the separate maps.

The application of the 3x3 km floating merger shows in comparison with the 1x1 km grid merger improved results at the urban stations and all stations of the combined final map, but worsened results at the rural stations for both RMSE and MPE. Since the uncertainty reductions are not unambiguous positive at both area types we prefer not using this 3x3 km floating grid merger, but just the standard EEA 1x1 km grid resolution.

The conclusion is to recommend the use for $PM_{10}$of 1x1 km grid resolution as merger from now on as part of the routine mapping methodology.

**Ozone**

Similar analysis is executed for ozone and the uncertainty results are presented in Table 6.5 for RMSE and in Table 6.6 for MPE.

*Table 6.5 Uncertainty estimates, expressed as RMSE (in µg.m$^{-3}$, resp. µg.m$^{-3}$.d) of the interpolation results for the ozone indicators 26$^{th}$ highest daily maximum 8-hour mean and SOMO35 for 2006 at rural and urban station points, and at all station points for the final European map (with no weighting criteria) and constructed from the separate rural and urban maps and merger at different grid resolutions.*

| Ozone | | 26$^{th}$ highest daily max. 8-h | | | SOMO35 | | |
|---|---|---|---|---|---|---|---|
| | | rural stations | urban stations | final map no weight | rural stations | urban stations | final map no weight |
| A (10-10-10) | 10x10 sep. maps, 10x10 merger | 11.02 | 10.74 | 10.84 | 2062 | 1717 | 1840 |
| B (10-fl3-1) | 10x10 sep. m., 3x3 float. merger | 10.95 | 10.22 | 10.47 | 2009 | 1401 | 1631 |
| C (10-1-1) | 10x10 separate, 1x1 merger | 10.77 | 10.34 | 10.49 | 1996 | 1393 | 1622 |
| D (1-1-1) | 1x1 separate maps, 1x1 merger | 10.66 | 10.44 | 10.51 | 1924 | 1562 | 1693 |
| E (1-fl3-1) | 1x1 sep. m., 3x3 floating merger | 10.80 | 10.36 | 10.51 | 1927 | 1524 | 1671 |

*Table 6.6 Uncertainty estimates, expressed as MPE (in µg.m$^{-3}$, resp. µg.m$^{-3}$.d) of the interpolation results for the ozone indicators 26$^{th}$ highest daily maximum 8-hour mean (left) and SOMO35 (right) for 2006 at rural and urban station points, and at all station points for the final European map (with no weighting criteria) and constructed from the separate rural and urban maps and merger at different grid resolutions.*

| Ozone | | 26$^{th}$ highest daily max. 8-h | | | SOMO35 | | |
|---|---|---|---|---|---|---|---|
| | | rural stations | urban stations | final map no weight | rural stations | urban stations | final map no weight |
| A (10-10-10) | 10x10 sep. maps, 10x10 merger | -0.19 | 1.40 | 0.86 | -149 | 342 | 176 |
| B (10-fl3-1) | 10x10 sep. m., 3x3 float. merger | -0.44 | 0.48 | 0.17 | -215 | -5 | -76 |
| C (10-1-1) | 10x10 separate, 1x1 merger | -0.36 | 0.71 | 0.35 | -212 | -9 | -77 |
| D (1-1-1) | 1x1 separate maps, 1x1 merger | -0.27 | 0.58 | 0.29 | -181 | 130 | 25 |
| E (1-fl3-1) | 1x1 sep. m., 3x3 floating merger | -0.35 | 0.39 | 0.14 | -174 | 93 | 3 |

Applying a finer resolution at both interpolation and merging provides a reduced uncertainty (i.e. reduced RMSE) of about 3-5 % at both rural and urban areas in the case of 26$^{th}$ highest daily 8-hour maximum, as well as for rural SOMO35, and for urban SOMO35 it is even about 10%. The observed improvements of the RMSE at both area types is consistent with improvements obtained at Table 6.2 with the comparison of interpolation of the separate maps on a 1x1 km versus a 10x10 km grid resolution.

In the case of MPE, the finer resolution brings some reduced uncertainty at the urban stations and slight worsening at the rural stations.

As with $PM_{10}$, the resolution of the merger plays the major role in the uncertainty reductions. Next to this, but contrary to $PM_{10}$, also the finer resolution of the separate rural ozone map brings improved (i.e. reduced) RMSE. Not for MPE. As motivated in Section 6.1, the more pronounced correlation between the fine resolutions of the altitude field with the ozone field, specifically at SOMO35, plays a role. Nevertheless, the uncertainty reductions involved at applying a 1x1 km merger at both the 10x10 and 1x1 km gridded separate maps are quite similar at both the rural and urban stations. This leads to the same conclusion as at $PM_{10}$: we advice to apply the finer resolution for the merger only, not for the separate maps.

The application of the 3x3 km floating merger shows quite similar results as at the 1x1 merger: in most cases both the RMSE and MPE are slightly better at the urban station points and slightly worse at the rural stations. Like at $PM_{10}$, for ozone the uncertainty reductions obtained with this 3x3 km floating merger do not significantly improve at both the rural and the urban stations. Therefore, we prefer as merger improvement the application of just the standard EEA 1x1 km grid resolution.

From the comparison of the uncertainties of the 1x1 km merger with those of the alternative population density class boundaries (Tables 5.16 and 5.17), one can conclude that the use of a merger on a finer grid resolution results in more uncertainty reductions than the use of alternative class boundaries (Section 5.3.1). This leads into the recommendation to use for ozone the 1x1 km grid resolution as merger from now on as part of the routine mapping methodology.

## 6.3 Population exposure

In addition to the uncertainty analysis of the separate and combined final concentration maps, this section examines the influence of the grid resolution on the population exposure estimates for both the population weighted concentration (i.e. concentration per inhabitant) and the population exposure classes (specifically to population living in the areas in exceedance of limit or target values). See this chapter's introduction and Section 2.4 for details on the combinations examined.

#### *Population weighted concentration*

Tables 6.7 and 6.8 present for $PM_{10}$ and ozone the resulting national and European population weighted concentrations obtained when applying these five different combinations A through D and F.

In Table 6.7 the highest $PM_{10}$ population weighted concentrations are obtained at merging with the 1x1 km resolution (combinations C and D). This was to be expected when it is generally assumed that $PM_{10}$ shows higher concentration levels in urbanised areas due to higher human activities (industry, domestic heating and traffic) leading to higher particulate matter emissions compared to the neighbouring rural areas. The finer resolution does distinguish the population density distribution more accurately; it better accounts for the population in the small cities and towns in mainly rural areas (Section 3.6). Comparable results are provided by Table 6.8 for ozone, but then with the lowest values at C and D. That is due to opposite characteristics of ozone: ozone concentrations in urban areas are considered to be lower than in neighbouring rural areas due to the higher ozone depleting effect of the higher nitrogen oxide emissions in urbanised areas.

Interesting is that the combinations C and D at both the $PM_{10}$ and ozone indicators provide almost the same results, despite their different resolutions in the separate input maps. It indicates that for determining the national population weighted concentrations mainly the resolution of the merger plays a role and not the resolution of the separate rural and urban maps. This is in agreement with the findings at Section 6.2. It leads to more realistic and more accurate European exposure estimates. At both $PM_{10}$ indicators combination C *(10-1-1)* shows, compared to the current default methodology, represented by combination A *(10-10-10)*, an increased European population averaged exposure of about 5 %. For ozone it decreases of about 1 % for 26$^{th}$ highest daily 8-hour maximum and almost 6 % for SOMO35.

The combination B which uses the 3x3 km floating merger shows in comparison with C and D for $PM_{10}$ slightly lower values and for ozone slightly higher values, both at a very strong correlation.

The lowest population weighted concentrations at both $PM_{10}$ indicators and the highest values at both ozone indicators are obtained at combination F, where separate 1x1 km maps are merged on a 1x1 km resolution into a combined final 1x1 km map which is subsequently spatial averaged into a 10x10 km resolution map. Comparison of the population weighted concentration results of combinations D and F shows which considerable adverse 'dilution'-effect the influence of the spatial aggregation from 1x1 km into 10x10 km has on the population weighted concentrations. The spatial averaging results for $PM_{10}$ in a decreased European population weighted concentration of about 10 % for both the annual average and the $36^{th}$ maximum daily average. Simultaneously, for ozone it results in an increased population weighted concentration of 2.5 % for the $26^{th}$ highest daily 8-hour maximum average and about 12 % for SOMO35. Spatial aggregations of the final concentration map should therefore at all time be avoided in air quality impact assessments.

The scatter plots in Figure 6.3 for $PM_{10}$ and Figure 6.4 for ozone are presented, where we compare the national population weighted concentrations as obtained by each resolution combination. In all cases, Andorra has been excluded as an outlier based on linear regression, as well as Monaco that does not result in data at all grid resolutions. Combination D, selected in Section 6.2 as one of the best estimators, is plotted against the other combinations A, B, C and F. The plots confirm the findings at Tables 6.7 and 6.8 quite systematically for all countries in the same linear order of magnitude: the method fit for determining the population weighted concentrations can consist of preparing the separate maps at just the 10x10 km resolution, if they are subsequently combined by the 1x1 km grid merger into a final concentration map on a 1x1 km resolution with successive population weighting on that resolution.

## $PM_{10}$

*Table 6.7 Population weighted concentration (in µg.m$^{-3}$) for $PM_{10}$ indicators annual average (left) and 36th maximum daily mean (right), for year 2006, for five different grid resolution combinations.*

| Country | $PM_{10}$ – annual average | | | | | $PM_{10}$ – 36$^{th}$ highest daily mean | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A. 10 - 10 - 10 | B. 10 - fl.3 - 1 | C. 10 - 1 - 1 | D. 1 - 1 - 1 | F. aggr.D-10 | A. 10 - 10 - 10 | B. 10 - fl.3 - 1 | C. 10 - 1 - 1 | D. 1 - 1 - 1 | F. aggr.D-10 |
| Albania | 30.3 | 31.5 | 31.8 | 31.8 | 27.5 | 51.3 | 53.5 | 54.0 | 53.5 | 46.1 |
| Andorra | 18.1 | 22.5 | 22.5 | 21.6 | 10.6 | 29.5 | 35.7 | 35.7 | 34.9 | 18.7 |
| Austria | 23.8 | 25.2 | 26.0 | 26.1 | 22.9 | 43.2 | 45.6 | 47.1 | 47.3 | 41.7 |
| Belgium | 30.9 | 31.1 | 31.3 | 31.2 | 30.1 | 50.8 | 51.1 | 51.3 | 51.2 | 50.0 |
| Bosnia-Herzeg. | 31.4 | 32.6 | 33.1 | 33.1 | 29.2 | 54.3 | 56.5 | 57.4 | 57.6 | 50.5 |
| Bulgaria | 36.7 | 40.3 | 41.6 | 41.8 | 31.9 | 64.5 | 71.7 | 74.2 | 74.5 | 55.2 |
| Croatia | 29.9 | 31.0 | 31.5 | 31.6 | 29.2 | 50.9 | 52.9 | 53.7 | 53.8 | 49.2 |
| Cyprus | 34.0 | 35.0 | 35.4 | 35.9 | 33.1 | 54.8 | 57.0 | 58.2 | 57.8 | 50.4 |
| Czech Republic | 31.7 | 33.0 | 33.5 | 33.5 | 30.6 | 54.6 | 56.7 | 57.5 | 57.7 | 53.2 |
| Denmark | 22.3 | 23.2 | 23.5 | 23.6 | 21.5 | 35.2 | 36.5 | 37.0 | 37.5 | 34.2 |
| Estonia | 18.8 | 19.6 | 19.7 | 19.7 | 17.4 | 32.5 | 34.0 | 34.1 | 34.4 | 30.6 |
| Finland | 16.6 | 16.9 | 17.0 | 16.6 | 16.0 | 28.5 | 29.2 | 29.5 | 29.5 | 28.4 |
| France | 20.1 | 20.3 | 20.4 | 20.3 | 19.6 | 32.6 | 32.8 | 32.9 | 32.7 | 31.7 |
| Germany | 23.1 | 23.8 | 24.2 | 24.3 | 22.6 | 39.7 | 40.7 | 41.3 | 41.5 | 38.8 |
| Greece | 31.7 | 33.0 | 33.6 | 33.6 | 29.8 | 51.1 | 53.3 | 54.3 | 54.0 | 47.8 |
| Hungary | 32.5 | 32.8 | 32.9 | 33.0 | 32.3 | 57.7 | 58.3 | 58.5 | 58.3 | 56.9 |
| Iceland | 15.4 | 17.0 | 17.4 | 17.4 | 13.2 | 23.4 | 26.6 | 27.2 | 29.1 | 21.7 |
| Ireland | 13.8 | 14.6 | 14.9 | 15.1 | 13.2 | 21.5 | 23.6 | 24.1 | 24.9 | 20.6 |
| Italy | 32.8 | 33.6 | 33.9 | 33.9 | 31.2 | 56.9 | 58.2 | 58.6 | 58.7 | 54.1 |
| Latvia | 21.5 | 21.9 | 21.9 | 22.1 | 20.9 | 39.0 | 39.9 | 40.0 | 40.7 | 38.3 |
| Liechtenstein | 21.6 | 24.8 | 24.9 | 24.9 | 20.7 | 42.4 | 47.2 | 47.5 | 49.0 | 42.4 |
| Lithuania | 22.3 | 22.4 | 22.5 | 22.4 | 21.8 | 39.1 | 39.4 | 39.7 | 40.0 | 38.6 |
| Luxembourg | 20.3 | 20.6 | 20.8 | 21.0 | 20.6 | 35.0 | 35.5 | 35.9 | 35.9 | 34.5 |
| Macedonia FYR | 37.5 | 39.0 | 39.3 | 38.4 | 28.6 | 66.6 | 69.4 | 69.9 | 69.1 | 50.3 |
| Malta | 28.7 | 29.5 | 29.4 | 30.5 | 26.7 | 44.0 | 44.9 | 44.8 | 45.2 | 38.9 |
| Monaco | | 36.7 | 36.7 | 37.5 | | | 59.7 | 59.7 | 64.3 | |
| Montenegro | 28.2 | 32.4 | 33.1 | 32.7 | 25.1 | 49.1 | 56.7 | 57.9 | 57.6 | 43.5 |
| Netherlands | 28.9 | 29.1 | 29.1 | 29.1 | 28.0 | 45.8 | 46.1 | 46.1 | 46.2 | 44.8 |
| Norway | 18.8 | 19.6 | 19.6 | 19.4 | 17.3 | 30.3 | 31.8 | 31.9 | 33.0 | 29.7 |
| Poland | 34.4 | 36.2 | 37.0 | 37.0 | 32.0 | 59.4 | 62.8 | 64.0 | 64.1 | 55.2 |
| Portugal | 27.3 | 28.1 | 28.4 | 28.8 | 27.0 | 46.1 | 47.7 | 48.3 | 48.8 | 44.7 |
| Romania | 36.2 | 37.8 | 39.1 | 39.3 | 33.6 | 61.0 | 63.5 | 65.4 | 65.5 | 57.3 |
| San Marino | 27.7 | 34.3 | 33.9 | 33.6 | 25.7 | 46.7 | 58.1 | 57.4 | 57.3 | 43.0 |
| Serbia | 38.7 | 40.8 | 41.8 | 41.2 | 32.8 | 67.3 | 71.3 | 73.1 | 72.4 | 56.6 |
| Slovakia | 30.8 | 32.7 | 33.8 | 33.7 | 29.0 | 53.0 | 56.6 | 58.5 | 58.3 | 49.5 |
| Slovenia | 26.9 | 28.3 | 29.0 | 29.2 | 25.9 | 45.6 | 48.0 | 49.2 | 49.4 | 43.4 |
| Spain | 28.0 | 30.9 | 31.4 | 31.6 | 24.2 | 44.7 | 48.6 | 49.3 | 49.6 | 39.3 |
| Sweden | 17.4 | 18.5 | 19.0 | 18.8 | 16.7 | 28.6 | 31.0 | 32.0 | 32.2 | 28.2 |
| Switzerland | 21.8 | 23.0 | 23.2 | 23.2 | 20.6 | 41.4 | 43.5 | 43.9 | 43.8 | 39.3 |
| United Kingdom | 22.5 | 23.2 | 23.2 | 23.3 | 20.8 | 34.4 | 35.5 | 35.5 | 35.7 | 32.1 |
| **Total** | **27.1** | **28.1** | **28.5** | **28.6** | **25.5** | **45.4** | **47.1** | **47.8** | **47.9** | **42.9** |

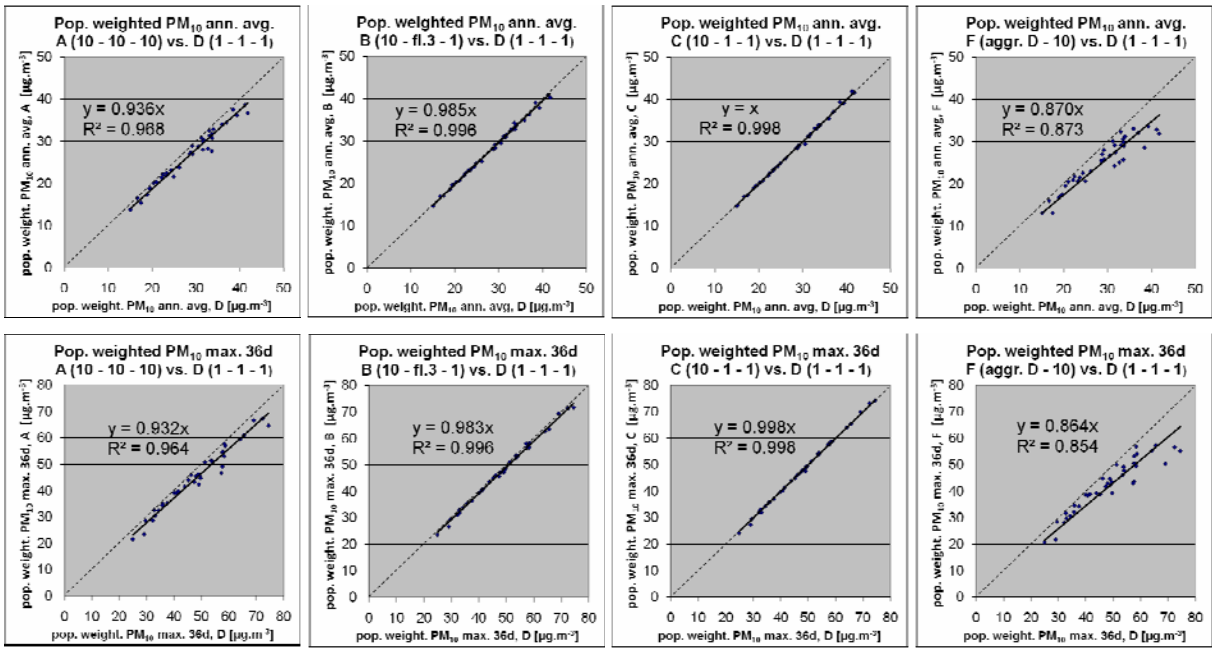


*Figure 6.3 Scatter-plots showing the correlation between the population weighted concentration as result of grid resolution combination D with about the highest estimates (x-axis), against the results of combinations A (left), B (centre left), C (centre right), and F (right) (y-axis) for $PM_{10}$ annual average (top) and $36^{th}$ maximum daily mean (bottom), across the European countries. Slope and $R^2$ should be as close to 1 as possible.*

## Ozone

*Table 6.8 Population weighted concentration (in µg.m$^{-3}$ resp. µg.m$^{-3}$.d) for ozone indicators 26$^{th}$ highest daily maximum 8-hour (left) and SOMO35 (right), for year 2006, for five different grid resolution variants.*

| Country | Ozone – 26$^{th}$ highest daily 8-h maximum | | | | | Ozone – SOMO35 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A. 10 - 10 - 10 | B. 10 - fl.3 - 1 | C. 10 - 1 - 1 | D. 1 - 1 - 1 | F. aggr.D-10 | A. 10 - 10 - 10 | B. 10 - fl.3 - 1 | C. 10 - 1 - 1 | D. 1 - 1 - 1 | F. aggr.D-10 |
| Albania | 119.3 | 118.2 | 117.9 | 117.4 | 121.8 | 7632 | 7273 | 7193 | 7307 | 8653 |
| Andorra | 123.7 | 119.1 | 119.1 | 120.1 | 133.3 | 7872 | 6587 | 6587 | 6784 | 10434 |
| Austria | 127.6 | 125.9 | 124.9 | 124.7 | 128.4 | 6975 | 6500 | 6237 | 6172 | 7171 |
| Belgium | 126.1 | 126.1 | 126.0 | 126.0 | 126.5 | 4066 | 4046 | 4017 | 4018 | 4145 |
| Bosnia-Herzeg. | 119.8 | 118.5 | 118.1 | 117.3 | 120.6 | 7120 | 6722 | 6571 | 6281 | 7352 |
| Bulgaria | 107.9 | 105.7 | 105.0 | 105.0 | 110.8 | 5611 | 5076 | 4896 | 4851 | 6244 |
| Croatia | 126.3 | 125.2 | 124.8 | 124.3 | 126.1 | 7337 | 7053 | 6928 | 6728 | 7233 |
| Cyprus | 105.0 | 103.1 | 102.1 | 102.1 | 109.0 | 6368 | 5979 | 5759 | 5660 | 7273 |
| Czech Republic | 127.5 | 126.7 | 126.5 | 126.1 | 127.5 | 6471 | 6196 | 6097 | 5999 | 6448 |
| Denmark | 108.6 | 105.8 | 104.9 | 104.9 | 110.6 | 4086 | 3711 | 3578 | 3583 | 4397 |
| Estonia | 106.4 | 105.2 | 105.1 | 105.3 | 107.0 | 3837 | 3621 | 3594 | 3512 | 4007 |
| Finland | 102.6 | 101.2 | 100.7 | 100.2 | 103.0 | 3454 | 3226 | 3141 | 3060 | 3529 |
| France | 122.7 | 122.2 | 122.0 | 121.8 | 123.1 | 5117 | 5016 | 4972 | 4948 | 5221 |
| Germany | 127.0 | 126.2 | 125.8 | 125.6 | 127.8 | 5104 | 4944 | 4860 | 4820 | 5207 |
| Greece | 117.3 | 116.3 | 115.8 | 114.7 | 117.4 | 7020 | 6782 | 6657 | 6635 | 7269 |
| Hungary | 123.4 | 122.1 | 121.7 | 121.6 | 124.1 | 6117 | 5842 | 5738 | 5654 | 6123 |
| Iceland | 95.4 | 93.7 | 93.3 | 93.9 | 98.3 | 2605 | 2327 | 2265 | 2419 | 3229 |
| Ireland | 91.3 | 90.4 | 90.2 | 90.0 | 92.3 | 2598 | 2484 | 2453 | 2370 | 2678 |
| Italy | 137.1 | 135.5 | 135.1 | 134.7 | 138.6 | 8678 | 8306 | 8205 | 8264 | 9145 |
| Latvia | 106.1 | 104.6 | 104.5 | 103.8 | 107.3 | 4047 | 3761 | 3734 | 3636 | 4273 |
| Liechtenstein | 131.1 | 127.4 | 127.3 | 126.2 | 130.9 | 8179 | 6306 | 6258 | 6045 | 7935 |
| Lithuania | 110.6 | 110.4 | 110.1 | 109.6 | 111.1 | 4678 | 4621 | 4535 | 4440 | 4816 |
| Luxembourg | 130.8 | 130.3 | 130.0 | 130.0 | 131.6 | 5279 | 5150 | 5090 | 5010 | 5546 |
| Macedonia FYR | 111.6 | 110.4 | 110.3 | 110.2 | 119.5 | 6627 | 6316 | 6297 | 6472 | 9007 |
| Malta | 116.8 | 115.5 | 115.6 | 115.8 | 120.7 | 7973 | 7784 | 7797 | 7898 | 8599 |
| Monaco | | 142.4 | 142.4 | 132.0 | | | 8903 | 8903 | 7541 | |
| Montenegro | 118.7 | 114.8 | 114.3 | 114.2 | 121.2 | 7802 | 6725 | 6554 | 6580 | 8622 |
| Netherlands | 116.1 | 116.1 | 116.1 | 116.2 | 116.7 | 3245 | 3245 | 3245 | 3259 | 3351 |
| Norway | 103.1 | 101.7 | 101.7 | 101.7 | 105.0 | 3750 | 3512 | 3496 | 3447 | 4128 |
| Poland | 122.2 | 120.9 | 120.4 | 120.3 | 123.8 | 5797 | 5524 | 5416 | 5391 | 6135 |
| Portugal | 121.0 | 119.9 | 119.4 | 119.3 | 123.2 | 5534 | 5337 | 5257 | 5247 | 5951 |
| Romania | 108.1 | 106.8 | 105.7 | 105.8 | 110.4 | 5330 | 5023 | 4798 | 4745 | 5734 |
| San Marino | 128.2 | 120.2 | 120.8 | 120.4 | 129.3 | 7841 | 6215 | 6321 | 6115 | 7881 |
| Serbia | 110.5 | 109.2 | 108.5 | 108.6 | 114.1 | 5724 | 5395 | 5239 | 5284 | 6586 |
| Slovakia | 124.6 | 123.1 | 122.2 | 122.2 | 125.7 | 6967 | 6506 | 6261 | 6220 | 7295 |
| Slovenia | 134.2 | 133.0 | 132.6 | 131.7 | 133.2 | 8124 | 7670 | 7480 | 7287 | 8081 |
| Spain | 118.0 | 116.5 | 116.2 | 116.4 | 121.1 | 6174 | 5875 | 5813 | 5824 | 6716 |
| Sweden | 108.1 | 105.5 | 104.5 | 104.3 | 109.0 | 4146 | 3773 | 3635 | 3605 | 4286 |
| Switzerland | 133.4 | 132.9 | 132.6 | 132.1 | 135.3 | 6701 | 6414 | 6321 | 6295 | 7428 |
| United Kingdom | 98.8 | 98.1 | 98.0 | 98.0 | 101.2 | 2799 | 2687 | 2676 | 2686 | 3126 |
| **Total** | **119.6** | **118.5** | **118.2** | **118.0** | **121.1** | **5485** | **5252** | **5167** | **5155** | **5785** |

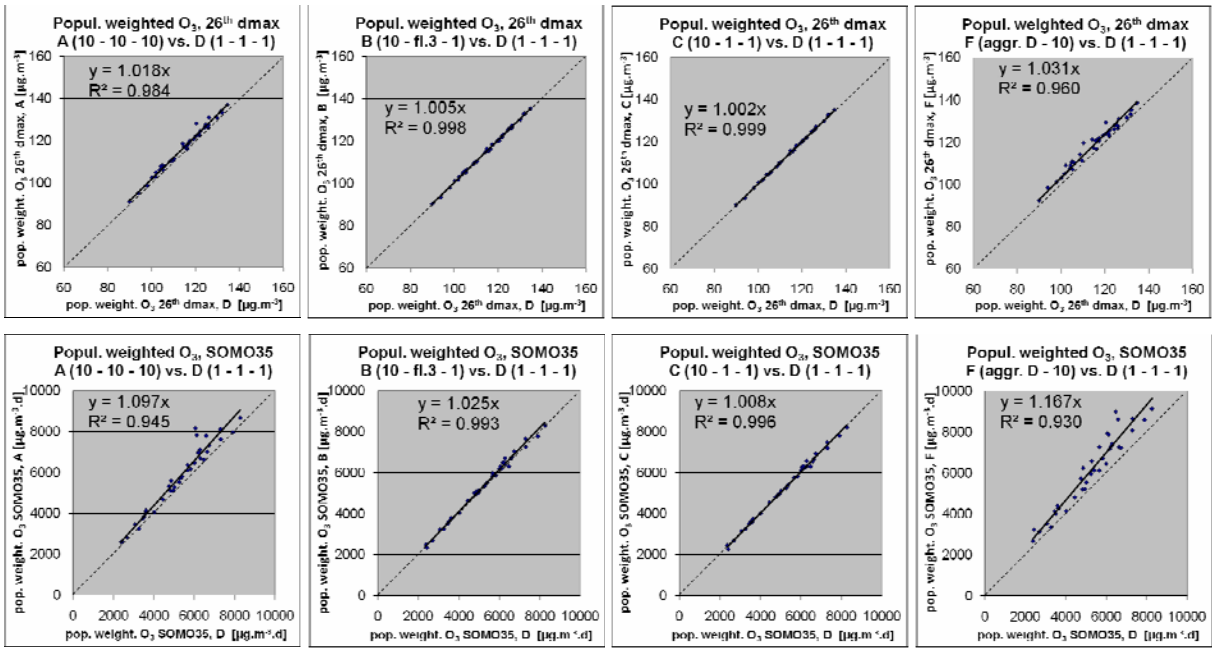


*Figure 6.4 Scatter-plots showing the correlation between the population weighted concentration as result of grid resolution combination D with about the lowest estimates (x-axis), against the results of combinations A (left), B (centre left), C (centre right) and F (right) (y-axis) for ozone indicators 26th highest daily maximum 8-hour value (top) and SOMO35 (bottom), across the European countries. Slope and $R^2$ should be as close to 1 as possible.*

### *Population exposure classes*

At examining the influence of the grid resolution on the population exposure class distributions, specific interest is there in the apportionment to classes of high concentrations, particularly those above the limit or target values.

Tables 6.9 and 6.10 present for $PM_{10}$ and Tables 6.11 and 6.12 for ozone the percentages of the population exposed to concentrations above the limit or target values for individual countries and for Europe as a whole. Combinations C and D provide quite similar results with more population exposed to $PM_{10}$ exceedances and less population exposed to ozone exceedances than at combination A and F. This higher $PM_{10}$ and lower ozone levels of apportionment of population in exceedance can be explained by the more accurate estimate of the population numbers when using the 1x1 km population density grid for merging at combinations C and D instead the 10x10 km population density grid used to merge at combination A or the aggregation at F. The fact that the percentages at C and D, and to a less extend at B, are of the same order of magnitude at most of the individual countries proves that the grid resolution of the merger plays a significant and a dominant role in the estimations, and not the resolution of the separate interpolated rural and urban maps.

The differences in the exceedances at the current default methodology, combination A *(10-10-10),* and the recommended combination C *(10-1-1)* show clearly that implementing the refined merging resolution does improve the accounting of the population of smaller cities and towns in predominantly rural areas. It leads to more realistic and more accurate European exposure estimates. The refined merger used at combination C shows at $PM_{10}$ an increased exceedance exposure of about 37 % for annual average and about 25 % for 36th highest daily mean. For ozone the decrease is some 7.5 % for 26th highest daily maximum 8-hour and some 21 % for SOMO35.

The lowest exposure exceedances at both $PM_{10}$ indicators and the highest at both ozone indicators are obtained at combination F with its 1x1 km separate maps merged on 1x1 km into a 1x1 km final map that is successively spatial averaged into a 10x10 km concentration map. Comparing its exceedance percentages with those of combination D shows the substantial adverse 'thinning'-effect the spatial aggregation from 1x1 km into 10x10 km has on the exposure estimates. The spatial averaging of the final concentration maps in air quality impact assessments should be avoid at all times, as it results for

$PM_{10}$ in a decreased European exceedance of about 56 % for the annual average and 32.5 % for the 36th maximum daily average. Simultaneously, for ozone it results in an increased exceedance of about 15% for the 26th highest daily 8-hour maximum average and about 50% for SOMO35.

### $PM_{10}$

*Table 6.9 Percentage of national population living in exceedance of $PM_{10}$ indicator annual average (Limit Value = 40 µg.m$^{-3}$) in 2006 at the five different grid resolution combinations A, B, C, D and F.*

| Country | Popul. x 1000 | $PM_{10}$ – annual average | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A. 10 - 10 - 10 | | B. 10 - fl.3 - 1 | | C. 10 - 1 - 1 | | D. 1 - 1 - 1 | | F. aggr.D - 10 | |
| | | 40 - 45 | > 45 | 40 - 45 | > 45 | 40 - 45 | > 45 | 40 - 45 | > 45 | 40 - 45 | > 45 |
| Albania | 3 961 | 1.1 | - | 2.1 | - | 3.0 | 0.1 | 2.7 | - | - | - |
| Andorra | 61 | - | - | - | - | - | - | - | - | - | - |
| Austria | 8 217 | - | - | - | - | - | - | - | - | - | - |
| Belgium | 10 594 | - | - | - | - | - | - | - | - | - | - |
| Bosnia-Herzeg. | 4 203 | 5.1 | - | 6.0 | - | 6.9 | - | 4.4 | - | - | - |
| Bulgaria | 8 011 | 4.8 | 27.5 | 3.7 | 40.9 | 4.6 | 45.3 | 4.6 | 45.1 | 10.1 | 6.4 |
| Croatia | 4 400 | - | - | 0.0 | - | 0.1 | - | - | - | - | - |
| Cyprus | 836 | - | - | - | - | - | - | 14.3 | - | 10.5 | - |
| Czech Republic | 10 163 | 3.2 | 7.4 | 4.5 | 8.3 | 5.0 | 8.8 | 4.8 | 8.7 | 2.4 | 6.2 |
| Denmark | 5 423 | - | - | - | - | - | - | - | - | - | - |
| Estonia | 1 364 | - | - | - | - | - | - | - | - | - | - |
| Finland | 5 184 | - | - | - | - | - | - | - | - | - | - |
| France | 58 437 | - | - | - | - | - | - | - | - | - | - |
| Germany | 82 028 | - | - | - | - | - | - | - | - | - | - |
| Greece | 10 876 | 0.4 | - | 2.6 | - | 3.6 | 0.1 | 3.1 | 0.0 | - | - |
| Hungary | 10 125 | 1.0 | - | 1.9 | - | 2.2 | - | 3.6 | - | - | - |
| Iceland | 183 | - | - | - | - | - | - | - | - | - | - |
| Ireland | 3 734 | - | - | - | - | - | - | - | - | - | - |
| Italy | 56 675 | 7.6 | 14.6 | 8.0 | 15.6 | 8.2 | 16.0 | 8.0 | 15.5 | 6.9 | 10.6 |
| Latvia | 2 390 | - | - | - | - | - | - | - | - | - | - |
| Liechtenstein | 35 | - | - | - | - | - | - | - | - | - | - |
| Lithuania | 3 479 | - | - | - | - | - | - | - | - | - | - |
| Luxembourg | 427 | - | - | - | - | - | - | - | - | - | - |
| Macedonia FYR | 2 297 | 16.2 | 35.5 | 17.7 | 42.9 | 17.4 | 43.9 | 17.6 | 40.3 | 17.5 | - |
| Malta | 395 | - | - | - | - | - | - | - | - | - | - |
| Monaco | 32 | | | - | - | - | - | 3.3 | - | | |
| Montenegro | 724 | - | - | 7.8 | - | 9.2 | 0.4 | 8.8 | - | - | - |
| Netherlands | 15 746 | - | - | - | - | - | - | - | - | - | - |
| Norway | 3 226 | - | - | - | - | - | - | - | - | - | - |
| Poland | 38 237 | 5.9 | 15.2 | 8.0 | 17.9 | 8.8 | 19.7 | 8.7 | 19.6 | 5.0 | 11.5 |
| Portugal | 9 910 | - | - | - | - | - | - | 0.5 | - | - | - |
| Romania | 22 341 | 10.4 | 21.4 | 12.9 | 27.2 | 16.2 | 30.8 | 14.9 | 33.2 | 6.8 | 7.1 |
| San Marino | 20 | - | - | - | - | - | - | - | - | - | - |
| Serbia | 10 821 | 10.7 | 38.7 | 13.5 | 46.4 | 14.3 | 51.7 | 17.1 | 46.5 | 13.4 | - |
| Slovakia | 5 295 | 1.1 | - | 7.8 | 4.6 | 9.9 | 6.4 | 9.6 | 7.4 | - | - |
| Slovenia | 2 047 | - | - | - | - | - | - | - | - | - | - |
| Spain | 39 046 | 2.4 | 0.2 | 5.4 | 1.1 | 6.3 | 1.2 | 5.4 | 1.1 | - | - |
| Sweden | 8 898 | - | - | - | - | - | - | - | - | - | - |
| Switzerland | 7 270 | - | - | 0.9 | - | 0.9 | - | - | - | - | - |
| United Kingdom | 59 051 | - | - | - | - | - | - | - | - | - | - |
| **Total** | **516 162** | **2.4** | **5.2** | **3.2** | **6.2** | **3.6** | **6.8** | **3.5** | **6.8** | **2.0** | **2.5** |

*Table 6.10 Percentage of national population living in exceedance of $PM_{10}$ indicator 36th maximum daily mean (Limit Value = 50 µg.m$^{-3}$) in 2006 at the five different grid resolution combinations A, B, C, D and F.*

| Country | Popul. x 1000 | $PM_{10}$ – 36th highest daily mean | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A. 10 - 10 - 10 | | B. 10 - fl.3 - 1 | | C. 10 - 1 - 1 | | D. 1 - 1 - 1 | | F. aggr.D - 10 | |
| | | 50 - 65 | > 65 | 50 - 65 | > 65 | 50 - 65 | > 65 | 50 - 65 | > 65 | 50 - 65 | > 65 |
| Albania | 3 961 | 58.7 | 4.5 | 57.6 | 11.8 | 56.9 | 13.7 | 57.8 | 13.3 | 44.0 | - |
| Andorra | 61 | - | - | - | - | - | - | - | - | - | - |
| Austria | 8 217 | 31.2 | - | 38.3 | - | 43.9 | - | 44.1 | - | 26.7 | - |
| Belgium | 10 594 | 69.2 | - | 72.0 | - | 73.1 | - | 73.3 | - | 64.4 | - |
| Bosnia-Herzeg. | 4 203 | 49.5 | 20.1 | 54.2 | 21.8 | 57.1 | 22.9 | 63.2 | 18.2 | 65.4 | - |
| Bulgaria | 8 011 | 31.4 | 34.3 | 27.9 | 49.0 | 27.3 | 54.5 | 28.6 | 54.9 | 48.2 | 16.5 |
| Croatia | 4 400 | 67.1 | - | 74.9 | 1.9 | 77.7 | 2.5 | 79.6 | 1.9 | 58.3 | - |
| Cyprus | 836 | 67.6 | - | 76.2 | - | 81.5 | - | 81.8 | - | 58.5 | - |
| Czech Republic | 10 163 | 49.3 | 13.4 | 57.1 | 15.8 | 59.7 | 16.9 | 61.1 | 16.5 | 48.0 | 11.4 |
| Denmark | 5 423 | - | - | - | - | - | - | 0.1 | - | - | - |
| Estonia | 1 364 | - | - | - | - | - | - | - | - | - | - |
| Finland | 5 184 | - | - | - | - | - | - | - | - | - | - |
| France | 58 437 | 1.6 | - | 1.6 | - | 1.7 | - | 1.1 | 0.0 | 1.1 | - |
| Germany | 82 028 | 1.1 | - | 1.6 | - | 2.0 | - | 2.0 | - | - | - |
| Greece | 10 876 | 62.1 | 2.8 | 69.7 | 4.8 | 72.3 | 6.3 | 72.7 | 6.4 | 48.3 | 0.4 |
| Hungary | 10 125 | 84.4 | 10.1 | 84.1 | 11.9 | 84.4 | 12.5 | 88.8 | 7.6 | 92.1 | - |
| Iceland | 183 | 0.1 | - | 0.1 | - | 0.1 | - | 3.7 | - | - | - |
| Ireland | 3 734 | - | - | - | - | - | - | - | - | - | - |
| Italy | 56 675 | 26.6 | 26.5 | 29.5 | 27.7 | 30.1 | 28.4 | 33.7 | 28.4 | 20.6 | 24.6 |
| Latvia | 2 390 | - | - | - | - | - | - | - | - | - | - |
| Liechtenstein | 35 | - | - | - | - | - | - | 37.3 | - | - | - |
| Lithuania | 3 479 | - | - | - | - | - | - | - | - | - | - |
| Luxembourg | 427 | - | - | - | - | - | - | - | - | - | - |
| Macedonia FYR | 2 297 | 3.0 | 64.9 | 5.7 | 67.9 | 6.3 | 68.3 | 6.4 | 67.1 | 20.5 | 17.5 |
| Malta | 395 | - | - | - | - | - | - | - | - | - | - |
| Monaco | 32 | | | 100.0 | - | 100.0 | - | 80.2 | 19.8 | | |
| Montenegro | 724 | 19.6 | 24.9 | 28.7 | 38.1 | 27.3 | 42.1 | 30.3 | 41.1 | 21.6 | - |
| Netherlands | 15 746 | 2.0 | - | 3.5 | - | 3.9 | - | 3.8 | - | 0.3 | - |
| Norway | 3 226 | - | - | - | - | - | - | 0.0 | - | - | - |
| Poland | 38 237 | 33.2 | 26.6 | 38.1 | 33.3 | 38.8 | 36.4 | 39.3 | 36.2 | 28.4 | 19.1 |
| Portugal | 9 910 | 50.0 | - | 54.4 | - | 57.2 | - | 57.4 | - | 38.8 | - |
| Romania | 22 341 | 43.2 | 39.2 | 38.2 | 49.0 | 34.5 | 56.8 | 35.1 | 56.8 | 66.8 | 15.5 |
| San Marino | 20 | - | - | 92.5 | - | 84.8 | - | 84.1 | - | - | - |
| Serbia | 10 821 | 28.4 | 52.1 | 21.2 | 64.9 | 17.1 | 70.4 | 18.4 | 68.6 | 62.0 | 15.0 |
| Slovakia | 5 295 | 56.7 | 9.7 | 58.0 | 20.4 | 58.8 | 25.0 | 58.2 | 25.2 | 51.7 | 3.0 |
| Slovenia | 2 047 | 42.8 | - | 53.8 | - | 63.3 | - | 64.3 | - | 16.9 | - |
| Spain | 39 046 | 39.9 | 0.8 | 50.7 | 2.1 | 53.1 | 2.4 | 52.2 | 2.5 | 11.9 | - |
| Sweden | 8 898 | - | - | - | - | - | - | - | - | - | - |
| Switzerland | 7 270 | 3.4 | 0.7 | 5.5 | 2.1 | 6.2 | 2.1 | 4.3 | 1.9 | 0.9 | 0.7 |
| United Kingdom | 59 051 | - | - | - | - | - | - | - | - | - | - |
| **Total** | **516 162** | **19.2** | **9.2** | **22.9** | **11.3** | **23.4** | **12.3** | **24.0** | **12.3** | **18.8** | **5.7** |

## Ozone

*Table 6.11 Percentage of population living in exceedance of ozone indicators 26th highest daily 8-hour maximum (Target Value = 120 µg.m$^{-3}$) in 2006 at the five different grid resolution combinations A, B, C, D and F.*

| Country | Popul. x 1000 | $O_3$ – 26th highest daily 8-h maximum | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A. 10 - 10 - 10 | | B. 10 - fl.3 - 1 | | C. 10 - 1 - 1 | | D. 1 - 1 - 1 | | F. aggr.D - 10 | |
| | | 120 - 140 | > 140 | 120 - 140 | > 140 | 120 - 140 | > 140 | 120 - 140 | > 140 | 120 - 140 | > 140 |
| Albania | 3 961 | 31.6 | 0.6 | 25.4 | 0.5 | 24.4 | 0.5 | 22.3 | 1.6 | 52.6 | 2.0 |
| Andorra | 61 | 33.3 | - | 27.2 | - | 26.8 | - | 21.4 | 6.6 | 94.7 | 5.3 |
| Austria | 8 217 | 93.7 | 1.7 | 87.9 | 1.1 | 84.0 | 0.8 | 86.3 | 0.5 | 98.7 | 1.3 |
| Belgium | 10 594 | 95.0 | - | 94.2 | - | 94.1 | - | 93.7 | - | 94.5 | - |
| Bosnia-Herzeg. | 4 203 | 50.7 | - | 39.0 | - | 34.9 | - | 25.3 | 0.0 | 50.6 | - |
| Bulgaria | 8 011 | 1.6 | - | 0.9 | - | 0.8 | - | 1.0 | - | 1.7 | - |
| Croatia | 4 400 | 88.2 | 0.6 | 82.0 | 0.5 | 79.2 | 0.4 | 74.8 | 0.4 | 83.4 | 0.3 |
| Cyprus | 836 | 1.8 | - | 1.7 | - | 1.2 | - | 1.5 | - | 2.0 | - |
| Czech Republic | 10 163 | 98.6 | 0.0 | 96.6 | 0.0 | 95.6 | 0.0 | 96.0 | 0.0 | 100.0 | 0.0 |
| Denmark | 5 423 | 0.3 | - | 0.3 | - | 0.2 | - | 0.1 | - | 0.0 | - |
| Estonia | 1 364 | - | - | - | - | - | - | - | - | - | - |
| Finland | 5 184 | - | - | - | - | - | - | - | - | - | - |
| France | 58 437 | 62.3 | 2.1 | 60.6 | 1.7 | 59.8 | 1.6 | 58.0 | 1.6 | 63.3 | 2.6 |
| Germany | 82 028 | 89.8 | 0.0 | 88.6 | 0.0 | 88.0 | 0.0 | 88.0 | 0.0 | 90.8 | 0.0 |
| Greece | 10 876 | 38.6 | - | 35.6 | - | 34.6 | - | 29.3 | 0.0 | 36.9 | - |
| Hungary | 10 125 | 85.2 | - | 73.3 | - | 69.3 | - | 68.7 | - | 96.4 | - |
| Iceland | 183 | - | - | - | - | - | - | - | - | - | - |
| Ireland | 3 734 | - | - | - | - | - | - | - | - | - | - |
| Italy | 56 675 | 57.1 | 36.9 | 55.2 | 34.5 | 55.1 | 33.6 | 55.4 | 33.1 | 57.0 | 40.2 |
| Latvia | 2 390 | - | - | - | - | - | - | - | - | - | - |
| Liechtenstein | 35 | 100.0 | - | 98.5 | 1.5 | 99.6 | 0.4 | 99.4 | 0.6 | 100.0 | - |
| Lithuania | 3 479 | 0.0 | - | 0.0 | - | 0.0 | - | 0.0 | - | 0.0 | - |
| Luxembourg | 427 | 100.0 | - | 100.0 | - | 100.0 | - | 100.0 | - | 100.0 | - |
| Macedonia FYR | 2 297 | 16.6 | 1.2 | 13.8 | 1.1 | 13.9 | 1.1 | 14.4 | 1.6 | 39.9 | 1.6 |
| Malta | 395 | 14.2 | - | 4.5 | - | 4.9 | - | 5.2 | - | 22.6 | - |
| Monaco | 32 | | | - | 100.0 | - | 100.0 | 99.8 | 0.2 | | |
| Montenegro | 724 | 45.2 | - | 26.2 | - | 23.7 | - | 23.0 | - | 54.9 | - |
| Netherlands | 15 746 | 38.3 | - | 38.8 | - | 38.8 | - | 36.8 | - | 35.9 | - |
| Norway | 3 226 | 0.0 | - | 0.0 | - | 0.0 | - | 0.1 | - | 0.1 | - |
| Poland | 38 237 | 66.2 | - | 56.9 | - | 53.0 | - | 53.5 | 0.0 | 79.7 | - |
| Portugal | 9 910 | 58.1 | 0.5 | 49.6 | 0.3 | 46.3 | 0.3 | 46.1 | 0.4 | 65.0 | 0.4 |
| Romania | 22 341 | 1.3 | - | 0.8 | - | 0.6 | - | 0.8 | - | 1.2 | - |
| San Marino | 20 | 100.0 | - | 21.3 | - | 22.9 | - | 19.2 | - | 100.0 | - |
| Serbia | 10 821 | 8.8 | - | 6.5 | 0.0 | 6.3 | 0.0 | 5.5 | 0.1 | 9.6 | 0.0 |
| Slovakia | 5 295 | 86.4 | - | 74.0 | - | 66.5 | - | 67.1 | 0.0 | 98.3 | - |
| Slovenia | 2 047 | 86.8 | 13.2 | 90.1 | 9.9 | 91.6 | 8.4 | 95.1 | 4.9 | 89.9 | 10.1 |
| Spain | 39 046 | 50.0 | 0.1 | 43.8 | 0.0 | 42.5 | 0.0 | 41.3 | 0.0 | 55.4 | 0.1 |
| Sweden | 8 898 | 0.3 | - | 0.2 | - | 0.1 | - | 0.1 | - | 0.1 | - |
| Switzerland | 7 270 | 92.4 | 7.6 | 93.5 | 6.5 | 93.9 | 6.1 | 94.5 | 5.4 | 91.2 | 8.8 |
| United Kingdom | 59 051 | 0.0 | - | 0.0 | - | 0.0 | - | 0.1 | - | 0.0 | - |
| **Total** | **516 162** | **51.0** | **4.5** | **48.3** | **4.2** | **47.3** | **4.1** | **46.7** | **4.0** | **53.3** | **4.9** |

*Table 6.12 Percentage of population living in high SOMO35 exposure classes (µg.m$^{-3}$.d; no LV or TV defined) in 2006 at the five different grid resolution combinations A, B, C, D and F.*

| Country | Popul. x 1000 | $O_3$ – SOMO35 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A. 10 - 10 - 10 | | B. 10 - fl.3 - 1 | | C. 10 - 1 - 1 | | D. 1 - 1 - 1 | | F. aggr.D - 10 | |
| | | 6000 - 10000 | > 10000 | 6000 - 10000 | > 10000 | 6000 - 10000 | > 10000 | 6000 - 10000 | > 10000 | 6000 - 10000 | > 10000 |
| Albania | 3 961 | 67.3 | 16.0 | 65.1 | 12.1 | 63.8 | 11.5 | 66.9 | 12.8 | 72.3 | 20.6 |
| Andorra | 61 | 66.7 | 33.3 | 23.4 | 13.1 | 15.7 | 13.6 | 12.5 | 17.5 | 69.4 | 30.6 |
| Austria | 8 217 | 61.7 | 3.7 | 46.9 | 2.4 | 38.3 | 1.7 | 37.6 | 1.2 | 80.1 | 2.9 |
| Belgium | 10 594 | 0.5 | - | 0.4 | - | 0.3 | - | 0.2 | - | 0.1 | - |
| Bosnia-Herzeg. | 4 203 | 63.8 | 6.7 | 56.2 | 4.4 | 51.6 | 3.9 | 43.4 | 3.2 | 84.4 | 6.3 |
| Bulgaria | 8 011 | 50.4 | 0.4 | 35.0 | 0.2 | 28.0 | 0.2 | 25.9 | 0.3 | 64.3 | 0.5 |
| Croatia | 4 400 | 94.5 | 0.8 | 88.6 | 0.6 | 85.2 | 0.5 | 81.8 | 0.5 | 90.1 | 0.7 |
| Cyprus | 836 | 45.6 | 1.8 | 30.6 | 1.7 | 24.5 | 1.2 | 24.2 | 1.4 | 80.2 | 1.7 |
| Czech Republic | 10 163 | 67.1 | 0.1 | 52.1 | 0.0 | 47.2 | 0.0 | 41.6 | 0.0 | 66.8 | 0.0 |
| Denmark | 5 423 | 0.0 | - | 0.0 | - | 0.0 | - | 0.0 | - | 0.0 | - |
| Estonia | 1 364 | - | - | - | - | - | - | - | - | - | - |
| Finland | 5 184 | - | - | 0.0 | - | 0.0 | - | 0.0 | - | 0.0 | - |
| France | 58 437 | 21.4 | 0.4 | 19.3 | 0.3 | 18.1 | 0.2 | 17.5 | 0.3 | 25.2 | 0.6 |
| Germany | 82 028 | 16.4 | 0.0 | 11.2 | 0.0 | 8.2 | 0.0 | 5.8 | 0.0 | 15.4 | 0.0 |
| Greece | 10 876 | 82.4 | 1.2 | 77.4 | 0.9 | 73.7 | 0.9 | 71.5 | 1.1 | 91.9 | 1.4 |
| Hungary | 10 125 | 63.4 | - | 43.9 | - | 36.3 | - | 34.9 | - | 68.1 | - |
| Iceland | 183 | - | - | - | - | - | - | 0.0 | - | - | - |
| Ireland | 3 734 | - | - | - | - | - | - | 0.0 | - | - | - |
| Italy | 56 675 | 84.4 | 13.7 | 87.7 | 8.9 | 88.2 | 7.9 | 88.1 | 8.2 | 78.9 | 21.1 |
| Latvia | 2 390 | - | - | - | - | - | - | - | - | - | - |
| Liechtenstein | 35 | 100.0 | - | 56.1 | 1.5 | 51.0 | 0.4 | 16.9 | 1.2 | 100.0 | - |
| Lithuania | 3 479 | 0.0 | - | 0.0 | - | 0.0 | - | 0.0 | - | 0.0 | - |
| Luxembourg | 427 | 2.1 | - | 1.7 | - | 1.2 | - | 1.3 | - | 0.9 | - |
| Macedonia FYR | 2 297 | 31.2 | 11.4 | 24.1 | 9.4 | 23.1 | 9.6 | 22.5 | 11.1 | 72.2 | 27.8 |
| Malta | 395 | 100.0 | - | 100.0 | - | 100.0 | - | 100.0 | - | 100.0 | - |
| Monaco | 32 | | | 100.0 | - | 100.0 | - | 100.0 | - | | |
| Montenegro | 724 | 51.6 | 16.5 | 30.5 | 9.8 | 26.3 | 9.2 | 24.6 | 11.7 | 80.2 | 19.8 |
| Netherlands | 15 746 | - | - | - | - | - | - | - | - | - | - |
| Norway | 3 226 | 3.0 | - | 2.9 | - | 2.9 | - | 3.9 | 0.0 | 4.5 | - |
| Poland | 38 237 | 44.5 | - | 33.1 | 0.0 | 27.3 | 0.0 | 27.1 | 0.0 | 60.1 | - |
| Portugal | 9 910 | 33.6 | 0.4 | 26.8 | 0.3 | 24.6 | 0.2 | 23.6 | 0.6 | 40.7 | 0.8 |
| Romania | 22 341 | 35.7 | 0.1 | 26.8 | 0.0 | 19.5 | 0.0 | 17.0 | 0.1 | 37.3 | 0.1 |
| San Marino | 20 | 100.0 | - | 28.4 | - | 22.9 | - | 19.2 | - | 100.0 | - |
| Serbia | 10 821 | 41.7 | 0.7 | 31.6 | 0.4 | 26.7 | 0.4 | 24.9 | 0.7 | 65.0 | 0.9 |
| Slovakia | 5 295 | 81.5 | 0.1 | 62.3 | 0.0 | 51.8 | 0.0 | 54.8 | 0.1 | 99.8 | 0.2 |
| Slovenia | 2 047 | 92.1 | 7.9 | 94.4 | 4.7 | 94.0 | 4.6 | 97.6 | 2.1 | 96.3 | 3.7 |
| Spain | 39 046 | 59.2 | 0.5 | 52.2 | 0.3 | 50.3 | 0.3 | 47.6 | 0.4 | 67.7 | 0.7 |
| Sweden | 8 898 | 0.0 | - | 0.0 | - | 0.0 | - | 0.1 | - | 0.1 | - |
| Switzerland | 7 270 | 49.6 | 5.8 | 42.5 | 2.9 | 38.1 | 2.4 | 39.8 | 1.4 | 89.3 | 8.0 |
| United Kingdom | 59 051 | 0.2 | - | 0.0 | - | 0.0 | - | 0.1 | - | 0.2 | - |
| **Total** | **516 162** | **35.4** | **2.1** | **30.8** | **1.3** | **28.3** | **1.2** | **27.3** | **1.3** | **39.6** | **3.1** |

## 6.4 Conclusions and recommendation

The following advices can be given with regard to the most optimal grid resolution to use at each process step to attain accurate spatial air quality concentrations fields and related exposure estimates, against acceptable computational demand.

***Separate rural and urban concentration maps***

Section 6.1 leads to the finding that only at rural areas for SOMO35 some improvement of significance (>5%) was obtained at the finer grid resolution, due to a more pronounced relationship of ozone with the finer resolution of the altitude field. Altitude is used as supplementary data source at the rural areas. Considering these rather limited improvements alongside the increased computational demand we advice to stick at the 10x10 km resolution for deriving the separate rural and urban maps from the station measurements. However, in the future such refinement of (specifically rural) ozone maps could be reconsidered, bearing in mind the fact that the AOT40 maps are routinely calculated at the 2x2 km grid resolution for more accurate estimate of ozone vegetation impact indicators used in assessments by EEA.

***Combined final concentration maps***

Applying the finer 1x1 km merger improves for both $PM_{10}$ and ozone the accuracy of the final concentration maps and their consecutive exposure estimates. The 3x3 km floating merger shows quite similar uncertainty results as with the 1x1 km merger. Since this additional floating grid calculation step increases computational demand, something we aim to avoid, we prefer the use of the 1x1 km merger. Section 6.2 shows that using the finer 1x1 km merger better accounts for the small towns and cities in predominantly and sparsely populated areas than using the alternative population density class boundary of Section 5.3.1. The merger resolution plays a major role in improving the final map, not the resolution of the separate interpolated maps. We recommend applying the 1x1 km merger on the 10x10 km separate rural and urban maps from now on as part of the routine $PM_{10}$ and ozone concentration mapping.

***Population exposure***

The highest estimates for the population weighted concentrations and the exceedance exposures for $PM_{10}$ - and the lowest for ozone - are obtained when using the final maps resulting from the 1x1 km merger, disregard the 1x1 km or 10x10 km grid resolution of the separate maps. Section 6.3 concludes that at the population weighted concentration and population exposure exceedances the resolution of the merger mainly plays a role and not the resolution of the separate rural and urban maps. This is in line with findings of Section 6.2. Spatial aggregations of the final concentration map should be avoided in air quality impact assessments: it causes considerable to substantial adverse 'dilution'-effects. Recommended is to examine in more detailed the sub-grid variability, including the effect of spatial averaging on impact assessments.

*Table 6.13 Spatial resolutions recommended as methodological improvements to be implemented as part of the default mapping methodologies as listed in Table 7.1 of Horálek et al. (2008).*

| **Combinations of spatial grid resolution: Interpolation - Merging - Population exposure estimates** | **Improves health exposure estimates** | **Advice to implement** | |
|---|---|---|---|
| | **($PM_{10}$ and Ozone )** | **Yes/No** | **Remark** |
| A (10-10-10) *(Current methodology)* | *Point of departure* | *n/a* | *Replace with C* |
| B (10-fl3-1) | Yes | No | C is more efficient and as accurate. |
| C (10-1-1) | Yes | **Yes** | **Implement!**; As accuracate as D, more efficient as D. |
| D (1-1-1) | Yes, but limited compared to C | No | C is more efficient and as accurate. |
| E (1-fl3-1) | Yes, but limited compared to B | No | C is more efficient and as accurate. |
| F (aggr.D-10) | No, far worse than A | No | Avoid at all times! |

# 7 Summary of conclusions and recommendations

This chapter summarises which tested methodological alternatives contributed in uncertainty reductions and as such are suitable and efficient enough to be implemented as part of the default spatial interpolation methodology to produce European air quality maps and exposure estimates.

In Chapter 4 we investigated three options for potential improvements of the interpolation methodology of the separate rural and urban maps. Chapter 5 focused on expected potential improvements of the current merging method and of alternative merging options. In Chapter 6 we examined the potential of introducing higher grid resolutions at the consecutive process steps. Below summarises the conclusions and recommendations.

***Log-normal concentration transposition, altitude transformation, introduction of land cover data***

Three different potential improvements of the used mapping methodology were examined for $PM_{10}$ and ozone indicators: log-normal transposition, altitude transformation and land cover inclusion. The comparison with the presently used methodology was done, separately for rural and urban areas. Log-normal transposition and altitude transformation were tested for 2006 and 2007 data, for 10x10 km resolution. Land cover inclusion was tested on resolution 1x1 km for 2006 data only and subsequently at the most promising results for both 2006 and 2007 years on the 10x10 km resolution.

*Log-normal transposition* provides for $PM_{10}$ slight improvement of the interpolation quality compared to the current methodology, specifically at rural areas and to less extend at urban areas. Also fits the log-normal distribution the $PM_{10}$ data better than the normal distribution. In conclusion, the log-normal transposition is recommended for implementing in the default interpolation methodology at the human health $PM_{10}$ indicators annual average and $36^{th}$ maximum daily average. For ozone the log-normal transposition brings no improvement and should not be implemented.

*Altitude transformation* brings no significant improvement for rural maps at both $PM_{10}$ and ozone. (Altitude is not applied as supplementary data in the urban areas). We conclude that no implementation of additional altitude data treatment should be done for any of the $PM_{10}$ and ozone indicators.

*Land cover inclusion* is the most complex of the considered potential improvements. For $PM_{10}$ the use spatial data of forested area seems to be helpful for rural mapping, but at the limit of significance. The observed improvements at $PM_{10}$ indicate that it is worth considering the inclusion of land cover data in the default methodology. However, as the land cover data is lacking for some countries, the implementation in the routine procedure complicates the mapping of these areas. This leads to the conclusion that land cover is not worth for inclusion in the $PM_{10}$ indicator mapping as long as it does not cover the entire mapping domain. When the land cover data will cover the completed European mapping domain, inclusion should be reconsidered in the routine $PM_{10}$ mapping methodology. For ozone no significant interpolation improvements are obtained. No implementation of land cover data into the default spatial interpolation methodology for ozone indicators should take place.

See Table 4.11 for the summary on the concluding recommendations.

***Rural and urban merger***

The current methodology for merging the separate rural and urban map interpolations into one combined final map for Europe has been scrutinized on it limitations and causes of uncertainties (Section 5.1). Based on the results of that and other projects and literature three alternative merging options (Section 5.2) as well as improvements of the current method (Section 5.3) have been tested on their potential for uncertainty reductions in the combined final concentration map.

In analysis of the current merging methodology we detected some weak elements:

- The type of the measurement station as reported by the countries, which we use to select the stations for the interpolation of the separate rural and urban map, appeared not to match always that consistent with the population density class of corresponding grid cell in the population density map. This map we used to select, i.e. to merge, either the rural or urban map grid cell values into the final map. This mismatch in characteristics causes in the combined final map overestimations for $PM_{10}$ and underestimations for ozone, specifically in some lower to moderately populated rural areas.
- At some urban/suburban stations in moderately to lower populated areas the grid cells of the combined final map show lower concentration values than the corresponding station measurement. Here the grid resolution of the final map plays a role.
- Additionally, we suspect that incorrectly reported classification or coordinates of stations may play a negative role in the accuracy of the spatial interpolations and the subsequent final maps.

The most problematic seem to be those cases where urban/suburban stations are sited in areas with low or moderate population densities. It concerns here most likely stations at small towns with relative high pollution situated in predominantly rural areas with low air pollution.

These findings lead to our recommendation to increase verification efforts on a correct reporting of the station classification and coordinates. Also an improvement in the spatial layout of monitoring networks will contribute positively. Both would lead to increased consistency between station types and the corresponding population density type, resulting in reduced uncertainties in the final European maps and its exposure estimates.

*Alternative merging options of rural and urban concentration levels*

With regard to certain weak points of the current methodology, three alternative merging methods have been tested and compared in Section 5.2 on the uncertainty levels (RMSE and MPE statistics) observed in the final map attained by merging. The alternative merging options that use either all stations together irrespective their type, or apply some weighting factor in the linear regressions before executing spatial interpolation on the common residuals, do not provide reductions in the uncertainties. They lead to an overestimation of $PM_{10}$ (resp. underestimation of ozone) at rural areas in the final map. In the current default merging process the application of the additional joint rural/urban interpolation field at specific conditions next to the separate rural and urban interpolation fields appears to reduce uncertainties more than when not applied. This all leads to the overall conclusion that the alternative approaches do not improve the mapping results. The current merging method used so far gives in general already the best results, therefore, we advice to stick to this methodology.

*Improvements of the current merging method*

Section 5.3.1 concludes that an alternative lower value for the rural class boundary of $\alpha_1 = 50$ inhbs.km$^{-2}$, instead the current 100 inhbs.km$^{-2}$, provides improvement of the interpolated map and can provide subsequently improved impact assessments. Specifically improved interpolations will be obtained at the mapping of small towns and cities in predominantly and more sparsely populated rural areas. It reduces the uncertainties caused by 'overlooking' small urbanisations in more extended rural areas which had been detected as most likely one of the weakest points in the current mapping methodology. Therefore, the use of $\alpha_1 = 50$ inhbs.km$^{-2}$ could be recommended to be introduced as part of the default merging methodology at both $PM_{10}$ and ozone, if the 10x10 km merger would still be applied in the mapping methodology. However, Section 6.4 concludes that the use of a merger with finer 1x1 km grid resolution will be considerably more effective to resolve the overlooking the small cities in lower and moderately populated areas than the use of an alternative population density class boundary for the rural areas. Nevertheless, this alternative class boundary could be used occasionally in mapping processes when the application of the 1x1 km grid resolution merger would become too computational and time demanding. For example at occasions where 'quick and dirty' draft map impressions or presentations are preferred over the more accurate maps.

Furthermore, Section 5.3.2 specifies more precisely the merging conditions for selecting from the separate interpolated maps the grid cell values for the combined final map at those areas. It concerns cases where the rural $PM_{10}$ map has higher grid cell values (resp. lower at ozone) than the corresponding grid cells in the urban map. This more specified merging function is recommended for routine use in mapping of the $PM_{10}$ and ozone health indicators.

See Table 5.18 for the summary on the concluding recommendations.

***Grid resolution***

The influence of the grid resolution on the interpolation of the separate rural and urban maps, on the merging of these maps into the final combined map and on its subsequent exposure estimates leads to the following conclusions:

Refining the grid resolution from 10x10km to 1x1 km at the *interpolation* of the separate rural and urban maps shows improvement of some significance at rural areas for SOMO35 only. However, considering the increase in involved computational demand we advice to stick at the 10x10 km resolution (see Table 5.18). Nevertheless, bearing in mind that the AOT40 maps are already routinely calculated at the 2x2 km grid resolution for more accurate ozone impact estimates of the EEA CSI indicators, one could for consistency reasons consider to introduce this refinement in the near future.

Refining the grid resolution for *merging* the separate maps helps improving the accuracy of the obtained combined final concentration map both for $PM_{10}$ and ozone, rather indifferent from resolution of the separate interpolated maps. It better accounts for smaller urbanisations in predominantly rural areas. The resolution of the merger plays a substantial role in these improvements, not the resolution of the separate interpolated maps. This, and the increase in computational demands at interpolations on a 1x1 km grid for the separate maps, leads to the recommendation to apply the finer resolution for merging only, not for the separate maps for both $PM_{10}$ and ozone.

Additionally, we examined empirically what effect *spatial aggregation* of a 1x1 km to a 10x10 km the final concentration map has on subsequent population weighted concentration and exposure exceedance estimates. Spatial aggregation causes a considerable decreased accuracy in the accounting for population averaged concentrations and exceedance exposures. This methodological adverse 'thinning'-effect should be avoided at all times.

For population *health exposure* assessments the findings lead to the concluding advice that it is most optimal to use combination C (10-1-1) instead A (10-10-10) used so far, meaning:

- calculate the separate urban and rural concentration maps on a 10x10 km grid resolution,

- subsequently merge these separate maps on a 1x1 km grid resolution into a 1x1 km combined final concentration map, and

- derive population averaged concentrations and exceedance exposures using the population density map with the 1x1 km grid resolution.

See Table 6.13 for the summary on the concluding recommendations.

***Recommended improvements on the default methodology of mapping $PM_{10}$ and ozone***

Based on the findings of this paper, the following improvements of the default mapping methodology are recommended for routine production of European wide air quality maps, and their derivates:

- Use a log-normal transposition at the human health $PM_{10}$ indicators annual average and $36^{th}$ maximum daily average at the interpolation methodology. Not for the ozone indicators.
- Use at both the $PM_{10}$ and ozone health indicator mapping the more precisely specified merging criteria, as represented at Equation 5.1.

- Use the merger on a 1x1 km grid resolution instead of the currently used 10x10 km grid resolution, and stick to the computational less demanding interpolations of the separate rural and urban maps on the 10x10 km grid resolution.
- Reconsider the introduction of land cover for mapping $PM_{10}$ as supplementary data once it covers the entire European mapping domain. Not for ozone.

The combined final map in 1x1 km grid resolution should serve in particular more accurate population exposure estimates. For presentational purposes on European scale, this 1x1 km grid resolution could be spatially aggregated into a 10x10 km grid resolution. However, this spatially aggregated map can in no case serve as input to the population exposure estimates on national or European scale.

Table 7.1 summarizes the overall recommended implementations.

*Table 7.1 Overall recommended methodological improvements to be implemented as part of the default mapping methodologies as listed in Table 7.1 of Horálek et al. (2008).*

| Implement at current methododology | $PM_{10}$ | Ozone |
|---|---|---|
| **at interpolation** | | |
| Keep current method (with rural, urban and joint urban/rural map) | **Yes** | **Yes** |
| Additional lognormal transposition of concentrations | **Implement** | -- (1) |
| Altitude transformation | -- (1) | -- (1) |
| Land cover inclusion | (Reconsider in future) (2) | -- (1) |
| Improve quality reported classification & coordinates of stations | (Implement) (3) | (Implement) (3) |
| Improve quality spatial layout of monitoring networks | (Implement) (3) | (Implement) (3) |
| Interpolate at higher resolution (1x1 km instead of 10x10 km) | Optional (4) | Optional (4) |
| **at merging** | | |
| Apply pop. dens. class boundary $\alpha_1$ = 50 inhbs.km$^{-2}$ (instead 100) | (Impl. occasionally) (5) (6) | (Impl. occasionally) (5) |
| Apply refined merging criteria (apply Eq. 5.1 instead of Eq. 2.3) | **Implement** | **Implement** |
| Apply high merging resolution (combination C (10-1-1)) | **Implement (7)** | **Implement (7)** |
| **at exposure estimates** | | |
| Apply high resol. of merger (combin. C (10-1-1)) and pop. density | **Implement (8)** | **Implement (8)** |

*(1)No significant improvement found; do not implement.*

*(2)Land cover inclusion should be reconsidered when it covers the entire European mapping domain.*

*(3)Implementation is primarily at control and responsibility of the countries and DG ENV.*

*(4)At the higher interpolation resolution quite similar end results against more computational demand are obtained. (Improvement of some significance is reached at rural areas for SOMO35 only).*

*(5)To apply occasionally in mapping processes in case application of a finer 1x1 km grid merger would be too computational and/or time demanding. Nevertheless, the use of a finer 1x1 km grid as merging resolution is a better alternative to this.*

*(6)Verify at application of a log-normal transposition at $PM_{10}$ if indeed improvements are obtained. (Not tested in this paper).*

*(7) Combination C (10-1-1) is used primarily for exposure estimates. For presentational purposes of European map pictures it should be spatially aggregated. (Note that if 1x1 km maps would be needed for presentational purposes, then at combination C square patterns of 10x10 km would 'show through'. These originate from the 10x10 km interpolated maps. To avoid this one should use then combination D (1-1-1)).*

*(8)For exposure estimates never spatially aggregate concentration or population density maps.*

# References


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

Beelen R, Hoek G, Pebesma E, Vienneau D, de Hoogh K, Briggs DJ (2009). Mapping of background air pollution at a fine spatial scale across the European Union. Science of the Total Environment 407, pp. 1852-1867.

Benedictow A et al (2009). Transboundary Acidification,Eutrophication and Ground Level Ozone in Europe in 2007. EMEP Report 1/2009. http://www.emep.int/publ/reports/2009/status_report_1_2009.pdf

Brabec M, Horálek J, Fiala J (2005). Statistical Approach to Pooling Empirical Data and Air Pollution Physical Modeling Output. Meteorological Journal 8, 177-186.

Briggs, D. et al. (2005). APMoSPHERE – Air Pollution Modelling for Support to Policy on Health and Environmental Risks in Europe. (APMoSPHERE Final Report. - 6. Detailed Report, Related To Overall Project Duration) http://www.apmosphere.org/

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

Denby B, Horálek J, Walker SE, Eben K, Fiala J (2005). Interpolation and assimilation methods for European scale air quality assessment and mapping. Part I: Review and recommendations. ETC/ACC Technical paper 2005/7. http://air-climate.eionet.europa.eu/docs/ETCACC_TechnPaper_2005_7_SpatAQ_Interpol_Part_I.pdf

Denby B, Schaap m, Segers A, Builtjes P, Horálek J (2008a). Comparison of two data assimilation methods for assessing $PM_{10}$ exceedances on the European scale. Atmospheric Environment 42, Issue 30, 7122-7134.

Denby B, Horálek J, de Smet P, de Leeuw F, Kurfürst P (2008b) European scale exceedance mapping for PM10 and ozone based on daily interpolation fields. ETC/ACC Technical Paper 2007/8. http://air-climate/eionet.europa.eu/reports/ETCACC_TP_2007_8_spatAQmaps_dly_interpol

Denby B, de Leeuw F, de Smet P, Horálek J (2009). Sources of uncertainty and their assessment in spatial mapping. ETC/ACC Technical Paper 2008/20. http://air-climate.eionet.europa.eu/docs/ETCACC_TP_2008_20_sources_mapping_uncertainty.pdf

De Smet P, Horálek J, Coňková M, Kurfürst P, de Leeuw F, Denby B (2009). European air quality maps of ozone and $PM_{10}$ for 2006 and their uncertainty analysis. ETC/ACC Technical paper 2008/8. http://air-climate.eionet.europa.eu/reports/ETCACC_TP_2008_8_spatAQmaps_2006

De Smet P, Horálek J, Coňková M, Kurfürst P, de Leeuw F, Denby B (2010). European air quality maps of ozone and $PM_{10}$ for 2007 and their uncertainty analysis. ETC/ACC Technical paper 2009/9. http://air-climate.eionet.europa.eu/docs/ETCACC_TP_2009_9_spatialAQmaps_2007.pdf

ECMWF: Meteorological Archival and Retrieval System (MARS). It is the main repository of meteorological data at ECMWF (European Centre for Medium-Range Weather Forecasts; http://www.ecmwf.int/).

EEA (2007). Corine land cover 2000 (CLC2000). lceugr100_00, 100x100m gridded version 9/2007. http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-clc2000-100-m-version-9-2007

EEA (2009a). Spatial assessment of PM10 and ozone concentrations in Europe (2005). EEA Technical report 1/2009. http://air-climate.eionet.europa.eu/reports/EEA_TR_1_2009_Spatial_AQ_assessment_2005

EEA (2009b). Corine land cover 2000 (CLC2000). lceugr100_00, 100x100m gridded version 12/2009. http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-clc2000-100-m-version-12-2009

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

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

Horálek J, Kurfürst P, Denby P, de Smet P, de Leeuw F, Brabec M, Fiala J (2005). 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. http://air-climate.eionet.europa.eu/docs/ETCACC_TechnPaper_2005_8_SpatAQ_Part_II.pdf

Horálek J, Denby B, de Smet PAM, de Leeuw FAAM, Kurfürst P, Swart R, van Noije T (2007). Spatial mapping of air quality for European scale assessment. ETC/ACC Technical paper 2006/6. http://air-climate.eionet.europa.eu/docs/ETCACC_TechnPaper_2006_6_Spat_AQ.pdf

Horálek J, de Smet PAM, de Leeuw FAAM, Denby B, Kurfürst P, Swart R, (2008). European air quality maps including uncertainty analysis. ETC/ACC Technical paper 2007/7. http://air-climate.eionet.europa.eu/reports/ETCACC_TP_2007_7_spatAQmaps_ann_interpol

Janssen S, Dumont G, Fierens F, Mensink C (2008). Spatial interpolation of air pollution measurements using CORINE land cover data. Atmospheric Environment 42, pp. 4884-4903.

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

Mareckova K, Wankmueller R, Anderl M, Poupa S, Wieser M (2009). Inventory Review 2009; Emission Data reported under the LRTAP Convention and NEC Directive, Technical Report CEIP 1/2009. http://www.ceip.at/fileadmin/inhalte/emep/pdf/2009/Inventory_Review_2009.pdf

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

Rao CR (1973). Linear statistical inference and its applications. Wiley. New York (2nd ed., Jan. 2002).

Sandnes Lenschow H, Tsyro S (2002). Meteorological input data for EMEP/MSC-W air pollution models. EMEP MSC-W Note 2/2000. http://www.emep.int/publ/reports/2000/dnmi_note_2_2000.pdf

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

Tarrasón L, Nyíri Á. (2008). Transboundary acidification, eutrophication and ground level ozone in Europe in 2006. EMEP Status Report 1/2008. http://www.emep.int/publ/common_publications.html

Van de Kassteele J (2006). Statistical air quality mapping. Doctoral Thesis Wageningen University.

Vestreng V, Mareckova K, Kakareka S, Malchykhina A, Kukharchyk T (2007). Inventory Review 2007; Emission Data reported to LRTAP Convention and NEC Directive, MSC-W Technical Report 1/07 http://www.emep.int/mscw/mscw_publications.html#2007

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