ETC/ACM air quality mapping method and its evaluation

Jan Horálek (ETC/ACM, CHMI)
Nina Benešová (ETC/ACM, CHMI)
Peter de Smet (ETC/ACM, RIVM)

European Topic Centre on Air Pollution and Climate Change Mitigation

CZECH HYDROMETEOROLOGICAL INSTITUTE

National Institute for Public Health and the Environment

1. Mapping methodology

2. Routine evaluation (especially cross-validation)

3. Evaluation using Delta tool (first attempt)

ETC/ACM mapping methodology

Developed (in 2005-2007) with the objective of the European Environmental Agency of having interpolated maps ***primarily based on air quality measurements***.

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 primarily 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.

Primarily data – measurement data
Supplementary data – chemical transport model output,
other proxy data (altitude, meteorology, popul. density)

ETC/ACM mapping methodology – continuation

Linear regression model followed by kriging of its residuals (residual kriging)

The supplementary data for ***linear regression model*** were selected based on their relation with measured AQ data.

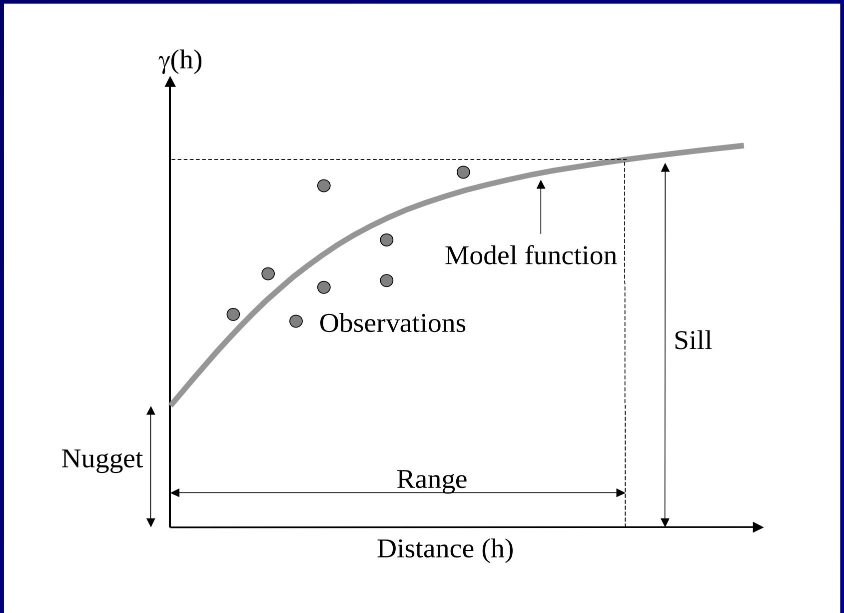
In the case of PM_{10} and $PM_{2.5}$, both measured data and dispersion model output are ***logarithmically transformed***, due to the lognormal distribution of these data.

kriging – spatial interpolation geostatistical method (i.e. knowledge of the spatial structure of air quality field is utilized, using variogram)

Mapping method – continuation

variogram - measure of a spatial correlation

parameters: ***sill, nugget, range***



Empirical variogram fitted by an analytical function – in our case spherical.

The method is routinely used for ***annual data*** (i.e. the monitoring and modelling and other data combined for annual indicators.) For sensitivity analysis (and comparison with the results based on daily data), see ETC/ACM Technical Paper 2012/8.

Mapping methodology – continuation

Separate mapping of rural and urban air quality

– due to different character of urban and rural air quality

PM_{10}, $PM_{2.5}$, NO_2 – urban/suburban concentrations are in general higher than the rural concentrations

Ozone – rural concentrations are higher than urb/sub

Rural and urban background maps are created separately,

rural maps – based on rural background stations

urban background maps – based on urban and suburban background stations

Final maps are created by merging of rural and urban background maps, using ***population density***.

Mapping methodology – continuation

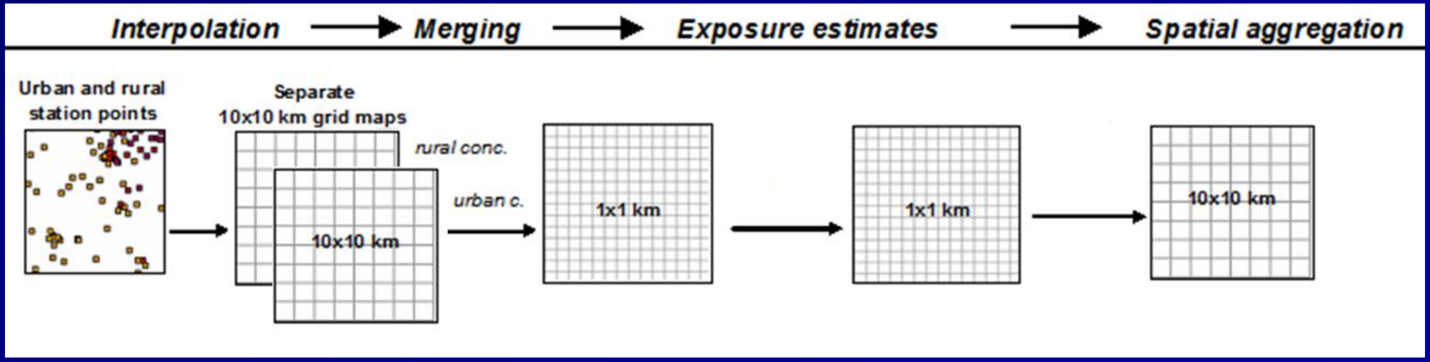
Grid resolution of the health-related indicators

Separate rural and urban background maps
– created in ***10x10 km resolution***

These maps are merged using population density (in 1x1 km)
– into ***1x1 km resolution***

Exposure estimates – based on these 1x1 km maps.

Presentation – final maps are spatial aggregated into 10x10 km resolution. (Plus urban background maps.)



Mapping methodology – continuation

Pollutants and indicators mapped

Regularly:

PM_{10} – annual average [µg.m^{-3}]
– 36th maximum daily average value [µg.m^{-3}]

$PM_{2.5}$ – annual average [µg.m^{-3}].

Ozone – 26th highest daily max. 8-hourly mean [µg.m^{-3}]
– SOMO35 [µg.m^{-3}.day]
– AOT40 for crops [µg.m^{-3}.hour]
– AOT40 for forests [µg.m^{-3}.hour]

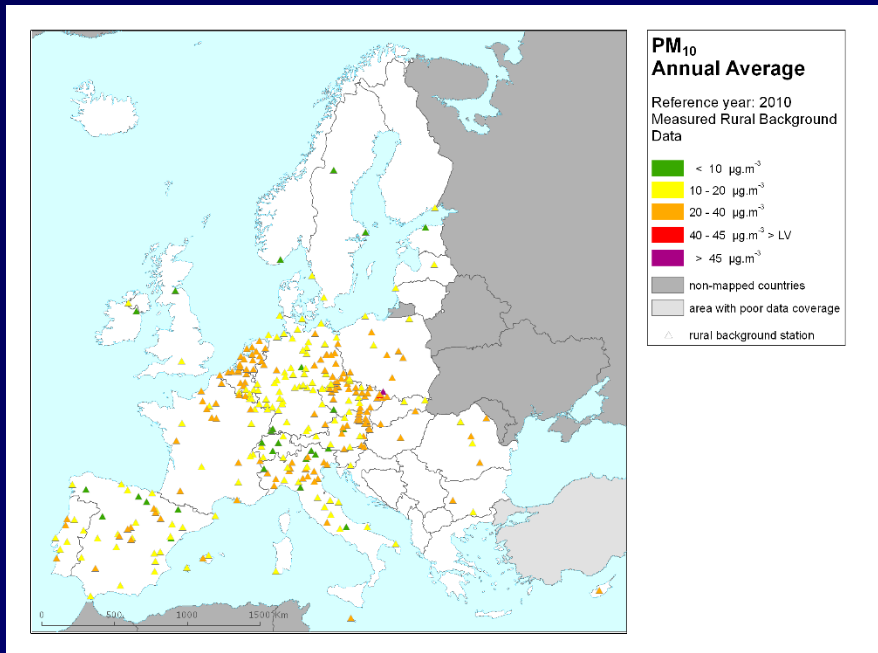
Repetitively:

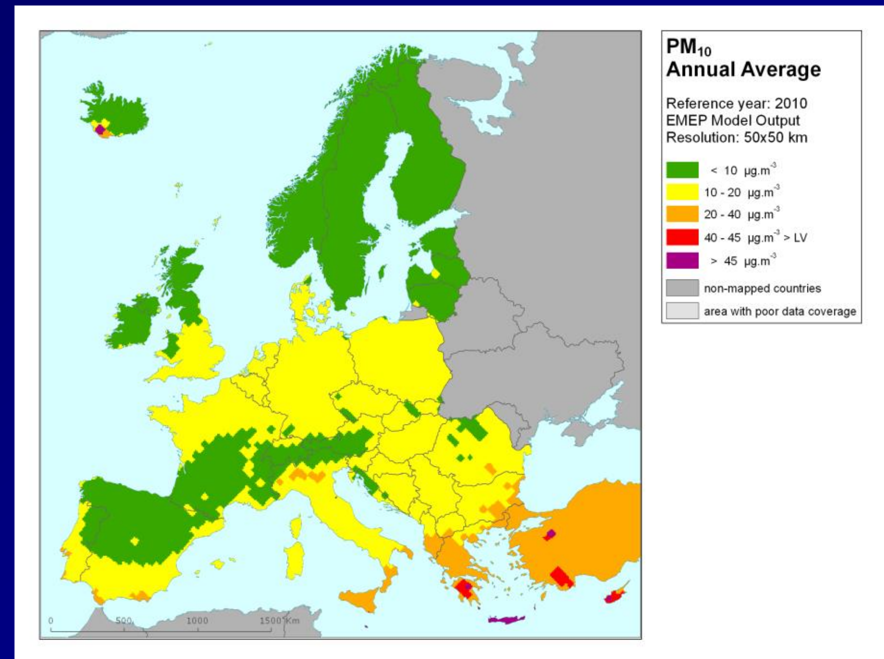
NO_2 – annual average [µg.m^{-3}]
NO_x – annual average [µg.m^{-3}]
SO_2 – annual average [µg.m^{-3}]

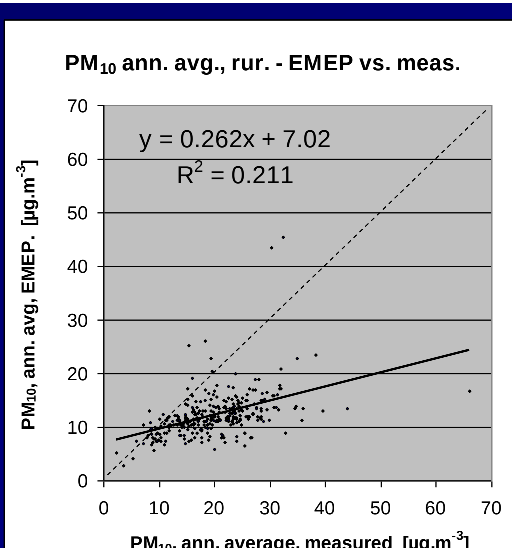
Newly: **BaP** – annual average [µg.m^{-3}]

PM_{10} annual average, 2010 – rural areas

measured data / **EMEP model**



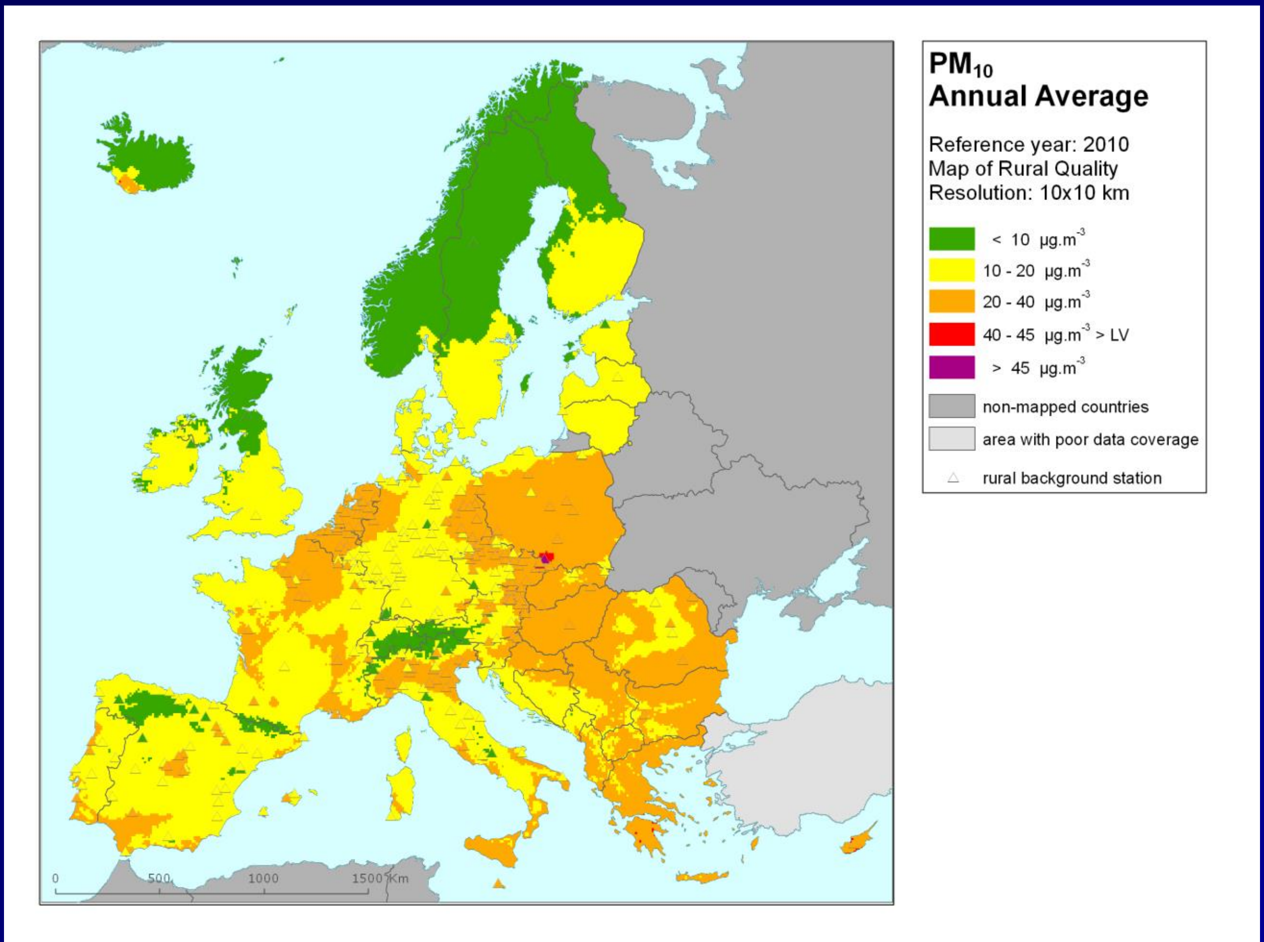




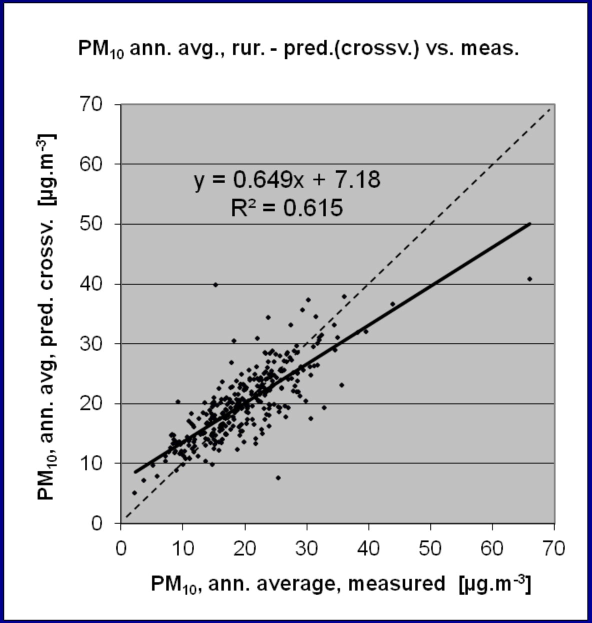
Linear regression model (log. transformed):

	adj. R^2	*SEE*
EMEP model	*0.33*	*0.324*
EMEP model, altitude	*0.41*	*0.306*
EMEP m., altitude, wind speed	*0.44*	*0.295*

PM_{10} annual average, 2010 – rural areas
rural map (applicable for rural areas only)



cross-validation

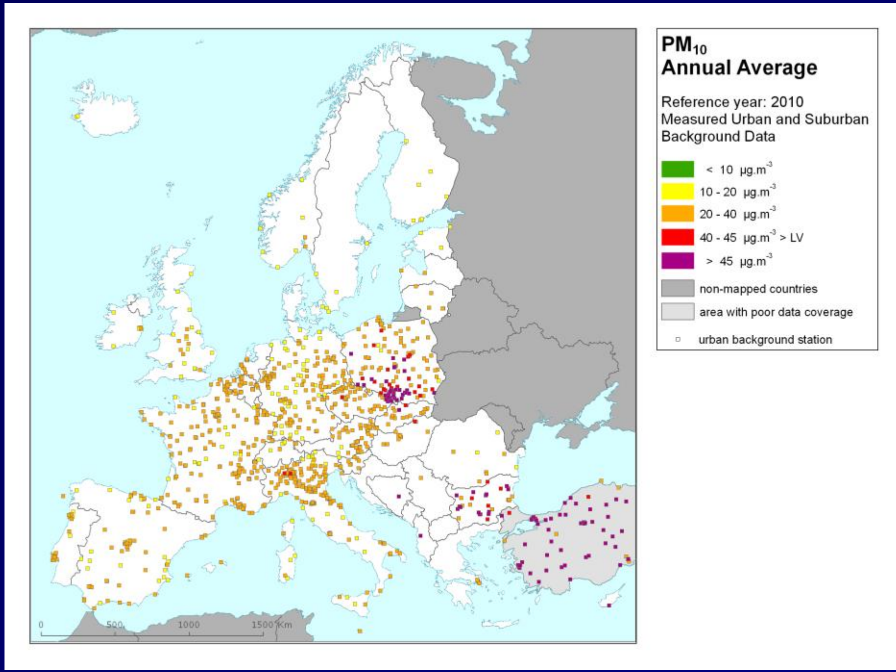


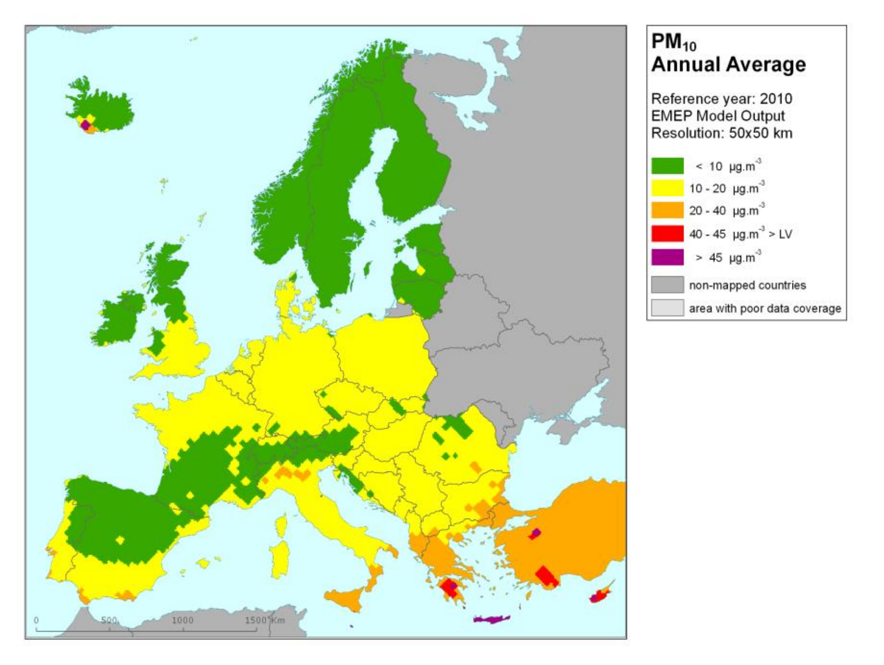
RMSE = 4.5 µg. m^{-3}

Bias = 0.2 µg. m^{-3}

PM_{10} annual average, 2010 – urban areas

measured data | **EMEP model**

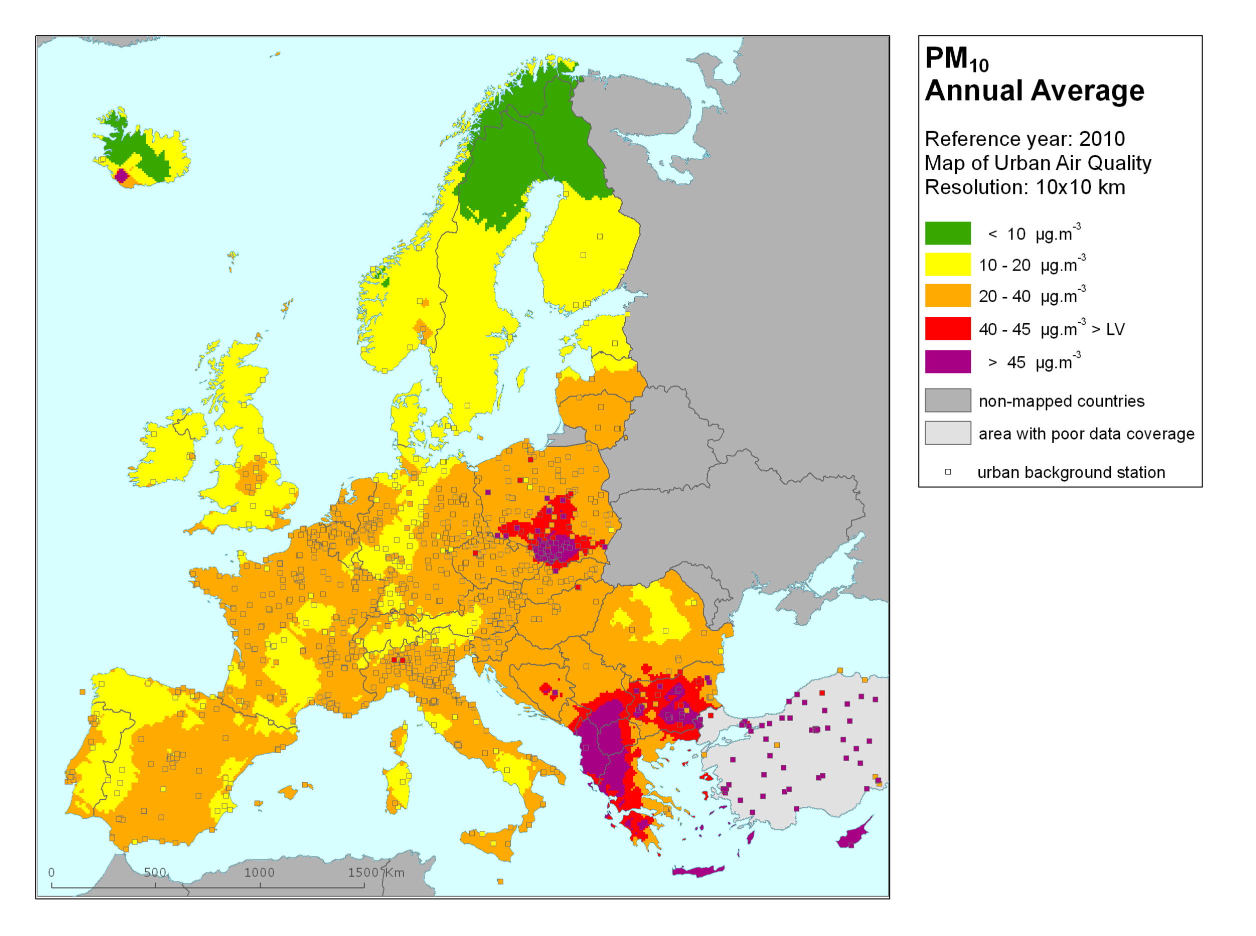




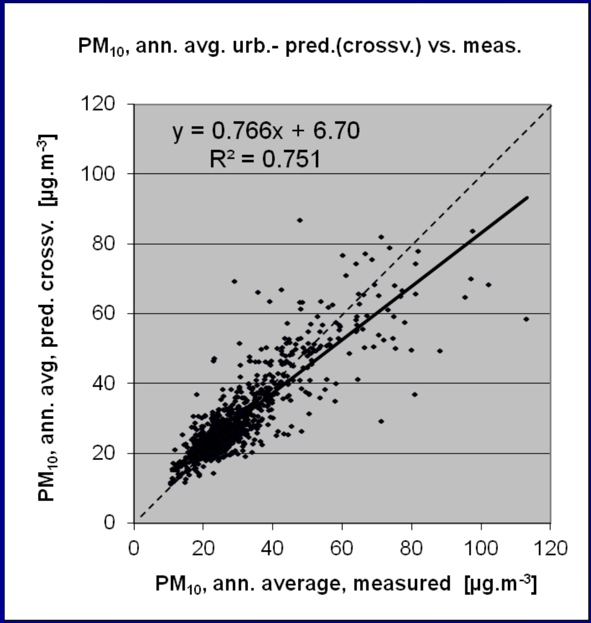
Linear regression model (log. transformed):

	adj. R^2	SEE
EMEP model	*0.38*	*0.292*

PM_{10} annual average, 2010 – urban areas
urban background map (applicable for urban areas only)



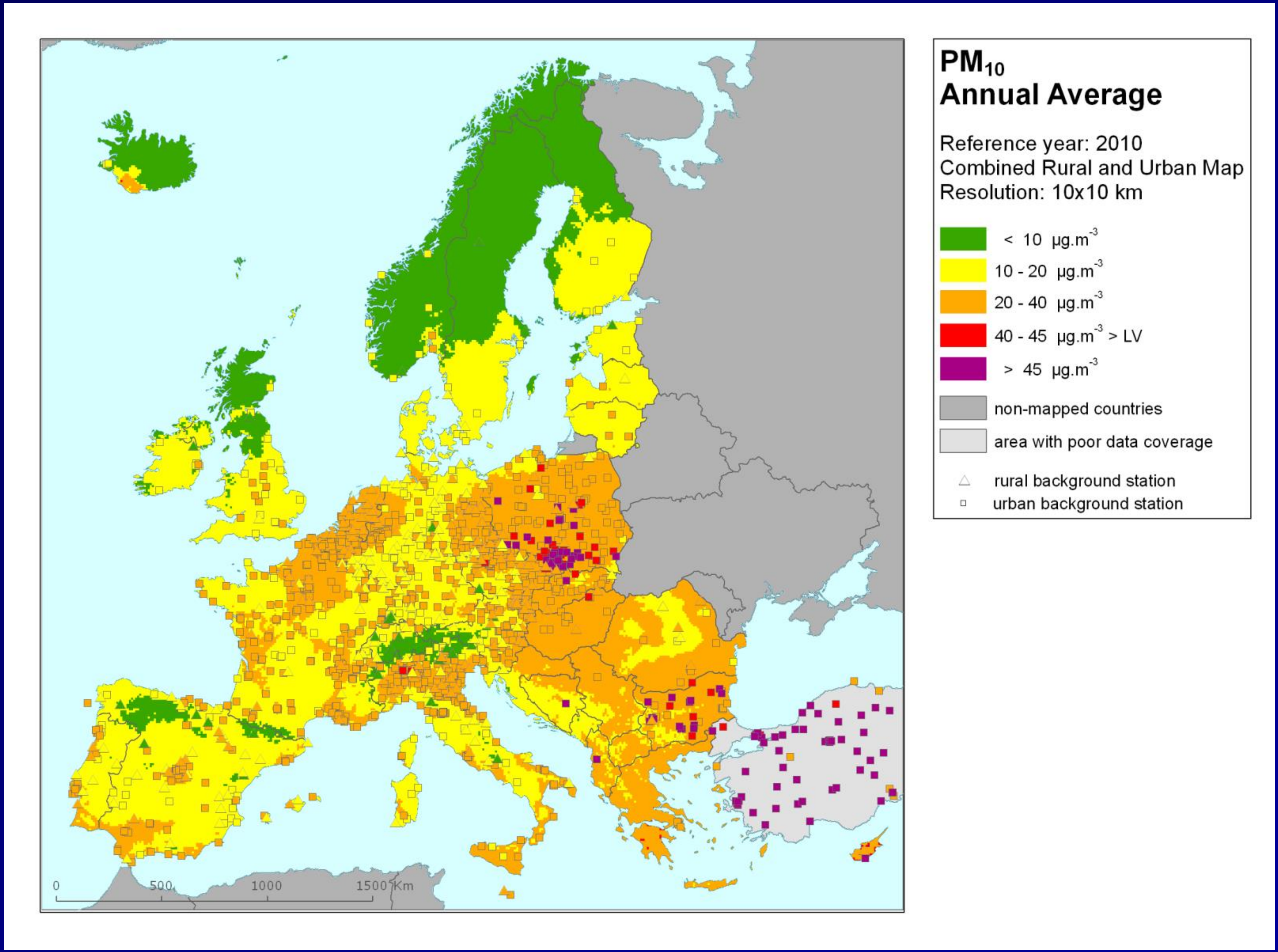
cross-validation



RMSE = 6.6 µg. m^{-3}

Bias = -0.1 µg. m^{-3}

PM_{10} annual average, 2010
final merged map



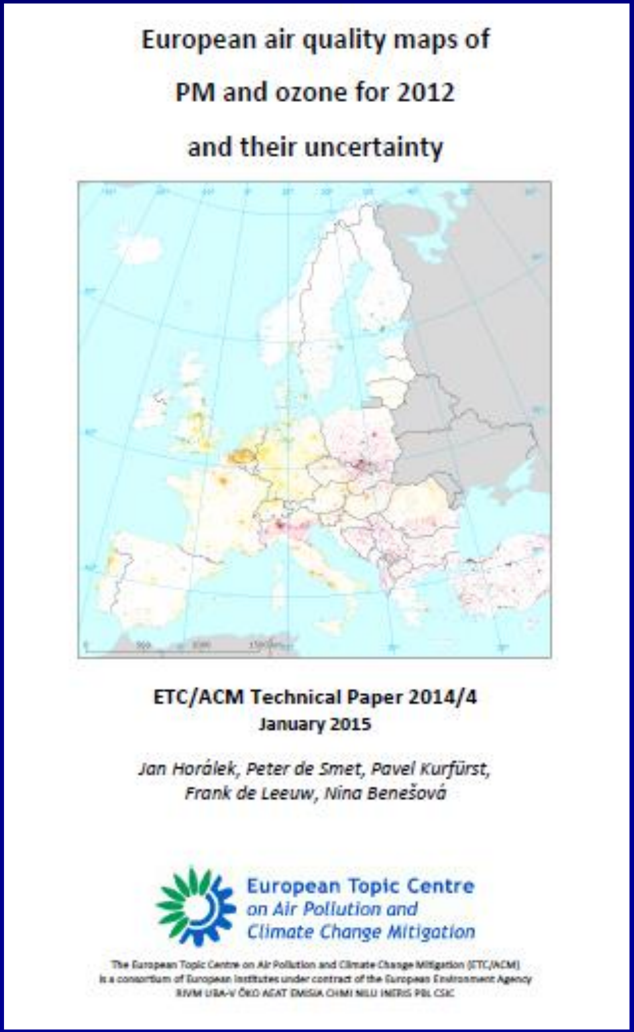
Actual air quality maps

Regular annual product:
ETC/ACM Technical Paper „European air quality maps of PM and ozone and their uncertainty“

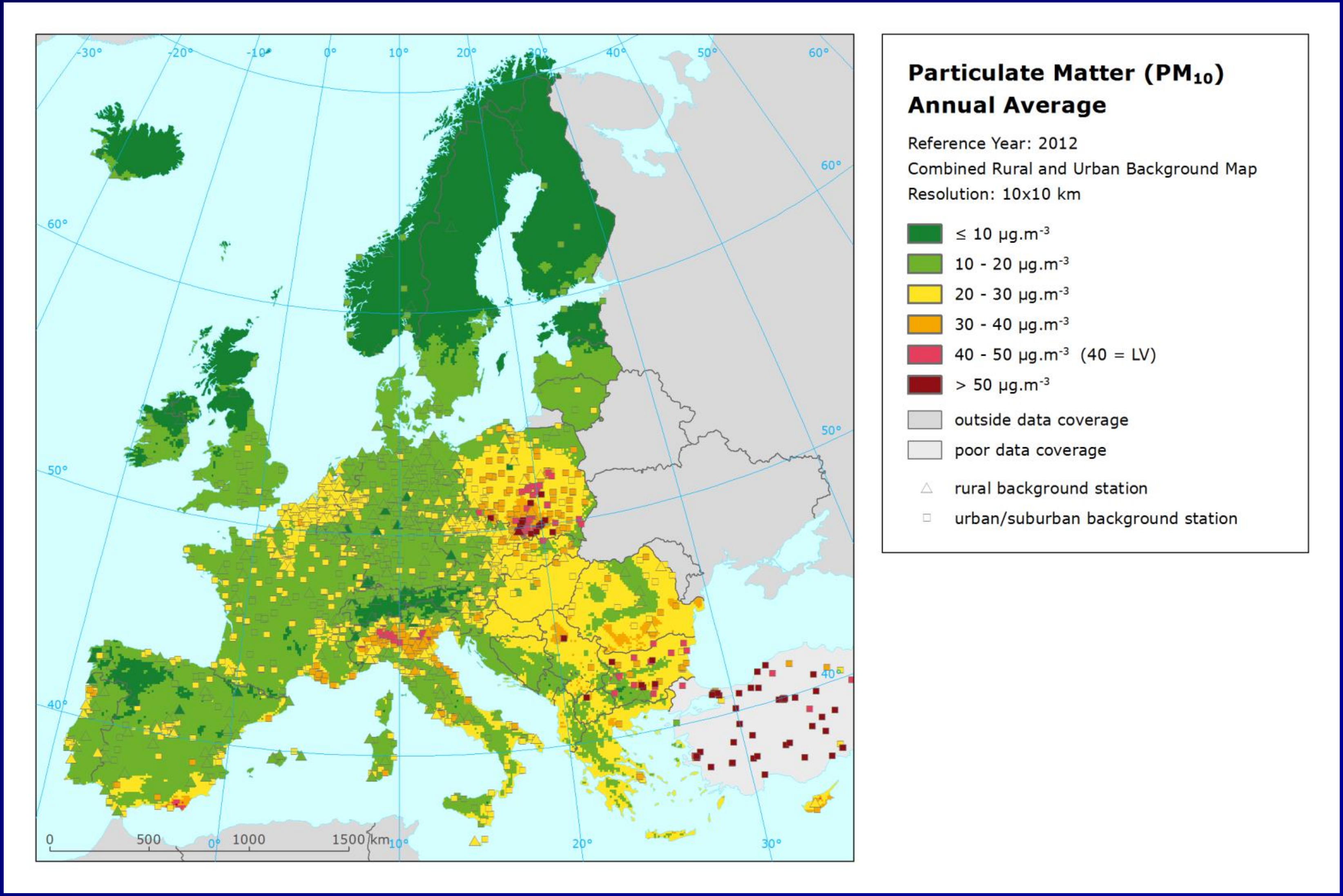
Concentration maps, inter annual difference maps, exposure tables, uncertainty analysis.

Most recent : ***ETC/ACM TP 2014/4***, maps for 2012

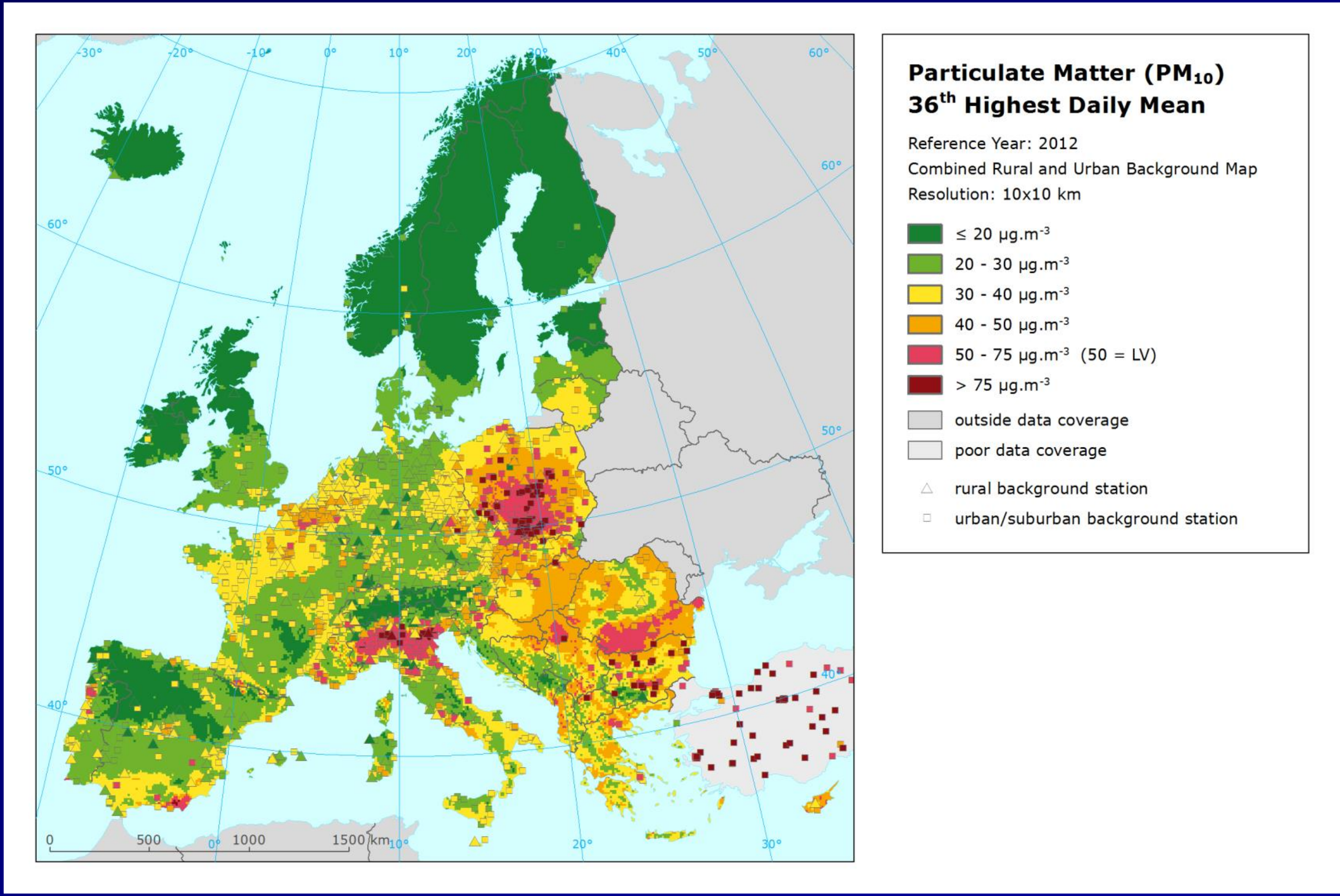
http://acm.eionet.europa.eu/reports/



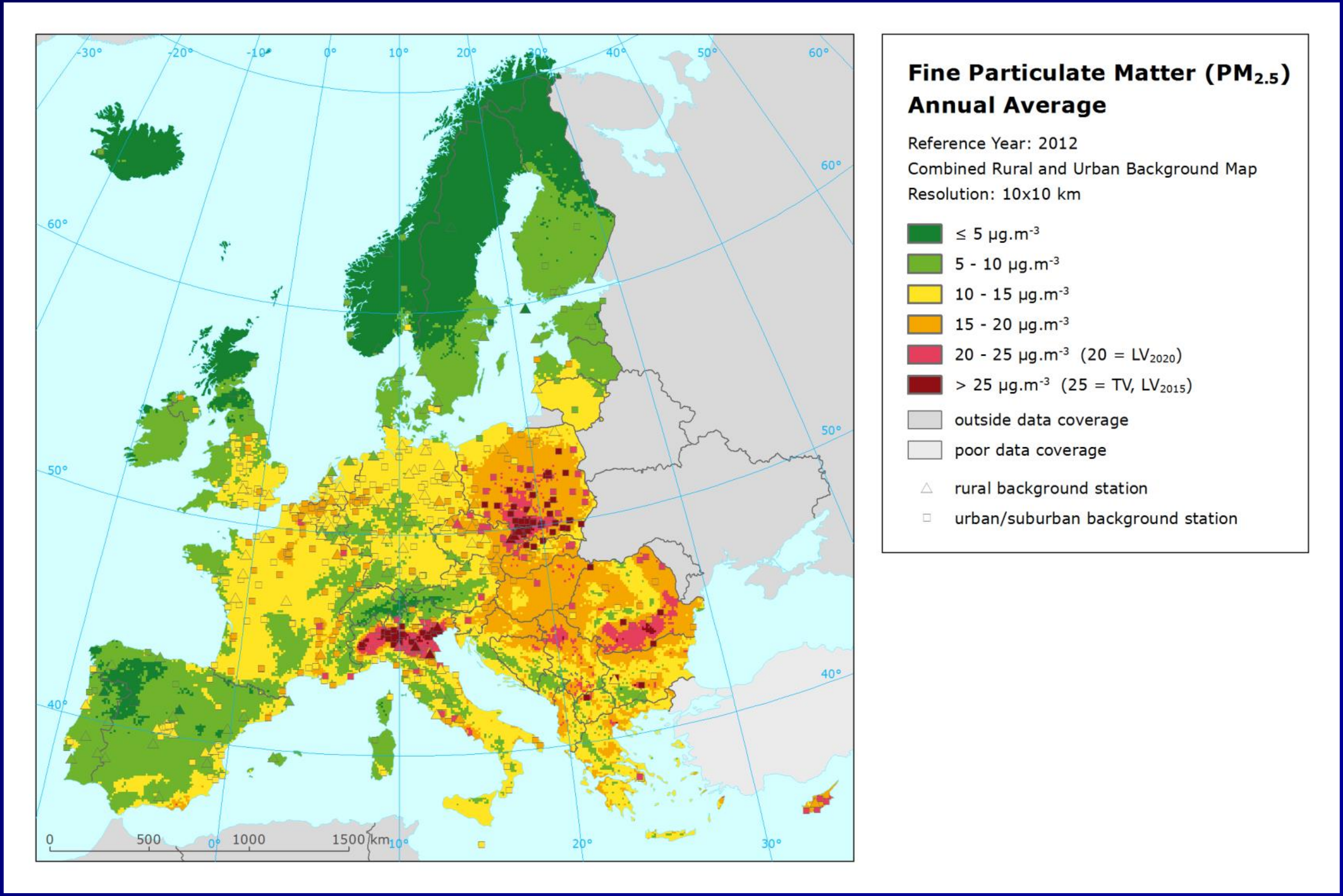
PM_{10} – annual average, 2012



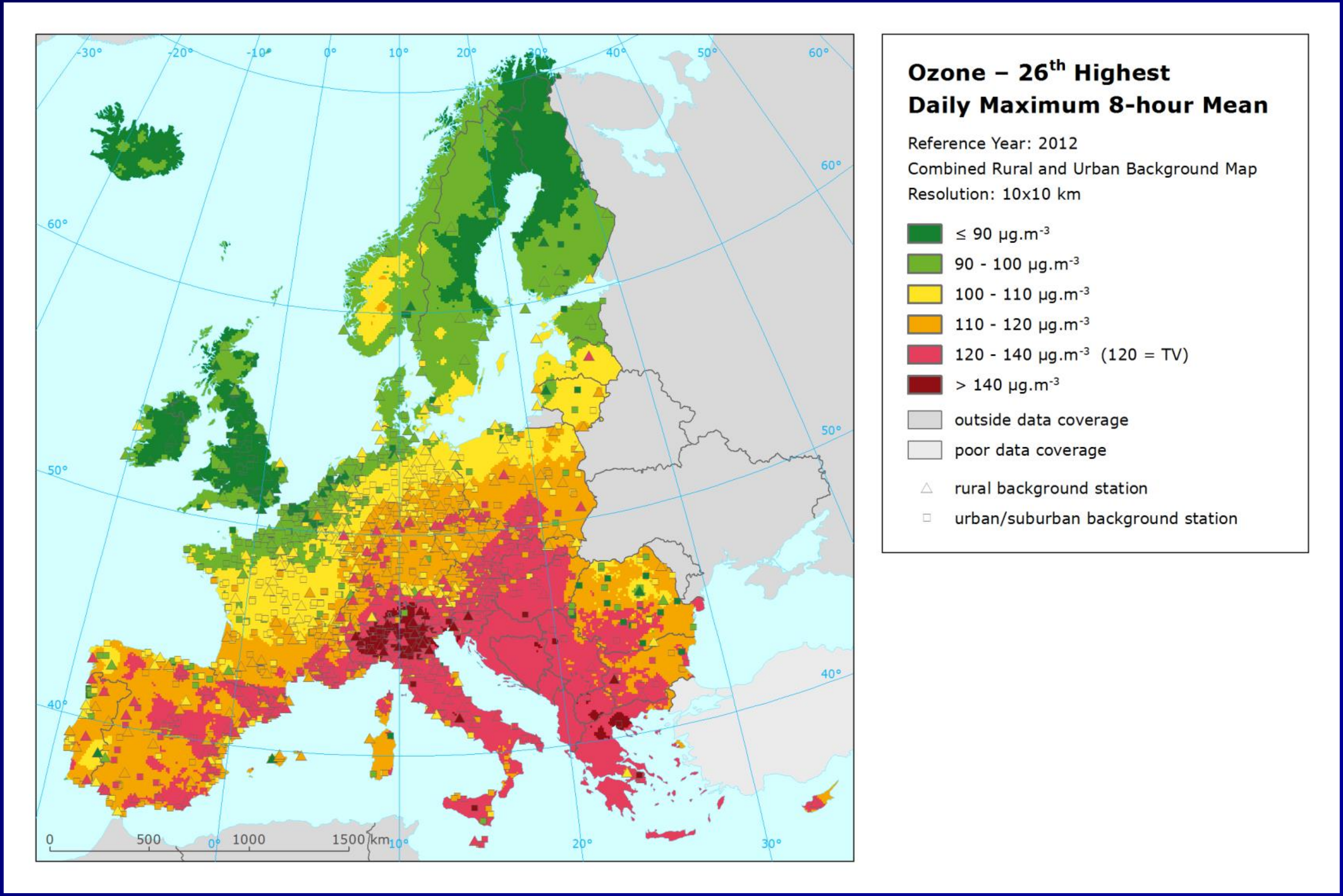
PM_{10} – 36th highest daily mean, 2012



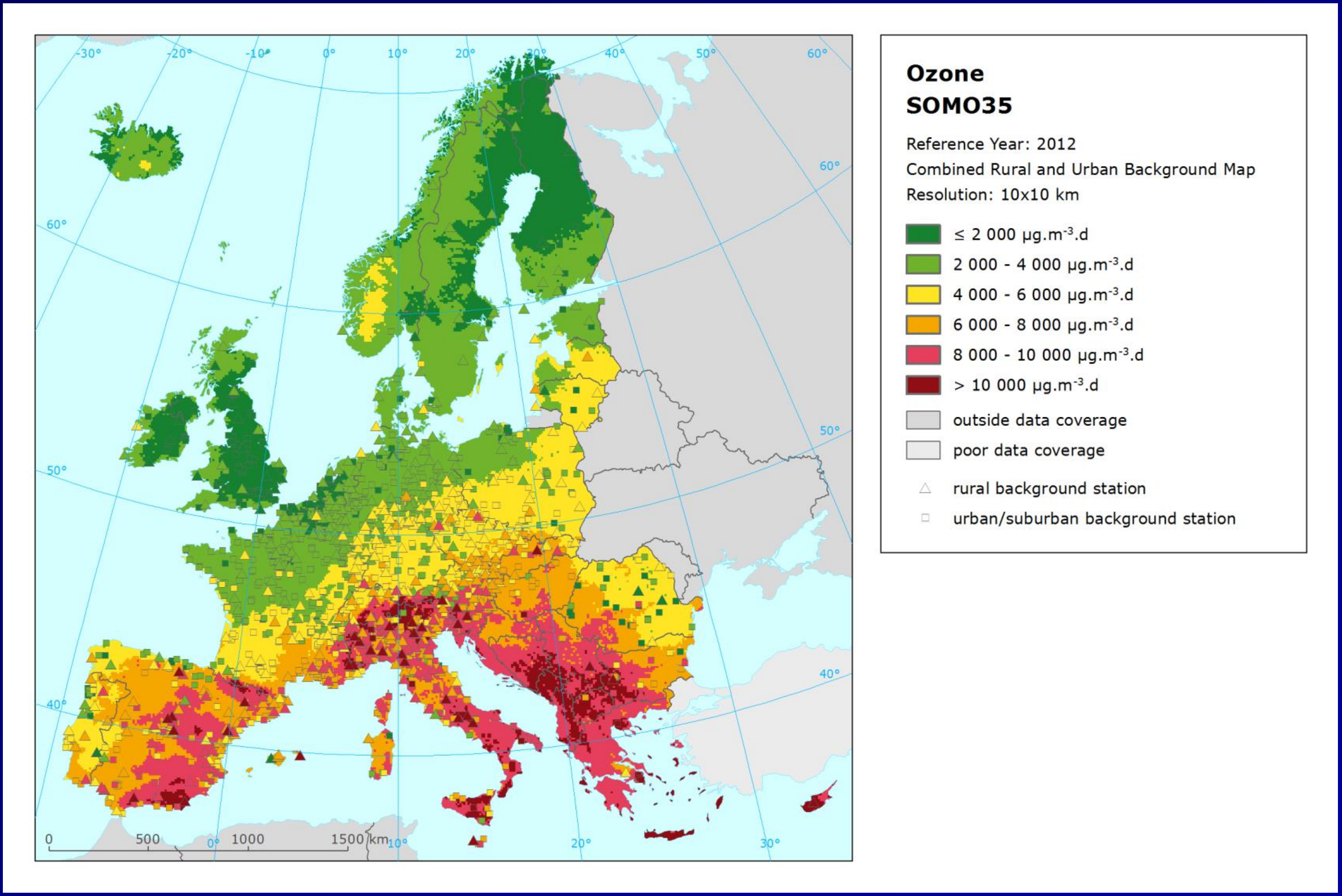
$PM_{2.5}$ – annual average, 2012



O_3 – 26th highest daily max. 8-hourly mean, 2011



O_3 – SOMO35, 2011



1. Mapping methodology

2. Routine evaluation (especially cross-validation)

3. Evaluation using Delta tool (first attempts)

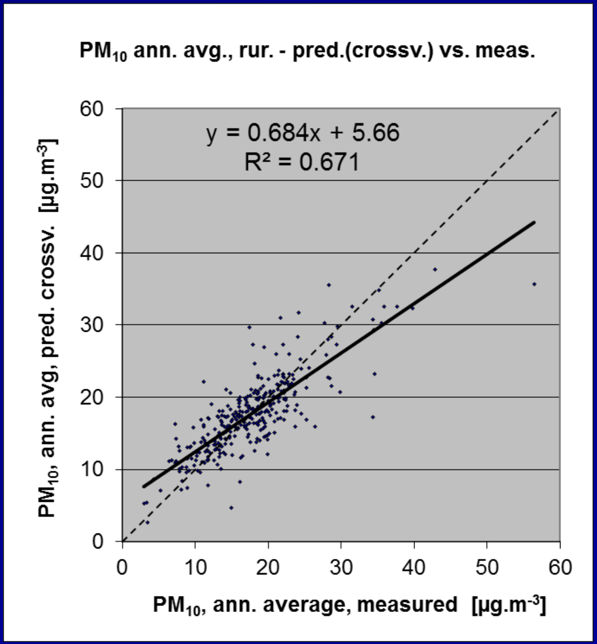
Routine evaluation

cross-validation – the spatial interpolation is calculated for every measurement point based on all available information except from the point in question. These estimated values are compared with the measured ones by *scatter-plot* (including *R^2* and *regression equation*) and by statistical indicators, espec. *RMSE* and *bias (MPE)*. Occasionally also *MAE* and other ones.

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

$$bias(MPE) = \frac{1}{N}\sum_{i=1}^{N}(\hat{Z}(s_i) - Z(s_i))$$

$$MAE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|Z(s_i) - \hat{Z}(s_i)\right|}$$



where $Z(s_i)$ is the measured value in point s_i

$\dot{Z}(s_i)$ is the estimation in the point s_i using other points

N is the number of the stations

Routine evaluation – continuation

Next to this: RMSE in relative terms

$$RRMSE = \frac{RMSE}{Z}.100$$

Z is the mean of the air pollution indicator value for all stations

Cross-validation evaluates of the quality of the predicted values at locations without measurements.

It also enables to validate the quality of the uncertainty map , i.e. kriging standard error (or standard deviation) map, created based on the geostatistical theory.

Routine evaluation – continuation

Comparison of the point measured and interpolated grid values – *the linear regression equation* and its R^2 , by *RMSE* and *bias*.

Simple comparison evaluates the quality of the map at locations of measurements. (Variability – due to interpolation smoothing, spatial averaging into 10x10 km cells , and eventually rural/urban merging).

Validation done separately for urban and rural areas

- for rural maps (using rural backround stations)
- for urban backrgound maps (using urban backround stations)
- for final merged maps (using rural and urban backround stations, separatelly)

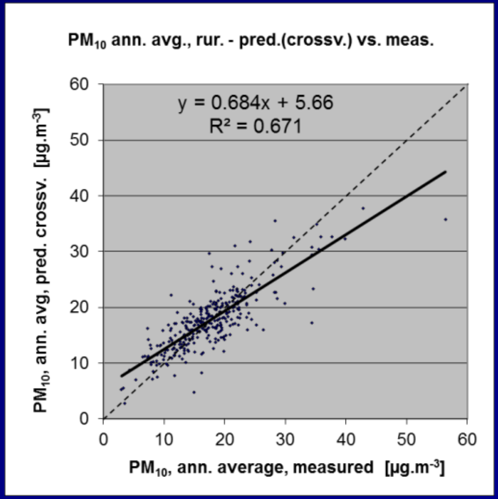
Routine evaluation – continuation

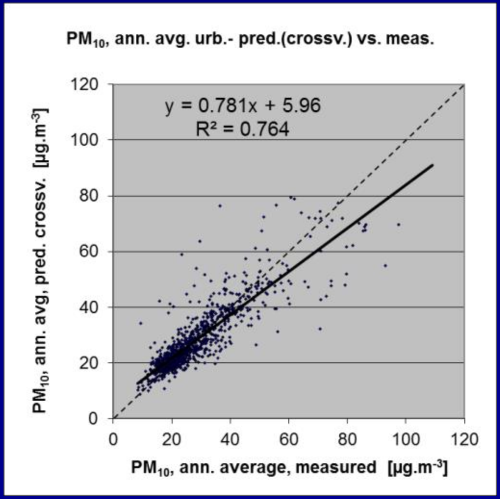
Separate for rural and urban background maps

PM_{10}, annual average, 2012

cross-validation

linear regr. model + OK of its residuals	rural areas parameter values	urban areas parameter values
RMSE [µg.m^{-3}]	3.8	6.1
Relative RMSE [%]	21.4	22.1
bias (MPE) [µg.m^{-3}]	0.1	0.0



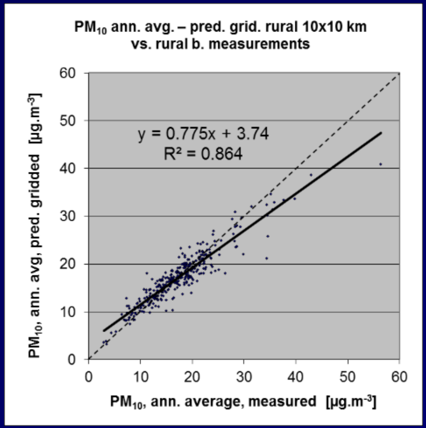


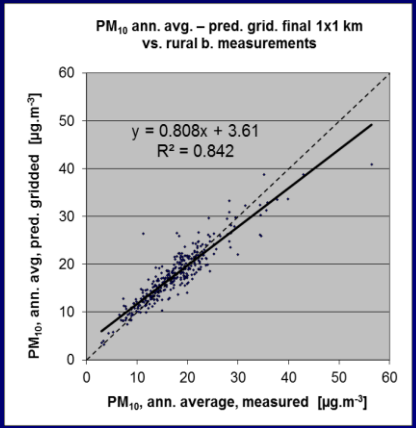
Level of underestimation in areas without measurement can be estimated.

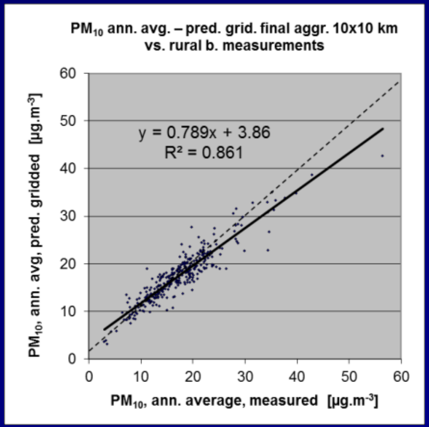
Analysis of rural/urban areas in final map

PM_{10}, annual average, 2012

simple comparison – rural areas







Good representation in both 1x1 km and 10x10 km maps.

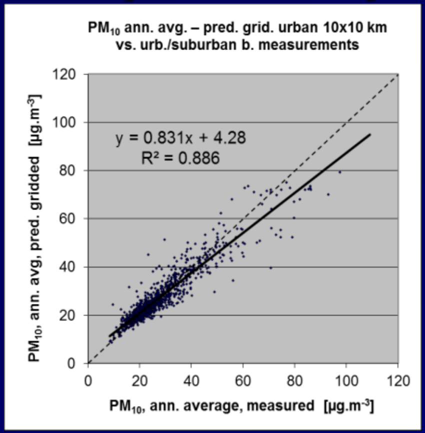
rural 10x10 *final merged 1x1* *final, aggr. 10x10*

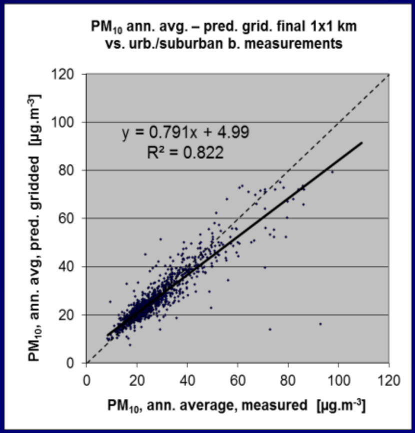
	rural backgr. stations			
	RMSE	bias	R^2	equation
cross-valid. prediction, separate (r or ub) map	3.8	0.1	0.67	y = 0.684x + 5.66
grid prediction, 10x10 km separate (r or ub) map	2.5	-0.2	0.86	y = 0.775x + 3.74
grid prediction, 1x1 km final merged map	2.6	0.3	0.84	y = 0.808x + 3.61
grid prediction, aggr. 10x10 km final merged map	2.5	0.2	0.86	y = 0.789x + 3.86

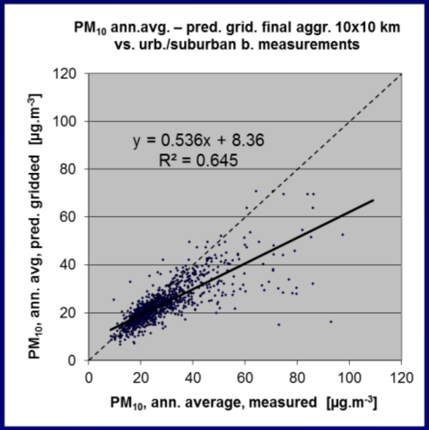
Analysis of rural/urban areas in final map – cont.

PM_{10}, annual average, 2012

simple comparison – urban background areas







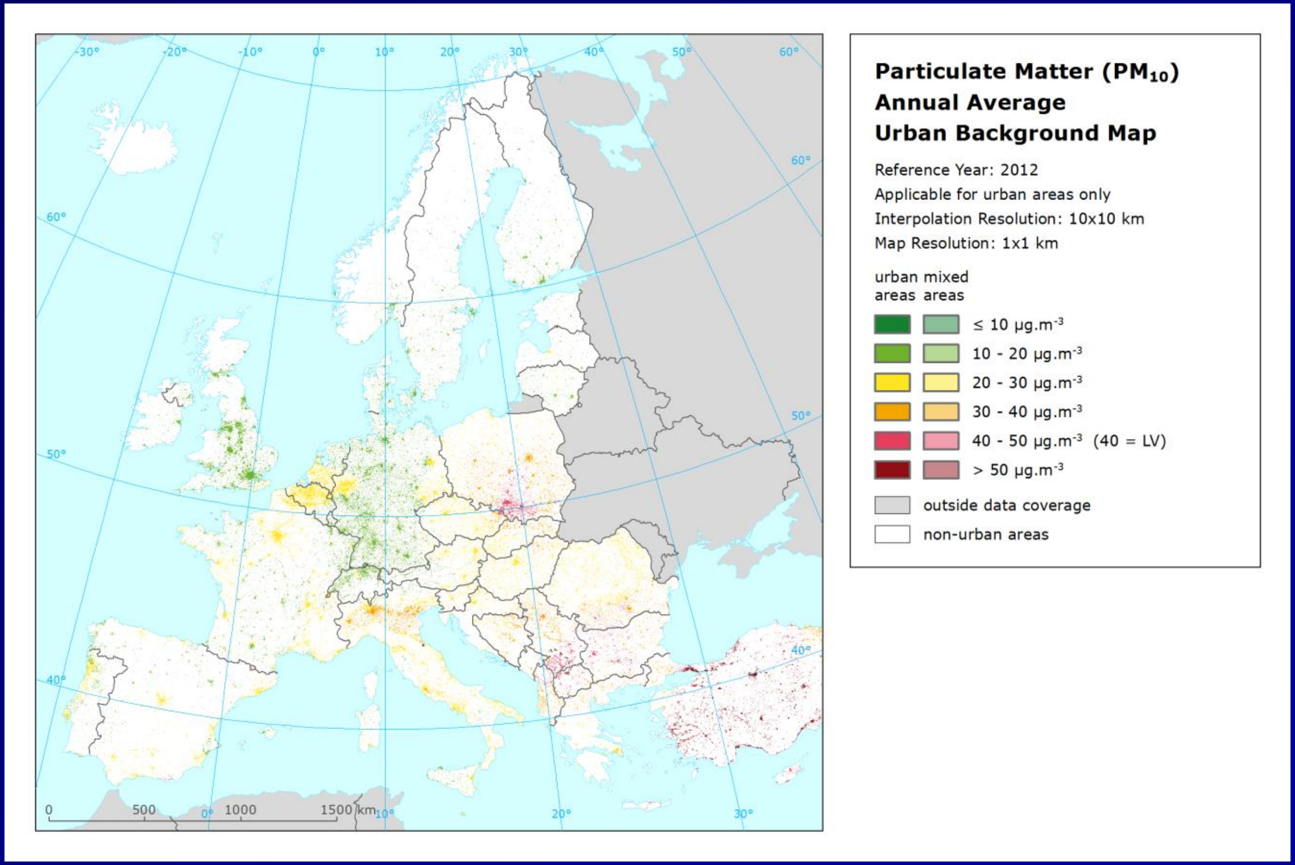
Good representation in1x1 km map, but not in 10x10 km map (bias, RMSE, R^2).

rural 10x10 *final merged 1x1* *final, aggr. 10x10*

	urb./suburban backgr. stations			
	RMSE	bias	R^2	equation
cross-valid. prediction, separate (r or ub) map	6.1	0.0	0.76	y = 0.781x + 6.0
grid prediction, 10x10 km separate (r or ub) map	4.3	-0.3	0.89	y = 0.831x + 4.3
grid prediction, 1x1 km final merged map	5.3	-0.7	0.82	y = 0.791x + 5.0
grid prediction, aggr. 10x10 km final merged map	8.7	-4.2	0.65	y = 0.536x + 8.4

Analysis of rural/urban areas in final map – cont.

Action: to present ***separate urban background AQ map*** to illustrate the difference with the aggregated final map.

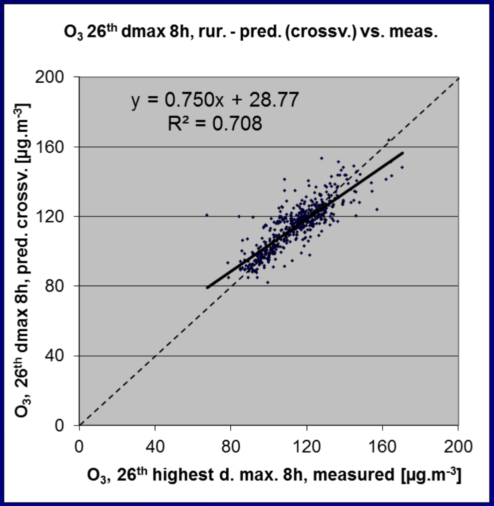


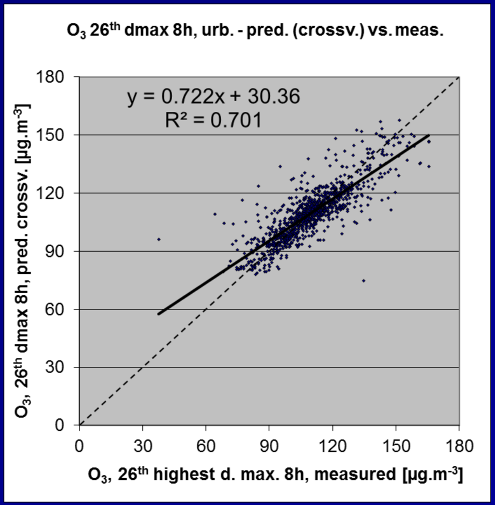
Routine evaluation – continuation

O_3, 26th highest daily maximum 8-hourly mean, 2012

cross-validation

linear regr. model + OK on its residuals	rural areas	urban areas
	parameter values	parameter values
RMSE [µg.m^{-3}]	8.49	9.06
realtive RMSE [%]	7.4	8.3
bias (MPE) [µg.m^{-3}]	0.18	-0.07



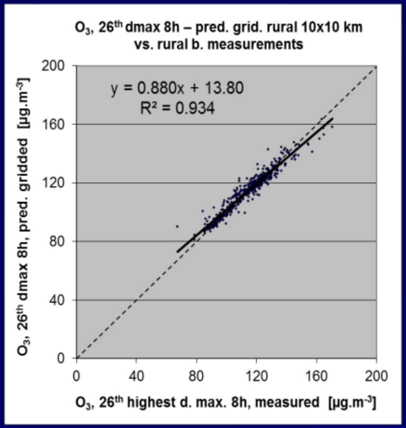


Level of underestimation in areas without measurement can be estimated.

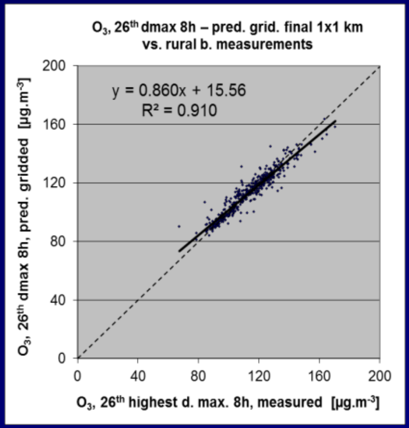
Analysis of rural/urban areas in final map

O_3 26th highest daily maximum 8-hourly mean, 2012

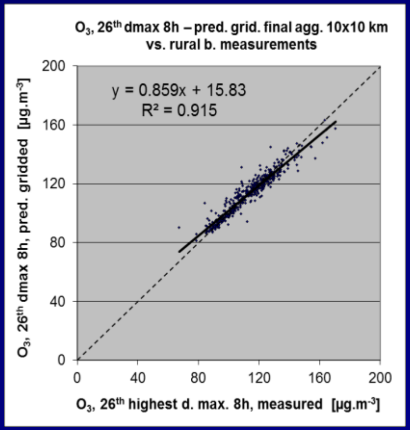
simple comparison – rural areas



rural 10x10



final merged 1x1



final, aggr. 10x10

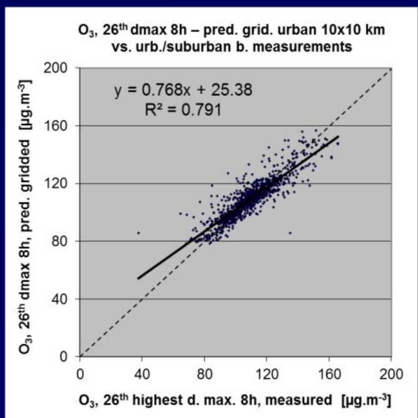
Good representation in both 1x1 km and 10x10 km maps.

	rural backgr. stations			
	RMSE	bias	R^2	equation
cross-valid. prediction, separate (r or ub) map	8.5	0.2	0.71	y = 0.750x + 28.8
grid prediction, 10x10 km separate (r or ub) map	4.1	0.1	0.93	y = 0.880x + 13.8
grid prediction, 1x1 km final merged map	4.8	-0.4	0.910	y = 0.860x + 15.6
grid prediction, aggr. 10x10 km final merged map	4.7	-0.3	0.915	y = 0.859x + 15.8

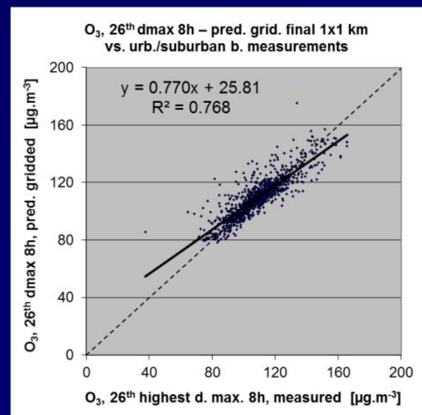
Analysis of rural/urban areas in final map

O_3 26th highest daily maximum 8-hourly mean, 2012

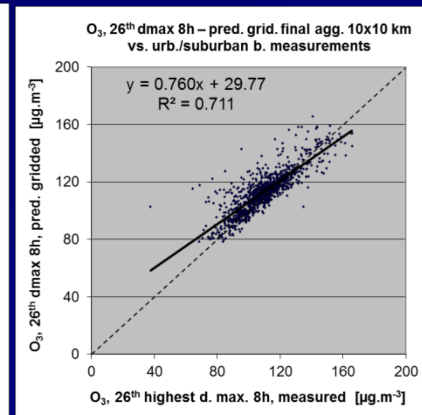
simple comparison – urban background areas



rural 10x10



final merged 1x1



final, aggr. 10x10

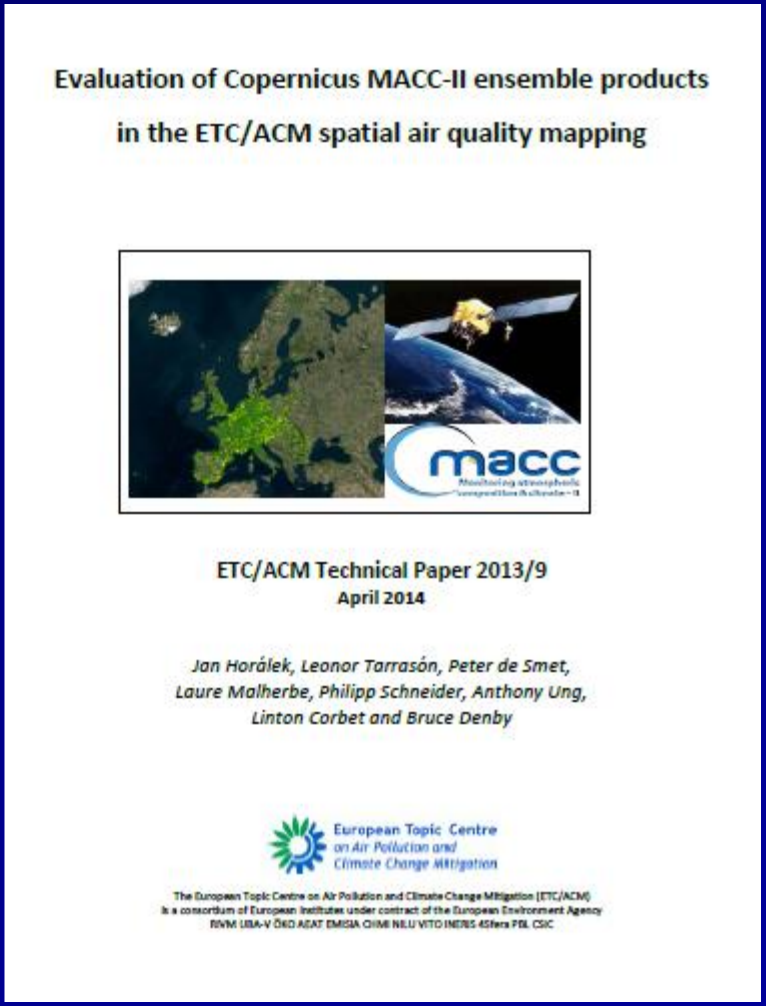
Good representation in1x1 km map, but not in 10x10 km map (bias, RMSE, R^2).

	urb./suburban backgr. stations			
	RMSE	bias	R^2	equation
cross-valid. prediction, separate (r or ub) map	9.1	-0.1	0.70	y = 0.722x + 30.4
grid prediction, 10x10 km separate (r or ub) map	7.6	0.0	0.79	y = 0.768x + 25.4
grid prediction, 1x1 km final merged map	8.0	0.7	0.77	y = 0.770x + 25.8
grid prediction, aggr. 10x10 km final merged map	9.6	3.5	0.71	y = 0.760x + 29.8

Analysis of different CTMs use

Detailed analysis presented in ***ETC/ACM Technical Paper 2013/9 „Evaluation of Copernicus MACC-II ensemble products in the ETC/ACM spatial air quality mapping“***

Comparison of the use of EMEP, MACC-II Ensemble and CHIMERE-EC4MACS (in two different resolution) in ETC/ACM mapping.
Aditionally, comparison of ETC/ACM mapping and the model results


Evaluation of Copernicus MACC-II ensemble products

in the ETC/ACM spatial air quality mapping

macc

ETC/ACM Technical Paper 2013/9
April 2014

Jan Horálek, Leonor Tarrasón, Peter de Smet, Laure Malherbe, Philipp Schneider, Anthony Ung, Linton Corbet and Bruce Denby

European Topic Centre
on Air Pollution and Climate Change Mitigation

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

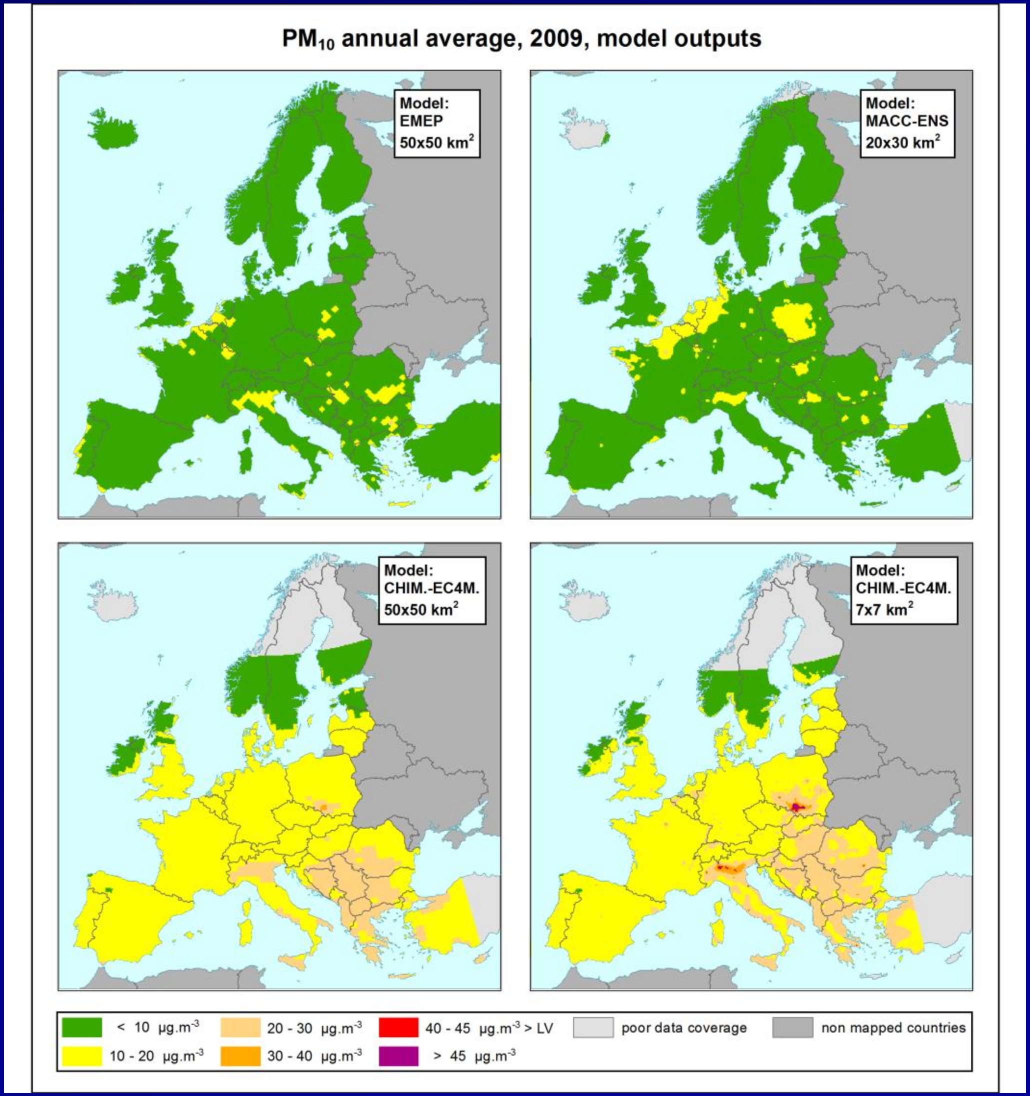
Analysis of the use of other models - continuation

Outputs of different models. PM_{10}, annual average, 2009

Statistical indicators against measured data at rural stations.

chemical transport model	rural			
	RMSE	bias	R^2	regr. equation
EMEP, 50x50	13.05	-11.63	0.254	y = 0.182x + 4.17
MACC-ENS, 20x30	12.49	-11.02	0.261	y = 0.192x + 4.60
CHIMERE-EC4M., 50x50	6.45	-3.67	0.348	y = 0.258x + 10.67
CHIMERE-EC4M., 7x7	5.36	-2.12	0.474	y = 0.414x + 9.21

Different results for different models.



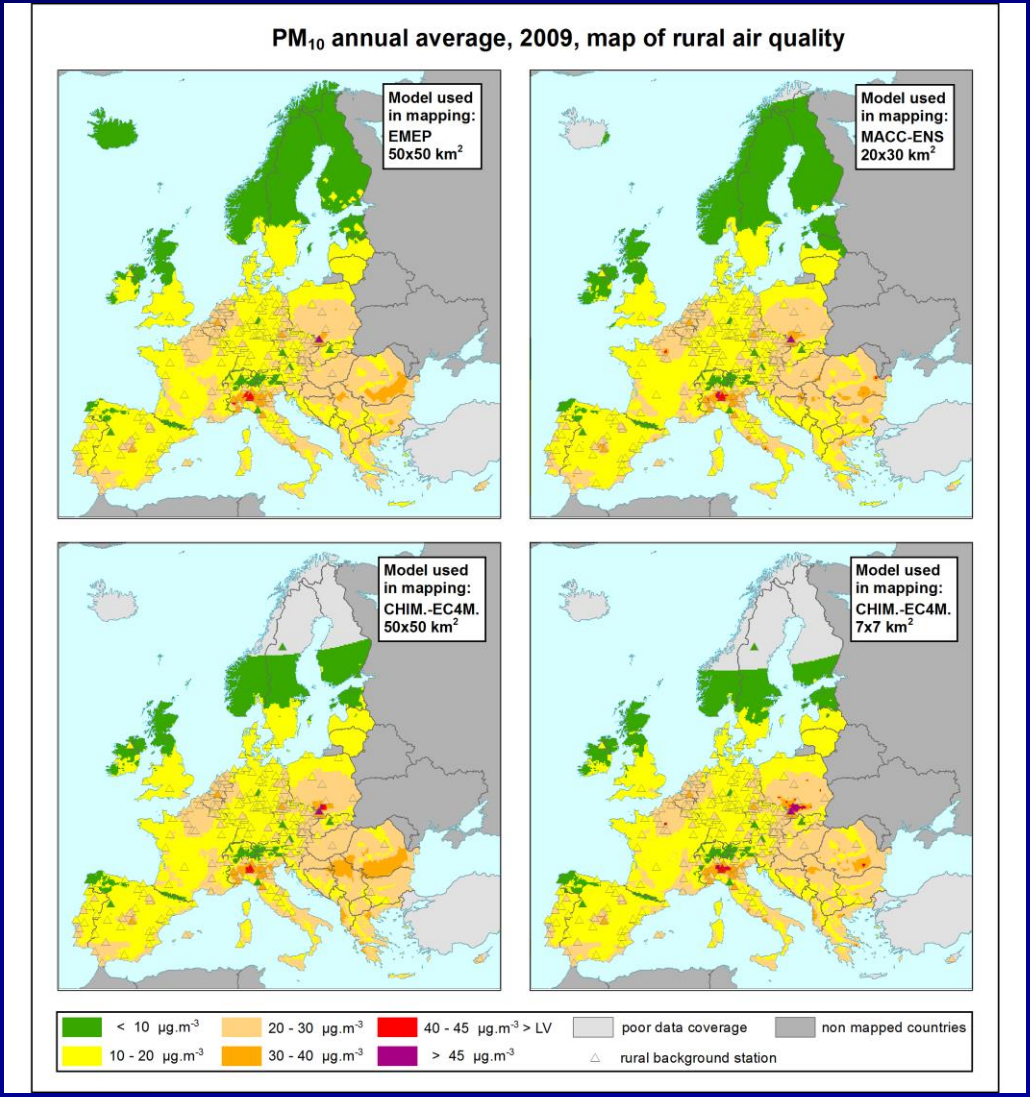
Analysis of the use of other models - continuation

ETC/ACM mapping using different models, rural map, PM_{10}, ann. average, 2009

Statistical indicators using cross-validation at rural stations.

model used in ETC/ACM mapping	rural				
	RMSE		bias	R^2	regr. equation
EMEP, 50x50	4.70	24.3%	0.16	0.518	y = 0.565x + 8.58
MACC-ENS, 20x30	4.55	23.5%	0.16	0.554	y = 0.620x + 7.51
CHIMERE-EC4M., 50x50	4.43	22.9%	0.14	0.568	y = 0.577x + 8.32
CHIMERE-EC4M., 7x7	4.21	21.8%	0.15	0.613	y = 0.645x + 7.01

Similar bias for mapping using different models.



1. Mapping methodology

2. Routine evaluation (especially cross-validation)

3. Evaluation using Delta tool (first attempt)

Evaluation using Delta tool (first attempt)

Obstacles: Annual, not daily data.
Especially: Monitoring data used in the result.

Is Delta tool suitable for the mapping methods based on the combination of monitoring and modelling data?

Approach: To test
- mapping using full set of the stations, against the same set of the stations
- mapping using the assimilation-subset of the stations, against the validation-subset of the stations

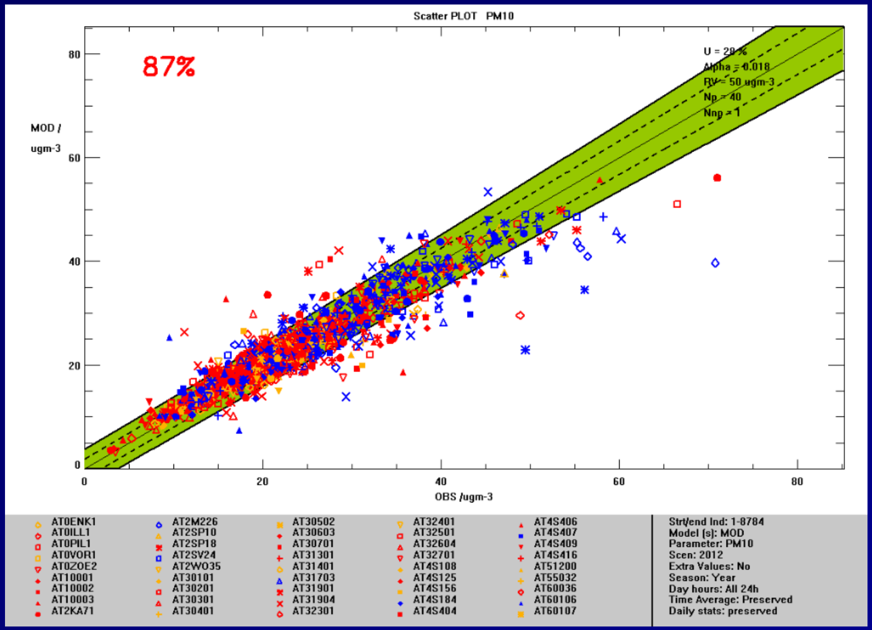
The subsets – received by INERIS, as used in MACC.

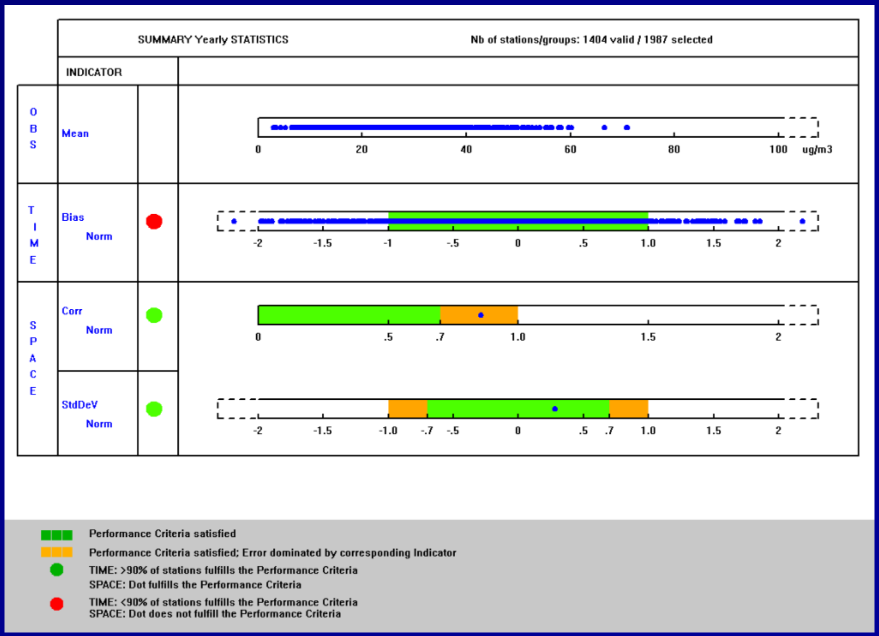
Evaluation using Delta tool (first attempt) – contin.

Preliminary results

PM_{10} annual average, 2012

full set of the stations, all types of the stations



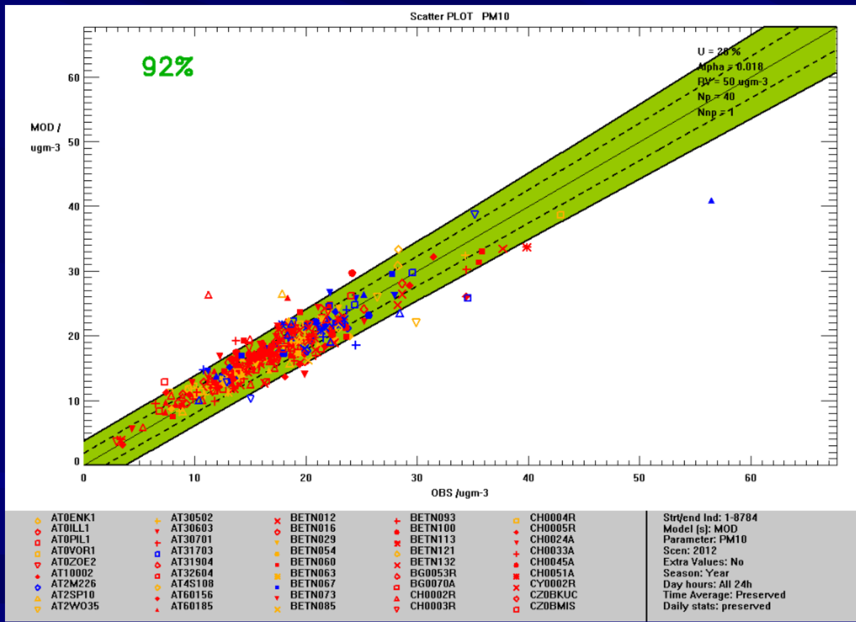


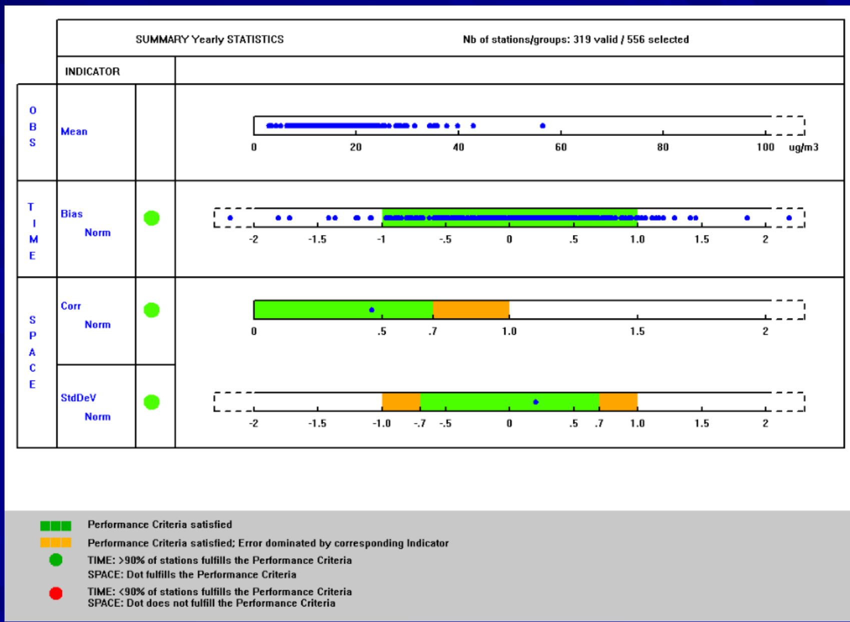
Evaluation using Delta tool (first attempt) – contin.

Preliminary results

PM_{10} annual average, 2012

full set of the stations, rural background stations



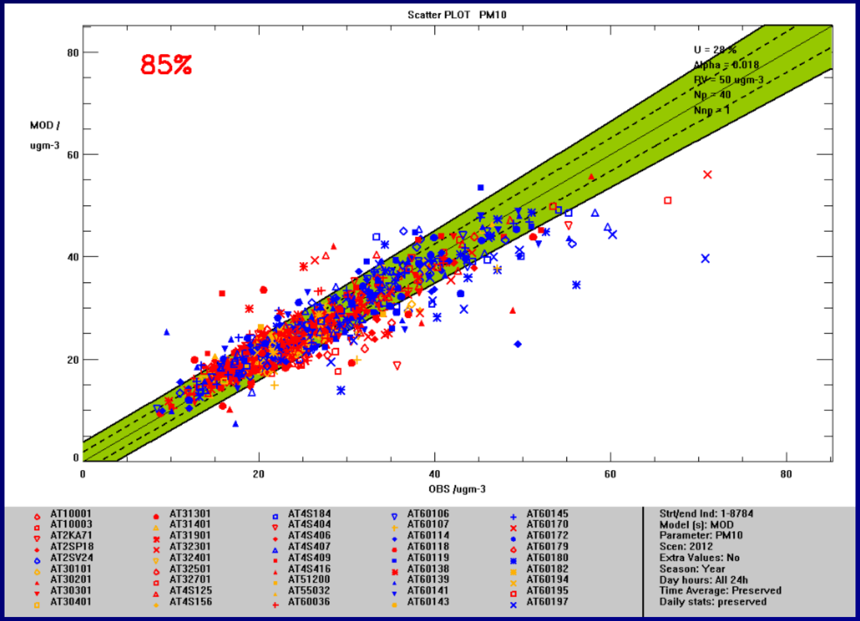


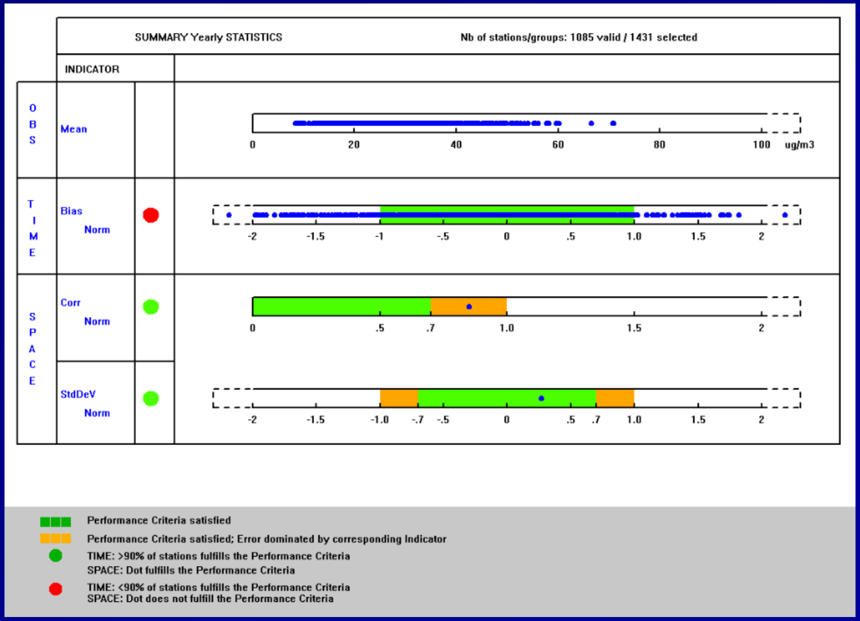
Evaluation using Delta tool (first attempt) – contin.

Preliminary results

PM_{10} annual average, 2012

full set of the stations, urban/suburban backgr. stations



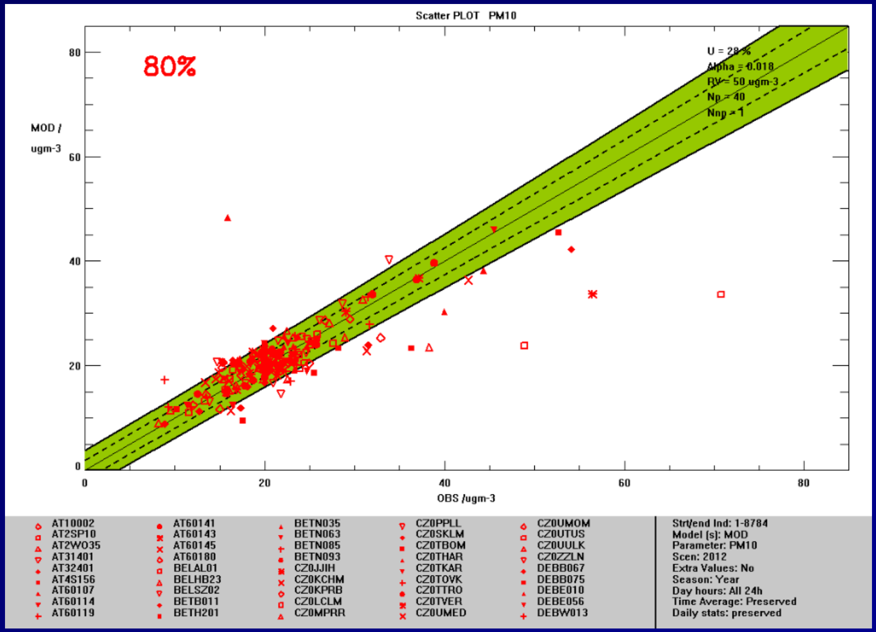


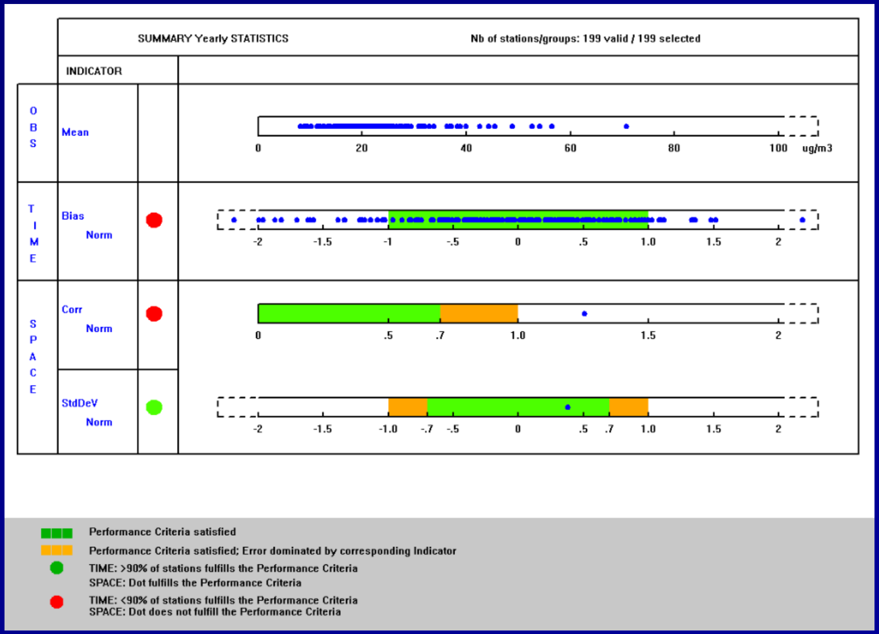
Evaluation using Delta tool (first attempt) – contin.

Preliminary results

PM_{10} annual average, 2012

mapping using assimilation subset of the stations, against validation subset of the stations, all types



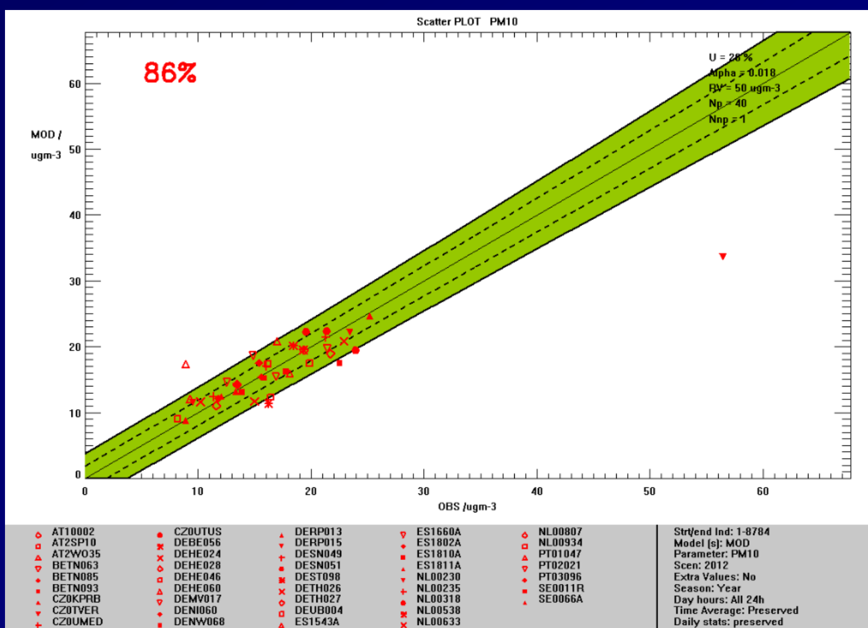


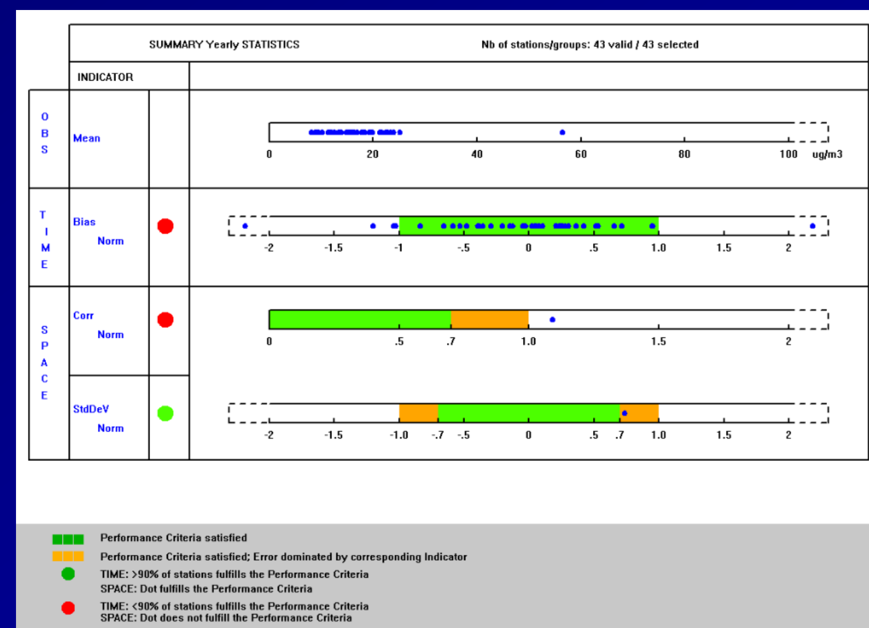
Evaluation using Delta tool (first attempt) – contin.

Preliminary results

PM_{10} annual average, 2012

mapping using assimilation subset of the stations, against validation subset of the stations, rural background stations



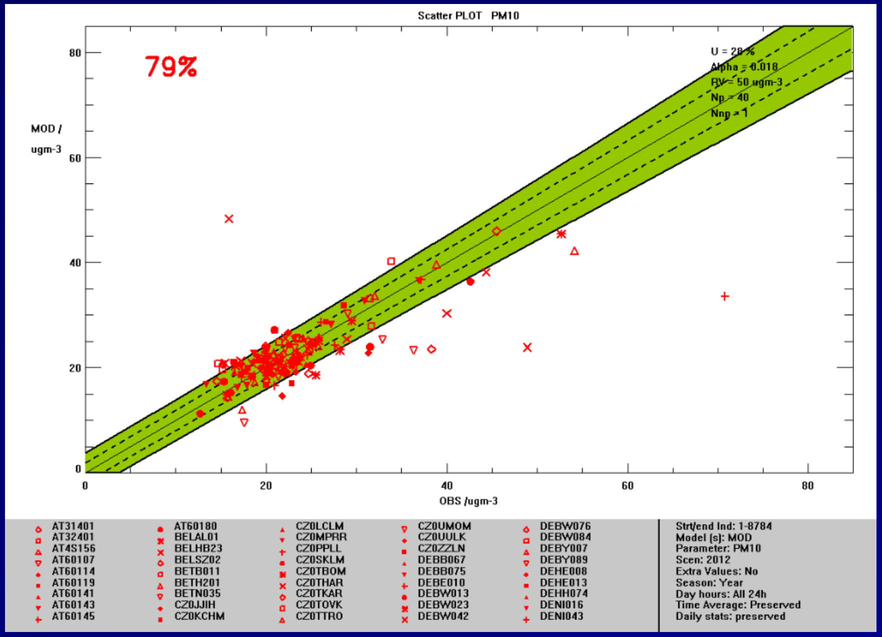


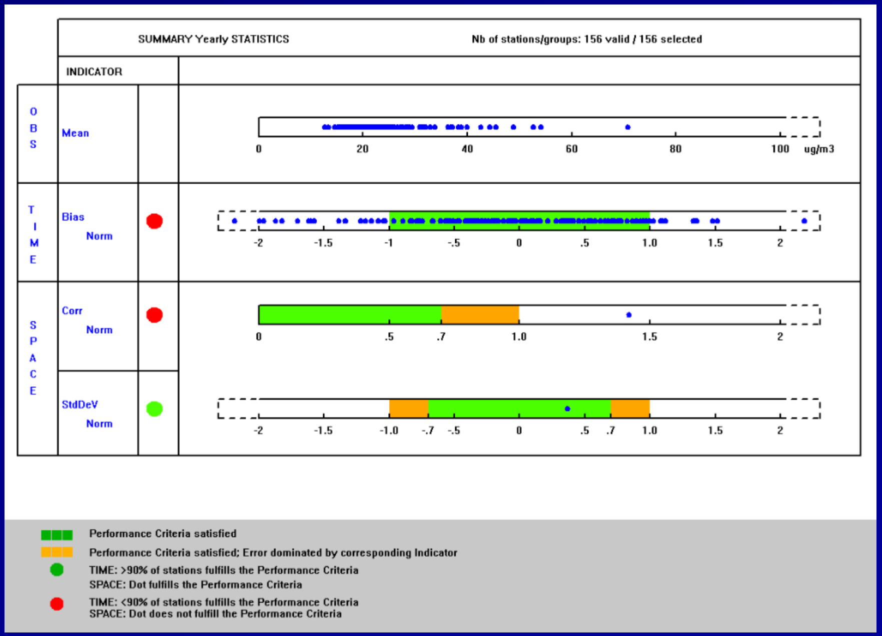
Evaluation using Delta tool (first attempt) – contin.

Preliminary results

PM_{10} annual average, 2012

mapping using assimilation subset of the stations, against the validation subset, urban/suburb. background stations



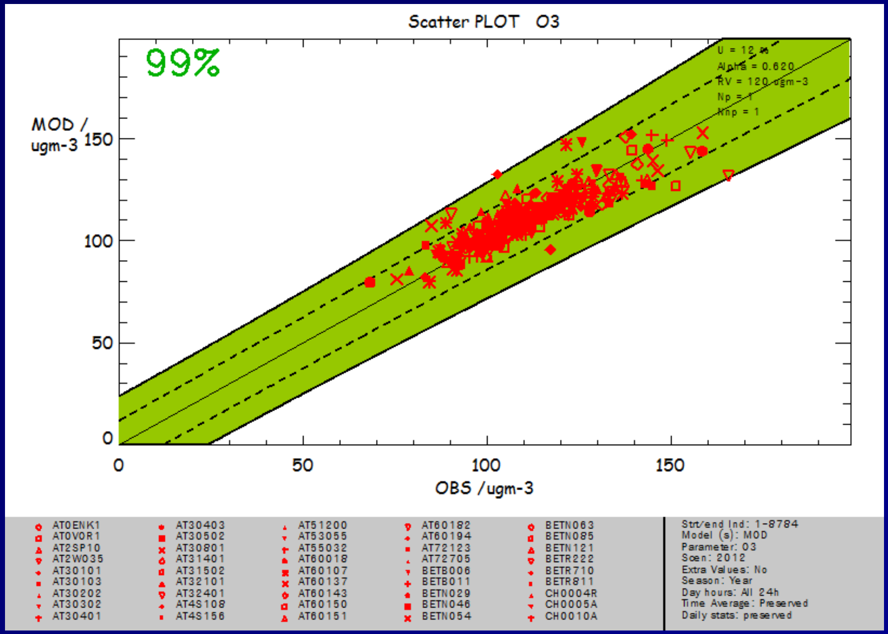


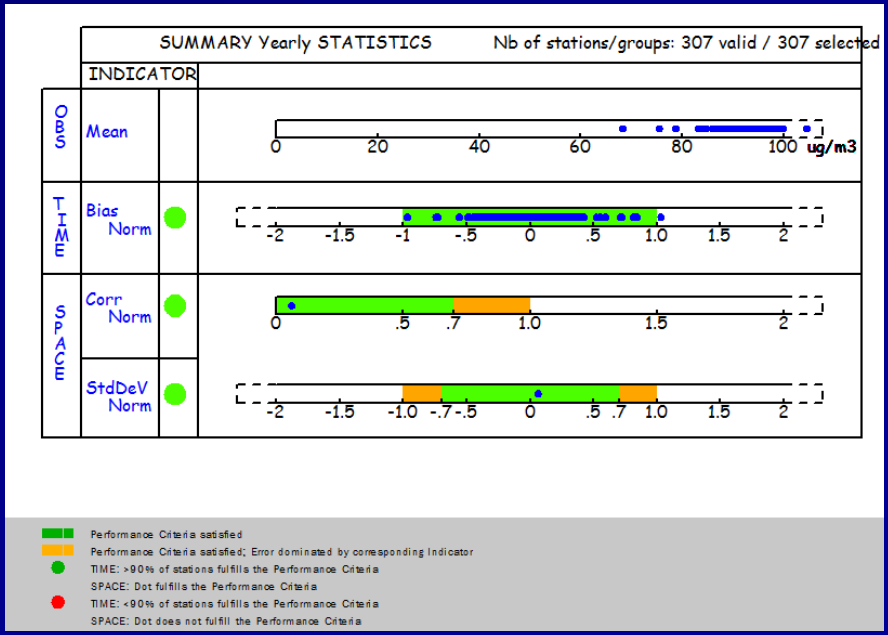
Evaluation using Delta tool (first attempt) – contin.

Preliminary results

Ozone, 26th highest daily max. 8-hourly daily mean, 2012

mapping using assimilation subset of the stations, against validation subset of the stations, all types



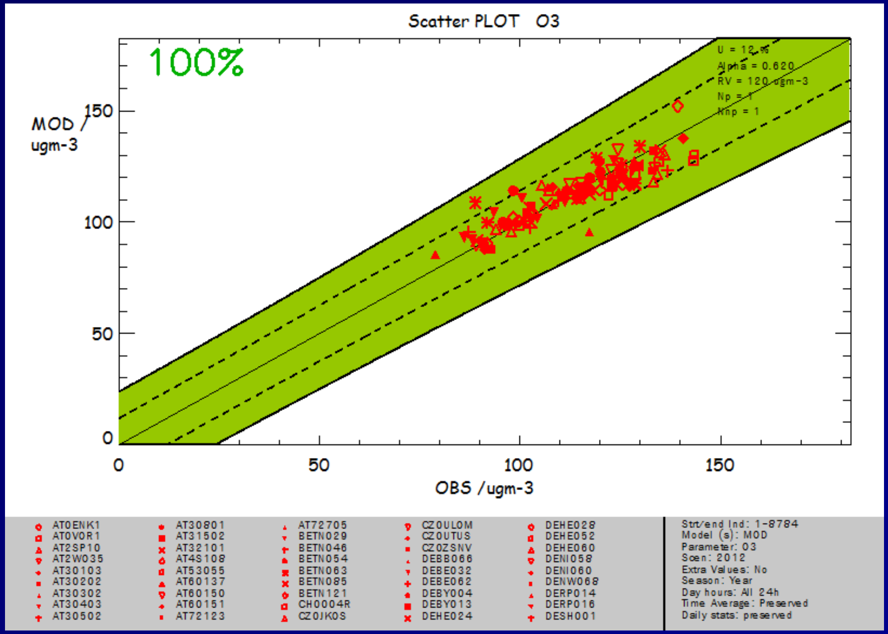


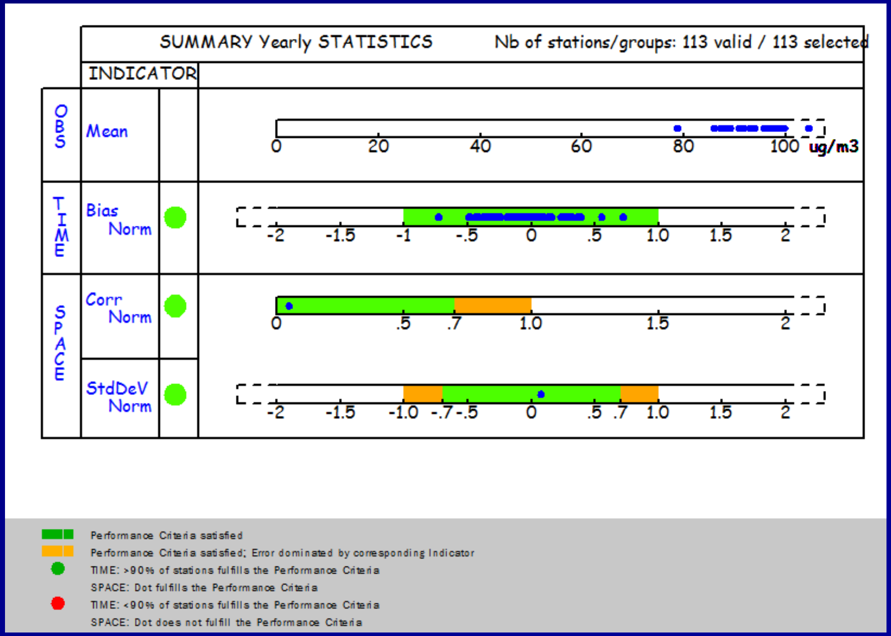
Evaluation using Delta tool (first attempt) – contin.

Preliminary results

Ozone, 26th highest daily max. 8-hourly daily mean, 2012

mapping using assimilation subset of the stations, against validation subset of the stations, rural background stations



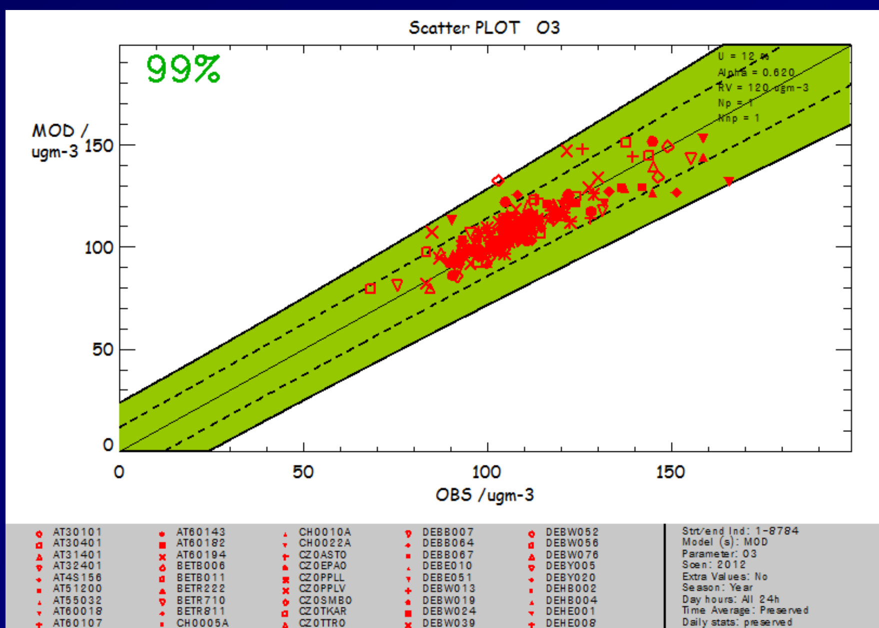


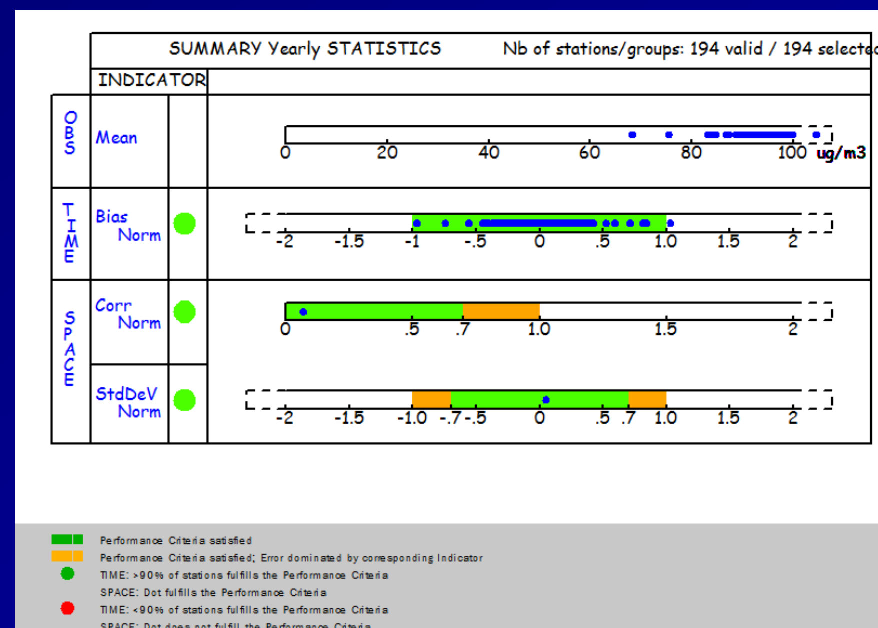
Evaluation using Delta tool (first attempt) – contin.

Preliminary results

Ozone, 26th highest daily max. 8-hourly daily mean, 2012

mapping using assimilation subset of the stations, against the validation subset, urban/suburb. background stations





Conclusions

ETC/ACM spatial interpolation mapping is based primarily on the measured data. Secondary data – chemical transport model data, altitude, meteorology, population density. Linear regression model plus kriging on its residuals.

Urban and rural areas are maped separately, merged together using population density.

Routine evaluation – cross-validation, simple comparison of monitoring and mapped data. Separatelly for rural and urban/suburban backround station.

Evaluation using Dela tool – first attempt, preliminary results. Is Delta tool suitable for the combined monit.-modelled map?

Thank you for your attention.