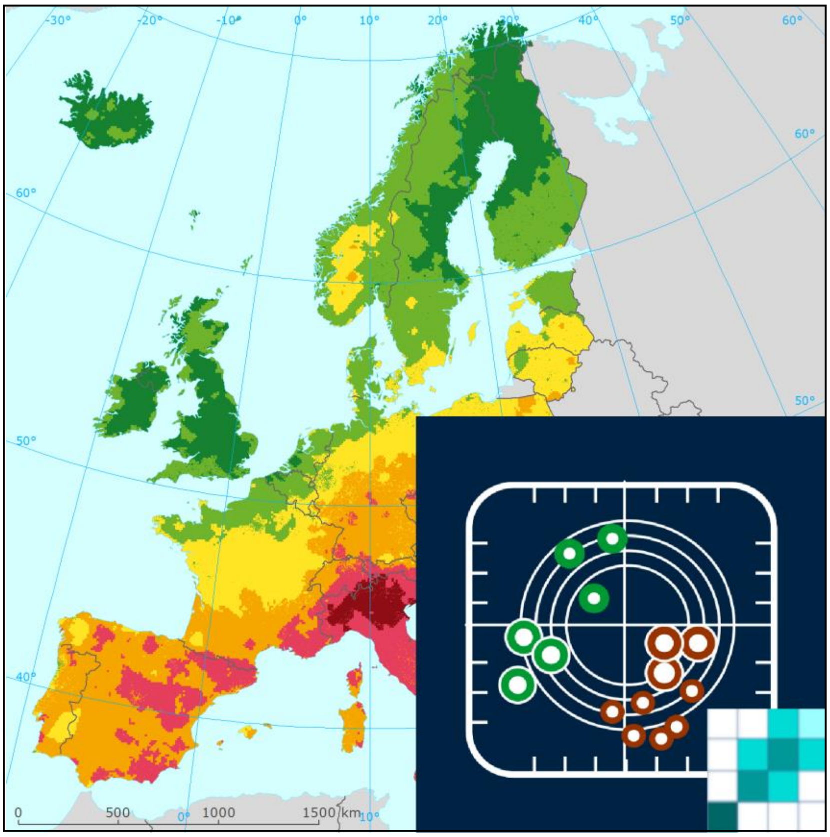
Application of FAIRMODE Delta tool to evaluate interpolated European air quality maps for 2012



ETC/ACM Technical Paper 2015/2
November 2015

Jan Horálek, Nina Benešová, Peter de Smet

European Topic Centre
on Air Pollution and
Climate Change Mitigation

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

Front page picture:
Main picture: Combined rural and urban background concentration map of ozone indicator 26th highest daily maximum 8-hour mean value for the year 2012. Units: µg.m^{-3}. (Figure 3.2 of this paper.)
Bottom right quarter: Diagram representing the Delta tool. Taken from Thunis et al. (2015), front page picture.
Bottom right: Logo of FAIRMODE.

Author affiliation:
Jan Horálek, Nina Benešová: Czech Hydrometeorological Institute (CHMI), Prague, Czech Republic
Peter de Smet: National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

Refer to this document as:
Horálek J, Benešová N, De Smet P (2015). Application of FAIRMODE Delta tool to evaluate interpolated European air quality maps for 2012. ETC/ACM Technical paper 2015/2. http://acm.eionet.europa.eu/reports/ETCACM_TP_2015_2_Delta_Evaluation_AQMaps2012

DISCLAIMER

ETC/ACM Technical Paper 2015/2
European Topic Centre on Air Pollution and Climate Change Mitigation
PO Box 1
3720 BA Bilthoven
The Netherlands
Phone +31 30 2748562
Fax +31 30 2744433
Email etcacm@rivm.nl
Website http://acm.eionet.europa.eu/

Contents

Acknowledgements

The authors would like to acknowledge Laure Malherbe, Anthony Ung and the INERIS team for providing the assimilation and validation sets of stations as routinely used in the MACC-II (resp. MACC) project.

The authors would also like to thank the reviewer of this report Stijn Janssen (VITO), leader of FAIRMODE Working group 1 – Assessment.

1 Introduction

The Directive 2008/50/EC on Ambient Air Quality and Cleaner Air for Europe requires that air quality should be assessed throughout the territory of each member state. It requires that the fixed measurements should be used as a primary source of information for such assessment in the polluted areas. Those measurement data may be supplemented by modelling techniques to provide adequate information on the spatial distribution of the air quality. Under the same directive the countries report their national measurement data annually to the EU. EEA has collected these data over the years covering by now a large time series with data from all over Europe (*AirBase*). In the aim to obtain a consistent European wide picture of the air quality, spatially interpolated maps primarily based on these air quality measurements have been routinely constructed by the ETC/ACM on an annual basis since 2007. These maps were developed as a means to make use of the unique amount of measurement data as reported by the countries, and to allow for impact and exposure assessments throughout Europe in a consistent way. The mapping methodology combines monitoring data, chemical transport model results and ancillary information through spatial interpolation methods (Horálek et al., 2007, 2015).

The aim of this paper is to present a sensitivity analysis of the spatial interpolation mapping methodology using FAIRMODE Delta tool. This evaluation could aim for recommendations for improvements on spatial interpolation AQ assessments within the context of the FAIRMODE Delta tool, which has been designed for model benchmarking.

FAIRMODE Delta tool supports the evaluation of different model applications performed in the frame of the Air Quality Directive. Based on paired modelled and monitored data, it offers rapid diagnostic of model performance in terms of various statistical indicators and diagrams. Some of these statistical indicators are evaluated against bound values (model performance criteria) to facilitate the benchmarking of the user's model application against agreed quality standards.

In this paper, the Delta tool is used to do a sensitivity analysis of the spatial interpolated mapping methodology by means of a validation-set and test-set of stations as is used in MACC, as well as applying the full set of stations as is used in the spatial interpolation mapping methodology. The sensitivity analysis is applied on the maps of the following human health related pollutant statistics: PM_{10} annual average and 36th highest daily mean, $PM_{2.5}$ annual average, and ozone 26th highest daily maximum 8-hour mean. All the analysis is done for the 2012 data.

This should allow for comparison of the evaluation results with those of MACC and other model evaluations. However, one should bear in mind that the spatial interpolation mapping methodology applies the stations as primary input and specified to their type, which is a different approach compared to the MACC ensemble and other chemical transport models. This will be discussed in this paper, together with the extent to which the Delta tool is actually compatible or suitable to apply for the evaluation of the current spatial interpolation methodology.

Chapter 2 describes the methodology used. Chapter 3 documents the input data and parameters. Chapter 4 presents the analysis, i.e. the evaluation of the maps using the Delta tool. Chapter 5 discusses the results and summarizes the conclusions.

2 Methodology

2.1 Spatial mapping method

The methodology used for the creation of the concentration maps is documented in Horálek et al. (2015). In the method, the air quality measurements are taken as the primary data and the results from chemistry transport modelling and other auxiliary data (altitude, meteorology) as the secondary sources. The mapping method consists of a linear regression model followed by kriging of the residuals produced from that regression model (residual kriging):

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

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

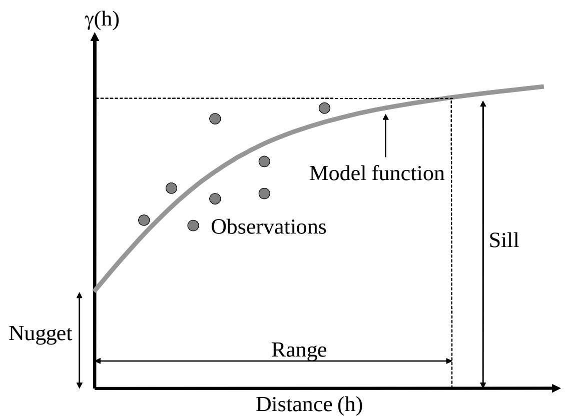
The spatial interpolation of the regression's residuals is carried out using ordinary kriging, according to

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

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

The empirical variogram is fitted by a spherical function. For the illustration of the spherical function and the parameters of variogram (called *nugget*, *sill*, and *range)*, see Figure 2.1.

Figure 2.1 Diagram showing the parameters of variogram function



In the case of PM_{10} and $PM_{2.5}$, prior to linear regression and interpolation, a logarithmic transformation to measured and modelled data is applied. After the interpolation, a back-transformation is used.

The maps of the health related indicators are created for the rural and urban background areas separately on a grid at 10x10 km resolution. Subsequently, we merge these rural and urban background maps into one combined air quality map using a population density grid at 1x1 km resolution. For details, see Horálek et al. (2015). The final merged maps in 1x1 km resolution are used for exposure estimates. For presentational purposes of these European-wide maps, we aggregate them from the 1x1 km grid resolution into maps at a 10x10 km grid resolution.

For different pollutants and area types, different supplementary data are used in the mapping. Table 2.1 presents the supplementary data used for the mapping of pollutants presented in this paper.

Table 2.1 Supplementary data used for the mapping of PM_{10}, $PM_{2.5}$ and ozone

Pollutant	Area type	EMEP model output	Altitude	Surface solar radiation	Wind speed	Population density
PM_{10}	Rural	+	+	+	+	-
	Urban background	+	-	-	-	-
$PM_{2.5}$	Rural	+	+	+	+	+
	Urban background	+	-	-	-	-
Ozone	Rural	+	+	+	-	-
	Urban background	+	-	+	+	-

The method is routinely used for *annual data* (i.e. the monitoring and the modelling are aggregated into the annual statistics at first). For sensitivity analysis and comparison with the results based on the daily data, see Gräler et al. (2012).

In the case of $PM_{2.5}$, the measurement data are supplemented by so-called *pseudo $PM_{2.5}$ data*, which are the estimates of $PM_{2.5}$ concentrations at the locations of PM_{10} stations with no $PM_{2.5}$ measurement. These estimates are based on PM_{10} measurement data and different supplementary data, using linear regression. For details, see Horálek et al. (2015) and Denby et al. (2011).

The uncertainty estimation of the mapping results is routinely based on 'leave one out' *cross-validation* method. This 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 in the available set. The results of the cross-validation are expressed by statistical indicators and scatter plots. The main indicators used are *root mean squared error* (RMSE) and *bias*:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{Z}(s_i) - Z(s_i))^2} \qquad (2.3)$$

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(\hat{Z}(s_i) - Z(s_i)) \qquad (2.4)$$

where $Z(s_i)$ is the air quality indicator value derived from the measured concentration at the i^{th} point,
$\hat{Z}(s_i)$ is the air quality estimated indicator value at the i^{th} point using other information, without the indicator value derived from the measured concentration at the i^{th} point,
N is the number of the measuring points.

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

$$RRMSE = \frac{RMSE}{\frac{1}{N}\sum_{i=1}^{N} Z(s_i)} .100 \qquad (2.5)$$

The *relative RMSE* (RRMSE) is expressed in percent.

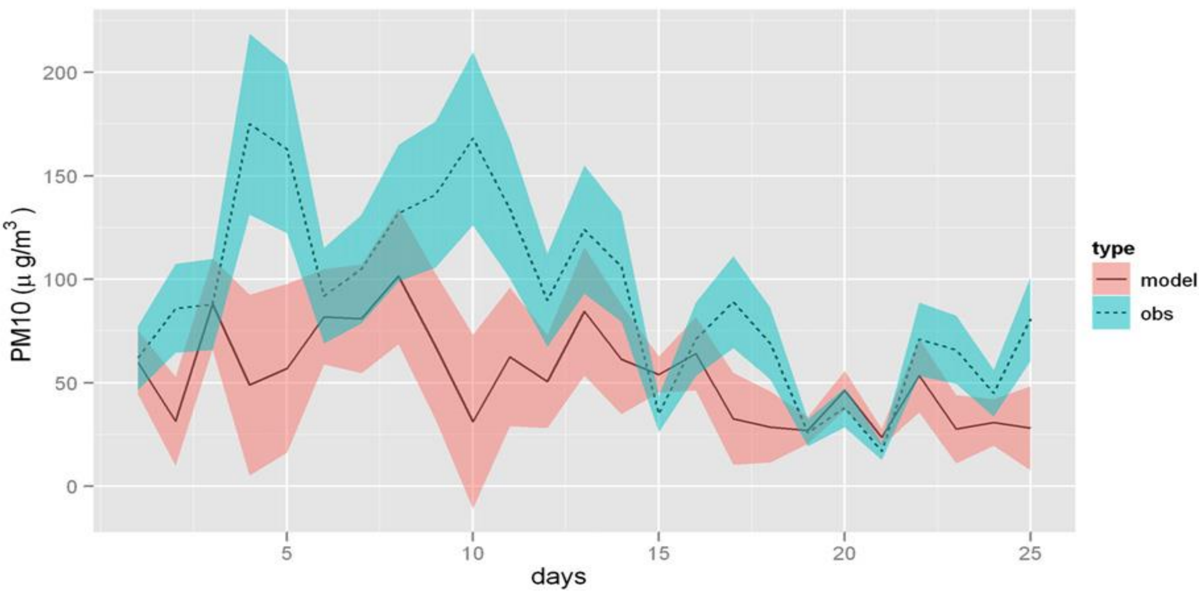
2.2 FAIRMODE Delta tool

Within the framework of FAIRMODE a methodology was proposed to evaluate model performance for policy applications (especially those related to the Air Quality Directive, EC, 2008). It was developed primarily for the time-dependent (e.g. hourly, daily) data, but can be used also for the annual statistics. The methodology is described in Thunis et al. (2012, 2013). The motivation for the methodology is as follows: Although the Air Quality Directive (EC, 2008) defines model quality objectives for the main pollutants, their practical application remains open to interpretation (EEA, 2011).

2.2.1 Model Quality Objective

Thunis et al. (2012, 2013) propose a statistic called *Model Quality Objective (MQO)* as an indicator to identify whether the model results are “good enough”. The basic concept of the methodology is based on the idea to use observation uncertainty *U* as a margin of tolerance also for model results. That is, we have observations with uncertainty *U*, we allow a model to have a similar uncertainty, and thus we call a model “good enough” when it differs from the observed value by *2U* or less. (For the observation uncertainty, the expanded uncertainty is used, see Section 2.2.3.) Figure 2.1 illustrates this principle of equal tolerance for model and measurement for a modelled (solid line) and an observational (dashed line) time series of PM_{10} at a given station.

Figure 2.2 Example of PM_{10} time series (measured and modelled concentrations) for a single station, together with a coloured area representative of the model and observed uncertainty ranges



Source: Thunis et al. (2012).

The *MQO* is built on the conditions that the model and the measurement confidence interval do overlap between each other and the model uncertainty should not exceed the measurement uncertainty. According to this, the successful models should fulfil relation

$$MQO = \frac{\sum_{i=1}^{N}(M_i - O_i)^2}{2\sum_{i=1}^{N}U(O_i)^2} \le 1, \qquad (2.6)$$

where O_i is the observation (or measurement) value in the *i*-th time step at the given station,
M_i is the modelled value in the *i*-th time step at the given station,
$U(O_i)$ is the expanded uncertainty of the observation O_i,
N number of the time steps.

For the definition of the expanded uncertainty, see Section 2.2.3.

In case of average annual values, the *MQO* reduces to

$$MQO = \frac{|\overline{M} - \overline{O}|}{2U(\overline{O})} = \frac{|bias|}{2U(\overline{O})} \le 1, \qquad (2.7)$$

where $\overline{O}$ is the annual average of the observation values at the given station,
$\overline{M}$ is the annual average of the modelled values at the given station,
$U(\overline{O})$ is the expanded uncertainty of the annual average of the observation values.

For percentile values (e.g. 26th highest daily 8h maximum in 365 days), the *MQO* is calculated similarly as in Equation 2.7, i.e. using the observed and the modelled values corresponding to the selected percentile and the uncertainty of the relevant observed value:

$$MQO = \frac{|M_{perc} - O_{perc}|}{2U(O_{perc})} \le 1, \qquad (2.8)$$

where O_{perc} is the relevant percentile of the observation values,
M_{perc} is the relevant percentile of the modelled values,
$U(O_{perc})$ is the expanded uncertainty of the observation value corresponding to the relevant percentile.

Any parameter *x-th highest value* is by its definition just the percentile value. (E.g., the 26th highest value of the set of 365 values is equal to the 93.2nd percentile of this set.) Thus, Equation 2.8 is directly applicable for such a parameter as well.

When comparing model results to a group of station observations the question arises how many stations should be required for the criteria to be fulfilled, as 100% would be probably too strict. It was decided that a successful model should fulfil the *MQO* criteria for at least 90% of stations (Thunis et al., 2015). The decision was motivated by the Air Quality Directive (EC, 2008), which defines the modelling uncertainty as the maximum deviation of the measured and calculated concentration levels for just 90% of individual monitoring points.

2.2.2 Performance criteria for correlation and standard deviation

Next to the *MQO* statistics, several other statistical indicators are used to derive so-called *performance criteria*, namely the bias, the correlation coefficient and the standard deviation. The performance criteria for the three indicators are derived from the *MQO* for the time-dependent data. Both the temporal and the spatial dimension are taken into account. In the case of annual averages (and other annual statistics), only the criteria for correlation and standard deviation are used for the spatial dimension:

$$N_R = \frac{(1-R)\cdot\sigma_O^2}{2\cdot RMS_U^2} \leq 1, \qquad (2.9)$$

$$N_\sigma = \frac{|\sigma_M - \sigma_O|}{2\cdot RMS_U} \leq 1, \qquad (2.10)$$

where N_R, N_σ are the normalized correlation and the normalized standard deviation,
σ_O is the standard deviation of the observed annual data across the set of the stations,
σ_M is the standard deviation of the modelled annual data across the set of the stations,
R is the correlation coefficient of the observed and modelled annual data,
RMS_U is quadratic mean of the expanded uncertainty of the observed annual data.

During the analysis, we realized that under the Delta 5.0 software, RMS_U is calculated differently than described in Thunis et al. (2015) and presented here, leading to slightly different results of N_σ. Next to this, N_R is calculated by this software differently than in Equation (2.9), leading to substantially different results. Our findings were confirmed by the Delta tool team. We were informed this inconsistency should be fixed in the next Delta release.

It should be noted that these performance criteria are constructed as complementary to the *MQO*.

2.2.3 Measurement uncertainty expression

The measurement uncertainty is a key input to the *MQO* and to the performance criteria (Thunis et al., 2013). To evaluate the uncertainty a simple approach was proposed (Thunis et al. 2013, Pernigotti et al., 2013). This approach is based on the assumption that the uncertainty of each measurement is composed of a component proportional to the concentration level and a non-proportional component as in

$$u^2(O_i) = u_p^{\,2}(O_i) + u_{np}^{\,2}(O_i) = (1-\alpha)(u_r^{RV} O_i)^2 + \alpha(u_r^{RV} RV)^2, \qquad (2.11)$$

where $u(O_i)$ is the uncertainty of the observation O_i,
$u_p(O_i)$ is the proportional component of the uncertainty of the observation O_i,
$u_{np}(O_i)$ is the non-proportional component of the uncertainty of the observation O_i,
α is the non-proportional fraction of the uncertainty around a reference value RV,
u_r^{RV} is the estimate of the relative uncertainty around a reference value RV,
RV is the reference concentration value.

Once the uncertainty $u(O_i)$ is estimated, the expanded uncertainty $U(O_i)$ is calculated by multiplying $u(O_i)$ by a so-called *coverage factor*. Each value of the coverage factor gives a particular confidence level that the true value lays within the interval $O_i \pm U(O_i)$. Usually, the level of confidence of 95% is used. The expanded uncertainty can be expressed as

$$U(O_i) = k u_r^{RV} \sqrt{(1-\alpha)\cdot O_i^2 + \alpha\cdot RV^2}, \qquad (2.12)$$

where $U(O_i)$ is the expanded uncertainty of the observation O_i,
k is the coverage factor.

The uncertainty for a time averaged concentration is expected to be reduced compared to the uncertainties associated to the raw measurements. To cover this aspect, the proportional and non-proportional components of the uncertainty are divided by parameters N_p and N_{np}, respectively. For models producing annual averages, the expanded uncertainty is expressed as

$$U(\overline{O}) = k u_r^{RV} \sqrt{\frac{(1-\alpha)}{N_p}\overline{O}^2 + \frac{\alpha \cdot RV^2}{N_{np}}}, \qquad (2.13)$$

where $U(\overline{O})$ is the expanded uncertainty of the annual average of the observation values,
N_p is the estimated parameter for the proportional component,
N_{np} is the estimated parameter for the non-proportional component.

The uncertainty for a percentile value is calculated according to Equation 2.12, where the relevant percentile of the observation values is used as the observation O_i. The uncertainty of the percentile is considered in the Delta tool just as the uncertainty of the observation value corresponding to the relevant percentile.

In the Delta tool, all the parameters used in Equations 2.11 – 2.13 are already predefined (see Section 3.4.) based on Thunis et al. (2012, 2013, 2015) and Pernigotti et al. (2013). Those values can be changed according to user’s needs.

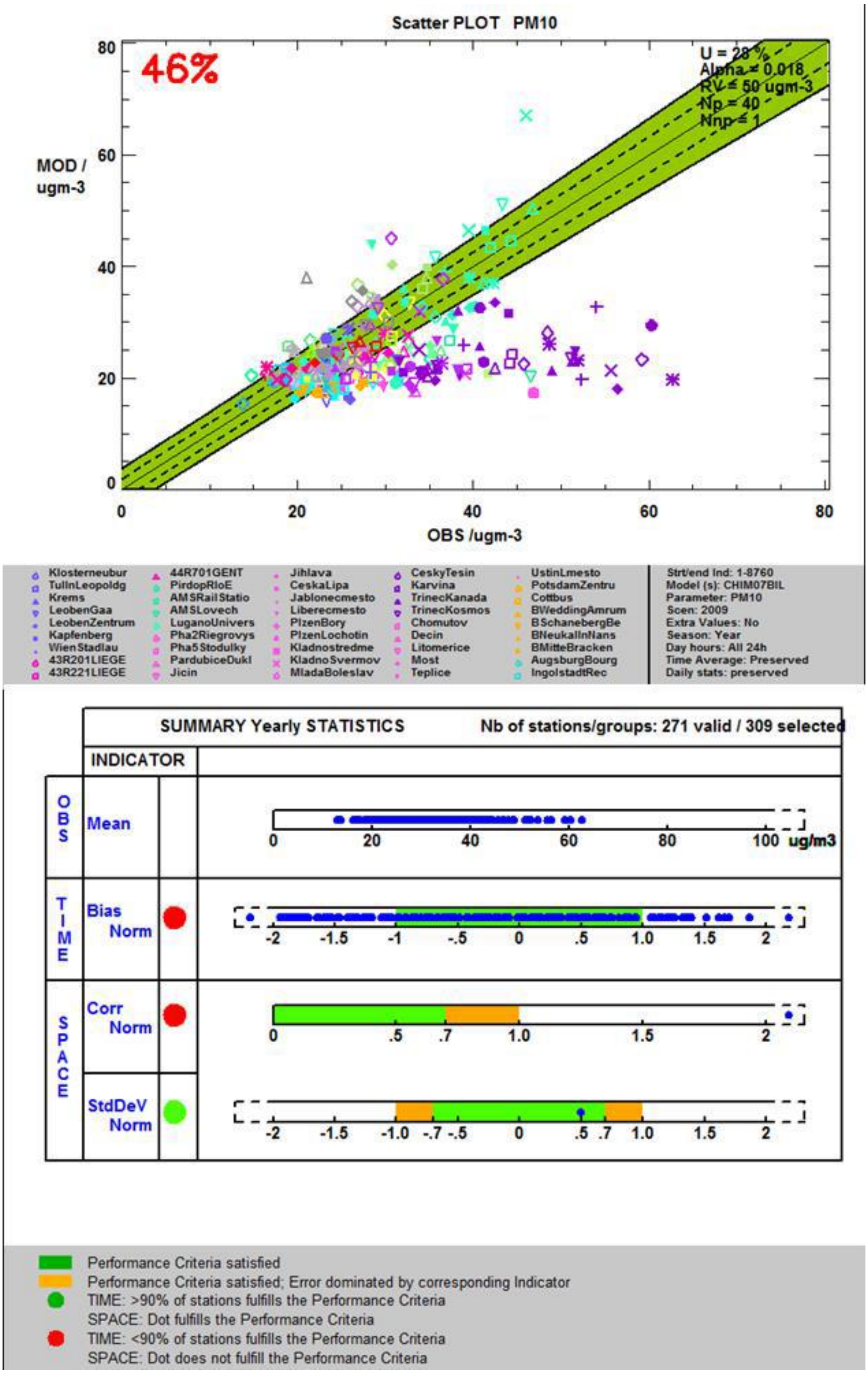
2.2.4 Benchmarking report of Delta tool

Delta tool is software provided by FAIRMODE to the modelling community in an effort to provide a simple tool to evaluate modelling results against measurement data and to support the idea of a common benchmarking approach within the modelling community. Besides the common plots, such as time series, bar plots and scatter plots, it provides a possibility to evaluate and visualise the model performance in terms of *MQO* in so-called *benchmarking reports*. Each report consists of a diagram and a summary statistics table. Different reports are provided for the time-dependent data and for the annual values. These reports are not mutually comparable. For example, the results are visualised as a *target plot* in the case of time-dependent data and as a *scatter plot* in the case of annual values. In this paper, we present results for the annual values only. For the annual data, the benchmarking report consists of the scatter plot and a summary statistics table. For example of such a report, see Figure 2.3.

The *scatter plot* represents the modelled outcomes at observational points on the vertical axis versus the observational values on the horizontal axis. The stations located in the green area (*MQO* < 1) fulfil the target criterion. Dashed lines mark the zone where *MQO<0.5*. The percentage of stations fulfilling the target criterion (*MQO* < 1) is indicated in the upper left corner and is meant to be used as the main indicator in the benchmarking procedure. The percentage is green when it is higher than the 90% criterion, i.e. more than 90% of the stations fulfil the uncertainty criterion and the model result is considered to be ‘good enough’; when it is red then it is below 90%, i.e. not fulfilling the *MQO* criterion and the model result is ‘not good enough’.

In the upper right corner of the scatter plot the parameters values defined and set to evaluate the uncertainty are presented (see Section 3.4). In the scatter plot, each station is marked by a different coloured marker. Under the plot, the legend of these markers is given. However, usually a small part of the legend only is visualized in the report due to the large number of the stations. Thus, in this paper we drop this legend.

Figure 2.3 Example of benchmarking performance summary report of the Delta tool for models producing annual values consisting of scatter plot (upper) and summary statistics table (lower)



Source: Thunis et al. (2015).

In the lower part of the report a *summary statistics table* is presented. It is meant as a complementary source of information to the bias-based *MQO*. In the case of the annual average and other annual aggregated data, the summary statistics table consists of four rows. Row *1 (OBS)* shows the annual indicator (i.e. annual average or x-th percentile) of the measurement data for the relevant stations. The OBS section summarises the observation values and each dot represents one station. Row *2 (TIME)* shows the fulfilment of the bias-based *MQO* for all stations. The green shaded area represents range in which the *MQO* will be fulfilled. The red dot in the third column, indicates that the number of the stations with the *MQO* fulfilment is smaller than 90%. Note this row is redundant with the scatter plot. (A green dot indicates a percentage higher than 90 percent.) *MQO* is calculated according to Equation 2.7. Rows *3* and *4 (SPACE)* provide an overview of spatial statistics for the normalized correlation and the standard deviation calculated, according to Equations 2.9 and 2.10. Annual values are used to calculate the spatial correlation and standard deviation. Thus, only one point representing the relevant statistics is plotted in each row. If the point is located at coloured (green and orange) areas it indicates that the performance criterion is fulfilled. However, location in the orange area indicates that the error associated with this statistic is dominant. Points located in the white area do not meet the performance criteria. This is also expressed by the red/green colour (fulfils/fails) in the third column of this plot.

2.3 Evaluation approach

In this report, the evaluation is executed

(i) by applying the *full set* of the stations as used in the spatial interpolation mapping methodology,

(ii) by applying the *validation-set* and the *assimilation-set* of the stations as used in MACC-II.

In the first case, the map is constructed using all the available stations, see Horálek et al. (2015). The map is evaluated using the same full set of the stations.

In the second case, two different subsets of stations are used like in MACC-II, i.e. the so-called *assimilation* and *validation* subsets, see Section 2.4.2. The map is constructed using the assimilation subset of the stations only. Then, the map is evaluated with the validation subset of the stations.

In both cases, the evaluation is performed with

a. *rural background stations* of the relevant set,
b. *urban/suburban background stations* of the relevant set,
c. *all stations* of the relevant set, i.e. both rural and urban/suburban background stations.

For the evaluation the Delta tool is used, being *DELTA Version 5.0*. For details, refer toThunis et al. (2015). In all the cases, the benchmarking report is presented, consisting of the scatter plot and the summary statistics table as explained in Section 2.2.3.

The evaluation is performed for the following indicators: *PM_{10} annual average*, *PM_{10} indicator 36^{th} highest daily mean*, *$PM_{2.5}$ annual average* and *ozone indicator 26^{th} highest daily maximum 8-hour mean*.

The sets of the stations used for the mapping and for the evaluation are described in Section 3.2.

3 Data and parameters

3.1 Data used for the mapping

The data used for the creation of the concentration maps for 2012 are documented in detail in Horálek et al. (2015) and summarized here.

3.1.1 Measurement (observational) air quality data

The station monitoring data for 2012 were extracted from the European monitoring database AirBase (http://acm.eionet.europa.eu/databases/airbase/index_html), supplemented by several rural stations from the database EBAS (http://ebas.nilu.no/). Only data from stations classified as type *background* for the areas *rural*, *suburban* and *urban* are used. The following pollutants and their statistics are considered: *PM_{10} annual average*, *PM_{10} indicator 36^{th} highest daily mean*, *$PM_{2.5}$ annual average* and *ozone indicator 26^{th} highest daily maximum 8-hour mean*. All the aggregated data are expressed in $\mu g.m^{-3}$. Only the stations with an annual data coverage of at least 75 percent are considered.

The number of stations included in the evaluation were: for PM_{10}, 319 rural background and 1176 urban and suburban stations; for $PM_{2.5}$, 137 rural background and 464 urban and suburban stations; for ozone, 504 rural background and 1024 urban and suburban stations. The different numbers for PM_{10} and $PM_{2.5}$ stations in comparison with Horálek et al. (2015) are caused by the different counting of the data measured by the different measurement devices (e.g. gravimetric and TEOM) at the same location. In Horálek et al. (2015), such measurements are counted as different stations, while here (in agreement with the Delta tool) as one station only. However, the same set of the station like in Horálek et al. (2015) is used and the measurement data in the same location are handled by the same way, i.e. the average of the data measured at the same location by the different measurement devices is used in the analysis.

3.1.2 Chemical transport model and other supplementary data

The *chemical dispersion model* used is the EMEP MSC-W (version rv4.5), which is an Eulerian model, see Simpson et al. (2012). Its resolution is circa 50x50 km. Emissions for the relevant year (Mareckova et al., 2014) are taken and the model is driven by ECMWF (http://www.ecmwf.int/) meteorology. The same set of parameters as at the air quality observations is used.

The *altitude* data field (in m) with an original grid resolution of 15x15 arcseconds is taken from GTOPO30 (https://lta.cr.usgs.gov/GTOPO30).

The *meteorological parameters* used are *wind speed* (annual average, in $m.s^{-1}$) and *surface solar radiation* (annual average of daily sum, $MWs.m^{-2}$). The daily data in resolution 30x30 arc-seconds were extracted from the Meteorological Archival and Retrieval System (MARS) ERA-interim reanalyses of ECMWF.

Population density (in $inhabitants.km^{-2}$, census 2001) for the majority of countries is based on data provided by the European Commissions (EC) Joint Research Centre (JRC). The original resolution is 100x100 m. For countries and regions lacking JRC data, we use ORNL population data in the 1x1 km resolution. For details, see Horálek et al. (2015).

All the supplementary data are converted into the EEA reference projection ETRS89-LAEA5210 on a 10x10 km grid resolution. Population data, however, are converted into a 1x1 km resolution to allow the merge of the rural and the urban background maps, see Section 2.1. For details of the supplementary data and its aggregation, see Horálek et al. (2015).

3.2 Assimilation and validation subsets of the stations

The assimilation and validation subsets of stations (see Section 2.3) were provided by INERIS. The assimilation and validation sets have been routinely used in the MACC-II (resp. MACC, previously) project, see e.g. Rouïl et al. (2014). The validation set is kept the same for years under the MACC project (the only exception is in the case of the closure of the station), while the assimilation set is being actualized continuously.

Table 3.1 presents the comparison of the full set of the stations applied for the interpolation mapping (see Section 3.1.1) and the assimilation and validation sets provided by INERIS. As can be seen, the full set is not identical to the sum of the assimilation and validation sets. In principle, there are two reasons. First, the MACC project includes the full set of data from automatic stations only, not from station measurements through gravimetric method. Second, the ETC/ACM mapping method includes only stations with an annual data coverage of at least 75 percent. In Table 3.1 one can see that the full set consists of three subsets of the stations – the assimilation subset, the validation subset and the other stations. The assimilation subset includes the stations of the "MACC assimilation set" with enough (i.e. at least 75 percent) data and the validation subset includes the stations of the "MACC validation set" with enough data. The other stations of the full set are the stations with enough data, which are not included neither in the assimilation nor in the validation set of MACC, e.g. the gravimetric stations. In the right part of the Table 3.1, one can see the stations included in the assimilation and validation sets of MACC, but not used in the analysis due to their poor annual data coverage.

Table 3.1 Stations used for the mapping of PM_{10}, $PM_{2.5}$ and ozone

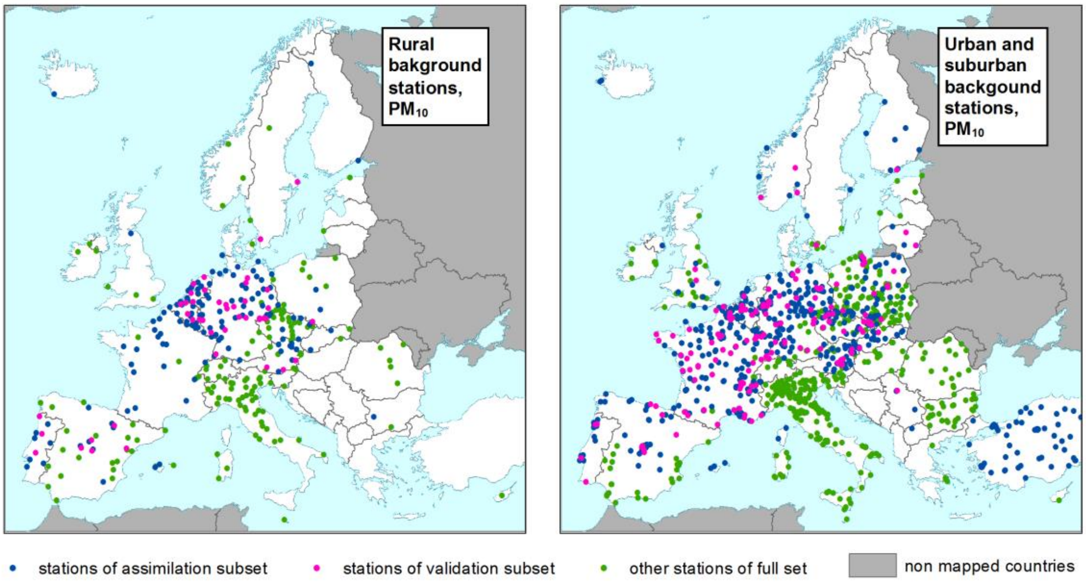
Pollutant	Area type	Full set				Not used	
		Assimilation subset	Validation subset	Other stations	In total	Assimilation set of MACC	Validation set of MACC
PM_{10}	Rural	131	43	145	319	13	8
	Urban backgr.	560	156	458	1176([a])	67	28
	All	691	199	605	1495([b])	80	36
$PM_{2.5}$	Rural	47	13	77	137	11	5
	Urban backgr.	174	66	224	464	45	7
	All	221	79	301	601	56	12
Ozone	Rural	370	113	21	504	22	12
	Urban backgr.	783	194	47	1024	72	49
	All	1153	307	68	1528	94	61

([a]) The number without Turkish stations is 1087.
([b]) The number without Turkish stations is 1406.

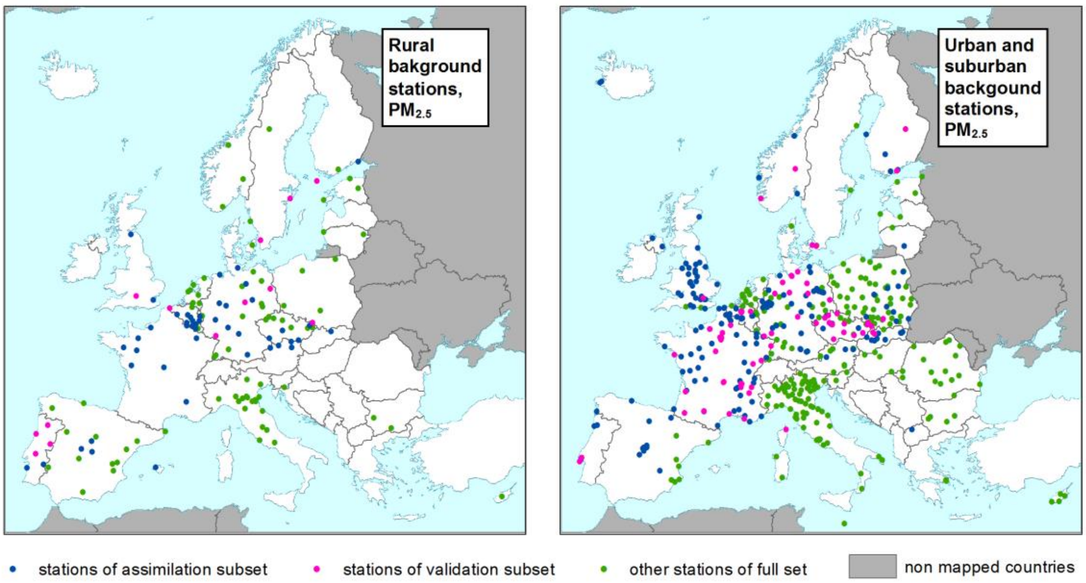
For the mapping, on the one hand the full set and on the other hand the assimilation subset is used. For the evaluation, on the one hand the full set and on the other hand the validation subset is used. It should be noted that Turkish stations are used for the mapping, but not for the evaluation. The reason is that only urban and suburban stations operated in Turkey in 2012, not the rural ones. Thus, only the urban background map, not the final map can be constructed for the area of Turkey, see Section 3.3.

Maps 3.1, 3.2 and 3.3 present the geographical distribution of the station used in the mapping, for different pollutants, for the rural and the urban background areas separately. The subsets are presented in different colours. The stations of the assimilation subset are marked by blue, the stations of the validation subset by pink and the other station s of the full set by green.

Map 3.1 **Geographical distribution of stations used in the mapping of PM_{10} for rural (left) and urban background (right) areas**

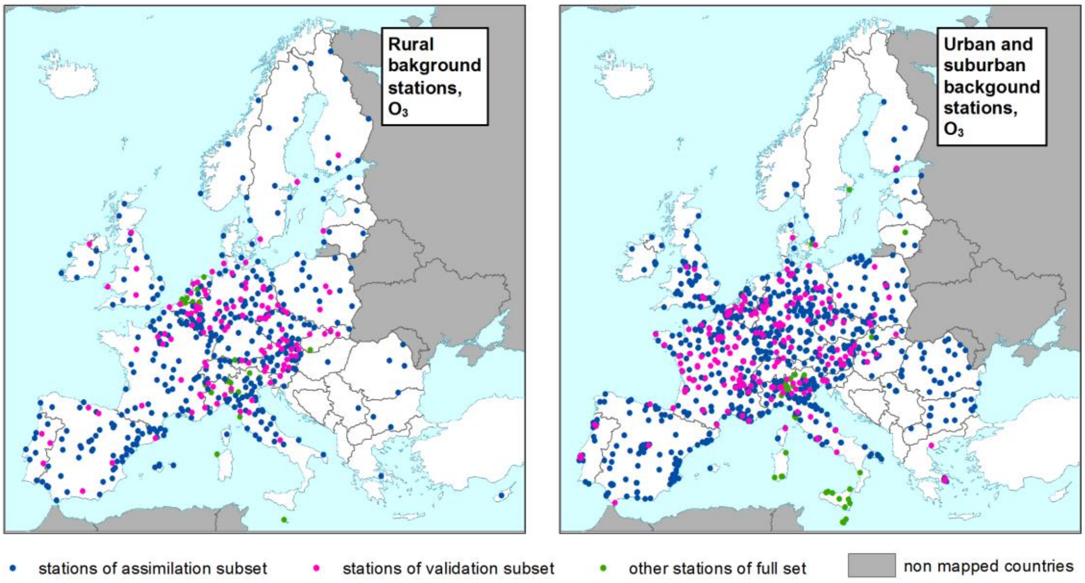


Map 3.2 **Geographical distribution of stations used in the mapping of $PM_{2.5}$ for rural (left) and urban background (right) areas**



One can see that both for PM_{10} and $PM_{2.5}$, in extended areas like Italy and Balkan countries operate only the stations, which are included neither in the assimilation nor in the validation subsets.

Map 3.3 **Geographical distribution of stations used in the mapping of ozone for rural (left) and urban background (right) areas**



3.3 *Air quality maps used in the analysis*

The maps for the analysis were created by the method described in Section 2.1. First, we created the maps based on the full set of stations, similar to the maps presented in Horálek et al. (2015). Second, we created the maps with the assimilation subsets of stations, as described in Section 3.2. Tables 3.2 and 3.3 present the parameters used in the mapping and the cross-validation uncertainty estimates.

Table 3.2 **Parameters used and uncertainty indicators estimated for map of PM_{10} annual average (left) and 36th highest daily mean (right) using full set and assimilation subset of stations**

Part of mapping method	Parameter	PM_{10} annual average				PM_{10}, 36th highest daily mean			
		Full set		Assimil. subset		Full set		Assimil. subset	
		Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Linear regr. model (Eq. 2.1)	c (constant)	1.71	1.63	1.47	1.27	2.06	1.80	1.58	1.41
	a1 (EMEP m.) ([a])	0.613	0.638	0.538	0.782	0.637	0.638	0.660	0.759
	a2 (altitude)	-0.00047	-	-0.00042	-	-0.0005	-	-0.0004	-
	a3 (wind speed)	-0.101	-	-0.058	-	-0.123	-	-0.084	-
	a4 (s. solar rad.)	0.014	-	0.040	-	-	-	0.025	-
Kriging of its residuals	Nugget	0.024	0.018	0.020	0.014	0.029	0.016	0.006	0.008
	Sill	0.068	0.075	0.048	0.060	0.065	0.095	0.042	0.072
	Range [km]	480	750	340	580	480	660	320	580
RMSE [μg.m^{-3}]		**3.8**	**6.1**	**4.4**	**7.4**	**7.7**	**11.9**	**8.7**	**14.9**
Relative RMSE		**21.4%**	**22.1%**	**25.0%**	**27.1%**	**24.5%**	**24.5%**	**27.5%**	**30.6%**
Bias [μg.m^{-3}]		**0.1**	**0.0**	**0.7**	**0.2**	**0.1**	**-0.1**	**1.0**	**-1.1**

([a]) EMEP modelled data are logarithmically transformed.

Table 3.3 Parameters used and uncertainty indicators estimated for map of $PM_{2.5}$ annual average (left) and ozone indicator 26th highest daily maximum 8-hour mean (right) using full set and assimilation subset of stations

Part of mapping method	Parameter	$PM_{2.5}$ annual average				O_3, 26th highest daily max. 8h			
		Full set		Assimil. subset		Full set		Assimil. subset	
		Rural	Urban	Rural	Urban	Rural	Urb.	Rural	Urb.
Linear regression model (Eq. 2.1)	c (constant)	1.31	1.28	1.61	1.30	-9.7	12.8	-7.0	1.80
	a1 (EMEP m.) ([a])	0.600	0.719	0.470	0.703	0.996	0.892	0.969	0.638
	a2 (altitude)	-0.00036	-	-0.00042	-	0.005	-	0.0055	-
	a3 (wind speed)	-0.082	-	-0.079	-	-	-2.62	-	-2.51
	a4 (s. solar rad.)	-	-	-	-	0.934	-	0.936	-
	a5 (popul. d.) ([a])	0.032	-	0.029	-	-	-	-	-
Kriging of its residuals	Nugget	0.032	0.019	0.034	0.014	30	53	38	61
	Sill	0.084	0.091	0.076	0.080	74	82	81	84
	Range [km]	620	900	620	750	100	360	400	310
RMSE [μg.m^{-3}]		**3.0**	**3.3**	**3.3**	**3.3**	**8.5**	**9.1**	**8.5**	**9.2**
Relative RMSE		**24.9%**	**18.7%**	**27.1%**	**18.7%**	**7.4%**	**8.3%**	**7.5%**	**8.4%**
Bias [μg.m^{-3}]		**-0.4**	**0.1**	**-0.7**	**-0.5**	**0.2**	**-0.1**	**-0.1**	**0.2**

([a]) In case of $PM_{2.5}$, EMEP modelled and population density data are logarithmically transformed.

For all the pollutants (and their respective aggregations), the rural and urban background maps were constructed in the 10x10 km resolution, using the parameters presented in Tables 3.2 and 3.3. The uncertainty estimates were calculated according to Equations 2.3 – 2.5. Comparing the uncertainty results for the maps constructed based on the full set and the assimilation subset, one can see the poorer uncertainty results of the maps based on the assimilation subset for both PM_{10} statistics and for $PM_{2.5}$ annual average. The reason is in the considerable higher number of the stations in the full set and in their better geographical coverage compared to the stations of the validation subset, see Table 3.1 and Maps 3.1 and 3.2. For ozone, the uncertainty results are almost the same for both map variants. The reason is mainly in quite high proportion and good geographical coverage of the assimilation subset, see Table 3.1 and Map 3.2.

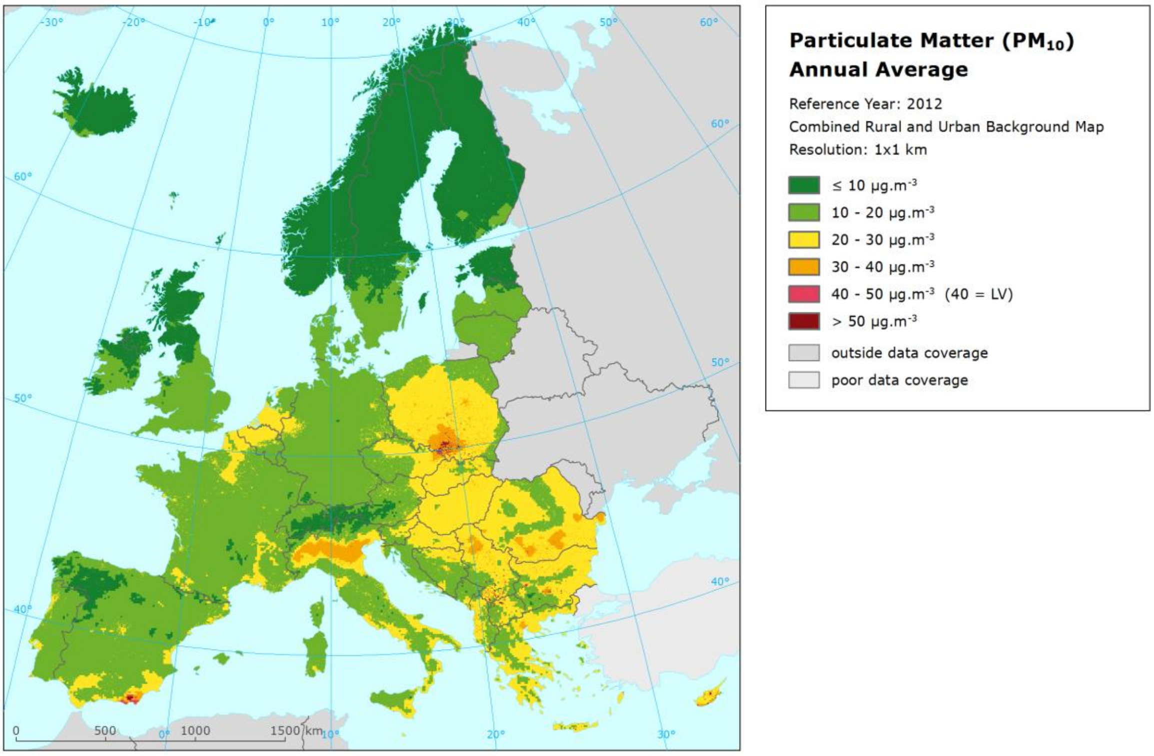
If compared the relative RMSE for PM_{10} statistics annual average and 36th highest daily value, one can see slightly better results for the maps of PM_{10} annual average, in all the cases.

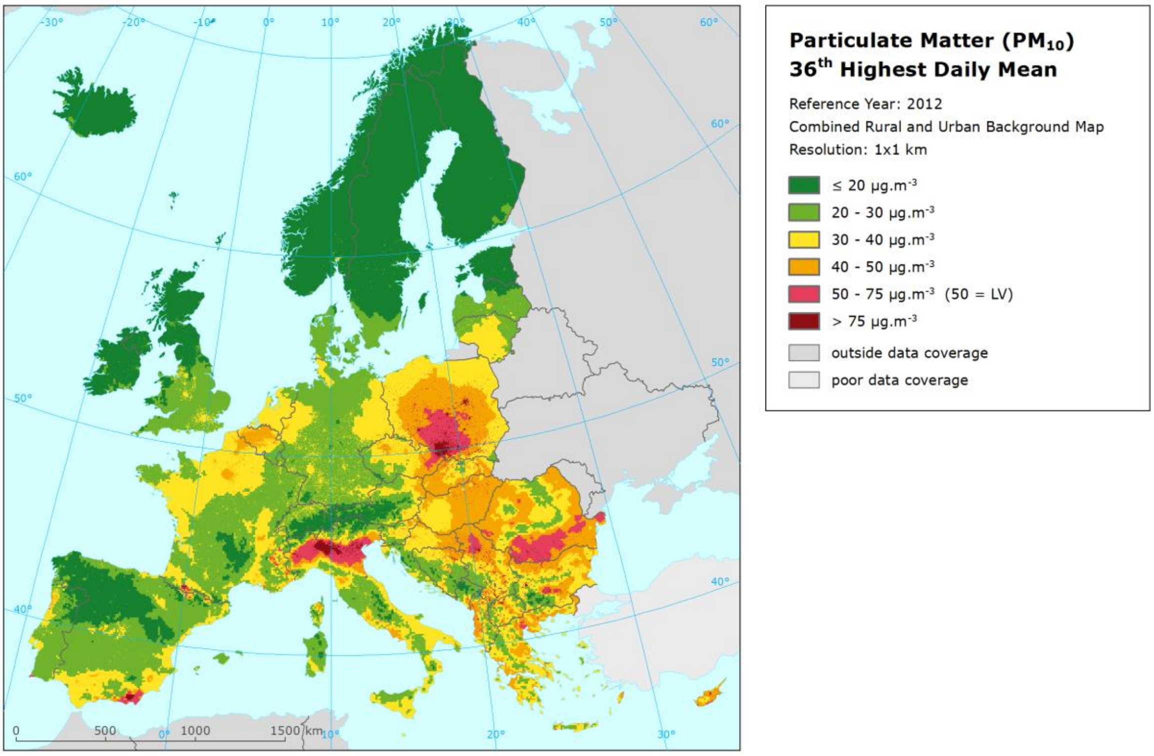
In all the cases, the rural and urban background maps were merged on basis of the population density in the 1x1 km resolution. The final merged maps in the 1x1 km resolution are further analysed in Chapter 4. Maps 3.4 and 3.5 present the mapping results based on the full set of stations.

Due to a lack of rural stations in Turkey, no interpolation results could be presented for this country in a rural map for all the indicators. Therefore, we excluded Turkey from the production of the final maps.

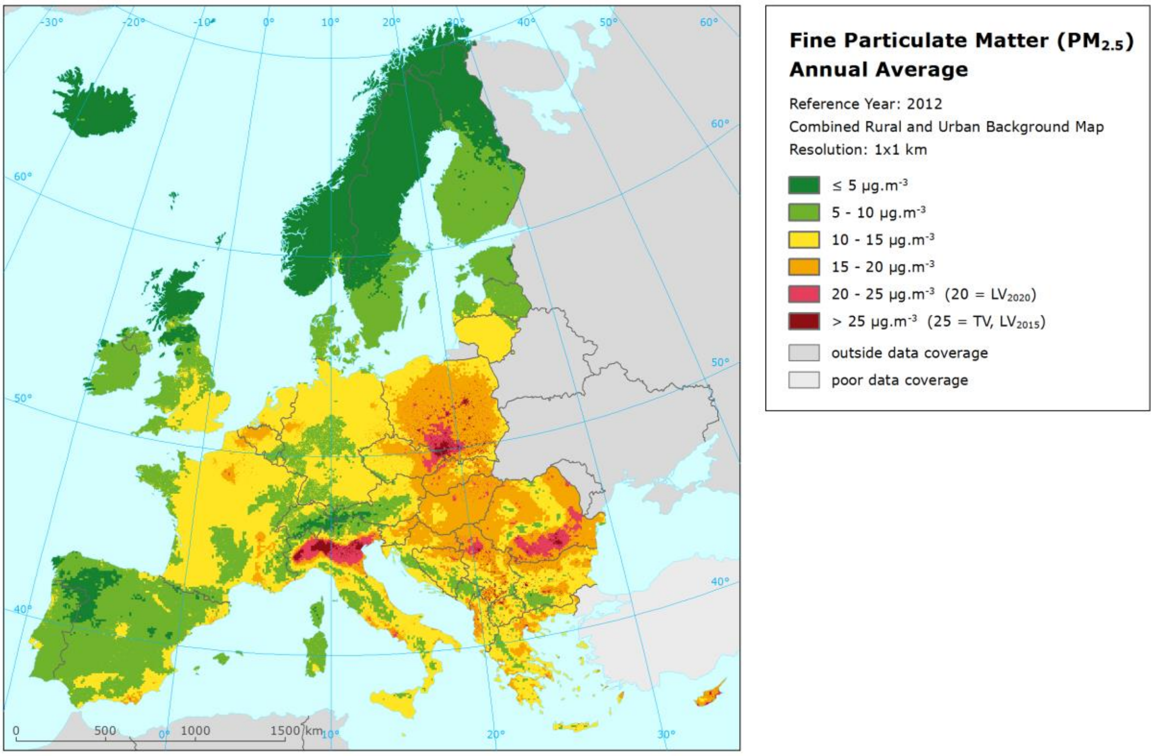
The mapping results are predominantly driven by the observations, leading in general to no bias, irrespectively to the concentration levels of the used chemical transport model (Horálek et al., 2014).

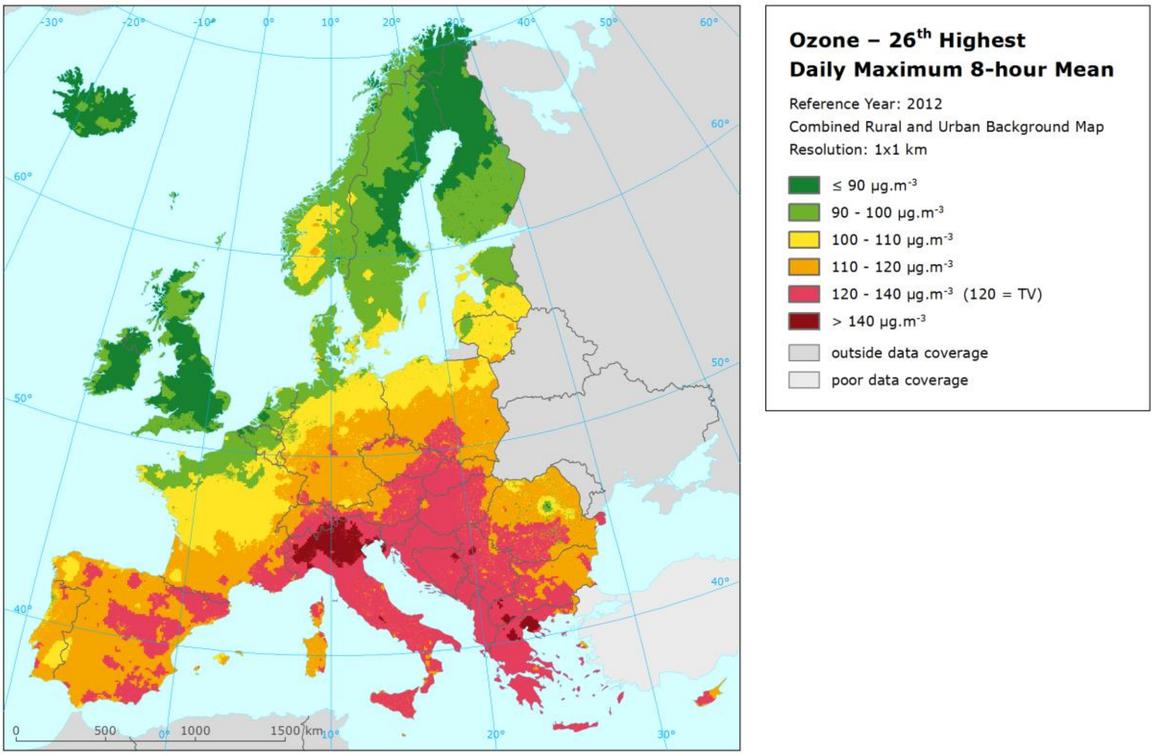
Map 3.4 Combined rural and urban background map of PM_{10} annual average (above) and 36th highest daily mean (below) for 2012, based on the full set of stations, in µg.m^{-3}





Map 3.5 **Combined rural and urban background map of $PM_{2.5}$ annual average (above) and ozone indicator 26th highest daily maximum 8-hour mean (below), 2012, based on the full set of stations, in µg.m^{-3}**





3.4 Uncertainty parameters used in the Delta tool

To evaluate the *MQO* (Equations 2.7 and 2.8), we need to define the measurement uncertainty, see Section 2.2.3, Equations 2.12 and 2.13. The Delta tool contains a predefined set of uncertainty parameters for the PM_{10} and $PM_{2.5}$ annual averages, which were used for this exercise. For percentile values like PM_{10} indicator 36th highest daily value or ozone indicator 26th highest daily maximum 8-hour mean, the uncertainty values for daily data types, such as daily mean or for 8-hour daily maximum, are recommended, see Thunis et al. (2015).

Table 3.4 gives the parameters currently defined and set in the Delta tool. Their values are estimated in Thunis et al. (2013) for ozone and in Pernigotti et al. (2013) for PM_{10}. The uncertainty u_r^{RV} for PM_{10} is based on the reference gravimetric method.

Going back to the case illustrated in Figure 2.3, one encounters in the upper right corner of the graph, different uncertainty parameters that are predefined for the evaluation. *U* stands for expanded uncertainty in percent and can be calculated as $U = k \,.\, u_r^{RV} \,.\, 100$. For PM_{10} it is set equal to 28%. N_p stands for the parameter of the proportional part of the uncertainty attributed to the time averaging of the annual average. For PM_{10} annual average; it is set equal to 40. *Alpha* is the non-proportional fraction of the uncertainty; it is set equal to 0.018 for PM_{10}.

Table 3.4 List of parameters used in the Delta tool to calculate the measurement uncertainty

Pollutant	Indicator	k	u_r^{RV}	RV	α	N_p	N_{np}
PM_{10}	Annual average	2.00	0.140	50 µg.m^{-3}	0.018	40	1
	Percentile ([a])	2.00	0.140	50 µg.m^{-3}	0.018	-	-
$PM_{2.5}$	Annual average	2.00	0.180	25 µg.m^{-3}	0.035 ([b])	40	1
O_3	Percentile ([a])	1.40	0.090	120 µg.m^{-3}	0.620	-	-

([a]) For percentiles, parameters relevant for daily mean (PM_{10}), resp. 8-hour daily maximum (O_3) are used.
([b]) In the Delta tool the value 0.035 is actually used, although in Thunis et al. (2015) the value 0.05 is given.

It should be noted that the measurement uncertainties are key input to the *MQO* equations 2.7 and 2.8, respectively. The *MQO* is highly sensitive to the parameter values defined as binding uncertainty criteria for model and map performance evaluations.

4 Analysis

This chapter assesses with the Delta Tool (Section 2.2) the performance of the mapping results of the linear regression model followed by kriging of its residuals for 2012. The maps in the evaluation are described in Section 3.3.

The assessment is executed for

(i) the maps constructed with the *full set* of stations as is routinely done in the spatial interpolation mapping methodology; the map evaluation uses the same full set of stations;

(ii) the maps constructed with the *assimilation subset* of stations; the map evaluation uses the *validation subset* of stations.

The *validation* and the *assimilation subsets* of stations as based on the similar default sets of stations as applied in the MACC-II project, see Section 3.2.

In both cases (i) and (ii), the evaluation is performed with

a. *rural background stations* of the corresponding (i.e. the full or the validation) set,
b. *urban/suburban background stations* of the corresponding set,
c. *all stations* (i.e. both rural and urban/suburban background stations) of the corresponding set.

The uncertainty parameters applied are given in Section 3.4. At all cases, the benchmarking report consists of a scatter plot and a summary statistics table. The scatter plot shows the percentage of stations fulfilling the criterion *MQO < 1* of Section 2.2.1. The summary statistics present *performance criteria* as a complementary source of information, see Section 2.2.2. Section 2.2.4 describes the benchmarking report.

Section 4.1 presents the results for the PM_{10} annual average, Section 4.2 for the PM_{10} indicator 36th highest daily mean, Section 4.3 for the $PM_{2.5}$ annual average and Section 4.4 for the ozone indicator 26th highest daily maximum 8-hour mean. Section 4.5 summarises results of this chapter.

The PM_{10} indicator 36th highest daily mean is presented additionally to be able to demonstrate the level of impact the uncertainty parameters settings have on the *MQO* criterion applied.

4.1 PM_{10} – annual average

4.1.1 Map constructed with the full set of stations

In this section we first evaluate with the Delta Tool the PM_{10} annual average map created with the full set of stations (Map 3.4) against the rural background stations of this full set. Figure 4.1 presents the benchmarking report of this evaluation, consisting of the scatter plot and the summary statistics table. For description of the plots see section 2.2.4.

Figure 4.1 shows that in the rural areas, the *MQO* criterion and both performance criteria are fulfilled (green dots). However, it should be noted the map is evaluated with stations, which were already used in the map construction itself. As such the evaluation cannot be considered objective.

Figure 4.1 Output of the Delta tool for the final merged map of PM_{10} annual average for 2012 in 1x1 km resolution based on the full set of stations, against rural background stations of this full set

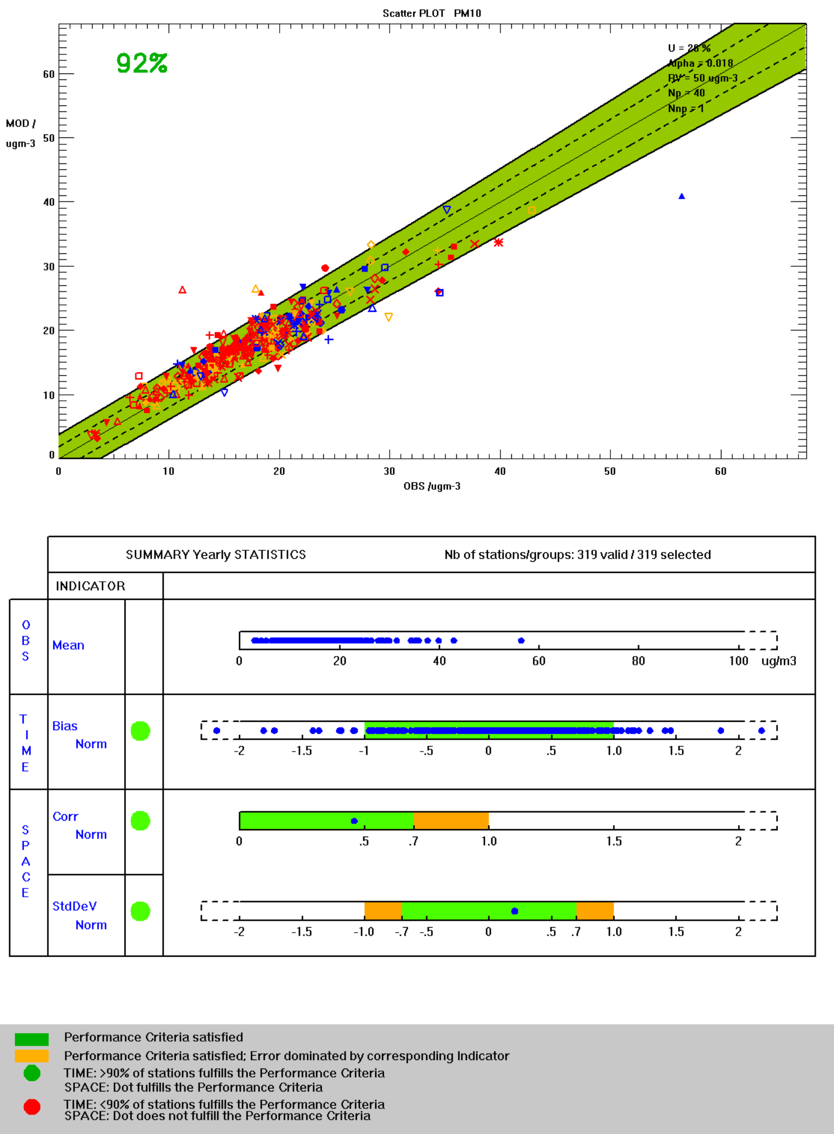


Figure 4.2 shows subsequently the benchmarking report for the same map based on the full set of stations (Map 3.4. However now evaluated against the urban and suburban background stations of this full set.

The red percentage in the upper left corner of the scatter plot and the red dot at Row 2 of the summary report indicate that the *MQO* criterion is not fulfilled in the urban areas. The performance criteria are fulfilled as the green dots indicate at the third and fourth row of the summary report. This somewhat poorer agreement of the mapping results with the measurements is attributable to several reasons. As stated in Horálek et al. (2015), differences are partly caused by the spatial averaging of the values in the 10x10 km grid cells having a smoothing effect, and partly by another smoothing effect the interpolation method in itself.

Concerning the *MQO* criterion, it should be noted it is quite stringent in the case of annual averages compared to its use in the case of percentile (i.e. the x-th highest) values, see Section 4.2.

Figure 4.2 **Output of the Delta tool for the final merged map of PM_{10} annual average for 2012 in 1x1 km resolution based on the full set of stations, against urban/suburban background stations of this full set**

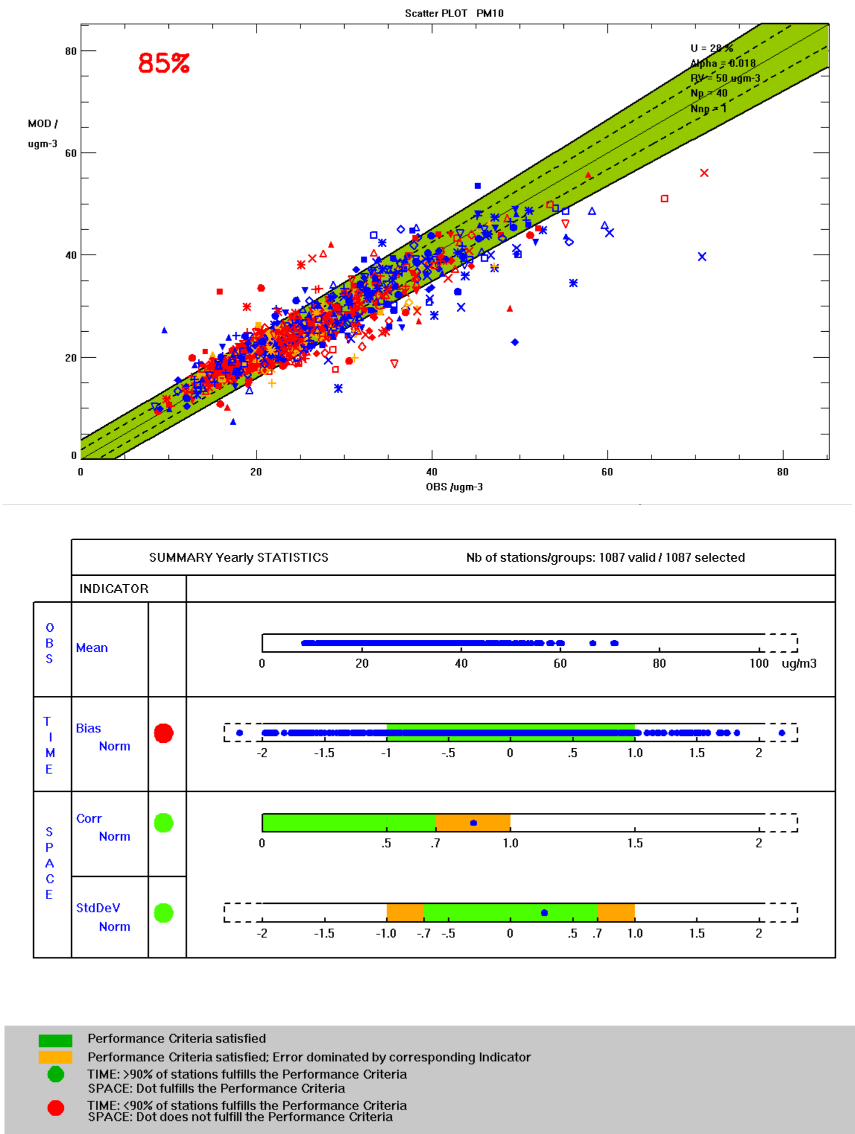
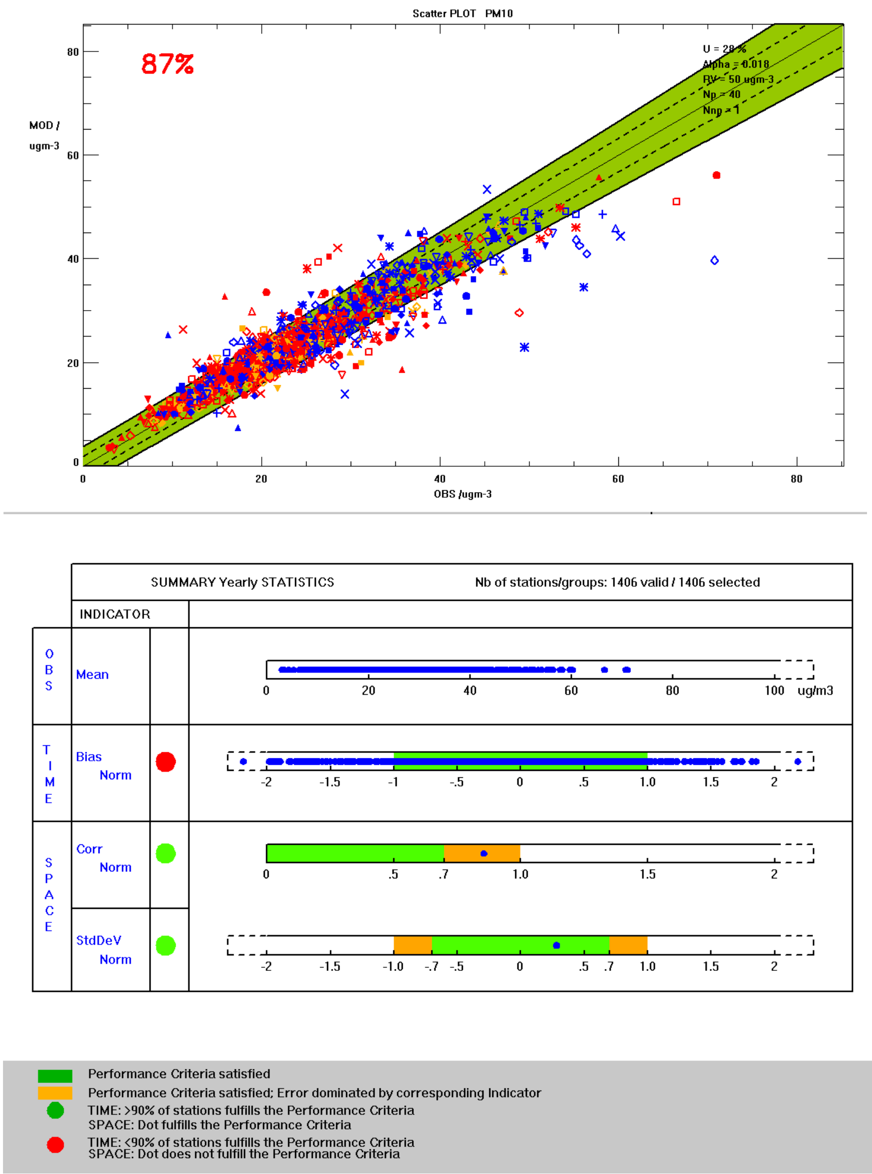


Figure 4.3 presents the benchmarking report for the same map created based on the full set of the stations (Map 3.4, but now evaluated against all stations (i.e. both rural and urban/suburban background) of this full set. Quite similar results as in Figure 4.2 are observed: non-fulfilment of the *MQO* criterion and the fulfilment of the performance criteria. This is caused mainly by the fact that the number of the rural stations is less than one quarter of all the stations used, see Table 3.1.

Figure 4.3 Output of the Delta tool for the final merged map of PM_{10} annual average for 2012 in 1x1 km resolution based on the full set of stations, against all stations of this full set



4.1.2 Map constructed with the assimilation subset of stations

In this section we evaluate the map created with the assimilation subset of stations and evaluate it against the validation subset as described in Section 3.2. The advantage of this approach is that the map evaluation happens with other stations than were used in the map construction itself. The disadvantage is the lower quality of this map due to the fact that it is constructed with about half of the available stations only, see Table 3.1.

Figure 4.4 shows the benchmarking report for the PM_{10} annual average map constructed with the assimilation subset of stations and evaluated against the rural background stations of the validation subset. In comparison with Figure 4.1, one can see worse results, caused both by the worse quality of the map created from the lower number of stations and by the validation with stations not used in the map construction. The *MQO* criterion is not fulfilled, although quite close to fulfilment with its 86%.

The performance criteria are fulfilled for the normalized standard deviation, but not for the normalized correlation.

Figure 4.4 Output of the Delta tool for the final merged map of PM_{10} annual average for 2012 in 1x1 km resolution based on the assimilation subset of stations, against rural background stations of the validation subset

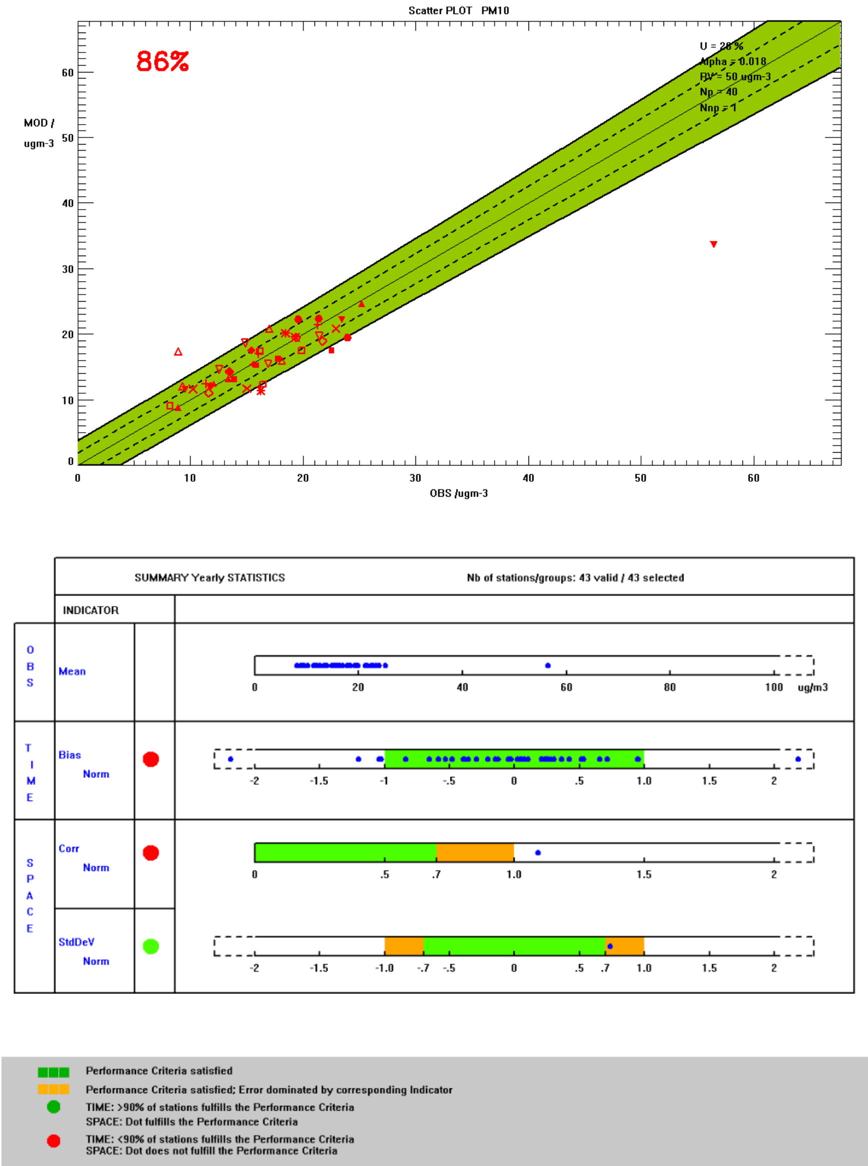


Figure 4.5 shows the benchmarking report for the PM_{10} annual average map constructed with the assimilation subset of stations and evaluated against the urban/suburban background stations of the validation subset. In comparison with Figure 4.2, one can see poorer results, caused by the same reason as in the case of the rural stations. The *MQO* criterion is with 79% not fulfilled. The performance criteria are fulfilled for the normalized standard deviation, but not for the normalized correlation, according to the benchmarking report. However, it should be noted that if calculated outside the Delta software, the normalized correlation criterion is fulfilled. For the discussion of the normalized correlation calculation in the Delta software, see Section 2.2.2.

Figure 4.5 **Output of the Delta tool for the final merged map of PM_{10} annual average for 2012 in 1x1 km resolution based on the assimilation subset of stations, against urban/suburban background stations of the validation subset**

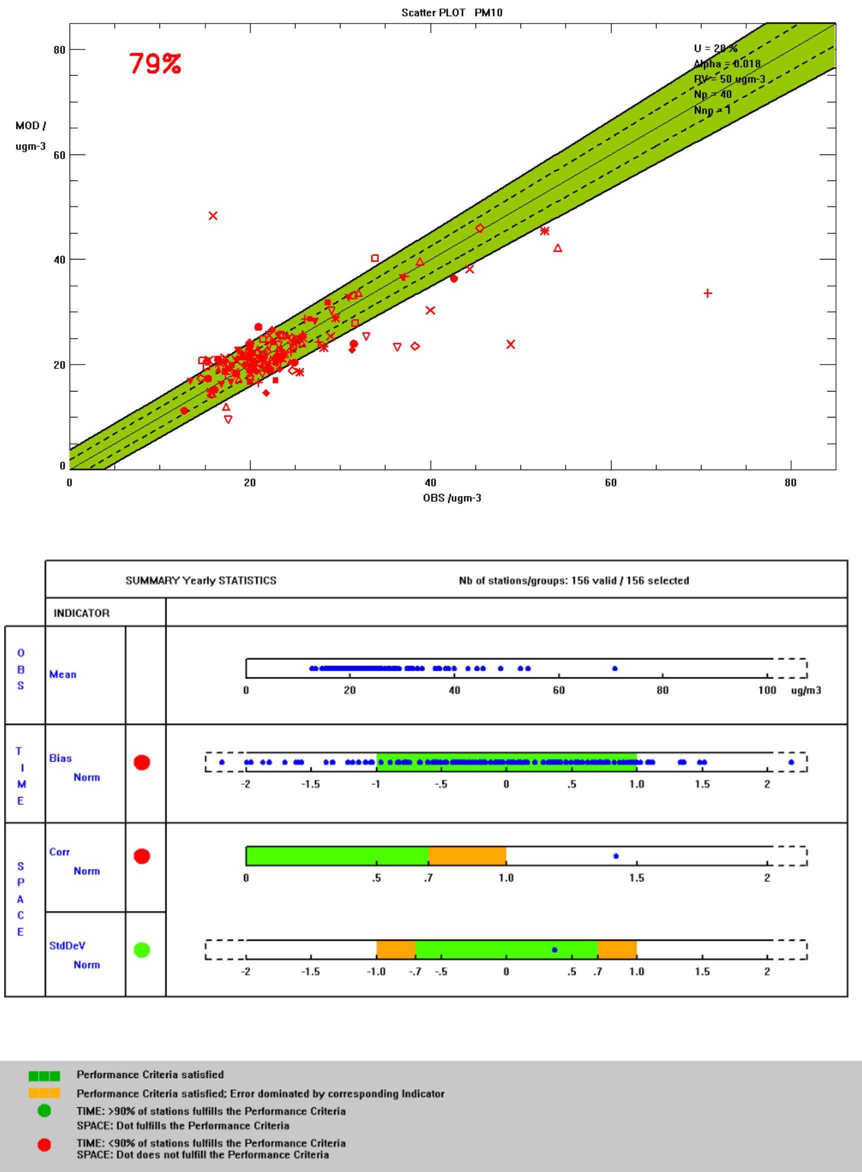
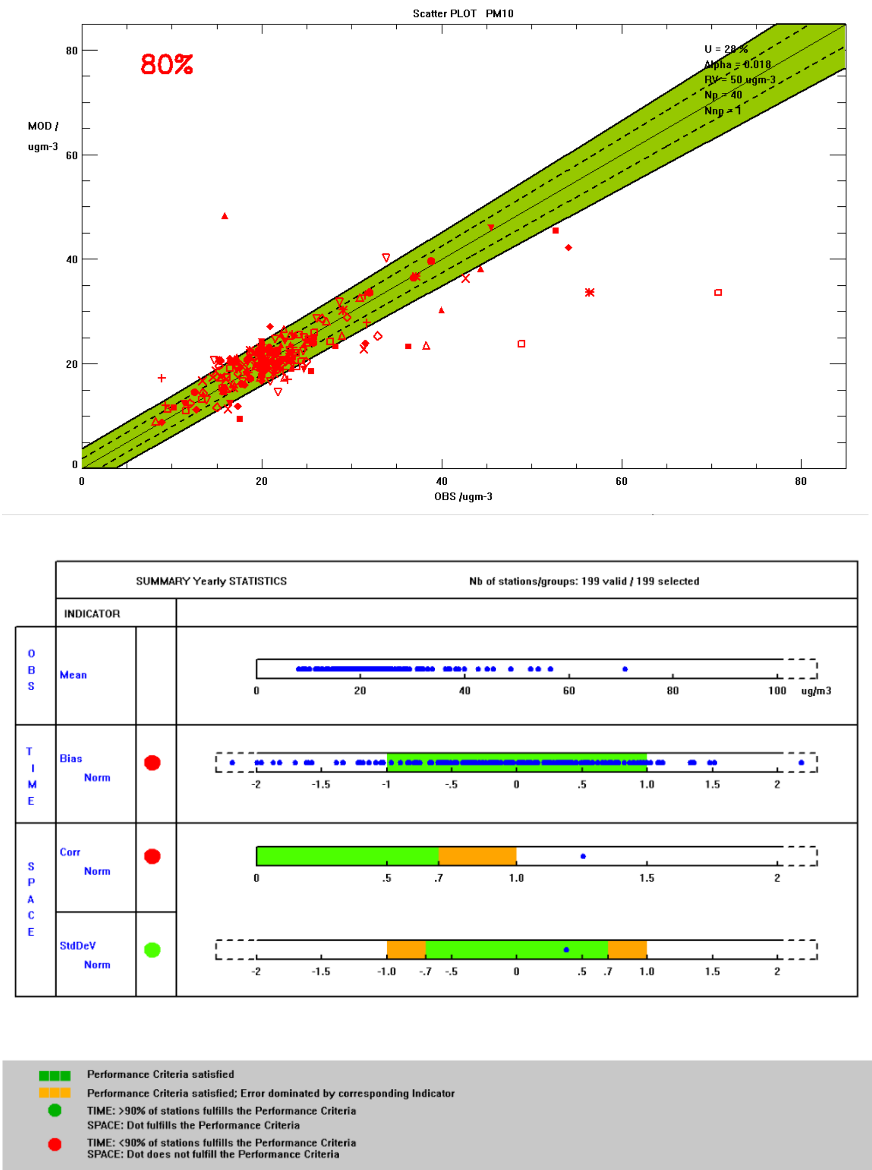


Figure 4.6 shows the benchmarking report for the PM_{10} annual average map constructed with the assimilation subset of stations and evaluated against all stations of the validation subset. The similar findings like for Figures 4.4 and 4.5 can be stated.

As ever, the results for the evaluation by all the stations are “in between” the results for the evaluation by the rural and the urban/suburban stations. The *MQO* criterion is with 80% closer to the results for the urban/suburban stations (i.e. 79%). This is caused mainly by the fact that the number of rural stations is lower than one quarter of all stations used, see Table 3.1.

Figure 4.6 Output of the Delta tool for the final merged map of PM_{10} annual average for 2012 in 1x1 km resolution based on the assimilation subset of stations, against all stations of the validation subset



4.2 PM_{10} – 36th highest daily mean

4.2.1 Map constructed with the full set of stations

The analysis for the PM_{10} indicator 36th highest daily mean is performed additionally in order to demonstrate the impact of the chosen uncertainty values for the proportional parameter N_p (see Table 3.4) for the Delta tool results. In the Delta tool performance, just the same uncertainty indicators as at the PM_{10} annual average map are used, with the exception of the N_p parameter, see Table 3.4. In fact, the measurement uncertainty for the annual average is calculated according to Equation 2.13, while the measurement uncertainty for the percentile values (including the 36th highest daily mean) is based on Equation 2.12. The difference is in the parameters N_p and N_{np}, which are not established under Equation 2.12, being in fact set equal to *1*. Whereas for the PM_{10} annual average the parameter N_p is set to 40 and

the parameter N_{np} equal to 1. Thus, the difference in the parameters used for the two PM_{10} indicators is just in the different values set for parameter N_p, which is set equal to 40 for the annual average and set equal to 1 for the 36th highest daily mean.

Figure 4.7 shows the benchmarking report for the PM_{10} indicator 36th highest daily mean map constructed with the full set of stations (Map 3.4 and evaluated against the rural background stations of this full set.

It can be seen that the *Model Quality Objective* criterion is fulfilled: $MQO < 1$ for 99 % of the stations (more than 90 %) and the performance criteria are fulfilled as well. In comparison with the results presented in Figure 4.1 for the PM_{10} annual average, one can see a wider green area in which the stations fulfil the target criterion (i.e. MQO < 1). This difference is caused by the setting of the uncertainty parameter N_p in the Delta tool being equal to 1, see above.

Figure 4.7 Output of the Delta tool for the final merged map of the PM_{10} indicator 36th highest daily mean for 2012 in 1x1 km resolution based on the full set of stations, against rural background stations of this full set

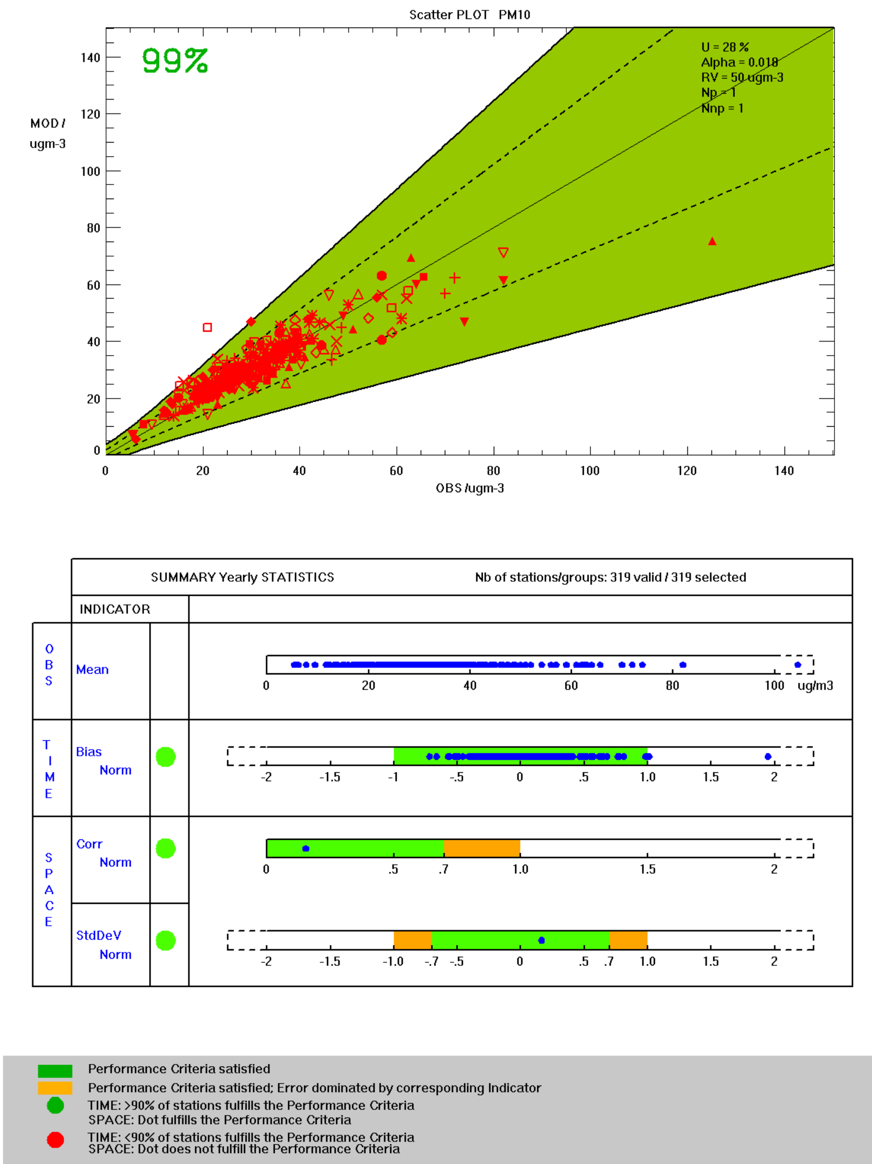


Figure 4.8 shows the benchmarking report for the same map created with the full set of the stations (Map 3.4, however now evaluated against the urban and suburban background stations of this full set. As can be seen with the 99 % of the stations complying with a *MQO < 1*, the *MQO* criterion is fulfilled in the urban areas. The performance criteria are fulfilled as well. Compared to the benchmarking report for the annual average (see Figure 4.2), one can see far better results of the *MQO* criterion at the 36th highest daily mean. The reason is the same as for the rural stations, see above.

Figure 4.8 Output of the Delta tool for the final merged map of the PM_{10} indicator 36th highest daily mean for 2012 in 1x1 km resolution based on the full set of stations, against urban/suburban background stations of this full set

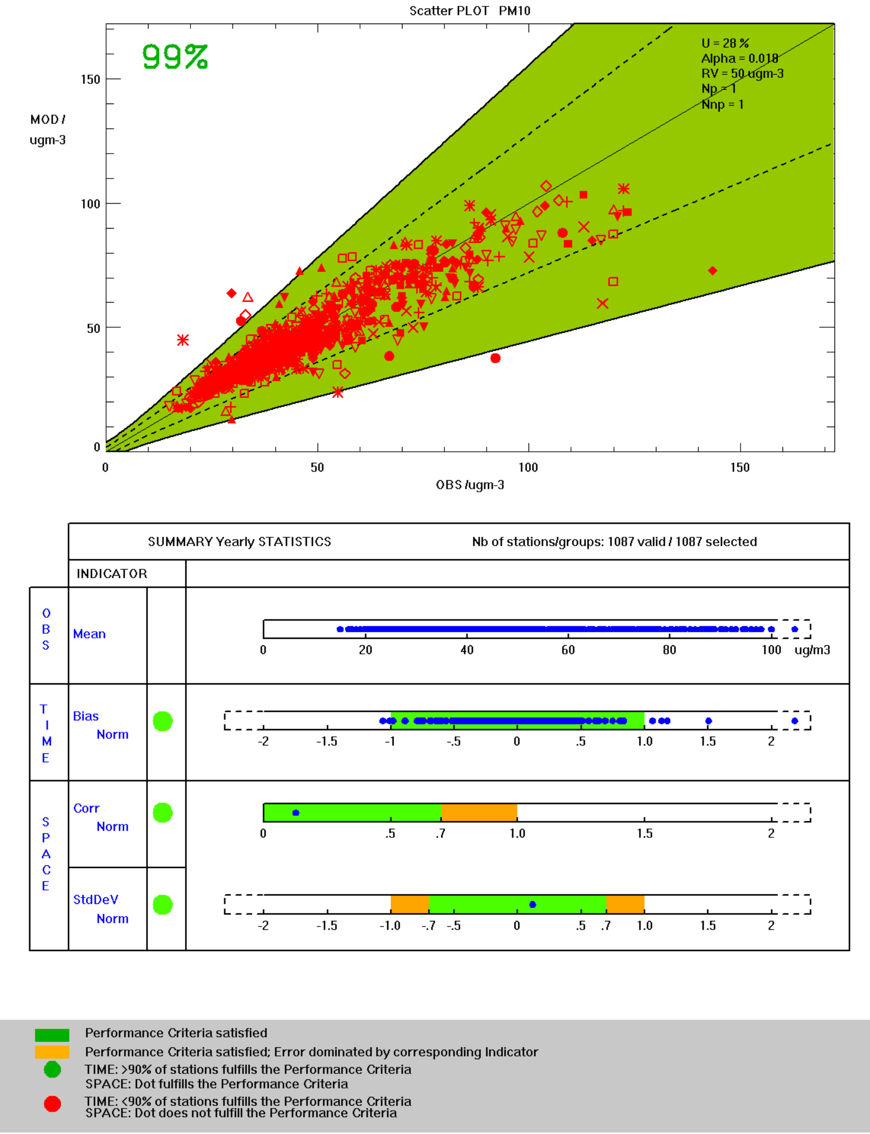
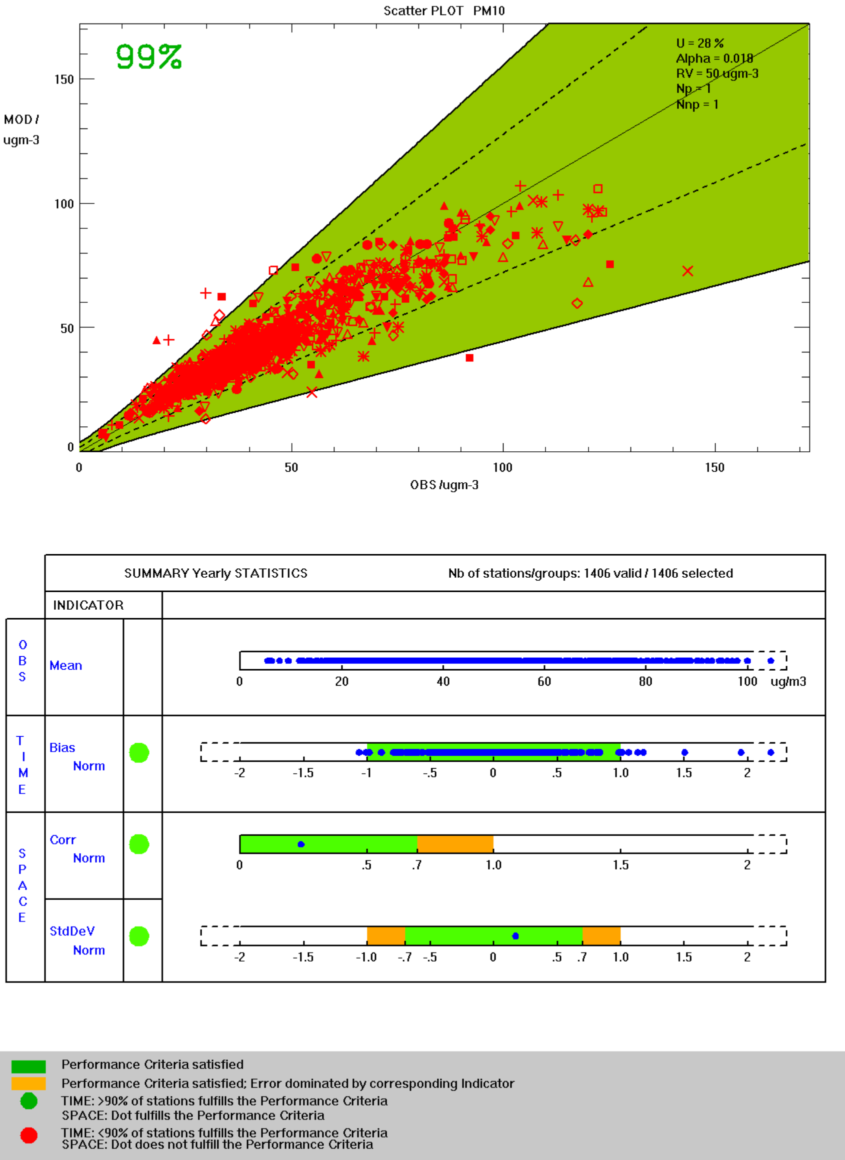


Figure 4.9 presents the benchmarking report for the same map created with the full set of stations (Map 3.4, which is evaluated against all stations (i.e. both rural and urban/suburban background) of this full set. The report shows similar results as in Figures 4.7 and 4.8. The *MQO* criterion is fulfilled: *MQO < 1* for 99 % of the stations. Both performance criteria are fulfilled as well.

Figure 4.9 **Output of the Delta tool for the final merged map of the PM_{10} indicator 36th highest daily mean for 2012 in 1x1 km resolution based on the full set of stations, against all stations of this full set**



4.2.2 Map constructed with the assimilation subset of stations

In this section we evaluate the map created with the assimilation subset of stations evaluated against the validation subset, see Section 3.2.

Figure 4.10 shows the benchmarking report for the map of PM_{10} indicator 36th highest daily mean constructed with the assimilation subset of stations and evaluated against the rural background stations of the validation subset. In comparison with Figure 4.7, one can see poorer results, caused both by the worse quality of the map created from the lower number of stations and by the validation with stations not used in the map construction. However, the *MQO* criterion is still fulfilled with the 97% of the stations complying with a $MQO < 1$. The performance criteria are fulfilled as well.

Figure 4.10 **Output of the Delta tool for the final merged map of the PM_{10} indicator 36th highest daily mean for 2012 in 1x1 km resolution based on the assimilation subset of stations, against rural background stations of the validation subset**

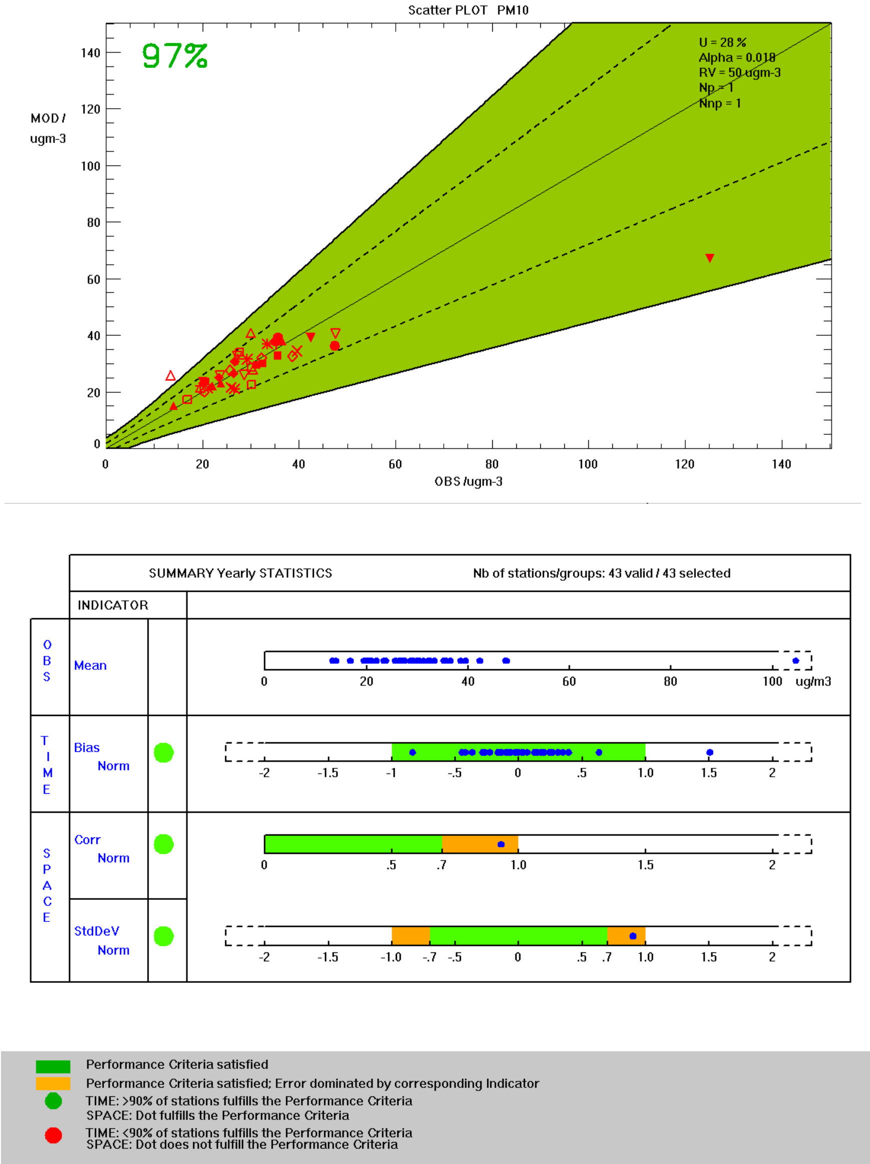


Figure 4.11 shows the benchmarking report for the same map created with the assimilation subset of the stations, now evaluated against the urban and suburban background stations of the validation subset. As can be seen with the 98 % of the stations complying with a $MQO < 1$, the *MQO* criterion is fulfilled. The performance criteria are fulfilled as well.

Figure 4.11 Output of the Delta tool for the final merged map of the PM_{10} indicator 36th highest daily mean for 2012 in 1x1 km resolution based on the assimilation subset of stations, against urban/suburban background stations of the validation subset

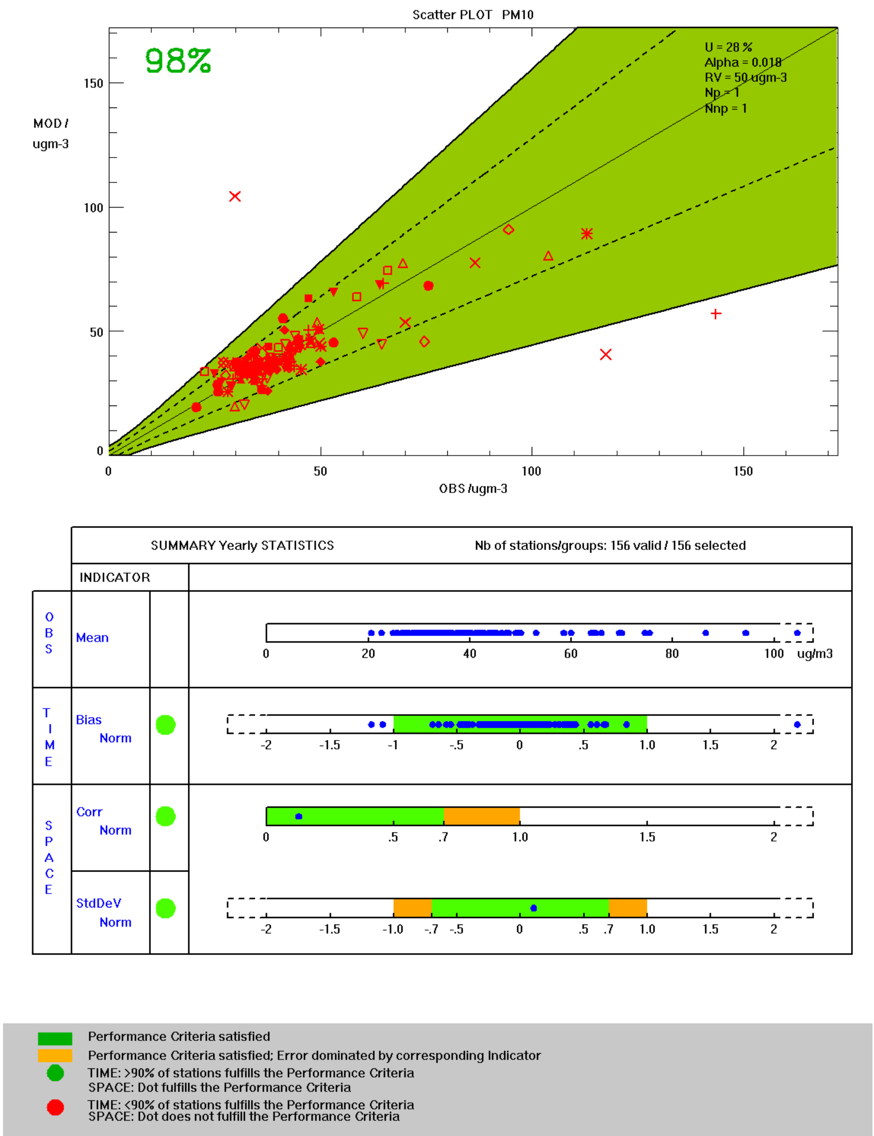
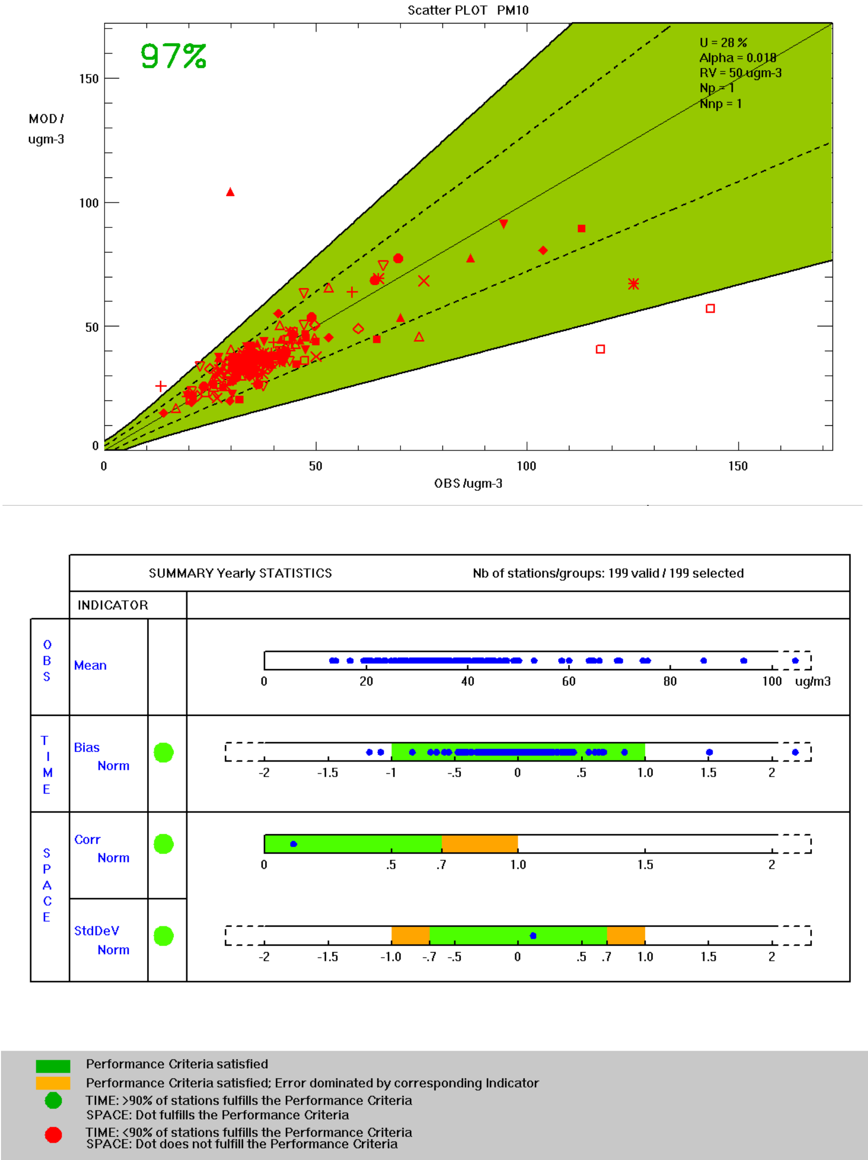


Figure 4.12 presents the benchmarking report for the same map created with the assimilation subset of stations, which is evaluated against all stations (i.e. both rural and urban/suburban background) of the validation subset. The similar results as in Figures 4.10 and 4.11 are visible. The *MQO* criterion is fairly fulfilled, as well as the both performance criteria.

Compared the Figures 4.10 – 4.12 to the benchmarking reports for the annual average (see Figures 4.4 – 4.6), one can see far better results of the *MQO* criterion at the 36th highest daily mean. The reason is the different setting of the value of the uncertainty parameter N_p in the Delta tool, see above.

Figure 4.12 Output of the Delta tool for the final merged map of the PM_{10} indicator 36th highest daily mean for 2012 in 1x1 km resolution based on the assimilation subset of stations, against all stations of the validation subset



4.3 $PM_{2.5}$ – annual average

4.3.1 Map constructed with the full set of stations

For $PM_{2.5}$ annual average, the similar analysis as for the PM_{10} annual average was done, using the Delta tool. In this section we first evaluated the map of the $PM_{2.5}$ annual average that was created using the full set of the stations (Map 3.5) against the rural background stations of this full set. Figure 4.13 presents the benchmarking report of this evaluation. At the upper right corner of the graph, the uncertainty parameters used for $PM_{2.5}$ (see Table 3.4) are presented. The expanded uncertainty *U* is set equal to 36% and the non-proportional fraction of the uncertainty *Alpha* is set equal to 0.035. The

parameter of the proportional part of the uncertainty attributed to the time averaging of the annual average N_p is set equal to 40, i.e. the same as for the PM_{10} annual average.

The results of the benchmarking report are quite similar to those the PM_{10} annual average in the rural areas (Figure 4.1). The *MQO* criterion is fulfilled and the percentage of the stations in fulfilment is slightly better than in the case of the PM_{10} annual average (94% for $PM_{2.5}$ annual average against 92% for PM_{10} annual average). Both performance criteria are fulfilled.

Figure 4.13 Output of the Delta tool for the final merged map of $PM_{2.5}$ annual average for 2012 in 1x1 km resolution based on the full set of stations, against rural background stations of this full set

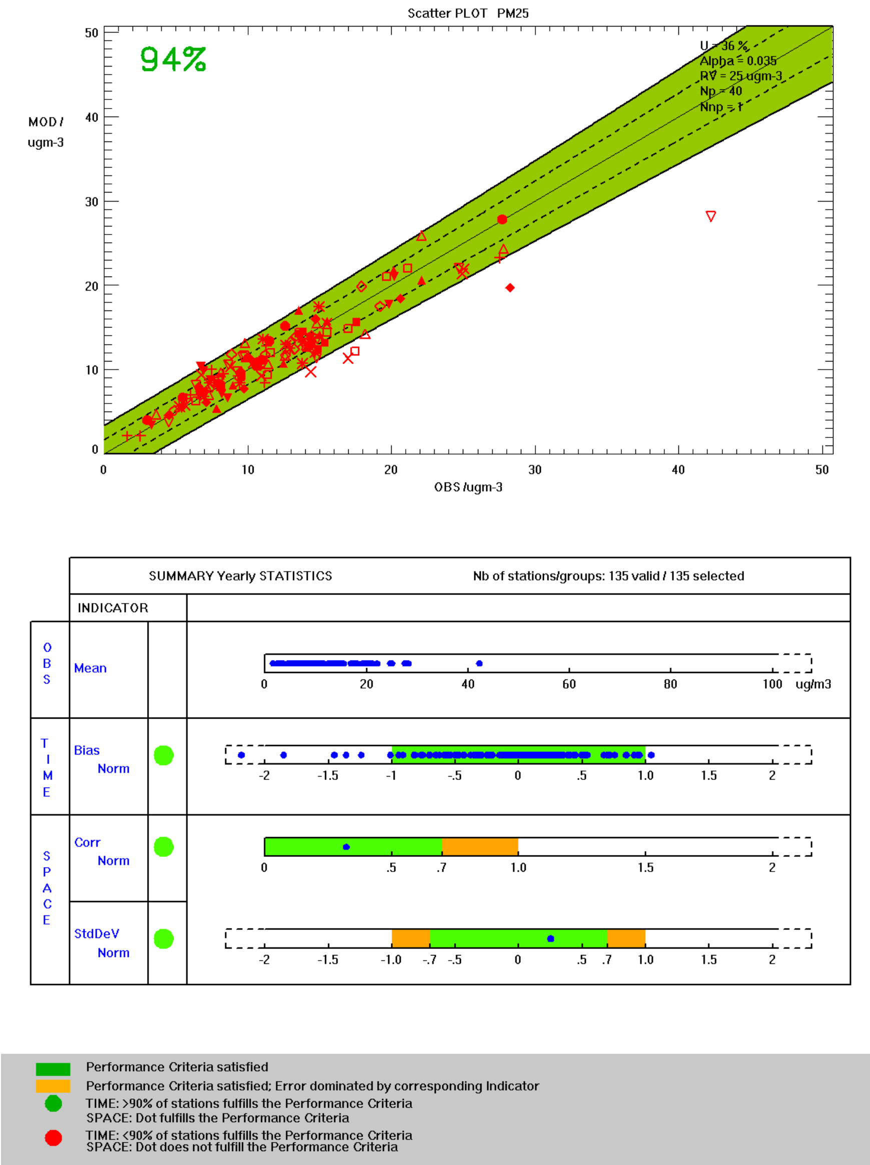
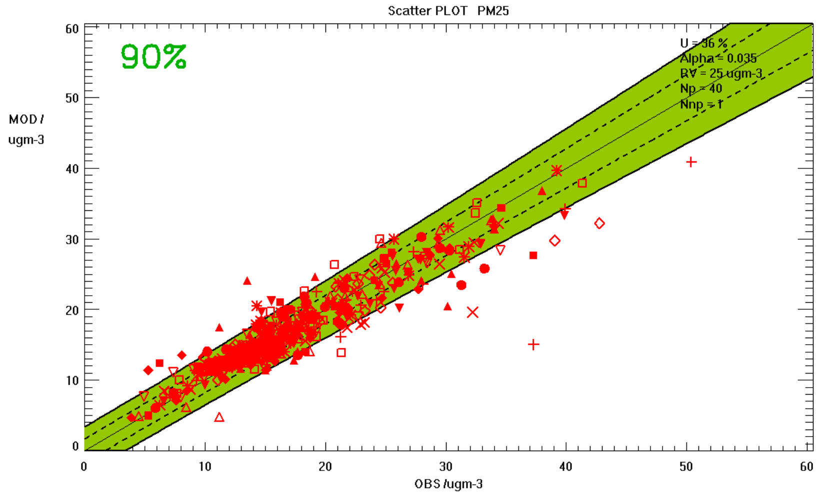
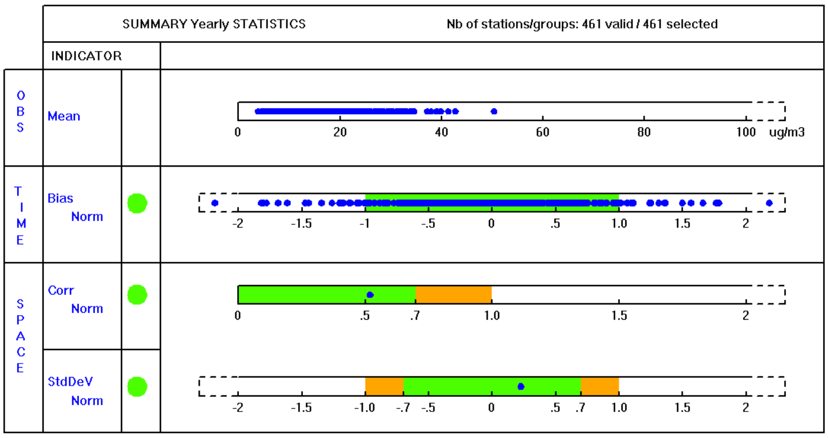


Figure 4.14 shows the benchmarking report for the same map created with the full set of the stations (Map 3.5), however now evaluated against the urban and suburban background stations of this full set. One can see the fulfilment of the *MQO* criterion: 90% of the stations are in the fulfilment, which is just the required minimum percentage. Both performance criteria are fulfilled as well.

Figure 4.14 Output of the Delta tool for the final merged map of $PM_{2.5}$ annual average for 2012 in 1x1 km resolution based on the full set of stations, against urban/suburban stations of this full set





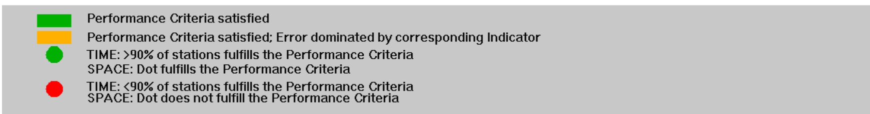
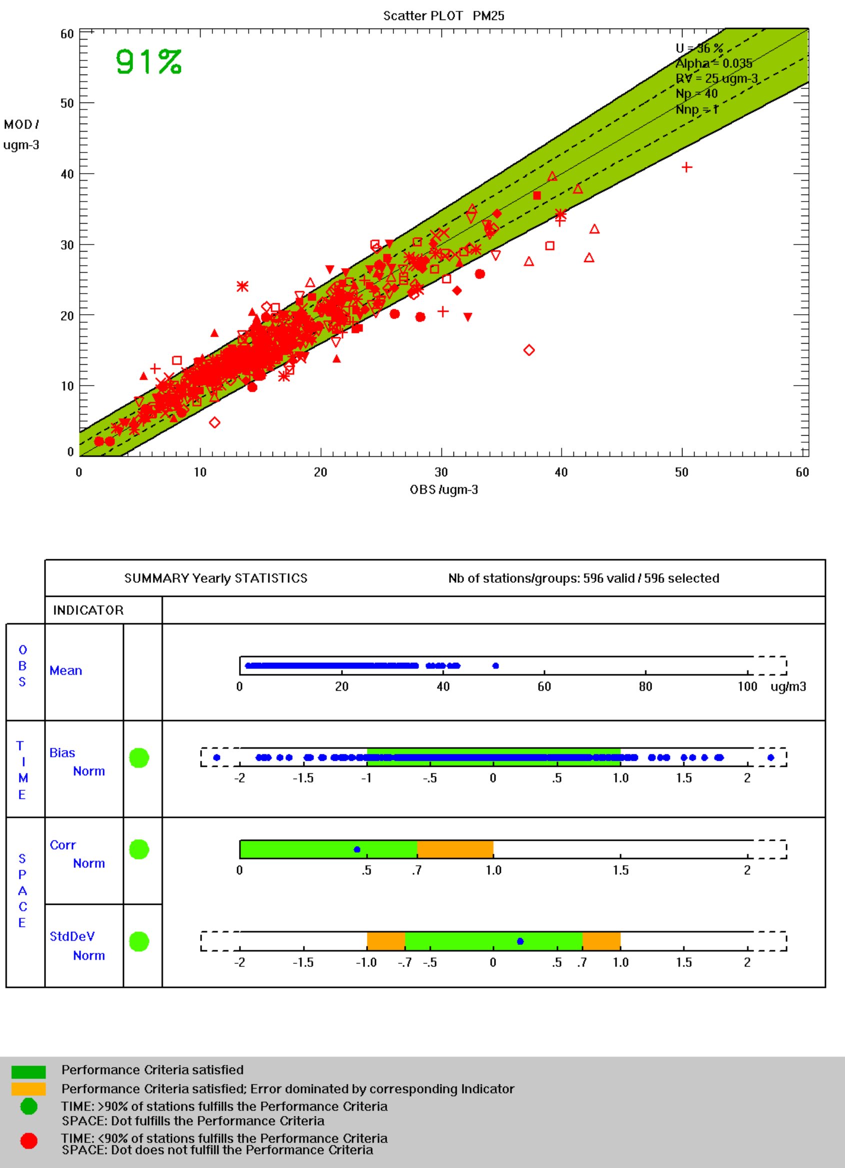


Figure 4.15 presents the benchmarking report for the same map created with the full set of the stations (i.e. Map 3.5), now evaluated against all stations (i.e. both rural and urban/suburban background) of this full set. Quite similar results are observed in Figure 4.15 compared to Figure 4.14.

The better results for $PM_{2.5}$ (Figure 4.1, 4.2, 4.3) compared to the results for the PM_{10} annual average (Figure 4.10, 4.11, 4.12) are probably given by the higher value set for the expanded uncertainty *U* in the case of $PM_{2.5}$ (36% instead 28% at PM_{10}), leading to a somewhat wider interval of fulfilment (i.e. wider green area in the graph).

Figure 4.15 Output of the Delta tool for the final merged map of $PM_{2.5}$ annual average for 2012 in 1x1 km resolution based on the full set of stations, against all stations of this full set



4.3.2 Map constructed with the assimilation subset of stations

In this section we evaluate the map created with the assimilation subset of stations evaluated against the validation subset, see Section 3.2.

Figure 4.16 shows the benchmarking report for the $PM_{2.5}$ annual average map constructed with the assimilation subset of stations and evaluated against the rural background stations of the validation subset. Compared to Figure 4.13, one can see worse results in this case: nor the *MQO* criterion, nor the performance criteria are fulfilled. This is caused both by the poorer quality of the map created with the lower number of stations and by the validation subset stations not being used in the map construction itself, which is in principle more objective. Note that 12 stations only are used in the validation subset as one can observe at the 2nd row with the limited number of stations plotted (blue dots). Ssee also the

list of all stations below the scatterplot. The non-fulfilment of the performance criteria is caused by one of these twelve stations (i.e. CZ0TVER) with a quite extreme high $PM_{2.5}$ value way out of the green area of the scatterplot.

Figure 4.16 Output of the Delta tool for the final merged map of $PM_{2.5}$ annual average for 2012 in 1x1 km resolution based on the assimilation subset of stations, against rural background stations of the validation subset

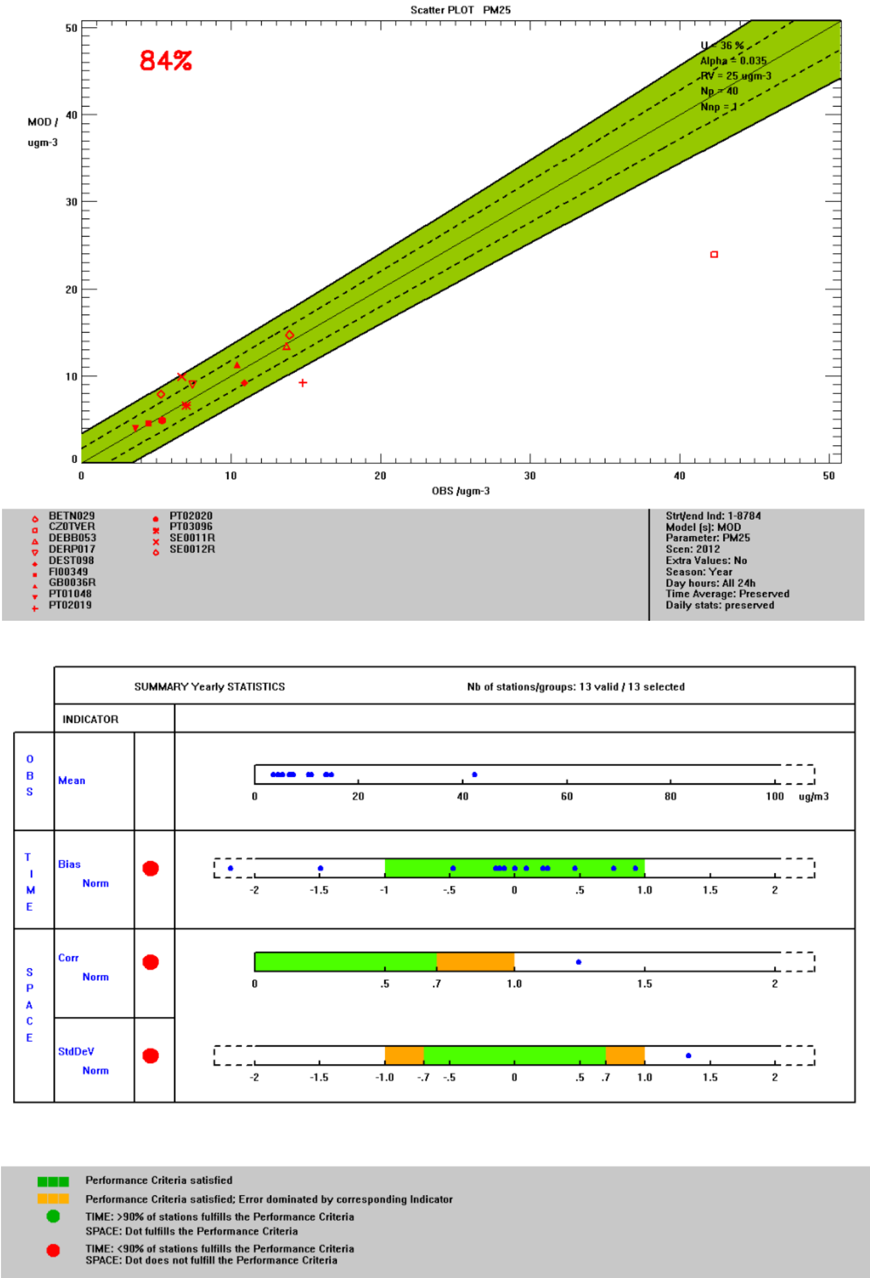


Figure 4.17 shows the benchmarking report for the $PM_{2.5}$ annual average map constructed with the assimilation subset of the stations and evaluated against the urban/suburban background stations of the validation subset. The *MQO* criterion is not fulfilled (the 84 % of urban and suburban background stations in fulfilment is equal to the percentage of the rural stations), while both performance criteria are fulfilled.

Figure 4.17 Output of the Delta tool for the final merged map of $PM_{2.5}$ annual average for 2012 in 1x1 km resolution based on the assimilation subset of stations, against urban/suburban background stations of the validation subset

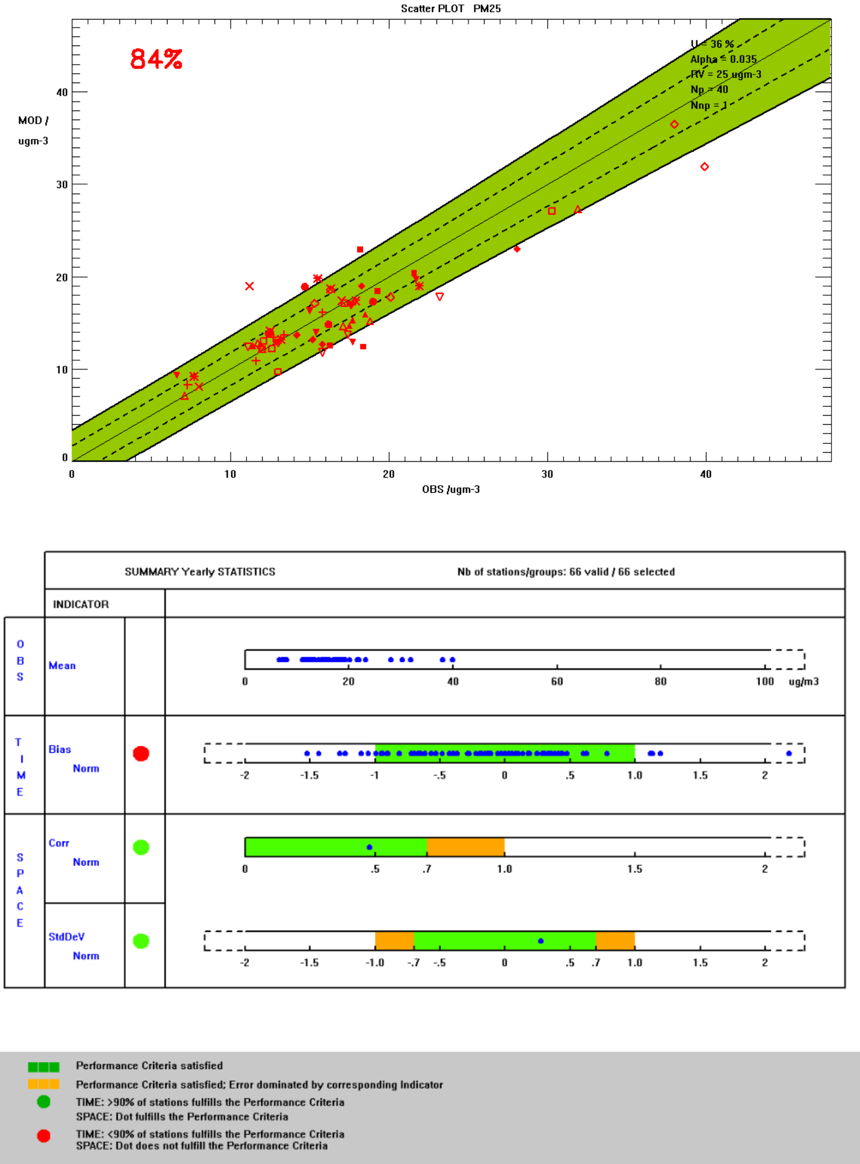
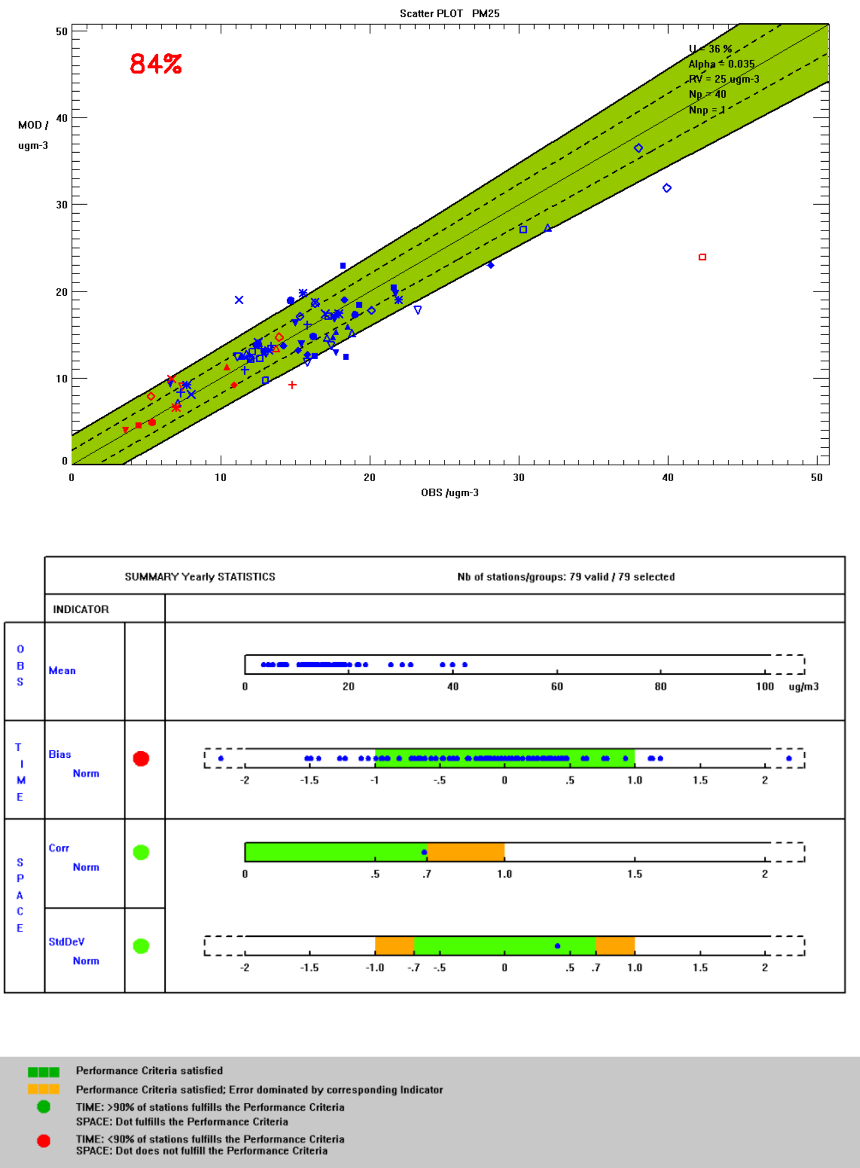


Figure 4.18 presents the benchmarking report for still the same $PM_{2.5}$ annual average map constructed with the assimilation subset of the stations and evaluated against all the stations of the validation subset. One can see the similar results as in Figure 4.17. The *MQO* criterion is not fulfilled (the percentage of the stations in fulfilment is quite the same as for the rural stations), while both performance criteria are fulfilled.

Figure 4.18 Output of the Delta tool for the final merged map of $PM_{2.5}$ annual average for 2012 in 1x1 km resolution based on the assimilation subset of stations, against all stations of the validation subset



4.4 Ozone – 26th highest daily maximum 8-hour mean

4.4.1 Map constructed with the full set of stations

For the ozone indicator 26th highest daily maximum 8-hour mean, similar analyses as for the PM_{10} and $PM_{2.5}$ annual averages has been executed with the Delta tool.

First, we evaluated the map the ozone indicator 26th highest daily maximum 8-hour mean using the full set of the stations (Map 3.5) against the rural background stations of this full set. Figure 4.19 presents the benchmarking report of this evaluation. At the upper right corner of the graph, the uncertainty parameters used for ozone (see Table 3.4) are presented. The expanded uncertainty *U* is set equal to 12% and the non-proportional fraction of the uncertainty *Alpha* is set equal to 0.62. As the indicator

26th highest daily maximum 8-hour mean represents a percentile value, the measurement uncertainty is calculated according Equation 2.12 in this case. Thus, the parameter N_p is not introduced (being in fact set equal to 1).

In Figure 4.19 one can see the *MQO* criterion is fulfilled well with 100% of the stations. Both performance criteria are fulfilled as well. Compared to the results for the annual average of both PM_{10} and $PM_{2.5}$ (see Figures 4.1 and 4.13), one can see a wider green area of the *MQO* fulfilment (i.e. *MQO* < 1). The difference is caused by the setting of the uncertainty parameter N_p in the Delta tool being equal to 1 for percentile values, while it is set to 40 for annual averages. Compared to the results of the PM_{10} indicator 36th highest daily mean, one can see the wider green area specifically at low concentrations. This is caused by the setting of the indicator *Alpha* showing the non-proportional uncertainty fraction, which is 0.62 for ozone, while 0.014 only for PM_{10}.

Figure 4.19 Output of the Delta tool for the final merged map of the ozone indicator 26th highest daily maximum 8-hour mean for 2012 in 1x1 km resolution based on the full set of stations, against rural background stations of this full set

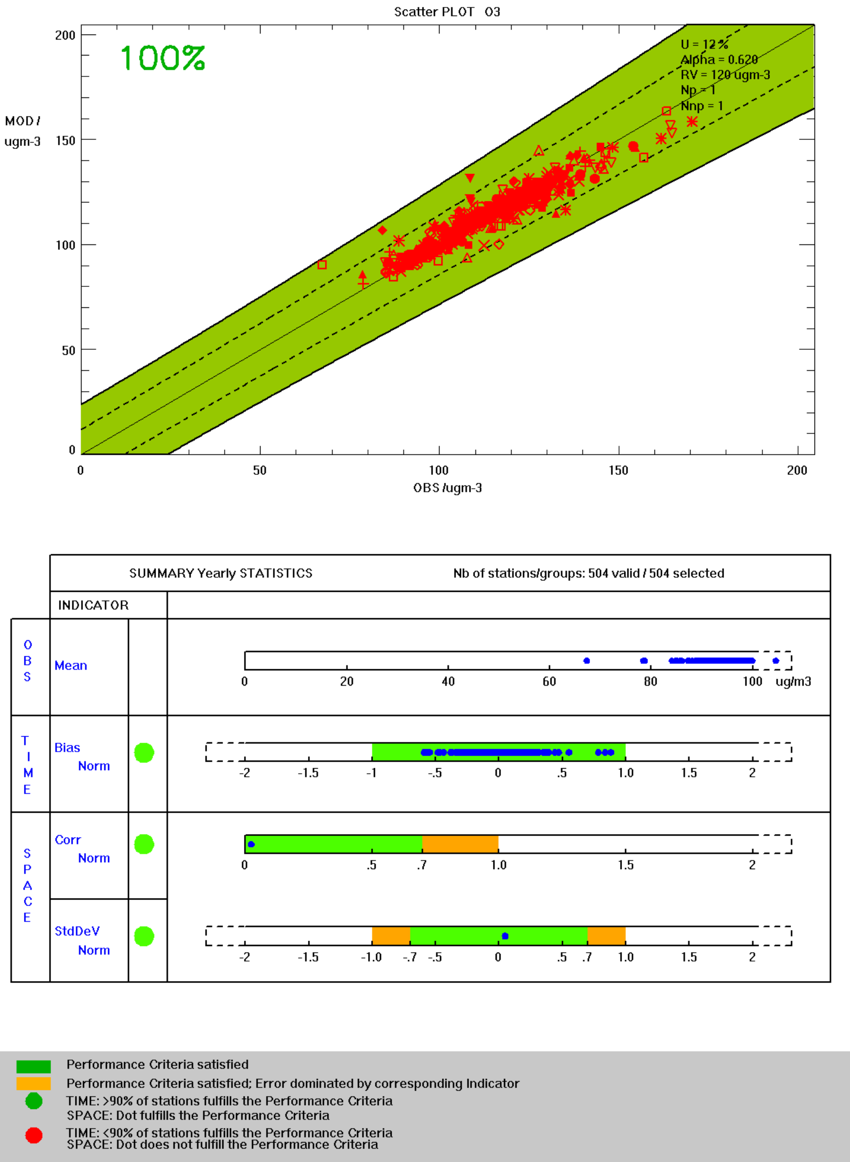


Figure 4.20 shows the benchmarking report for the same map created with the full set of the stations (i.e. Map 3.5), however now evaluated against the urban and suburban background stations of this full set. As can be seen, the *MQO* criterion is fulfilled in the urban areas with its 98 %. The performance criteria are fulfilled as well. The slightly poorer results of the *MQO* criterion compared to the rural areas in Figure 4.19 (98% for the urban areas against 100% for the rural areas) are caused probably partly by the spatial averaging of the values in the 10x10 km grid cells (see Section 2.1), partly by the higher nugget value in the case of the urban map (leading to smoothing at the station points), see Table 3.3.

Figure 4.20 Output of the Delta tool for the final merged map of the ozone indicator 26th highest daily maximum 8-hour mean for 2012 in 1x1 km resolution based on the full set of stations, against urban/suburban background stations of this full set

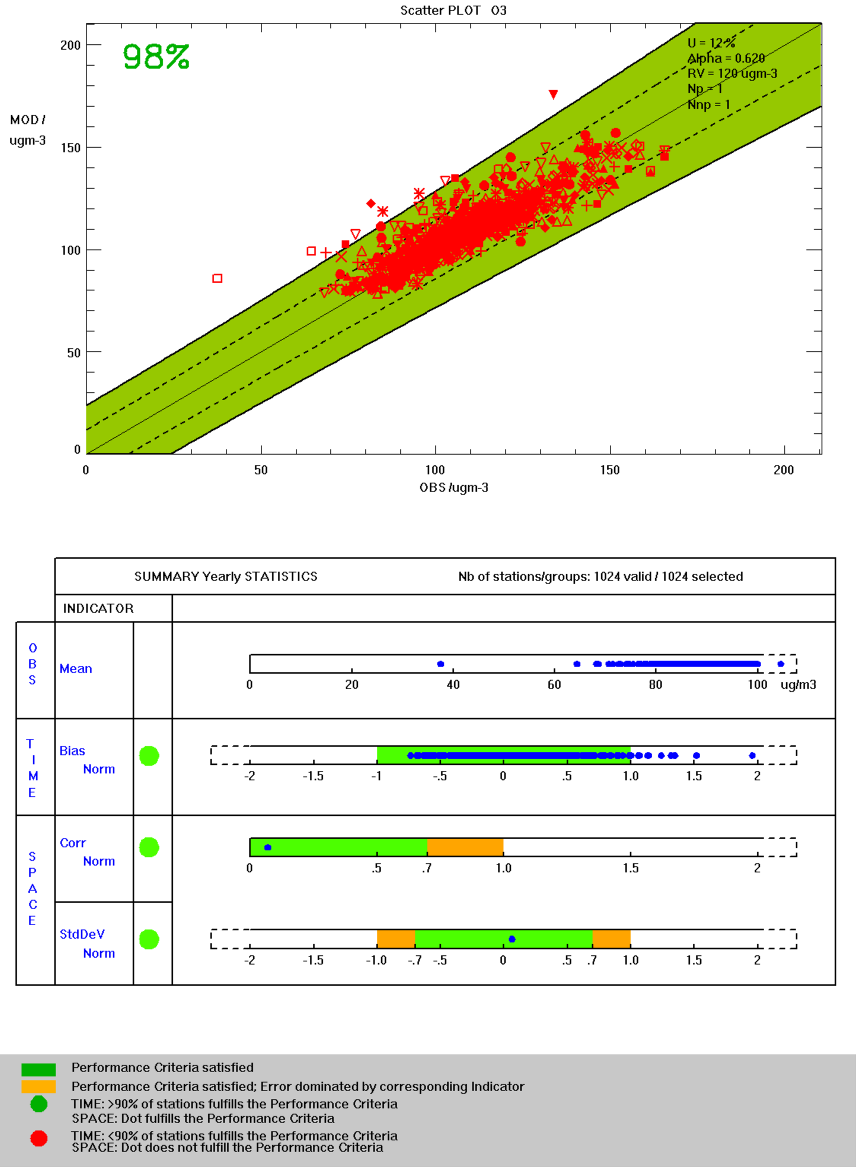
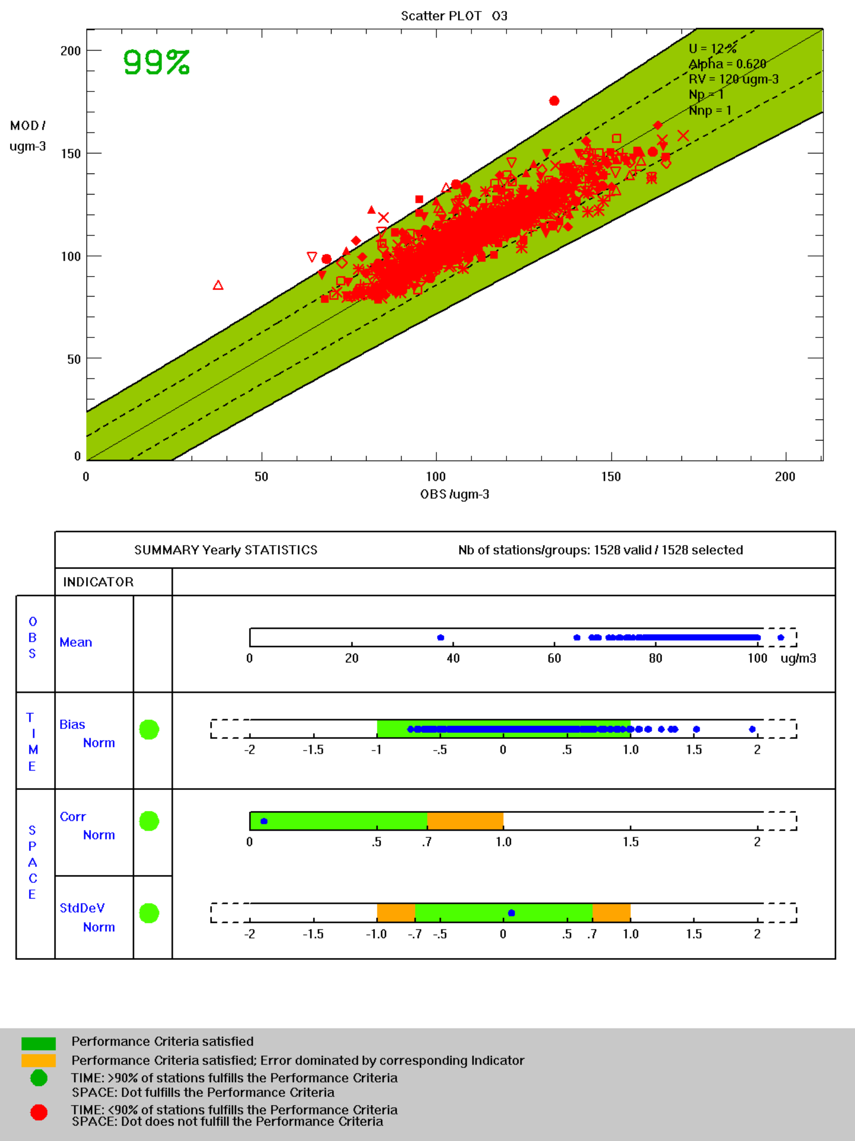


Figure 4.21 presents the benchmarking report for the same map created with the full set of the stations (Map 3.5), now evaluated against all stations (i.e. both rural and urban/suburban background) of this full set. The fulfilment of the *MQO* criterion, as well as both of performance criteria are given and are quite similar to those of Figure 4.20.

Figure 4.21 Output of the Delta tool for the final merged map of the ozone indicator 26th highest daily maximum 8-hour mean for 2012 in 1x1 km resolution based on the full set of stations, against all stations of this full set



4.4.2 Map constructed with the assimilation subset of stations

Ultimately, the map created with the assimilation subset of stations was evaluated against the validation subset, see Section 3.2.

Figure 4.22 shows the benchmarking report for the map of the ozone indicator 26th highest daily maximum 8-hour mean constructed with the assimilation subset of the stations and evaluated against the rural background stations of the validation subset. Contrary to both the PM_{10} and the $PM_{2.5}$ annual average results (see Figures 4.4 and 4.16), the *MQO* criterion is fulfilled although this map is created with a limited set of stations, compared to the map constructed with the full set of stations. The performance criteria are fulfilled as well.

Figure 4.22 Output of the Delta tool for the final merged map of the ozone indicator 26th highest daily maximum 8-hour mean for 2012 in 1x1 km resolution based on the assimilation subset of stations, against rural background stations of the validation subset

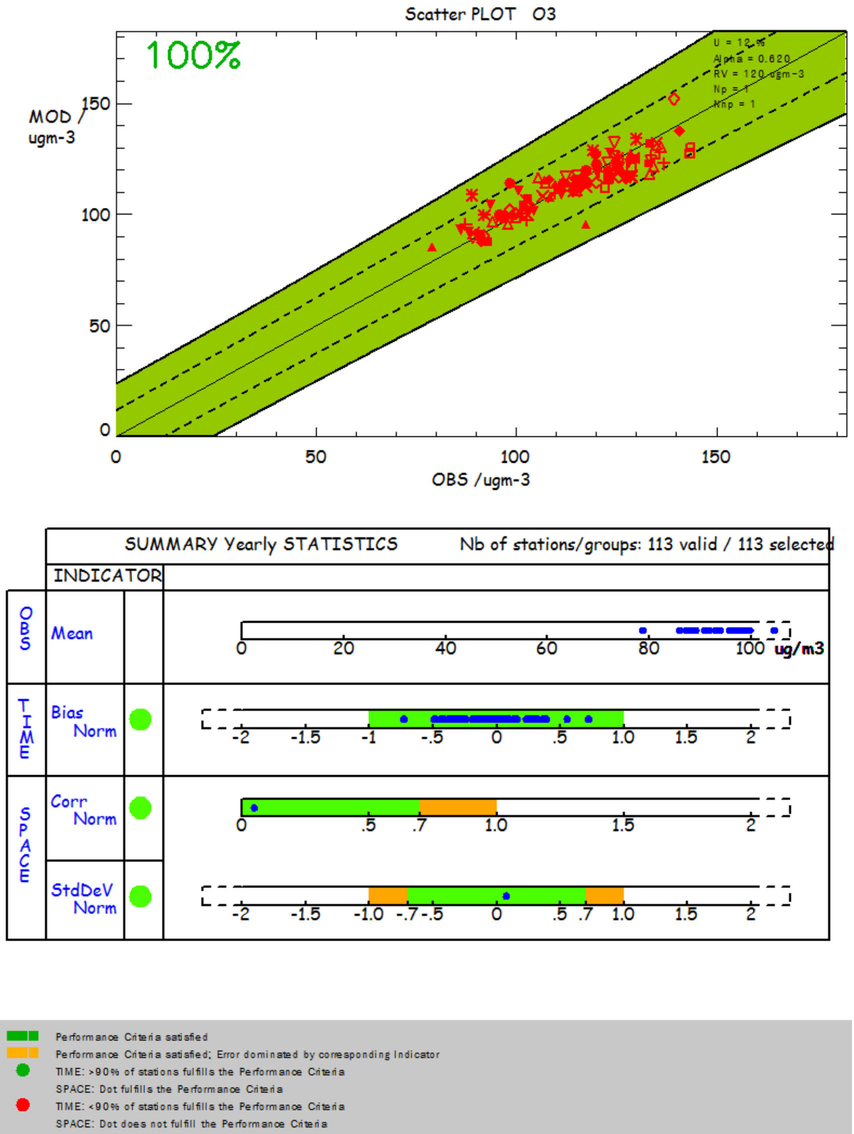


Figure 4.23 shows the benchmarking report for the same map constructed with the assimilation subset of stations and evaluated against the urban/suburban background stations of the validation subset. The *MQO* criterion and both performance criteria are well fulfilled.

Figure 4.23 Output of the Delta tool for the final merged map of the ozone indicator 26th highest daily maximum 8-hour mean for 2012 in 1x1 km resolution based on the assimilation subset of stations, against urban/suburban background stations of the validation subset

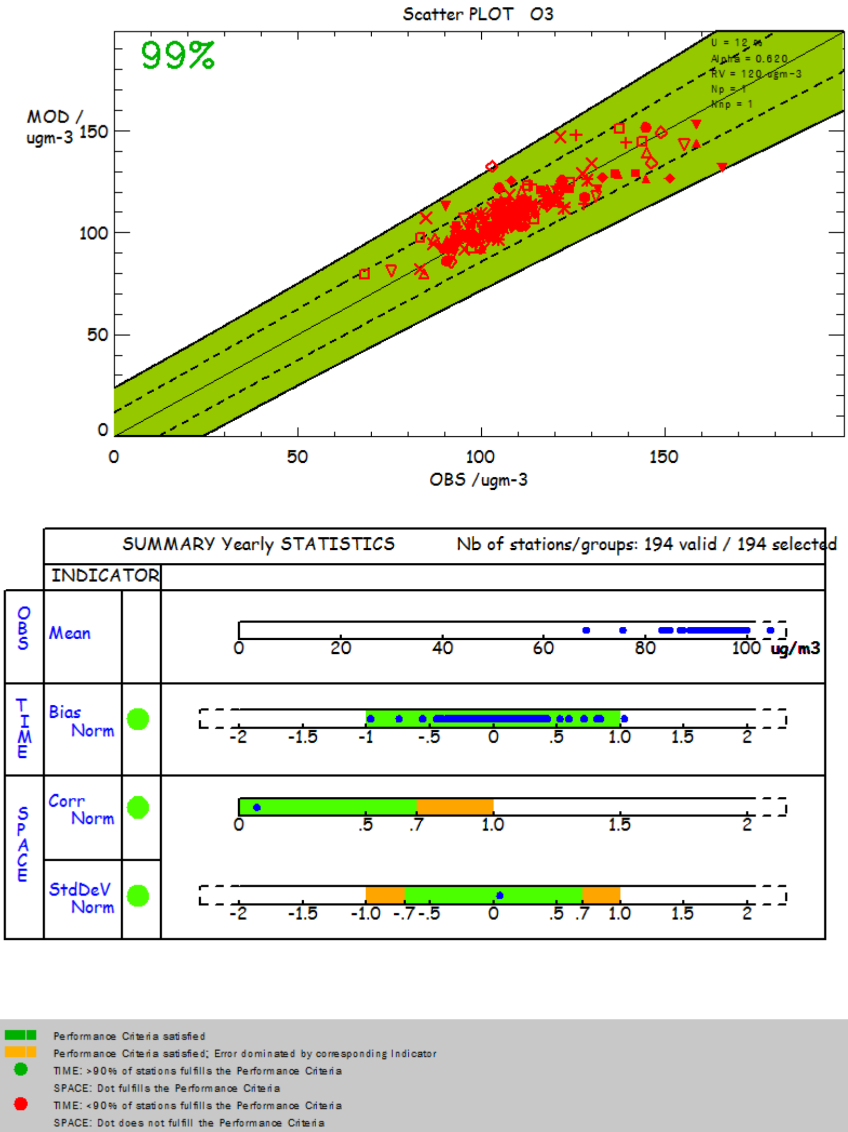
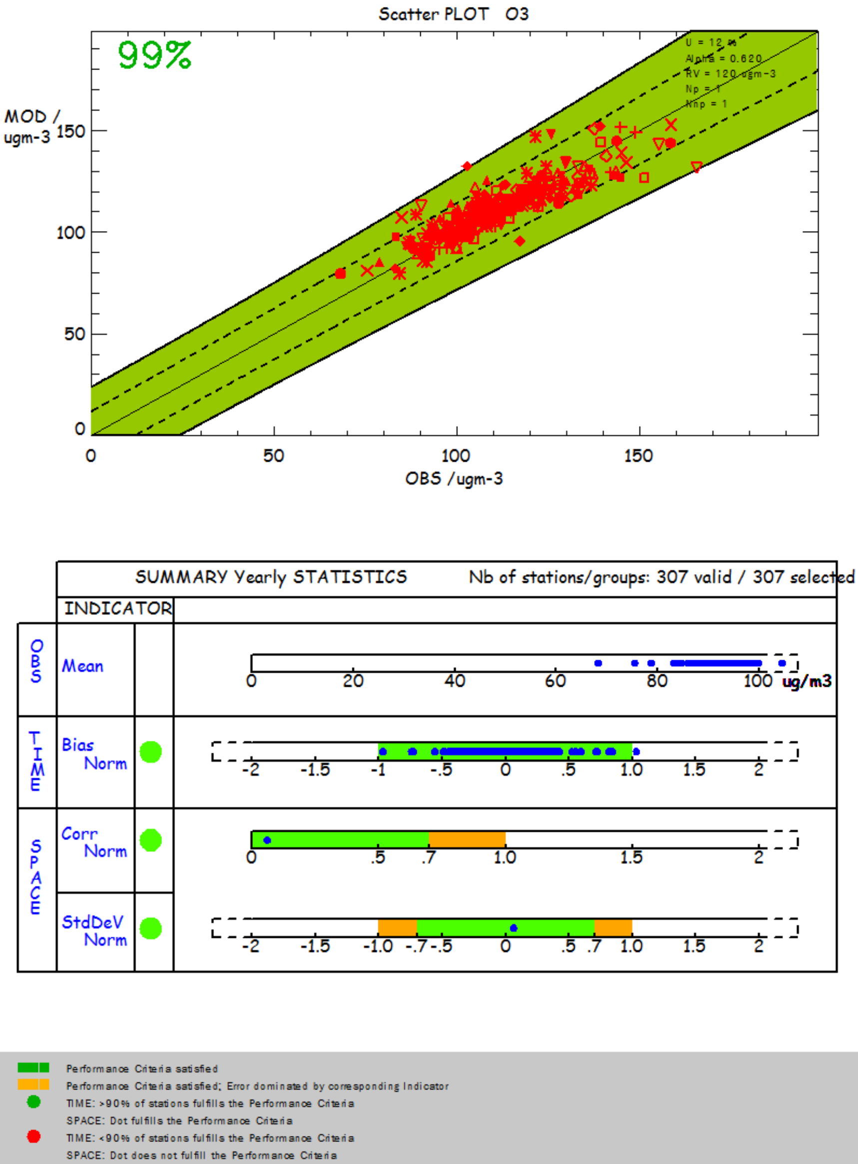


Figure 4.24 presents the benchmarking report for still the same map of the ozone indicator 26th highest daily maximum 8-hour mean constructed with the assimilation subset of stations, now evaluated against all stations (i.e. both rural and urban/suburban background) of this validation subset. The *MQO* criterion and both performance criteria are well fulfilled although the map is compared to the full set constructed with a rather limited number of the stations (i.e. 307, see Table 3.1), and that it is evaluated against stations that are not used in the map construction itself, which leads in most cases to lower fit but it is in itself more objective for evaluation purposes.

Figure 4.24 Output of the Delta tool for the final merged map of the ozone indicator 26th highest daily maximum 8-hour mean for 2012 in 1x1 km resolution based on the assimilation subset of stations, against all stations of the validation subset



4.5 Summary

The spatially interpolated maps of PM_{10} annual average, the PM_{10} indicator 36th highest daily mean, $PM_{2.5}$ annual average, and the ozone indicator 26th highest daily maximum 8-hour mean were evaluated with the Delta tool. The evaluation was executed on the one hand using the full set of stations for the maps constructed with the same full set of the stations. On the other hand, for the maps constructed with the assimilation subset of the stations the evaluation was done by using the validation subset of stations. All the maps were evaluated separately by the rural background stations, by the urban/suburban background stations and by all (i.e. both rural and urban/suburban background) stations of the involved set.

Table 4.1 summarises the performance results of the Delta tool evaluation. The columns of the *MQO* criterion show the percentage of the stations in fulfilment (i.e. *MQO* < 1). The percentage equal to or higher than 90% is in green indicating the *MQO* criterion fulfilment, while the percentage lower than 90% is in red indicating the non-fulfilment of the *MQO* criterion. The columns of the performance criteria show the fulfilment (indicated by '+' in green) resp. non-fulfilment (indicated by '-' in red) of the relevant performance criterion.

For PM_{10} annual average, the table summarizes the results presented in Figures 4.1 – 4.3 (for the map based on the full set) and Figures 4.4 – 4.6 (for the map based on the assimilation set). In the case of PM_{10} indicator 36th highest daily mean, it summarizes the results presented in Figures 4.7 – 4.9 (for the map based on the full set) and Figures 4.10 – 4.12 (for the map based on the assimilation subset). For $PM_{2.5}$, it summarizes the results presented in Figures 4.13 – 4.15, resp. in Figures 4.16 – 4.18. For ozone, it shows the results presented in Figures 4.19 – 4.21, resp. in Figures 4.22 – 4.24.

The numbers of the stations used in the maps' creation and evaluation are presented in Table 3.1.

Table 4.1 Summary results of the Delta tools evaluation. Green colour shows fulfilment, while red colour non-fulfilment of the relevant criterion.

Pollutant and indicator	Type of stations used for evaluation	Map based on full station set and evaluated by the same set			Map based on assimilation subset evaluated by validation subset		
		MQO criterion	Performance criteria		*MQO* criterion	Performance criteria	
			Corr.	St. dev.		Corr.	St. dev.
PM_{10} annual average	Rural background	92%	+	+	86%	- (a)	+
	Urban/suburb. b.	85%	+	+	79%	-	+
	All	87%	+	+	80%	-	+
PM_{10} 36th highest daily mean	Rural background	99%	+	+	97%	+	+
	Urban/suburb. b.	99%	+	+	98%	+	+
	All	99%	+	+	97%	+	+
$PM_{2.5}$ annual average	Rural background	94%	+	+	84%	-	-
	Urban/suburb. b.	90%	+	+	84%	+	+
	All	91%	+	+	84%	+	+
Ozone 26th highest d. max. 8-h.	Rural background	100%	+	+	100%	+	+
	Urban/suburb. b.	98%	+	+	98%	+	+
	All	99%	+	+	99%	+	+

(a) If calculated outside the Delta software according to Equation 2.9, the correlation criterion is fulfilled.

For the maps based on the full station set and evaluated by the same set, the *MQO* criterion is fulfilled in most the cases. The exception is PM_{10} annual average, for urban/suburban background stations and for all the stations. In the case of the percentile (resp. x-th highest) indicators, the *MQO* criterion is fulfilled also for the maps based on the assimilation subset and evaluated by the validation subset of the stations. Contrary to this, for annual average maps the *MQO* criterion is not fulfilled in the case of the interpolation based on the assimilation subset and evaluated by the validation subset of the stations.

The performance criteria are fulfilled for all the maps based on the full station set and evaluated by the same set. In the case of the maps based on the assimilation subset and evaluated by the validation subset of the stations, the performance criteria are fulfilled in most the cases. The exception is the correlation criterion for PM10 annual average, and also both the performance criteria for $PM_{2.5}$ in the rural areas.

5 Discussion of the results and conclusions

5.1 Discussion of the results

In this section we discuss the results of the Delta tool presented in Chapter 4.

First should be noted that the *Model Quality Objective (MQO)* introduced by Thunis et al. (2012) and used in the Delta tool (Thunis et al., 2015) is stricter (at least for the annual averages) than the requirements for the models established in the Air Quality Directive (EC, 2008). The basic concept of the *MQO* is that the model uncertainty should not exceed the measurement uncertainty, see Equation 2.7. Contrary to this, the requirement for the modelling uncertainty under the Air Quality Directive (EC, 2008) is 50% for both PM_{10} and $PM_{2.5}$ annual averages and also for ozone 8-hour averages, while for the measurement uncertainty it is 25% for both PM_{10} and $PM_{2.5}$ and 15% for ozone. (The modelling uncertainty is defined in the Air Quality Directive, EC, 2008, as the maximum deviation of the measured and calculated concentration levels for 90% of individual monitoring points.) Next to this, the values of the measurement uncertainty used in the Delta tool for PM_{10} and $PM_{2.5}$ are based on the reference gravimetric method (Pernigotti, 2013), which is many times lower than the uncertainty of the beta ray method (Wesseling, 2015). The issue is still under debate within FAIRMODE.

The output of the Delta tool is very sensitive to the used monitoring uncertainty. This is clearly seen in the comparison of the results for the PM_{10} statistics annual average (Section 4.1) and 36th highest daily mean (Section 4.3). The Delta tool results give highly different results of the *MQO* criterion for the two PM_{10} indicators, see Table 4.1 and Figures 4.1–4.3 and 4.7–4.8. The difference is caused by the different measurement expanded uncertainty (*U*) as defined in the *MQO* criterion for evaluating the annual averages and for the percentile (including 36th highest) values, see Equations 2.12 and 2.13. The reason is that the measurement uncertainty of the percentile value is considered as an uncertainty of the corresponding daily value, although this is probably not fully correct. (If X is the P-th percentile and U is the uncertainty of X, the value $X+U$ perhaps is no longer the P-th percentile.)

In fact, there is no reason for such a big difference of the Delta results for the two PM_{10} indicators. The maps for both the PM_{10} indicators are constructed similarly and the routinely used uncertainty estimates based on the cross-validation give quite similar results in relative numbers for both the maps, see Table 3.2 and Horálek et al. (2015). Thus, it seems the setting of the parameters for the *MQO* estimate under the Delta tool should be adjusted for the PM_{10} annual average and percentile values.

The different measurement uncertainty used for the annual averages and for the percentile values leads generally to the different results of the Delta tool for these two types of indicators. The percentage of stations with *MQO* fulfilment is higher at the percentile values, see Table 4.1. Thus, the recommendation for the tuning of the parameters for the *MQO* estimate in case of both annual average and percentile values can be extended to all pollutants. It looks like that presently the *MQO* criterion is set too strict at the annual averages and set too loose at the percentile values.

If compared the relevant (i.e. annual average or x-th highest value) results for different pollutants, one can see slightly worse results for PM_{10} in both annual average and x-th highest value cases. In the case of the x-th highest values (PM_{10} vs. ozone), slightly better results of ozone are caused probably by higher number of ozone stations compared to PM_{10} stations, see Table 3.1. In the case of the annual averages (PM_{10} vs. $PM_{2.5}$), somewhat better results of $PM_{2.5}$ are caused probably by two reason, i.e. by the different setting of the parameters used in the Delta tool for PM_{10} and $PM_{2.5}$ (Table 3.4) and by the additional information used in the $PM_{2.5}$ mapping by so-called pseudo $PM_{2.5}$ stations, see Section 2.1.

Another finding is that the maps created with the full set of the stations and evaluated by the stations of the same full set show in general better Delta tool results than the maps created with the assimilation subset of stations and evaluated by stations of the validation subset. One reason is that the first map is

created with a higher number of stations and therefore by nature being more representative. Another reason is that the first map is evaluated by stations, which have already been used in the map construction. These two reasons show the advantages and weaknesses of both the approaches: the use of all stations in the map construction is an advantage, but the use of the same set of stations in the evaluation is not fully objective. An option on how to overcome this difficulty is the use of the 'leave one out' cross-validation results in the Delta tool. This cross-validation method withholds one data point and then makes a prediction at the spatial location of that point. This procedure is repeated for all measurement points in the available set. The predicted and measured values at these points are mutually compared. The 'leave one out' cross-validation has been routinely used in the uncertainty estimates of the European air quality maps (Horálek et al., 2015). Moreover, this method is also mentioned in Viaene (2015) and it was also discussed in the FAIRMODE Technical meeting in Aveiro in June 2015.

Over a period of time preceding the FAIRMODE Technical Meeting of June 2015, different institutions participating in the FAIRMODE community compared available methodologies to evaluate model applications when monitoring data is assimilated. The methodologies compared are based on cross-validation and on the more complicated Monte Carlo based method. The Monte Carlo analysis has proven to have little added value compared to the simpler but contested 'leave one out' cross-validation approach. Concluding from this exercise the use of the cross-validation is considered adequate enough, compared to the much more demanding Monte Carlo method. Based on this outcome, the FAIRMODE Technical meeting agreed that for modelling based on the combination of measurement and modelling data, the use of the cross-validation is adequate in validations (FAIRMODE, 2015).

5.2 Conclusions

Several conclusions can be formulated:

As assumed, the Delta tool shows better results for the maps created with the full set of the stations and evaluated by the stations of the same full set than for the maps created with the assimilation subset of stations and evaluated by stations of the validation subset. The first type of maps is created with a higher number of stations and therefore it is by nature more representative. However, these maps are evaluated using the stations already involved in the map construction, which is not fully objective. It is recommended to use for evaluation purposes of the mapping performance in the future the 'leave one out' cross-validation results in the Delta tool, as is highlighted in section 5.1. In that case, the benefit of both approaches, one used in the map construction and one in the map evaluation, should be utilized: the maps will be created with all stations, but they will not be evaluated by stations already used in the mapping because of the 'leave one out' out cross-validation approach.

The Delta tool is highly sensitive to the parameters used for the measurement uncertainty calculation. The Delta tool gives highly different results for the maps of annual averages and for the maps of percentile (i.e. x-th highest) values. The reason is due to the large difference in the measurement uncertainty set in the Delta tool for the *MQO* criterion to evaluate the annual average and the percentile values. It is recommended to reconsider the measurement uncertainty values set in the Delta tool for the annual indicators. This is in agreement with the conclusions of the FAIRMODE Technical Meeting in Aveiro in June 2015 (FAIRMODE, 2015).

The benchmarking reports can be used for quick comparison of mapping results based on the different methods. In such cases, it is very important to compare the results then on basis of the same time step and type of indicator. More specifically, the reports for the time-dependent, i.e. hourly, daily data and the annual values are not mutually comparable and that the results of the Delta tool for the annual averages and for the percentile (i.e. x-th highest) values should not be mutually compared.

References

AirBase, 2014, 'AirBase — the European Air quality database' (http://air-climate.eionet.europa.eu/databases/airbase/#reporting) accessed 26 September 2014.

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

Denby, B., Horálek, J., de Smet, P., de Leeuw, F., 2011, 'Mapping annual mean $PM_{2.5}$ concentrations in Europe: application of pseudo $PM_{2.5}$ station data', *ETC/ACM Technical Paper* 2011/5. http://acm.eionet.europa.eu/reports/ETCACM_TP_2011_5_spatialPM2.5mapping accessed 15 September 2015.

EC, 2008. 'Directive 2008/50/EC of the European Parliament and of the Council on ambient air quality and cleaner air for Europe', *Official Journal L* 152 (11.6.2008), 1-44. (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF) accessed 15 September 2015.

EEA, 2011, 'The application of models under the European Union's Air Quality Directive: A technical reference guide', *EEA Technical report* 10/2011. (http://www.eea.europa.eu/publications/fairmode) accessed 15 September 2015.

FAIRMODE, 2015, 'Minutes of FAIRMODE Technical meeting, Aveiro (Portugal) 24-25/06/2015'. (http://fairmode.jrc.ec.europa.eu/document/fairmode/event/minutes/20150624_Minutes_FAIRMODE_TechnicalMeeting.pdf) accessed 15 September 2015.

Gräler, B., Rehr, M., Gerharz, L., Pebesma, E., 2012, 'Spatio-temporal analysis and interpolation of PM10 measurements in Europe for 2009', *ETC/ACM Technical paper* 2012/8. (http://acm.eionet.europa.eu/reports/ETCACM_2012_8_spatio-temp_PM10analyses) accessed 2 October 2015.

Horálek, J., Denby, B., de Smet, P., de Leeuw, F., Kurfürst, P., Swart, R., van Noije, T., 2007, 'Spatial mapping of air quality for European scale assessment', *ETC/ACC Technical paper* 2006/6. (http://acm.eionet.europa.eu/reports/ETCACC_TechnPaper_2006_6_Spat_AQ) accessed 3 July 2015.

Horálek, J., de Smet, P., Kurfürst, P., de Leeuw, F., Benešová, N., 2015, 'European air quality maps of PM and ozone for 2012 and their uncertainty', *ETC/ACM Technical Paper* 2014/4. (http://acm.eionet.europa.eu/reports/ETCACM_TP_2014_4_spatAQmaps_2012) accessed 3 July 2015.

Horálek, J., Tarrasón, L., de Smet, P., Malherbe, L., Schneider, P., Ung, A., Corbet, L., Denby, B., 2014, 'Evaluation of Copernicus MACC-II ensemble products in the ETC/ACM spatial air quality mapping', *ETC/ACM Technical Paper* 2013/9. (http://acm.eionet.europa.eu/reports/ETCACM_TP_2013_9_AQmaps_with_MACCproducts) accessed 3 July 2015.

Pernigotti, D., Thunis, P., Gerboles, M., Belis. C., 2013, 'Model quality objectives based on measurement uncertainty: Part II: PM10 and NO2', *Atmospheric Environment* 79, 869-878.

Rouïl, L., Emili, E., Beekman, M., Foret, G., Sofiev, M., Vira, J., Eskes, H., Meleux, F., Ung, A., Marécal, V., Valdebenito,, A., Carlin-Benedictow, A., Elbern, H., Friese, E., Strunk, A., Robertson, L., Segers, A. (2014), 'Assesment Report: Air quality in Europe in 2012', *MACC-II Deliverable* D113.4 & 'Validation report for 2012', *MACC-II Report* D113.5. (http://www.gmes-atmosphere.eu/documents/maccii/deliverables/eva/) accessed 3 July 2015.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyíri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, A., Wind, P., 2012, 'The EMEP MSC-W chemical transport model – technical description', *Atmospheric Chemistry and Physics*, 12, 7825–7865.

Thunis, P., Cuvelier, C., Pederzoli, A., Georgieva, E., Pernigotti, D., Degraeuwe, B., Marioni, M., 2015, *DELTA Version 5.0. Concepts / User's Guide / Diagrams*, Joint Research Centre, Ispra. (http://www.atmosys.eu/download/DELTA_UserGuide_V5_0.pdf) accessed 23 June 2015.

Thunis, P., Pederzoli, A., Pernigotti, D., 2012, 'Performance criteria to evaluate air quality modeling applications', *Atmospheric Environment* 59, 476-482.

Thunis, P., Pernigotti, D., Gerboles, M., 2013, 'Model quality objectives based on measurement uncertainty: Part 1: Ozone', *Atmospheric Environment* 79, 861-868.

Viaene, P., Janssen, S., Thunis, P., Cuvelier, C., Trimpeneers, E., Wesseling, J., Montero, A., Miranda, A., Stocker, J., Rørdam Olesen, H., Guerreiro, C., Sousa Santos, G., Vincent, K., Carnevale, C., Stortini, M., Bonafè, G., Minguzzi, E., Deserti, M., 2015, *Guidance Document on Model Quality Objectives and Benchmarking*, FAIRMODE. (http://fairmode.jrc.ec.europa.eu/document/fairmode/WG1/Guidance_MQO_Bench_vs1.1.pdf) accessed 23 June 2015.

Wesseling, J., 2015, 'Using measurement uncertainty in the MQO', presentation at the FAIRMODE Technical Meeting, Aveiro, 24-25 June 2015. (http://fairmode.jrc.ec.europa.eu/document/fairmode/event/presentation/20150624-Aveiro/WG1/WG1-S1.3%20-%20Using%20measurement%20uncertainties_21jun.pdf) accessed 15 September 2015.