Project No. 037005

CECILIA

Central and Eastern Europe Climate Change Impact and Vulnerability Assessment

Specific targeted research project

1.1.6.3.I.3.2: Climate change impacts in central-eastern Europe

D7.4: Analysis and evaluation of the results, comparison of the higher resolution runs (10x10) with the lower resolution runs (50x50) for the specific domain – D7.4

Due date of deliverable: 31st December 2009
Actual submission date: 21th January 2010

Start date of project: 1st June 2006 Duration: 43 months

Lead contractor for this deliverable: CUNI

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

D7.4 – Analysis and evaluation of the results, comparison of the higher resolution runs (10x10) with the lower resolution runs (50x50) for the specific domain

1 Introduction

The purpose of this deliverable D7.4 is to provide the analysis and evaluation of the air quality simulations in framework of the main objectives of the project, i.e. in this case the assessment of climate change impact on atmospheric chemistry. Within this tasks the final validation assessment as well as the comparison of the driving runs at 50 km resolution and at 10 km resolution for selected domains have been performed. For this purposes, in the course of the work for previous deliverables, the guidelines for such an validation and results analysis has been prepared (see D7.4 Appendix), which can be basically used for intercomparisons between the model realizations both comparing the model performance at different scales and for different periods of simulations as well.

The concentration of air pollutants depends on both anthropogenic and climate factors. However, in this study the anthropogenic emission are kept for all the time slices at the values of year 2000 to study climate effects only. Longer range transport to the target regions is taken into account from simulation for the whole Europe using Regional Climate Model (RCM) coupled to Air Quality Model (AQM) with the resolution of 50x50 km. These simulations are used to constrain nested higher resolution runs (10x10 km) focusing in CEE both for present and future climate. The key species are ozone, sulphur, nitrogen and PM, which have a central role in tropospheric chemistry as well as the strong health impacts.

It is now well established that climatically important (radiatively active) gases and aerosols can have substantial climatic impact trough their direct and indirect effects on radiation, especially on regional scales (Qian and Giorgi, 2000, Qian et al., 2001, Giorgi et al., 2002). To study these effects requires coupling of regional climate models with atmospheric chemistry/aerosols to assess the climate forcing to the chemical composition of the atmosphere and its feedback to the radiation, eventually other components of the climate system. In this study climate is calculated using model RegCM while chemistry is solved by model CAMx. The model RegCM was originally developed and further improved by Giorgi et al. (1999) or later see e.g. in Pal et al. (2007). For more details on the use of the model see Elguindi at al. (2006).

CAMx is an Eulerian photochemical dispersion model developed by ENVIRON Int. Corp. (Environ, 2006). In version 4.40 CAMx is used for air quality modeling here, with CB-IV gas phase chemistry mechanism option, wet deposition of gases and particles. It uses mass conservative and consistent transport numerics in parallel processing. It allows for integrated "one-atmosphere" assessments of gaseous and particulate air pollution (ozone, PM2.5, PM10, air toxics) over many scales ranging from sub-urban to continental. CAMx simulates the emission, dispersion, chemical reactions and removal of pollutants in the troposphere by solving the pollutant (eulerian) continuity equation for each chemical species on a system of nested three-dimensional grids. These processes are strongly dependent on the meteorological conditions, therefore CAMx requires meteorological input from a NWP model or RCM for successful run.

2 Model validation at WUT domain

At WUT the RegCM-EMIL-CAMx modelling system was implemented for the modelling domain, centred over Poland (52.00°N, 19.30°E) on a grid with 120 x 109 points and a resolution of 10 km (so-called WUT domain). The map projection choice was Lambert conformal. The high-resolution simulations were performed for the year 2000 in order to evaluate modelling results. RCM simulations have been completed at WUT as the added value within the project (WP2).

The RCM model applied is the improved version of RegCM3 for high resolution use – RegCM3-Beta. For reference year simulation, reanalysis ERA-40 meteorological fields were used to drive RegCM.

CAMx is a complex third-generation Eulerian Grid Model developed at ENVIRON International Corporation (Novato, California). Simulations have been carried out using CAMx v. 4.40. The chemistry mechanism invoked was Carbon Bond version 4 (Gery et al., 1989), including 117 reactions – 11 of which are photolytic – and up to 67 species (37 state gases, up to 18 state particulates and 12 radicals). The domain's vertical profile contained 12 layers of varying thickness, extending up to 450 hPa. The output fields from RegCM3-Beta were used to drive the CAMx model. All meteorological fields required by CAMx as well as biogenic emissions were calculated by a RegCM-CAMx pre-processor in a 6-hour basis. Top boundary conditions corresponded to concentrations of clean air. Lateral boundary concentrations were extracted as monthly mean concentrations, while initial conditions as mean of January concentrations from the results of the 50 km CAMx simulations for Europe performed by BOKU (Krüger et al., 2008). The emission input as well as the output of CAMx model have a resolution of one hour.

Anthropogenic emissions used in CECILIA project are based on the UNECE/EMEP database (http://webdab.emep.int/) for European emissions (Vestreng et al., 2005) for the year 2000 and are kept constant for all simulations. The detailed explanation of preparation of emission database for 50 km runs is given by Krüger et al. (2008). The so called zero-level emission database for the high resolution 10 km photochemical runs performed at WUT were compiled and provided by BOKU. The anthropogenic emissions were calculated with the emission model of BOKU-Met based on data from the UNECE/EMEP data base for the year 2000, available in 50 km x 50 km EMEP grid resolution. For the Pannonian countries (Austria, Czech Republic, Hungary, and Slovakia), a detailed 5 km x 5 km emissions inventory from the year 1995 (Winiwarter and Zueger, 1996) is used as data base for the spatial distribution of the emitters within the 50 km x 50 km EMEP grid cells for every sector from SNAP97 inventory. Next, for every emission sector the EMEP data are distributed to a spatial resolution of 10 km x 10 km. For the temporal disaggregation the BOKU emission model applies different distributions for the month, the day of the week and the hour of the day. The disaggregation factors are taken from the inventory by Winiwarter and Zueger (1996). They are available for the Pannonian countries. For all other countries the data for Austria have been used. In zero-level emission database all emissions were treated as surface area emissions.

At WUT the original emission model EMIL (EMIssion modeL) was developed during third reporting period (see Deliverable D7.3). The model applies for Poland and creates the so called second-level emission database for the high resolution 10 km photochemical runs.

One of the main added value to the project was development of a detailed emission and emission parameters database for a Large Combustion Plants (LCP) sources (with a stack height that is equal or above 100 m; $h \geq 100$ m) for PM_{10}, $PM_{2.5}$, SO_2, NO_x, NH_3 and NMVOC (see Deliverable D7.3 for details). Thus simulation of pollution plume from point sources was possible and constitutes the largest differences between air pollution modelling at WUT and at CUNI, AUTH and BOKU, where in CAMx simulations all sources were treated as area sources. In CAMx (WUT) emission database applied for other, than Poland, countries belonging to modelling domain (see Fig. 1) remains zero-level (based on EMEP inventory and without distinction for point sources).

We performed simulation for PM_{10} and SO_2 concentrations as well as for total deposition of oxidized sulphur, total deposition of oxidized nitrogen and total deposition of reduced nitrogen for the basic year of 2000. Due to lack of dry deposition measurements, evaluation of the results was possible only for concentration fields.

The statistical evaluation of Air Quality Models performance focuses on assessing the accuracy of the model predictions relative to observations. Several scientists, carried out discussion on the evaluation methods and criteria, e.g. Willmott (1982), Hanna et al., (1993), Brandt et al. (2001), Juda-Rezler (1986, 1989, 2010), Chang and Hanna (2004). However, standard evaluation procedures and performance standards still do not exist. Recently, Borrego et al. (2008) presented systematic description of the modelling uncertainty analysis methodologies as well as proposal of guidelines for uncertainty estimation.

The latest EU Directive (2008/50/EC) establish requirements for air quality modelling, including the definition of the modelling quality objectives, as a measure of modelling results acceptability. In this context, the uncertainty for modelling and the quality objective estimation is defined as the maximum deviation of the measured and calculated concentration levels, over the period for calculating the appropriate threshold, without taking into account the timing of the events. As was concluded by Flemming and Stern (2007) and Borrego et al. (2008) the quality indicators defined by EU directives are ambiguous and inadequate in several aspects, mainly concerning the error measures for hourly and daily indicators, as error measure is based on the highest observed concentration. The most robust measure – relative percentile error (RPE_p_LV) – as an alternative to relative maximum error (RME) was proposed by Flemming and Stern (2007).

During validation process of the Eulerian Grid Models, observed values of point measurements at a station, are compared against predicted values averaged for the grid cell area. Therefore, not all existing station data could be used for model validation purposes. The EU CAFE Directive (2008/50/EC) states only that "*the fixed measurements that have to be selected for comparison with modelling results shall be representative of the scale covered by the model*", without any recommendations concerning "representativeness". It can be concretized as follows: stations chosen for validation should be representative of the grid area climatic conditions (i.e. stations situated in specific conditions, for example sites of a high elevation, should be excluded) as well as for average air quality within the grid area (i.e. station should not be influenced by local sources). In addition, the usual requirement of temporal data completeness should be met as well as the requirement of a statistically sufficient number of stations, covering the entire area of interest.

For the purpose of WP7 of CECILIA and Deliverable D7.4 *Guidelines for operational evaluation of the AQ-CTMs under WP7 of the CECILIA project* have been prepared by a group leader, Katarzyna Juda-Rezler (see APPENDIX 1).

The paper addresses Air Quality – Chemical Transport Models (AQ – CTMs) evaluation in terms of operational evaluation, which is aiming on comparing model results with measurements of species concentrations for a specific time period. For CECILIA evaluation exercise, ten standard and well-accepted measures of model performance were proposed. Modelling Quality Objectives, defined (unfortunately very imprecisely) in the CAFE Directive (2008/50/EC), are interpreted and taken into account as well (see APPENDIX 1).

For final evaluation performance, a subset of parameters, which characterise the general uncertainties estimation, was applied. The subset consists of the following measures: NMB (Normalized Mean Bias), RMSE (Root Mean Square Error), with its systematic (RMSEs) and unsystematic (RMSEu) part, NMSE (Normalized Mean Square Error), correlation coefficient (r), IA (Index of Agreement) and FAC2 – a fraction of predictions within a factor 2 of observations. The formulas are given in APPENDIX 1.

RegCM-EMIL-CAMx modelling system was evaluated for PM_{10} and SO_2 concentrations using observations from EMEP (http://www.emep.int) and EIONET-Airbase (http://air-climate.eionet.europa.eu) databases. For PM_{10} evaluation, data from IfT (Leibniz Institute for Tropospheric Research, Germany) research station Melpitz were also used. The Melpitz research station is one of the European station with the longest PM observations. From EMEP and Airbase databases only rural background stations with annual data coverage of at least 75% has been considered for evaluation. Those rural background stations which are located near cities were excluded. For the reference year 2000, the above given requirements resulted with 30 stations for PM_{10} and 93 for SO_2 available for model evaluation. Unfortunately, in 2000 in Poland, there were only one such station for PM10 and 5 stations for SO_2. The location of chosen stations is presented in Fig. 1 and 2. Both qualitative analysis (scatter plots, time series) and quantitative analysis were performed.

For evaluation of annual mean model predictions, the scatter plots of the predicted versus observed annual mean PM_{10} and SO_2 concentrations in 2000 together with calculated values of statistical indicies are given in Fig. 3 and 4, respectively, whereas the statistical indices are given in the Tab. 1 as well.

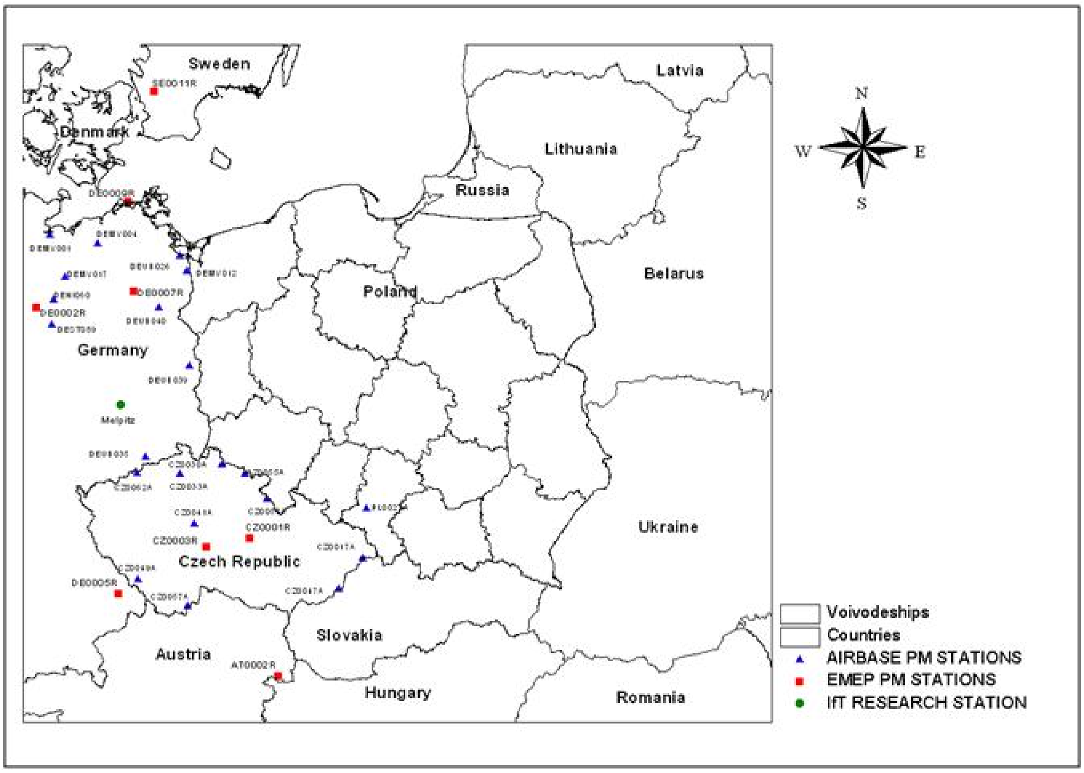

Figure 1. The WUT modelling domain with PM_{10} rural background monitoring stations.

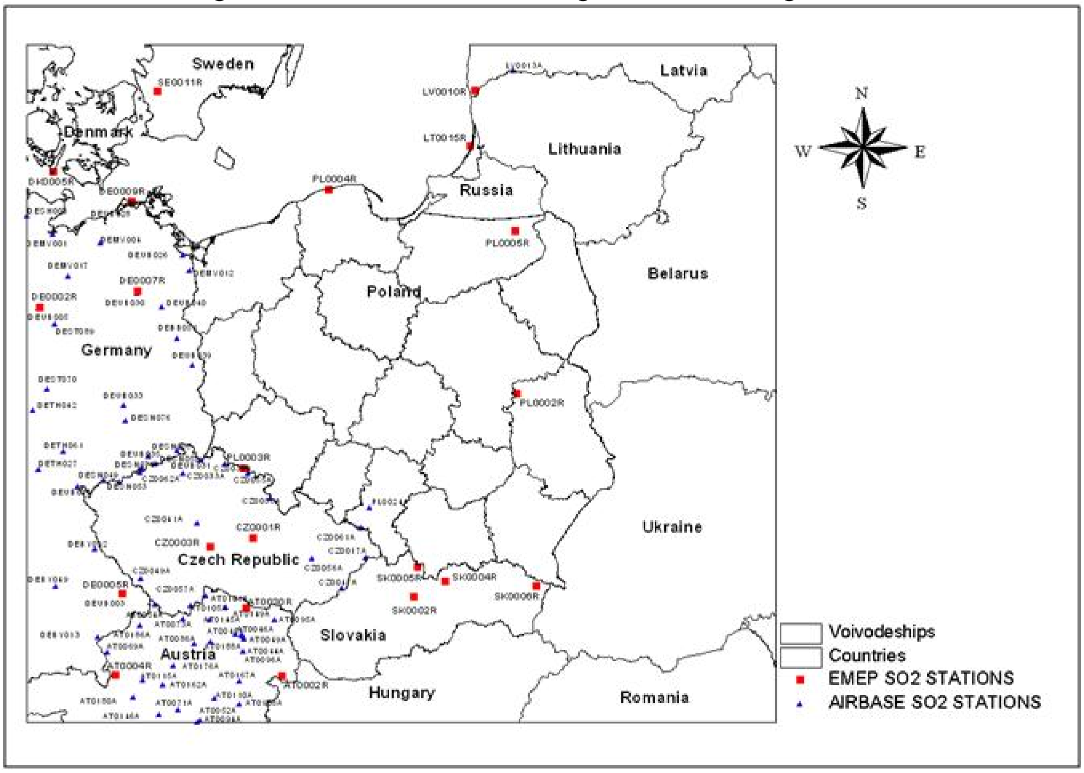

Figure 2. The WUT modelling domain with SO_2 rural background monitoring stations.

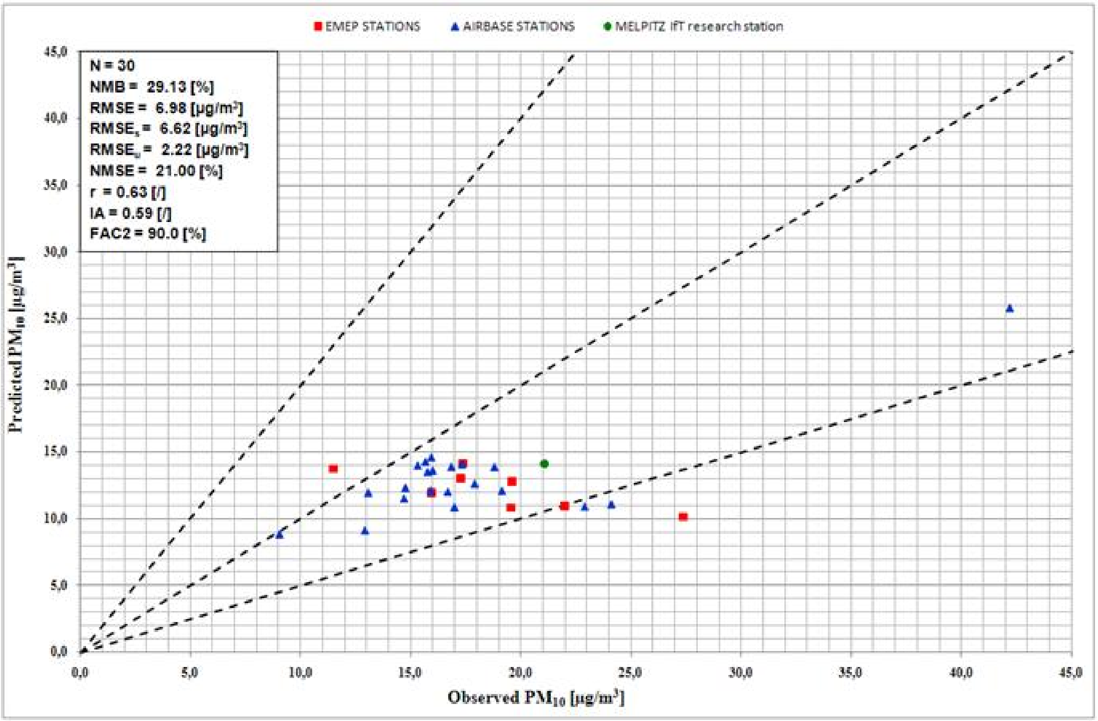

Figure 3. The scatter plot of the predicted and observed annual mean PM_{10} levels for WUT domain (year 2000). Dashed lines indicate perfect agreement (middle line) and a difference of a factor of 2.

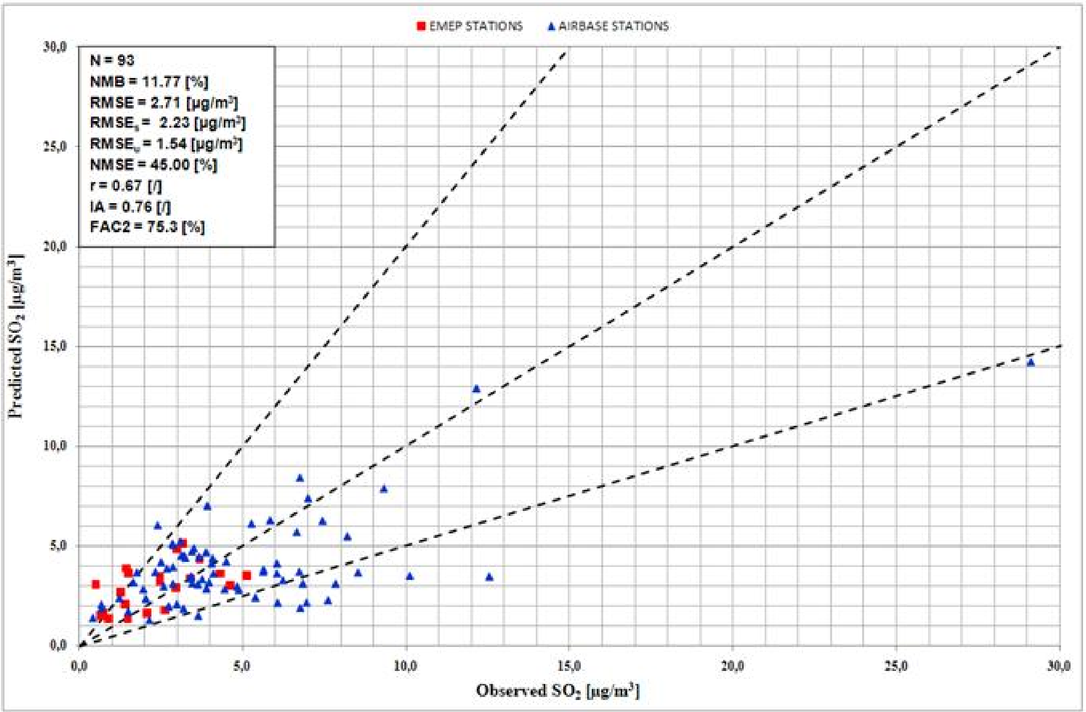

Figure 4. The scatter plot of the predicted and observed annual mean SO_2 levels for WUT domain (year 2000). Dashed lines indicate perfect agreement (middle line) and a difference of a factor of 2.

Table 1. Evaluation results of the RegCM–EMIL–CAMx modelling system for 2000 reference year.

Statistical measures	PM_{10} annual mean	PM_{10} daily mean	SO_2 annual mean
N – number of samples	30	10 188	93
NMB – Normalised Mean Bias [%]	29.13	27.20	11.77
RMSE – Root Mean Square Error [μg/m^3]	6.98	12.92	2.71
$RMSE_s$ – Systematic Root Mean Square Error [μg/m^3]	6.62	10.85	2.23
$RMSE_u$ – Unsystematic Root Mean Square Error [μg/m^3]	2.22	7.02	1.54
NMSE – Normalised Mean Square Error [%]	21.00	73.20	45.00
r – Correlation coefficient [/]	0.63	0.33	0.67
IA – Index of Agreement [/]	0.59	0.55	0.76
Predictions within a factor of 2 of the observations [%]	90.00	64.00	75.30

For annual means, evaluation results indicate a satisfactory model performance for both pollutants, however model performance for SO_2 is better than for PM_{10}. For PM_{10}, the observations are in general underestimated. The predicted spatial mean concentrations are very close to the measured means and the NMB bias is equal to 11.8% for SO_2 and to 29.3% for PM_{10} .The RMSE is also quite low; however, its systematic part is substantial for both pollutants, which means that there is a systematic error in the predictions. This is probably due to a specific of grid model. Also correlation measures (r, IA) show better model performance for SO_2. Model predictions satisfy FAC2 requirements for 90% of data points in the case of PM_{10} and for 75% data points for SO_2.
For the daily mean levels we performed evaluation analysis only for PM_{10} as priority species in the project. The scatter plot of the predicted versus observed daily mean PM_{10} levels in 2000 together with calculated values of statistical indices are given in Fig. 5, while the statistical indices are given in the Tab. 1 as well. Moreover the time series of daily mean PM_{10} values are presented for two stations: IfT research station in Melpitz, Germany (Fig. 6) and Kuznia, Poland (Fig. 7).

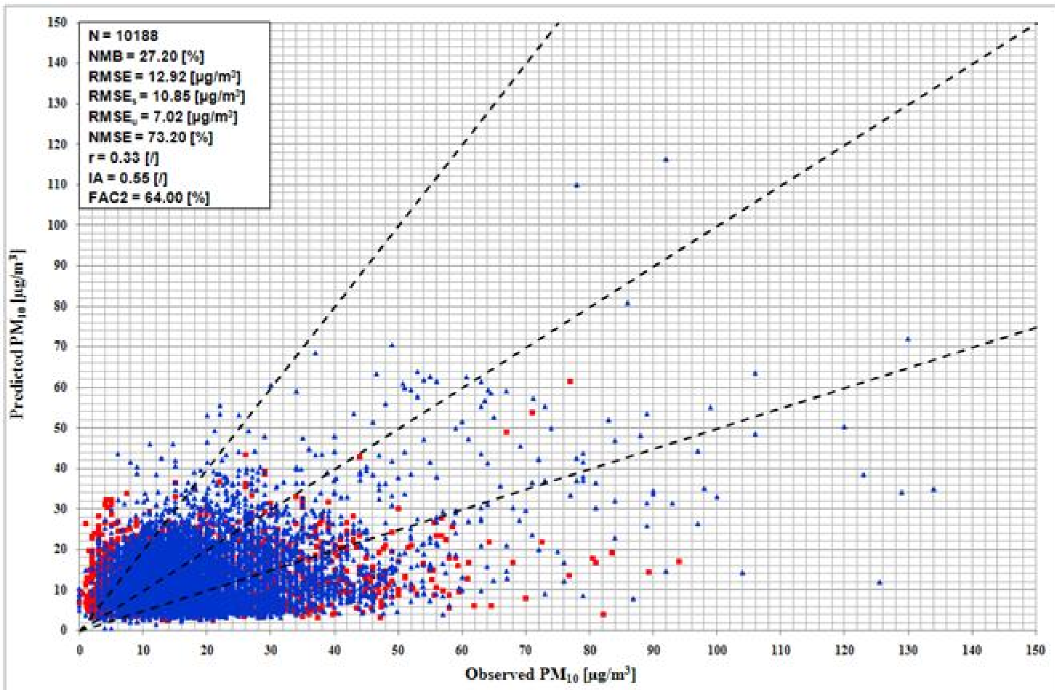

Figure 5. The scatter plot of the predicted and observed daily mean PM_{10} levels for WUT domain (year 2000). Dashed lines indicate perfect agreement (middle line) and a difference of a factor of 2.

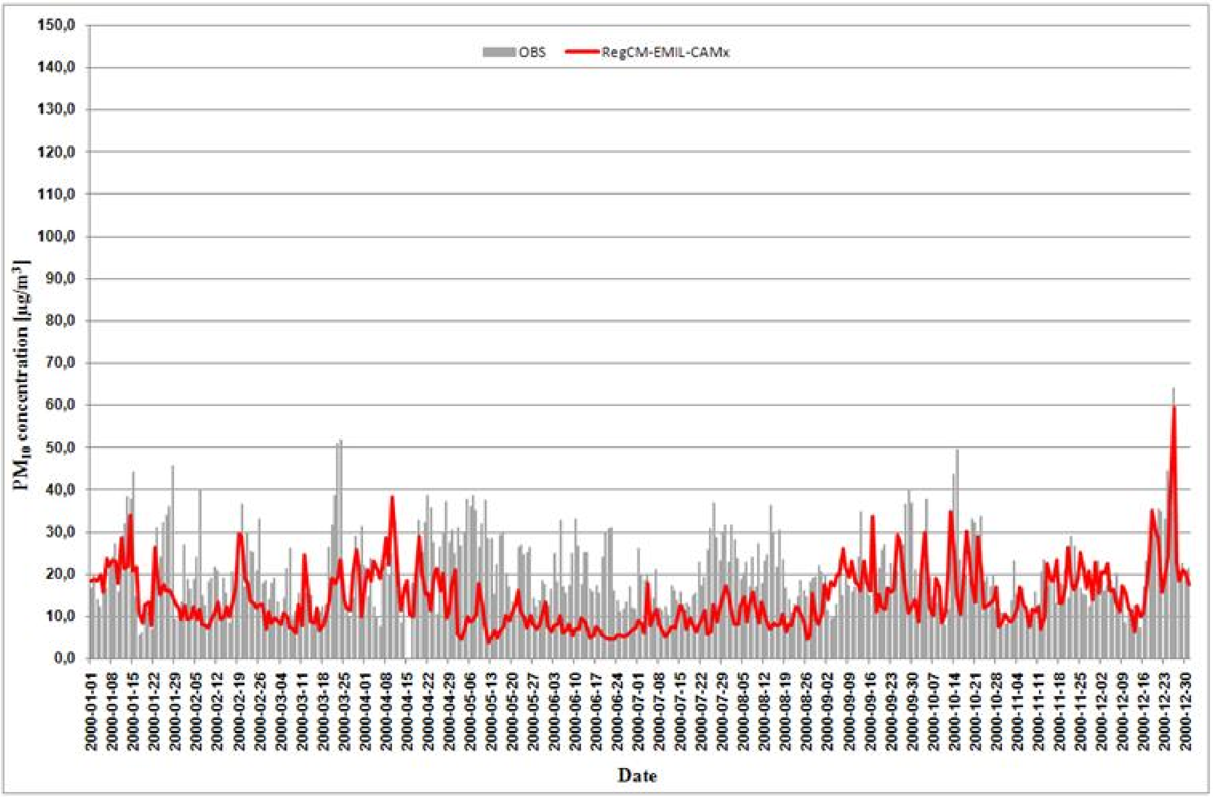

Figure 6. Observed and calculated PM_{10} time series of daily mean concentrations at Melpitz IfT research station (Germany) – year 2000. Observed data by courtesy of dr Gerald Spindler (IfT, Leipzig, Germany).

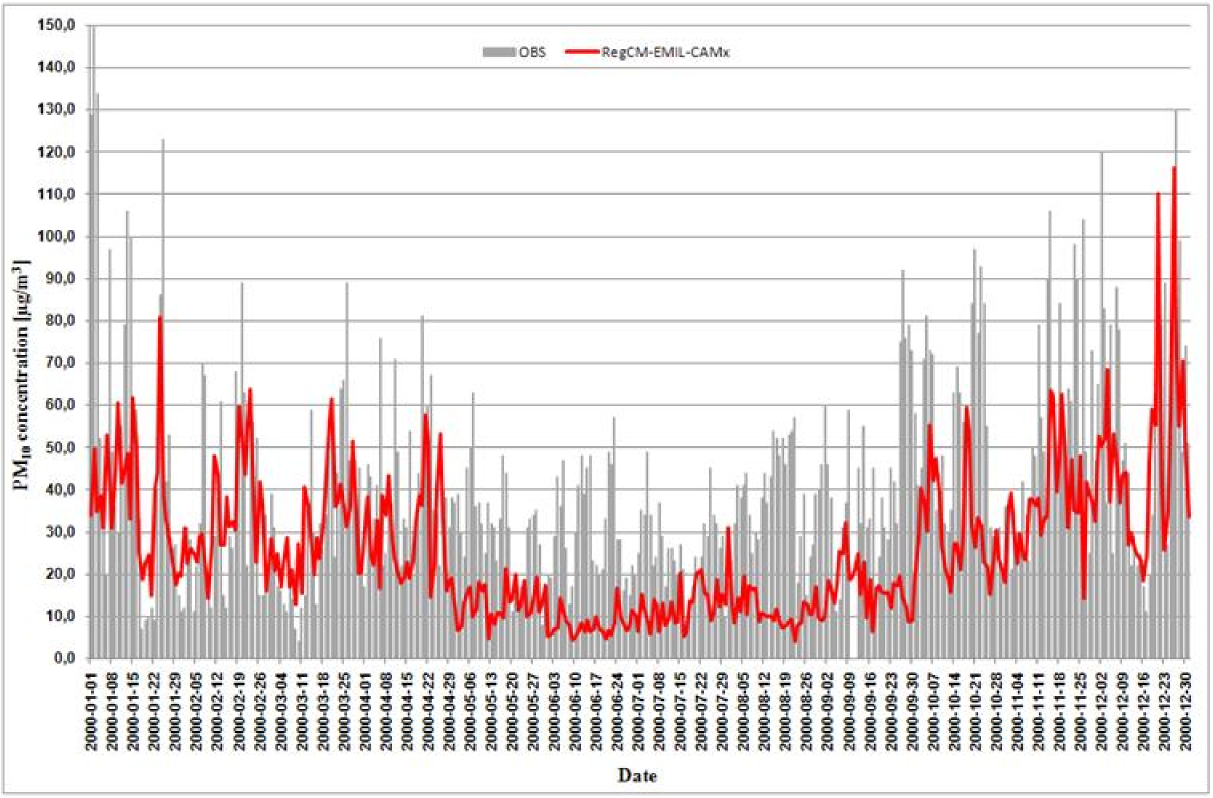

Figure 7. Observed and calculated PM_{10} time series of daily mean concentrations at Kuznia (Poland) – year 2000.

As it can be seen in Tab. 1 for daily mean predictions, evaluation results indicate poorer model performance than for annual means. This is quite understandable taking into account the specific of PM_{10} air pollution, which is due to both natural and anthropogenic sources, moreover it is both primary and secondary pollutant. Still, when comparing predicted and observed time-series it can be noted that during winter, the patterns of measured PM_{10} levels are fairly well simulated by the RegCM-EMIL-CAMx modelling system. For Melpitz research station winter PM_{10} levels, specially for January and October-December, are well predicted by model and peak values are well captured. During summer the model performance is worse. For Polish station Kuznia, model performance is better in winter than in summer as well. Overall, the predicted pattern of measured time series in Kuznia is poorer than in Melpitz, however model is also able to capture peak, up to 115 $\mu g/m^3$, values. The Polish station is categorized in Airbase database as rural background station, however it is situated in Upper Silesia region, which is the most industrialized and polluted region in Poland. In such region comparison of point measurement with 10 x 10 km grid average value can hardly give perfect agreement.

Summarizing, the ability of the modelling system to simulate PM_{10} and SO_2 concentrations in Central-Eastern Europe is promising. Detailed emission data provided by the emission model EMIL play important role in improving the model performance.

3 Comparison of 10 km- and 50 km-spatial resolution

Within the CECILIA project at BOKU photochemical model runs with CAMx were performed for three decades with two different spatial resolution, namely 50 km and 10 km. The decades were 1990-2000 with RegCM-data driven by reanalysis ERA40 for the validation of the model, 1990-2000 with meteorological data driven by the GCM ECHAM as the control run, and the end-century decade 2090-2100, also GCM-driven. The mid-century-decade 2040-2050 has only been calculated with 50 km resolution. Since differences in the results for ozone were small here, the calculation of this decade with 10 km resolution was skipped.

For a direct comparison of the two resolutions, the 50 km results were interpolated to the grid of the high resolution runs. Fig. 8 displays the difference between the control-runs for the decadal average (1991-2000) in the average ozone volume mixing ratio in summer (JJA). In most of the model domain the ozone in the 10 km runs is slightly higher by less than 5 ppbv. In the urban centers Prague, Vienna, and Budapest a decrease is observed. The industrial centers in the south of Poland and the north of Czech Republic show a remarkable increase of ozone with increasing resolution in the west of these regions and a decrease in the east. At the Adriatic Sea there is a decrease along the west coast and an increase at the east coast. For the other two calculated decades very similar differences are found.

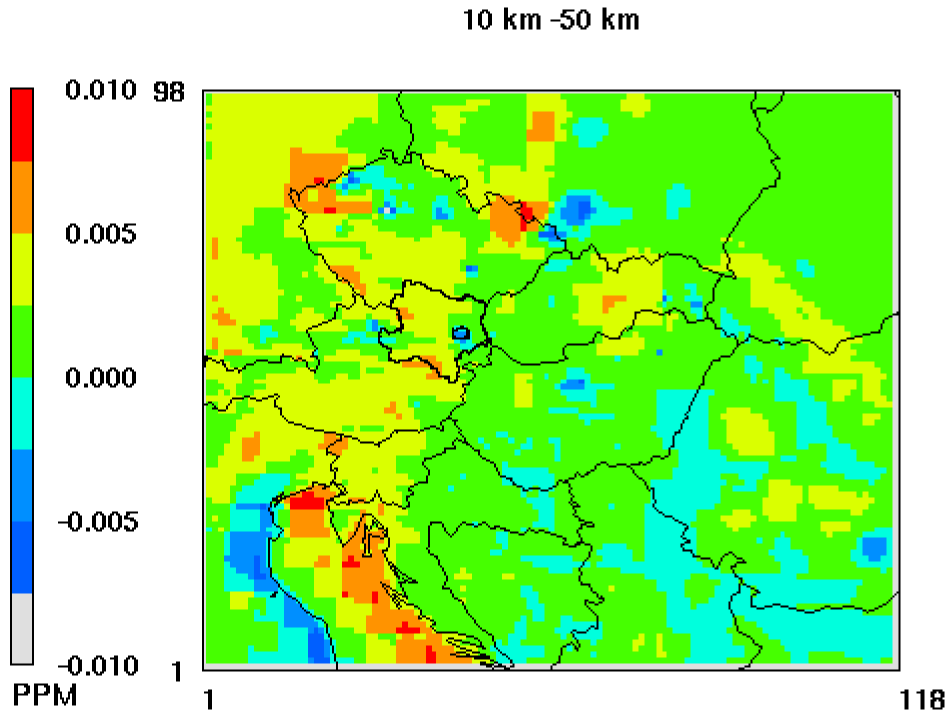

Figure 8. Difference in the average ozone volume mixing ratio for the control decade (1991-2000, RegCM driven by ECHAM) in summer (JJA) between in runs with 10 km and 50 km spatial resolution.

If the response of the ozone concentration to a changing climate is compared between the model runs with different spatial resolution, the found absolute differences of ±2 ppbv are much smaller. Fig. 9 compares the change of the ozone volume mixing ratio between the decade at the end of the century (2091-2100) with the control decade (1991-2000) for the two spatial resolutions. With the 10 km resolution the climate response is somewhat less than with 50 km in most of the model domain except of the very north. However, the 10 km calculation shows an effect of the model boundary in the results, which might be a reason for this discrepancy.

It had been observed, that the solar radiation, which drives the photochemical production of ozone, was generally lower in the 10 km calculations made by BOKU than in the 50 km calculations. It was caused by a higher cloud water content in the 10 km-RegCM-results used. This might be a reason for the smaller climate response in the high-resolution runs. It certainly is the reason for the decrease of ozone in the urban areas, since the nightly titration of ozone by nitrogen oxides is not effected by photolysis. Therefore a higher resolution leads to higher NO_x-concentrations at night near the emission sources and less ozone. On the other hand, an increase of urban plumes with high ozone concentrations due to the higher spatial resolution of the model is not observed, since this depends on photochemistry.

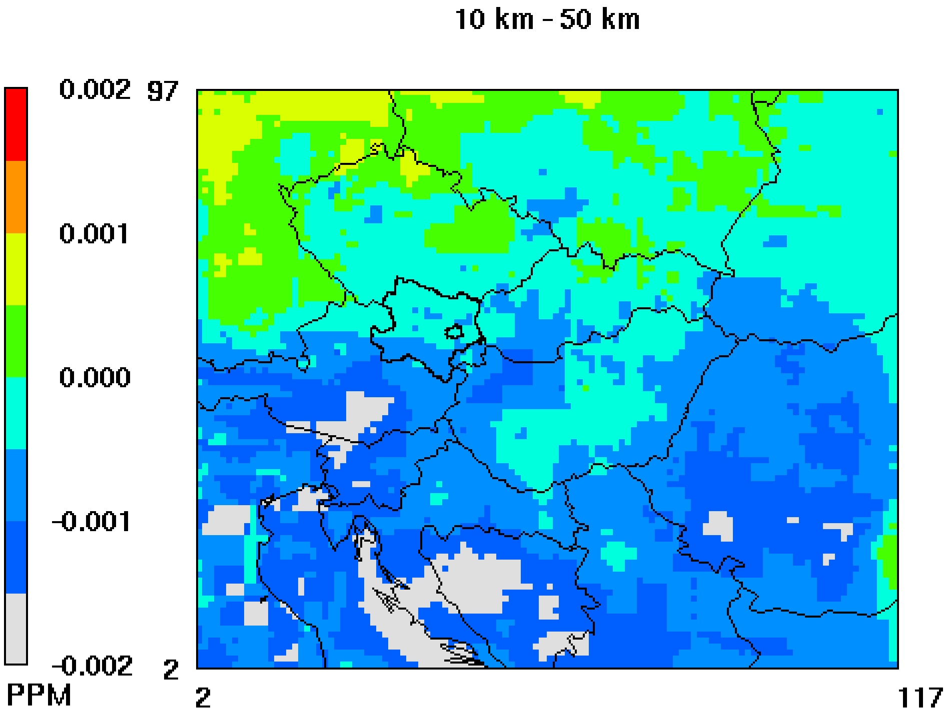

Figure 9. Difference of the climate response in the seasonal ozone average between 10 km and 50 km spatial resolution model runs in summer (JJA).

It was also compared, how the values for AOT40 (May-July) differ between the model runs with different spatial resolution. Again the data from the 50 km-run were interpolated to the 10 km-grid and the differences are displayed. Fig. 10 shows the difference in the calculated average of AOT40 for the control period (1991-2000) between the two runs. The highest difference is found over the Adriatic Sea. The 10 km calculation also leads to higher AOT40-values over the mountains of the Alps and the Carpathians and over Italy and western Croatia. Over Poland and Hungary the results of the two model calculations are very similar.

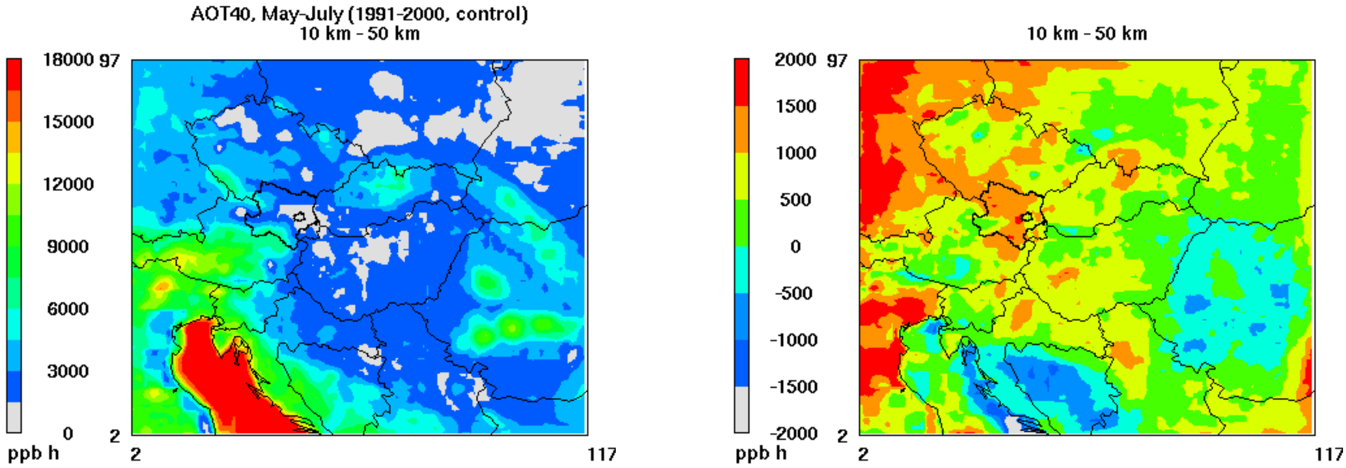

Figure 10. Left: Difference of the decadal average of AOT40 (May-July) for the control period (1991-2000) between the model runs with 10 km and 50 km spatial resolution. Right: Difference of the decadal climate response (difference of 2091-2090 to 1991 to 2000) in AOT40 (May-July) between the model runs with 10 km and 50 km spatial resolution.

The response to climate change differs much less between the model runs with different resolution. In Fig. 10 the difference in the climate response (difference of 2091-2090 to 1991 to 2000 in AOT40, May to July) between the runs is displayed. In most of the domain the values are between +1000 ppb h and –1000 ppb h with negative values over Romania, Bosnia-Herzegovina and western Croatia. Only at the western border of the domain the differences are higher, which might be caused partly by boundary effects of the high-resolution model.

It has been investigated, how the legal limit of 9000 ppb h for the AOT40-value in the growing season May to July is exceeded in the model calculations. AOT40 is the sum of the accumulated hourly ozone values above a threshold of 40 ppbv for the indicated months and the time between 8:00 and 20:00 hours. The data for the control period (1991-2000) and for the end-century decade (2091-2100) are shown in Fig. 11 from the BOKU-50 km-runs. A gradient with lower values in the north and higher values in the south is observed. The highest values are found over the Mediterranean Sea. With present day conditions (1991-2000) the value of 9000 ppb h is not exceeded north of the Alps and the Black Sea. In the calculation with future climate (2091-2100) the AOT40-values become higher and the limiting line at 9000 ppb h is shifted to the north. Fig. 12 displays the AOT40-values for the control period (1991-2000) and for the end-century decade (2091-2100) from the BOKU-10 km-runs. The values are generally higher than in the coarse resolution run. With present day conditions (1991-2000) only over Poland and Ukraine the limit of 9000 ppb h is not exceeded. The highest vales are found over the Adriatic Sea and over Italy. With future climate (2091-2100) a further increase is observed.

The differences between the end-century- (2091-2100) and the control-decade (1991-2000) in the AOT40 values for May to July are shown in Fig. 13 for the 50 km and the 10 km run. With the coarse resolution an increase over most of Europe is observed. Only over Africa, the north of the Atlantic Ocean and northern Russia there is a decrease. The highest increase occurs in northern Italy and over the Iberian Peninsula. In general, the calculated increase with climate change is similar in the 10 km resolution run. The high increase over southern Germany may be attributed to boundary effects of the model. In the target region of CECILIA the increase is strongest in a belt stretching from Austria over western Hungary and Serbia to Bulgaria.

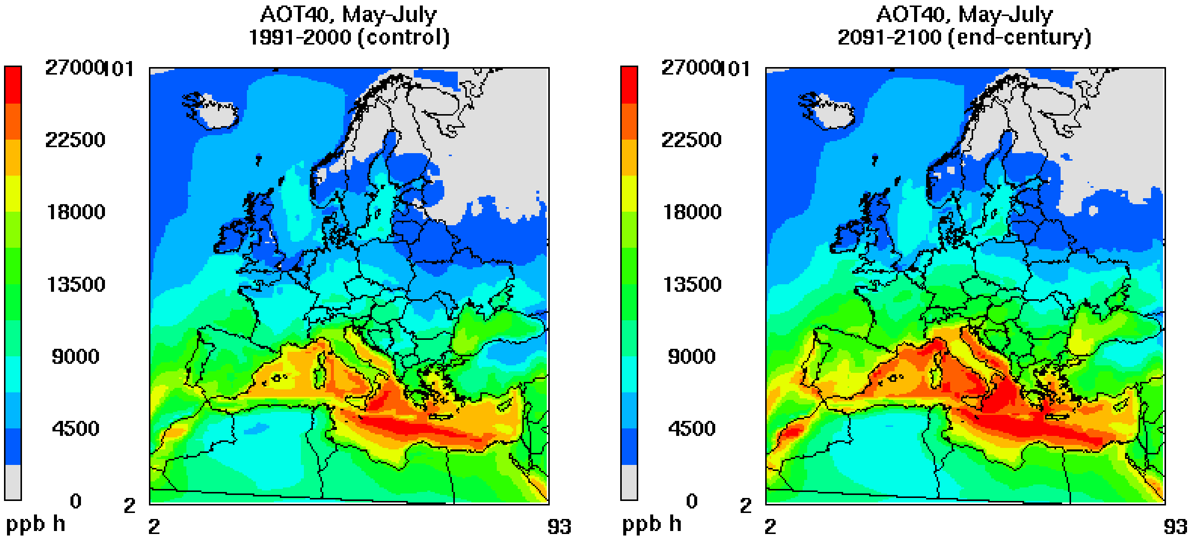

Figure 11. Average of the AOT40 for May to July for the control decade 1991-2000 (left) and the end-century decade 2091-2100 (right) in the calculation with 50 km spatial resolution.

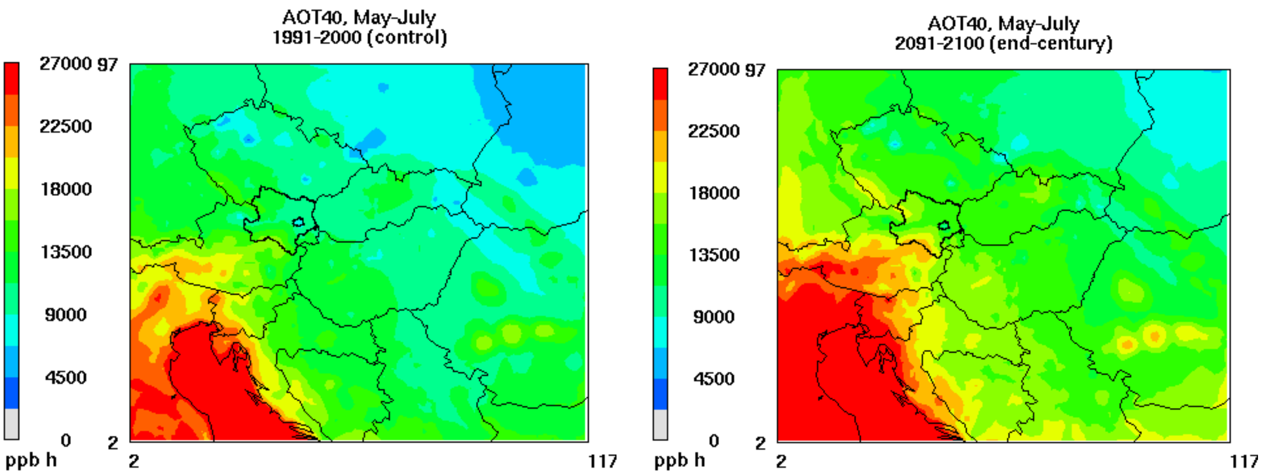

Figure 12. Average of the AOT40 for May to July for the control decade 1991-2000 (left) and the end-century decade 2091-2100 (right) in the calculation with 10 km spatial resolution.

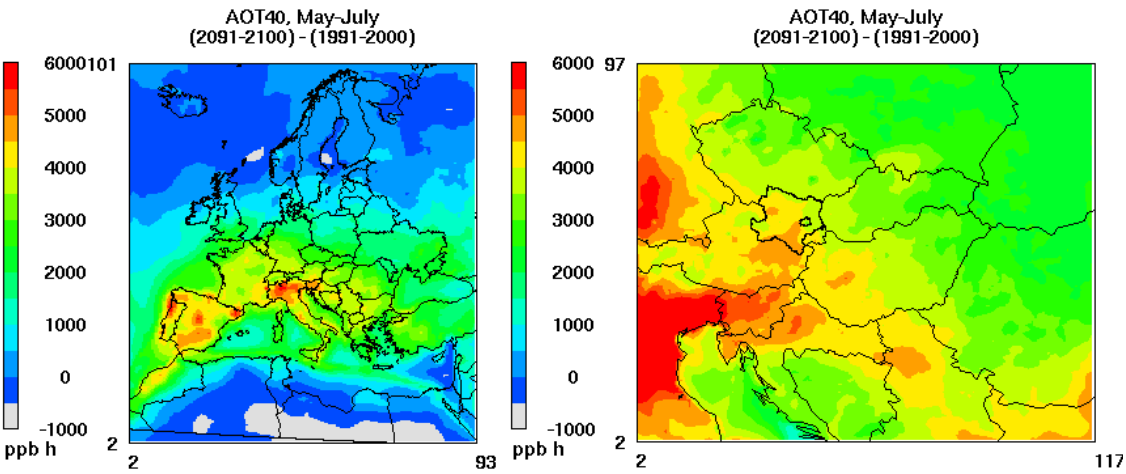

Figure 13. Difference in AOT40 (May-July) between the end-century run (2091-2100) and the control run (1991-2000) for the runs with 50 km (left) and 10 km (right) spatial resolution

References

Borrego C., Monteiro A., Ferreira J., Miranda A. I., Costa A. M., Carvalho A.C, Lopes M., 2008: Procedures for estimation of modelling uncertainty in air quality assessment, *Environ. Intern.*, **34**, 613–620.

Brandt J., Christensen J.H., Frohn L.M., Palmgren F., Berkowicz R., Zlatev Z., 2001: Operational air pollution forecasts from European to local scale, *Atmos. Environ.*, **35** (Suppl. 1), S91–S98.

Chang J. C., Hanna S.R., 2004: Air quality model performance evaluation, *Meteorology & Atmos. Physics*, **87**, 167-196.

Elguindi, N., X. Bi, F. Giorgi, B. Nagarajan, J. Pal, F. Solmon, S. Rauscher, A. Zakey, 2006: RegCM Version 3.1 User's Guide. PWCG Abdus Salam ICTP.

ENVIRON Corp., 2006: CAMx Users' Guide, version 4.40

Flemming J, Stern R., 2007: Testing model accuracy measures according to the EU directives—examples using the chemical transport model REM-CALGRID, *Atmos. Environ.*, **41**, 9206–16.

Gery, M.W., G.Z. Whitten, J.P. Killus, and M.C. Dodge. 1989: A Photochemical Kinetics Mechanism for Urban and Regional Scale Computer Modeling. J. Geophys. Res., 94, 925-956.

Giorgi, F., X. Bi, Y. Qian, 2002: Direct radiative forcing and regional climatic effects of anthropogenic aerosols over East Asia: A regional coupled climate-chemistry/aerosol model study. .J. Geophys. Res., 107, 4439, doi:10.1029/2001JD001066.

Giorgi, F., Y. Huang, K. Nishizawa and C. Fu, 1999: A seasonal cycle simulation over eastern Asia and its sensitivity to radiative transfer and surface processes. Journal of Geophysical Research, 104, 6403-6423.

Guenther, A.B., Zimmerman, P.R., Harley, P.C., Monson, R.K., and Fall, R., 1993: Isoprene and monoterpene rate variability: model evaluations and sensitivity analyses, J. Geophys. Res., 98, No. D7, 12609-12617.

Guenther, A. , Zimmerman, P., and Wildermuth, M., 1994: Natural volatile organic compound emission rate estimates for U.S. woodland landscapes, Atmospheric Environment, 28, 1197-1210.

Hanna S.R., Chang J.C., Strimaitis D.G., 1993: Hazardous gas model evaluation with field observations. *Atmos. Environ.,* **27A**, 2265–2285.

Juda K., 1986: Modelling of the air pollution in the Cracow area. *Atmos. Environ.*, **20**, 2449-2558.

Juda-Rezler K., 1989: Air Pollution Modelling, [In:] The Encyclopedia of Environment Control Technology, Gulf Publishing Company, Houston, 83-134. 4.

Juda-Rezler K., 2010: New challenges in air quality and climate modelling. *Archives of Environ. Prot.*, **36 (1)**, 3-28.

Katragkou, E., P. Zanis, I. Tegoulias, D. Melas, 2009: Tropospheric ozone in regional climate-air quality simulations over Europe: Future climate and sensitivity analysis. Proceedings 30th NATO/SPS International Technical Meeting on Air Pollution Modelling and its Application.

Krüger B.C., Katragkou E., Tegoulias I., Zanis P., Melas D., Coppola E., Rauscher S., Huszar P., Halenka T., 2008: Regional photochemical model calculations for Europe concerning ozone levels in a changing climate. *Q. J. Hung.Meteor. Serv.,* **112**, No. 3–4, 285–300.

Pal, J. S., F. Giorgi, X. Bi, N. Elguindi, F. Solmon, X. Gao, S. A. Rauscher, R. Francisco, A. Zakey, J. Winter, M. Ashfaq, F. S. Syed, J. L. Bell, N. S. Diffenbaugh, J. Karmacharya, A. Konaré, D. Martinez, R. P. da Rocha, L. C. Sloan, and A. L. Steiner, 2007: Regional Climate Modeling for the Developing World: The ICTP RegCM3 and RegCNET. Bull. Amer. Meteorol. Soc., 88, 9, 1395–1409.

Qian, Y., F. Giorgi, 2000: Regional climatic effects of anthropogenic aerosols? The case of Southwestern China, Geophys. Res. Lett., 27(21), 3521-3524, 10.1029/2000GL011942.

Qian, Y., F. Giorgi, Y. Huang, W.L. Chameides, and C. Luo, 2001: Simulation of anthropogenic sulfur over East Asia with a regional coupled chemistry/climate model. Tellus, Ser. B, 53, 171-191.

Vestreng V., Breivik K., Adams M., Wagener A., Goodwin J., Rozovskkaya O., Pacyna J.M., 2005: Inventory Review 2005, Emission Data reported to LRTAP Convention and NEC Directive, Initial review of HMs and POPs. Technical Report MSC-W 1/2005, ISSN 0804- 2446.

Willmott C. J., 1982: Some comments on the evaluation of model performance. Bull. Amer. *Meteor. Soc.*, **63**, 1309-1313.

Winiwarter W., Zueger, J., 1996: Pannonisches Ozonprojekt, Teilprojekt Emissionen. Endbericht. Report OEFZS-A-3817, Austrian Research Center, Seibersdorf.

D7.4 – Appendix
Guidelines for operational evaluation of the AQ-CTMs under WP7 of the CECILIA project

Katarzyna Juda-Rezler

version 1.2

1. Introduction

The paper addresses Air Quality – Chemical Transport Models (AQ – CTMs) evaluation in terms of *operational evaluation*, which is aiming on comparing model results with measurements of species concentrations for a specific time period. Several scientists carried out discussion on the operational evaluation methods and criteria (as e.g.; Willmot 1982; Juda, 1986; Chang & Hanna, 2004; Borrego et al., 2008), however, standard evaluation procedures and performance standards *still do not exist*. Herein we are proposing guidelines for operational evaluation of AQ – CTMs to be used by all partners of the CECILIA project WP7 group. Modelling Quality Objectives, defined (unfortunately very imprecisely...) in the CAFE Directive (2008/50/EC), are interpreted and taken into account as well.

2. Package of statistical indices

The number of statistical metrics that have been developed for use in operational AQM evaluation continues to expand and can be overwhelming. Depending on the measure of interest, bias for example, numerous variations exist (i.e. Mean Bias (MB), Normalized Mean Bias (NMB), Geometric Mean Bias (MG), Fractional Bias (FB), etc.) each with their own advantages and disadvantages.

For CECILIA evaluation, **ten standard and well-accepted measures** of model performance were selected. The first two measures selected are measures of model bias: the **MB** and **NMB**. Note to **calculate bias as observed (C_o) minus predicted (C_p) values** - a negative value of such bias indicates over-prediction of observations. Likewise, two accepted measures of model error: the Root Mean Square Error (**RMSE**) with its systematic (**$RMSE_s$**) and unsystematic (**$RMSE_u$**) part (see e.g. Juda, 1986; Appel et al., 2007) and Normalized Mean Square Error (**NMSE**), were selected. Next, we have two accepted measures of model correlation: correlation coefficient (**r**) and index of agreement (**IA**). IA, was proposed by Willmott (1982) as a more robust than (r) measure of agreement between predictions and observations. Then, one measure of model variance was selected – it is Explained Variance (**EXV**) proposed by Juda-Rezler (1986, 1989). EXV is a measure of how much of the observed variance is explained by the model. For a good model, the EXV should be greater than 30%. Finally, we have **FAC2** – a fraction of predictions within a factor 2 of observations.

These metrics, which provide both actual (i.e. measured in either ppb or $\mu g/m^3$) and normalized (%) measures of performance, are given below:

$$MB = \frac{1}{N} \cdot \sum_{i=1}^{N} \left(C_{oi} - C_{pi} \right) \tag{1}$$

$$NMB = \frac{\sum_{i=1}^{N} \left(C_{oi} - C_{pi} \right)}{\sum_{i=1}^{N} \left(C_{oi} \right)} \cdot 100\,\% \tag{2}$$

$$NMSE = \frac{\frac{1}{N} \cdot \sum_{i=1}^{N} \left(C_{oi} - C_{pi} \right)^2}{\overline{C}_o \cdot \overline{C}_p} \cdot 100\% \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(C_{oi} - C_{pi} \right)^2} \tag{4}$$

$$RMSE_s = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(C_{oi} - \hat{C}_{pi} \right)^2} \tag{5}$$

$$RMSE_u = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(C_{pi} - \hat{C}_{pi} \right)^2} \tag{6}$$

where:

$$\hat{C}_{pi} = a + b \cdot C_{oi}$$

$$a = \overline{C}_p - b \cdot \overline{C}_o$$

$$b = \frac{\sum_{i=1}^{N} C_{oi} \cdot C_{pi} - \frac{1}{N} \cdot \sum_{i=1}^{N} C_{oi} \cdot \sum_{i=1}^{N} C_{pi}}{\sum_{i=1}^{N} C_{oi}^2 - \frac{1}{N} \cdot \left(\sum_{i=1}^{N} C_{oi} \right)^2}$$

$$r = \frac{\sum_{i=1}^{N} \left(C_{oi} - \overline{C}_o \right) \cdot \left(C_{pi} - \overline{C}_p \right)}{\sqrt{\sum_{i=1}^{N} \left(C_{oi} - \overline{C}_o \right)^2 \cdot \sum_{i=1}^{N} \left(C_{pi} - \overline{C}_p \right)^2}} \tag{7}$$

$$IA = 1 - \frac{\sum_{i=1}^{N} \left(C_{oi} - C_{pi} \right)^2}{\sum_{i=1}^{N} \left(\left| C_{pi} - \overline{C}_o \right| + \left| C_{oi} - \overline{C}_o \right| \right)^2} \tag{8}$$

$$EXV = \frac{\left[\sum_{i=1}^{N} \left(C_{oi} - \overline{C}_o \right)^2 - \sum_{i=1}^{N} \left(C_{oi} - C_{pi} \right)^2 \right]}{\left[\sum_{i=1}^{N} \left(C_{oi} - \overline{C}_o \right)^2 \right]} \cdot 100\,\% \tag{9}$$

$FAC2$ = fraction (%) of the data for which

$$0.5 \le \frac{C_p}{C_o} \le 2 \qquad (10)$$

$$\overline{C}_p = \frac{1}{N} \cdot \sum_{i=1}^{N} C_{pi}$$

$$\overline{C}_o = \frac{1}{N} \cdot \sum_{i=1}^{N} C_{oi}$$

C_{pi} and C_{oi} are values of model **p**rediction and **o**bservation, respectively, at time and location ***i*** ; **N** is the total number of samples (by time and/or location).

As seen in Eqs. (2) and (3), the normalization is achieved by dividing by the sum of observed/predicted concentrations (as opposed to dividing by individual observations). This is avoiding the inflation that other metrics (i.e. Mean Normalized Bias (MNB) or Mean Normalized Error (MNE)) are susceptible to when applied to small concentrations (Eder & Yu, 2006).

3. Requirements of the latest EU Directive

The CAFE Directive (2008/50/EC) establishes requirements for air quality modelling, however – unfortunately – it gives only very brief definition of required modelling quality objectives (**MQO**). The uncertainty is defined as "the maximum deviation of the measured and calculated concentration levels for 90% of individual monitoring points over the period considered by the limit value (or target value in the case of ozone), *without taking into account the timing of the events*". According to the Directive the uncertainty measures defining the MQO are stipulated for those species and averaging intervals for which AQ limit/target values are given. For species considered under CECILIA WP7, **averaging times and thresholds** are given in **Table 1**. The MQO set by the CAFE is an "maximum deviation" < 30% for annual averages of SO_2 and NO_2, and < 50% for annual averages of PM_{10}, hourly and daily SO_2 averages, and 8-hourly O_3 averages. An MQO is not yet defined for daily PM_{10} averages (see **Table 2**).

The question is how to interpret this "maximum deviation" expressed in [%] ? Fleming & Stern (2007) interpreted it as the "relative maximum error" **RMaxE** which can be written by:

$$RMaxE = \frac{\max\limits_{per=1\ldots P}\left(\left|C_{o\ per} - C_{p\ per}\right|\right)}{C_{o\ per=per(\max)}} \qquad (11)$$

where *per*=1...P is percentile.

For annual means the question about timing is not relevant. The EU accuracy measure can be interpreted simply as the relative error of the annual averages (**RE**).

$$RE = \frac{\left|\overline{C}_o - \overline{C}_p\right|}{\overline{C}_o} \qquad (12)$$

For short term averages, the Eq. 11 will cause over-estimation of the model uncertainty, because at many stations the absolute maximum deviation is found at the highest percentiles, which could be an observed or calculated outlier!

Because of the disadvantages of **RMaxE** (Flemming & Stern, 2007) proposed the alternative and more robust accuracy measure – relative percentile error: **RP(LV)E**, which is based on the relative error at the percentile corresponding to the allowed number of exceedances of the AQ limit/target value (see **Table 1**). This measure also evaluates the model performance in the high concentration ranges but without being so sensitive to outliers:

$$RP(LV)E = \frac{\left| C_{o\,per} - C_{p\,per} \right|}{C_{o\,per}}, \quad per = per_{LV} \qquad (13)$$

Where per_{LV} is percentile that corresponds to the allowed number of exceedances of the AQ limit value.

4. Stations choice

During validation process of the CTMs, observed values of point measurements at a station, are compared against predicted values averaged for the grid cell area (in our case 50 x 50 km or 10 x 10 km). Therefore, in general only so-called **rural background** stations should be use for validation purposes. The EU CAFE Directive (2008/50/EC) states only that “the fixed measurements that have to be selected for comparison with modelling results *shall be representative* of the scale covered by the model”, without any recommendations concerning “representativeness”.

We can concretize here that stations chosen for model evaluation should be representative of:

- the grid area climatic conditions (i.e. stations situated in specific conditions, for example sites of a high elevation, or influenced by coastal circulation should be excluded)
- for average air quality within the grid area (i.e. station should not be influenced by local sources).

In addition, the usual requirement of temporal data completeness (75%, for ozone 90%) should be met as well as the requirement of a statistically sufficient number of stations, covering the entire area of interest.

However, in our case (year 2000) the number of rural background station with data completeness is relatively low, specially for PM_{10}. That is why, I’m proposing to perform our evaluation for 2 station selections and then we will decide on the final choice:

1) **SELECTION 1**: All available for the given domain rural background stations with adequate temporal data completeness (75%, for ozone 90%) from EMEP and AirBase databases. **Attention**: the station selection should be made individually for each of the pollutant under concern – O_3, PM_{10}, SO_2, NO_2!
2) **SELECTION 2**: As SELECTION 1, but with **exclusion** of not-representative stations i.e. mountain stations located above 1000 m a.s.l., stations influenced by coastal circulation, etc.

Step by step procedure of the AQ – CTM evaluation under CECILIA WP7

1. We are evaluating operational model performance for 4 species: O_3, PM_{10}, SO_2 and NO_2 for the reference year 2000 (see **Table 3** for details).
2. Each partner is preparing for its domain SELECTION 1 and then SELECTION 2 of stations for each pollutant.
3. Each partner is conducting its CTM evaluation **for annual/seasonal means** (O_3, PM_{10}, SO_2, NO_2) – see **Table 3** for details:
    a. **Qualitative analysis** – Model Performance Plots for predicted & observed levels as well as differences:
        i. Scatter plots, with CTM simulation results (ordinate) versus observations (abscissa) for each species, with factor of two reference lines. A distinction should be made between EMEP and AirBase stations. See **Fig. 1** as an exemplary scatter plot.
        ii. Time series plots (24h values) for selected stations.
        iii. Other graphical analysis – to be chosen by Partners (Box plots, Bar plots, "Soccer goal" plots, Bugle Plots).
    b. **Quantitative analysis**:
        i. Evaluation for each site i, i=1,...,NS1 or NS2 (number of sites in selection 1 or 2) – metrics 1÷10.
        ii. Overall evaluation over SPACE for NS1/NS2 stations – metrics 1÷10
        iii. For each station – metric (12) and % of stations fulfilling adequate AQO EU objective (see **Table 2**)
4. Model evaluation for **short-term averages** (O_3, PM_{10}, SO_2, NO_2) – see **Table 3** for details:
    a. **Qualitative analysis** – Model Performance Plots for predicted & observed levels as well as differences:
        i. Scatter plots, with CTM simulation results (ordinate) versus observations (abscissa) for each species, with factor of two reference lines. A distinction should be made between EMEP and AirBase stations.
        ii. Diurnal box plots averaged for all sites – see as an example **Fig. 2**.
        iii. Other graphical analysis – to be chosen by Partners (Bar plots, "Soccer goal" plots, Bugle Plots).
    b. **Quantitative analysis**:
        i. Evaluation for each site i, i=1,...,NS1 or NS2 – metrics 1÷10.

ii. For each station – metric (13) and % of stations fulfilling AQO EU objective (see **Table 2**)

iii. Overall evaluation over SPACE for all NS1/NS2 stations for N – the total, for given time period, number of model/observation pairs – metrics 1÷10.

5. Each Partner is preparing **evaluation tables** (the forms are given as **Tables 4÷7**) for each specie - with some basic statistics and & EU AQO metrics.
6. Each Partner is fulfilling **summary table** for SELECTION 1 and SELECTION 2 for its domain – the form of that table is given as **Table 8**.
7. Partners who are involved in **Taylor plots** – creates them as well.

REFERENCES

1. Appel K. W., Gilliland A. B., Sarwar G., R. C. Gilliam: Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model performance Part I—Ozone, *Atmos. Environ.,* 41, 9603–9615 (2007).
2. Borrego C., Monteiro A., Ferreira J., Miranda A. I., Costa A. M., Carvalho A.C, M. Lopes: Procedures for estimation of modelling uncertainty in air quality assessment, *Environ. Intern.,* **34**, 613–620 (2008).
3. Chang J. C., S.R. Hanna: Air quality model performance evaluation, *Meteorology & Atmos. Physics*, 87, 167-196 (2004).
4. Juda, K.: Modelling of the air pollution in the Cracow area. *Atmos. Environ*., **20**, 2449-2558 (1986).
5. Juda-Rezler K.: *Air Pollution Modelling*, [In:] The Encyclopedia of Environment Control Technology, Gulf Publishing Company, Houston 1989, 83-134.
6. Willmott C. J.: Some comments on the evaluation of model performance. Bull. Amer. Meteor. Soc., **63**, 1309-1313 (1982).
7. Flemming J, R. Stern: Testing model accuracy measures according to the EU directives—examples using the chemical transport model REM-CALGRID, Atmos. Environ., **41**, 9206–16 (2007).

TABLES

Table 1. EU air quality thresholds for different averaging intervals

Averaging interval	**PM_{10}**		**O_3**		**SO_2**		**NO_2**	
	Treshold ($\mu g/m^3$)	**Percentile**	**Treshold ($\mu g/m^3$)**	**Percentile**	**Treshold ($\mu g/m^3$)**	**Percentile**	**Treshold ($\mu g/m^3$)**	**Percentile**
Hourly	-	-	-	-	350	99.73	200	99.79
Daily	50	90.41	120 (8h max)	93.15	125	99.18	-	-
Annual	40	-	-	-	20	-	40	-
Winter (1.10-31.03)	-	-	-	-	20	-	-	-

Left: threshold value in $\mu g/m^3$, right: percentile value corresponding to the number of allowed exceedances of the AQ threshold.

For NO_2, PM_{10}, SO_2 threshold = limit value (LV). For O_3 threshold = target value (TV).

Table 2. Modelling Quality Objectives (MQO) established by the EU CAFE Directive (2008/50/EC).

Pollutant	**Quality indicator**	**MQO**	**Error to be used**
SO_2 & NO_2	Annual mean	30%	RE
	Daily mean	50%	RP(LV)E
	Hourly mean	50%	RP(LV)E
PM_{10}	Annual mean	50%	RE
PM_{10}	Daily mean	Not set	RP(LV)E
Ozone	8-hour daily mean	50%	RP(LV)E*

* Note that for ozone we are using Target Value (TV) instead of LV

Table 3. Criteria for model evaluation under CECILIA WP7.

SPECIES	**METRICS**	**NETWORKS**
PM_{10}	Annual, Daily	EMEP, AIRBASE
O_3	AOT40, Summer (JJA), Max 8h	EMEP, AIRBASE
SO_2	Annual, Winter, Daily, Hourly	EMEP, AIRBASE
NO_2	Annual, Hourly,	EMEP, AIRBASE

O3: Time period of AOT40 definition is 08:00 to 20:00 CET from 1 May to 31 July each year, for vegetation protection and from 1 April to 30 September each year for forest protection.

SO2: winter – from 1 October to 31 December and from 1 January to 31 March each year.

Table 4. PM_{10} – overall (for whole domain) performance with EU AQO.

Metric	Daily mean		Annual mean		Metric	Daily mean	Annual mean
	MODEL	OBS	MODEL	OBS			
Mean					**NS**		
Median					**N**		**NS**
SD					**RP(90.41)E range** (for selected sites)		
Max					**RP(90.41)E overall**		
90.41th					**% of sites with RP(LV)E< 50%**		
Min					**RE range** (for selected sites)		
					RE overall		
					% of sites with RE< 50%		

Table 5. O_3 – overall (for whole domain) performance with EU AQO.

	AOT40		Summer mean (JJA)		Max 8h mean		Metric	Max 8h mean	Summer mean
	MODEL	OBS	MODEL	OBS	MODEL	OBS			
Mean							**NS**		
Median							**N**		**NS**
SD							**RP(93.15)E range** (for selected sites)		
Max							**RP(93.15)E overall**		
93.15th							**% of sites with RP(LV)E< 50%**		
Min									

Table 6. SO_2 – overall (for whole domain) performance with EU AQO.

	Annual mean		**Winter mean**		**Daily mean**		**Hourly mean**		**Metric**	**Annual mean**	**Winter mean**	**Daily mean**	**Hourly mean**
	MODEL	**OBS**	**MODEL**	**OBS**	**MODEL**	**OBS**	**MODEL**	**OBS**					
Mean									**NS**				
Median									**N**	**NS**			
SD									**RP(99.18)E range** (for selected sites)				
Max									**RP(99.18)E overall**				
99.18^{th}									**% of sites with RP(LV)E< 50%**				
99.73^{th}									**RP(99.73)E range** (for selected sites)				
Min									**RP(99.73)E overall**				
									% of sites with RP(LV)E< 50%				
									RE range (for selected sites)				
									RE overall				
									% of sites with RE< 30%				

99.18^{th} is for hourly mean of SO_2
99.73^{th} is for daily mean of SO_2

Table 7. NO_2 – overall (for whole domain) performance with EU AQO.

Metric	**Annual mean**		**Hourly mean**		**Metric**	**Hourly mean**	**Annual mean**
	MODEL	**OBS**	**MODEL**	**OBS**			
Mean					**NS**		
Median					**N**		NS
SD					**RP(99.79)E range**		
Max					**RP(99.79)E overall**		
99.79th					**% of of sites with RE< 30%NS < 30%**		
Min					**RE range** (for selected sites)		
					RE overall		
					% of sites with RE< 30%		

Table 8. Summary performance evaluation of CTM for SELECTION 1 at AAAAA domain (here some Polish data of PM & SO2 evaluation for WUT domain as example).

	NS1	N	r		IA		MB [µg/m^3]		NMB [%]		RMSE [µg/m^3]		$RMSE_s$ [µg/m^3]		$RMSE_u$ [µg/m^3]		NMSE [%]		EXV [%]		FAC2 [%]	
			Site Range	Overall	Site Range	Overall	Site Range	Overall	Site Range	Overall	Site Range	Overall	Site Range	Overall	Site Range	Overall	Site Range	Overall	Site Range	Overall	Site Range	Overall
PM_{10} Annual	**33**	**33**		0.53		0.66		7.65		37.12		9.78						36		-70		81.8
PM_{10} Daily																						
O_3 AOT40																						
O_3 Summer																						
O_3 Max 8h																						
SO_2 Annual	**9**	**9**		0.39		0.58		-0.81		-33.94		1.41						26		-60		77.8
SO_2 Winter																						
SO_2 Daily																						
SO_2 Hourly																						
NO_2 Annual																						
NO_2 Hourly																						

NOTE:
NS – the total number of rural background monitoring station (NS1 – for SELECTION 1)
N – the total number of model/observation pairs (for the whole domain)
Summer – 1.06-31.08
Winter – 1.10-31.03

FIGURES

Figure 1. An example of scatter plot for PM performance under CECILIA (WUT domain).

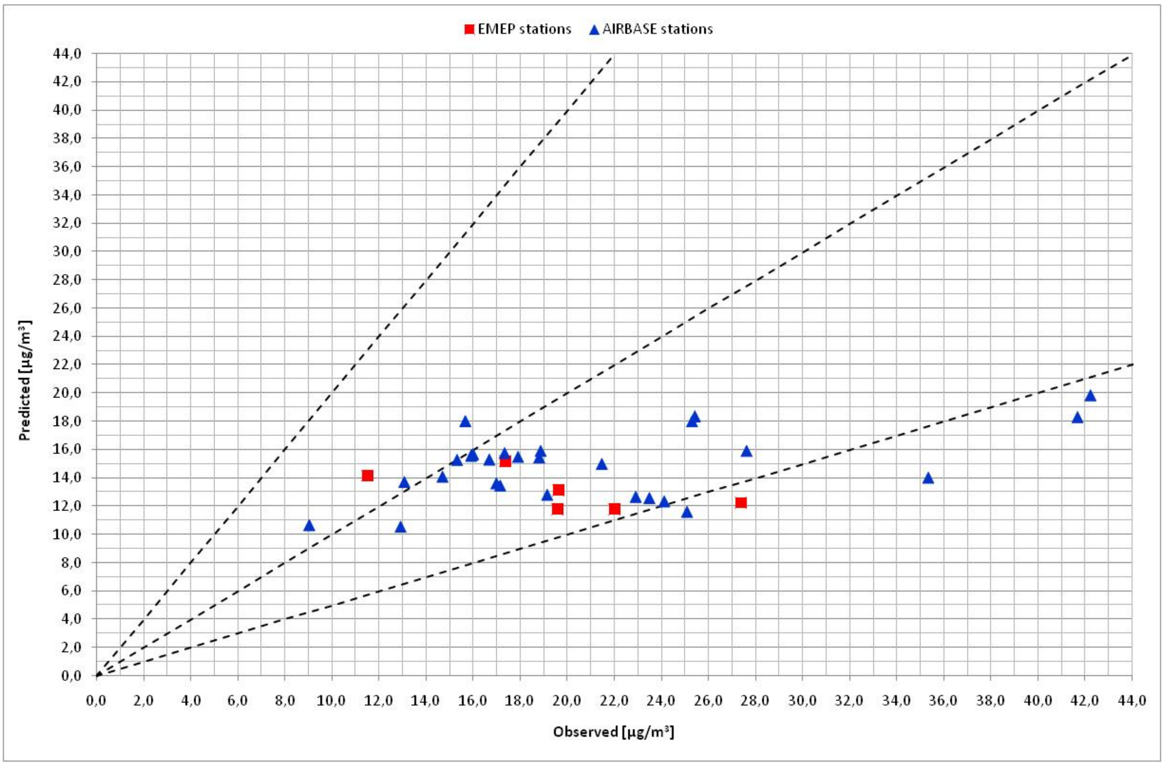

Figure 2. An example of diurnal box plot (Appel et al., 2007).