# Air quality modeling using the CMAQ model.

Dušan Štefánik[1,a)]

[1] *Slovak Hydrometeorological Institute, Jeséniova 17, 833 15 Bratislava, Slovakia*

[a)] dus.stefanik@gmail.com

**Abstract.** The time period of January 2017 with smog episodes was simulated by the recent version of Community Multiscale Air Quality modelling system using real meteorological data and constant emissions. The results show the important role of meteorology in the air pollution problem. Concentrations of $NO_2$, $O_3$ and $PM_{10}$ from model output were compared to the measurements at rural EMEP background stations.

**Keywords:** $NO_2$, $O_3$, $PM_{10}$, air quality modeling, concentrations, model comparison, background stations

## INTRODUCTION TO THE CMAQ MODEL

The Community Multiscale Air Quality (CMAQ) modeling system has been developed and maintained under the leadership of the EPA National Exposure Research Laboratory in Research Triangle Park, NC [1]. CMAQ is a third-generation air quality model which means that it could treat multiple pollutants simultaneously and at scales up to continental or larger. It is a three-dimensional Eulerian (i.e., gridded) atmospheric chemistry and transport modeling system that simulates, among others, ozone, particulate matter (PM), toxic airborne pollutants, visibility, and acidic and nutrient pollutant species throughout the troposphere. Mathematically, CMAQ computes the concentration change in each grid cell over time $\frac{\partial C}{\partial t}$ through the continuity equation, which is presented in simplified form below

$$\frac{\partial C}{\partial t} = \mathrm{ADV} + \mathrm{DIFF} + \mathrm{R}_i + \mathrm{E}_i - \mathrm{S}_i\ , \tag{1}$$

where ADV stands for the horizontal and vertical advection, DIFF represents the horizontal and vertical diffusion, $\mathrm{R}_i$ chemical transformation of species $i$, $\mathrm{E}_i$ emission of species $i$, and $\mathrm{S}_i$ = loss processes (deposition) for species $i$.

In CMAQ, the advection and emissions terms are calculated based on input files generated by the meteorology and emissions models, respectively. The diffusion, chemical transformation, and loss process terms are calculated within CCTM - which is the is the final program to be run in the CMAQ modeling sequence. There are three other main programs that prepare input data for CCTM (i.e., ICON, BCON, and MCIP). ICON generates a gridded binary netCDF file of the chemical conditions in the modeling domain for the first hour of a simulation. BCON generates a gridded binary netCDF file of the chemical conditions along the horizontal boundaries of the modeling domain. MCIP is the meteorological preprocessor which prepares meteorological input fields for the CMAQ. For more information about the model see CMAQv5.2 Operational Guidance Document [2].

## PRELIMINARY RESULTS OF MODELING FOR THE JANUARY 2017

### Grid definition and meteorological inputs

Experimental simulation was performed on the computational model domain with 169 x134 grid cells, spatial resolution of 14 km, lambert conformal conic projection - tangent case using offline hourly meteorological fields, generated by the Weather Research Forecasting (WRF) model version 3.9.1 using GFS meteorological reanalysis data (spatial resolution $0.25° \times 0.25°$ ) as an input. The parameterization of WRF included the following schemes: MilbrandtYau double-moment 7-class microphysics [3, 4], rapid radiative transfer model (RRTMG) longwave and shortwave scheme

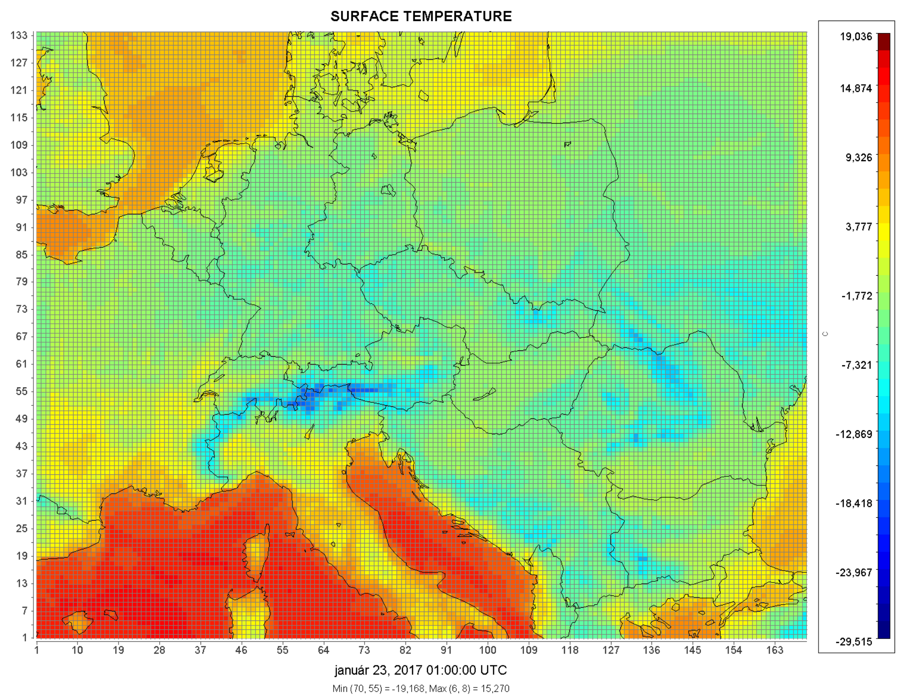


**FIGURE 1.** Example of the WRF output on the studied domain for the surface temperature at January 23, 2017 01:00:00 UTC.

[5], asymmetric convective model of the planetary boundary layer (PBL) [6], PleimXiu surface layer model [7], and the improved version of GrellDevenyi [8, 9] ensemble scheme for cumulus parameterization. Vertically, the domain was resolved in 32 layers, following the eta coordinate system.

## MCIP - meteorological preprocessor for the CMAQ and modification of the code

CMAQ ready meteorological input fields were prepared using the Meteorology-Chemistry Interface Processor (MCIP) [10]. Since we used the lambert conformal conic projection - tangent case for our simulation which is forbidden in recent MCIPv4.4 release we needed to slightly change the code. The next changes have been done:

- In line 73 of the file ll2xy_lam.f90, we allowed MCIP to take the Lambert Conformal Conic projection -Tangent Case by comment and comment out the following lines, respectively:

```
! CALL graceful_stop (pname)
  CALL ll2xy_lam_tan (phi, lambda, phi1, lambda0, xx, yy)
```

- In line 95 of the file xy2ll_lam.f90, we added the conditional for the case when the two standard parallel are equal (Tangent case). The equations were taken from [11]

```
IF ( phi1rad .EQ. phi2rad ) THEN
    sinphi0 = DSIN ( phi1rad )
ELSE
    sinphi0 = DLOG ( DCOS(phi1rad) / DCOS(phi2rad) )
    sinphi0 = sinphi0 / DLOG (term2 / term1)
ENDIF
```

The xy2ll_lam.f90 subroutine converts coordinates x y to coordinates LAT LON in Lambert Conformal Tangent Conic projection, but the original version of the subroutine returns NAN when two standard parallels were equal.

- The mapfac_lam.f90 file which represents subroutine for the calculation of the map scale factor was modified in the same way as xy2ll_lam.f90 subroutine.

## CMAQ configurations

CMAQ modelling system version 5.2 [12] with the CB06r3 gas-phase chemistry mechanism [13] and the AERO6 version of the aerosol module was used for the experimental simulation. CMAQ contains modules representing advection, eddy diffusion and in-cloud and below-cloud scavenging with precipitation, includes removal by dry deposition, in-cloud and precipitation processes and simulates aqueous chemistry and wet deposition by cloud droplets [14, 15]. The aerosol module is based on aerosol model ISORROPIA II [16] and determines concentrations of trimodal size-distributed particulate material with diameters less than 10 $\mu$m. It contains an inline windblown dust module.

## Emissions input

Emission input data, which was generated by emission model SMOKE EU [17] for the purpose of this experimental simulation, contains SNAP1 power generation sources (including the small heavy fuel oil combustion facilities) for a single day, which was used for the whole period of the simulation (31 days). The aim of the experimental simulation was to show the influence of meteorological conditions to time series and spatial distribution of gaseous pollutant concentrations and $PM_{10}$.

## Results and discusions

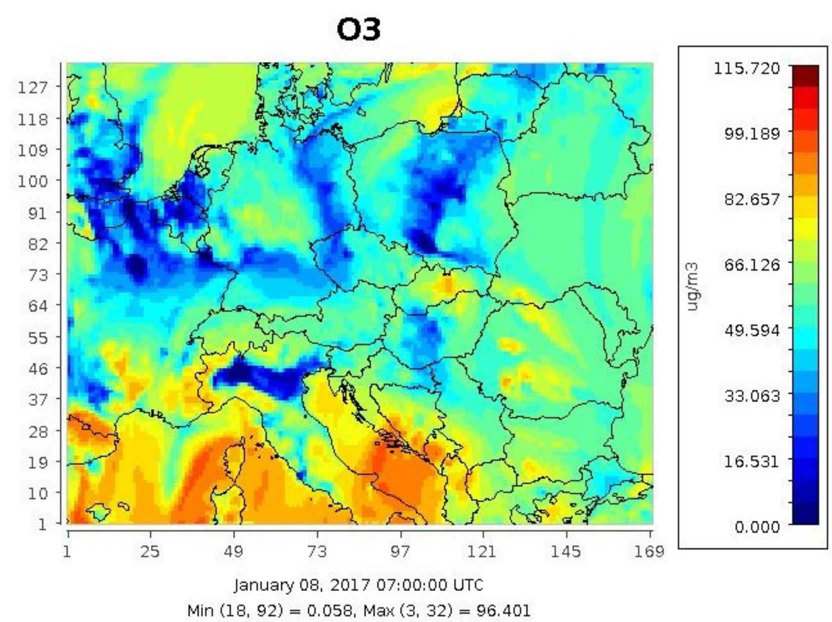


**FIGURE 2.** Example of the CMAQ CCTM output file for the avaraged concentration of $O_3$ near surface at January 08, 2017 07:00:00 UTC.

As the input data contained only SNAP 1 emission sources and the low grid spatial resolution (14 × 14 km) is used, the output is compared to the rural background EMEP air quality stations, situated far from the local air pollution sources.

There are 4 EMEP background stations in Slovakia with aviable measurement data during the considered time interval of January 2017:

- Topoľníky

- Chopok
- Stará Lesná
- Starina

**Results for the $O_3$ concentration**

Figure 3 shows plotted $O_3$ hourly and daily concentration series for the modeled and observed values at Topoľníky station. We can see that the model overestimates the surface ozone concentrations, but it more less reproduces the observed trends.

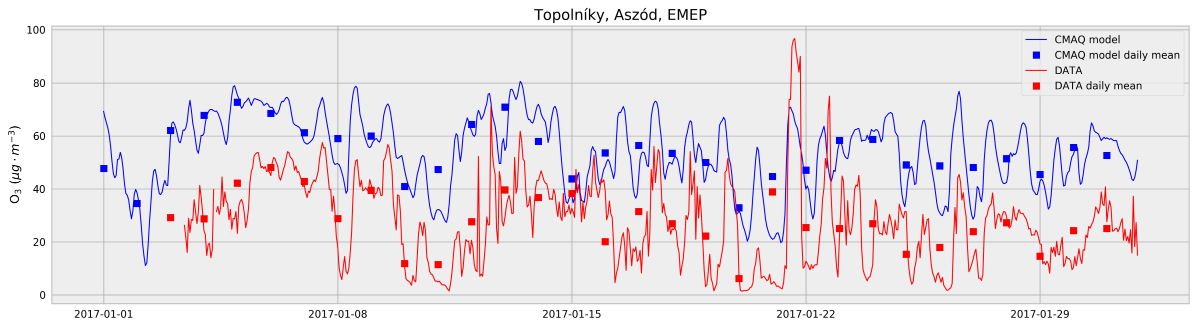


**FIGURE 3.** Comparison of the model output with measurement at Topoľníky EMEP background station for $O_3$ concentration.

For comparison of model against measurements we used the modStats function from the OpenAir package [18]. There is a very wide range of evaluation statistics that can be used to assess model performance. However, no single statistic that encapsulates all aspects of interest. For this reason it is useful to consider several performance statistics and also to understand the sort of information or insight they might provide [19].

**TABLE 1.** The modStats evaluation of the model comparison $O_3$ concentrations with the experimental data from the four EMEP rural background stations in Slovakia. The statistical functions are explained in text.

| n | FAC2 | MB | MGE | NMB | NMGE | RMSE | r | COE | IOA |
|---|---|---|---|---|---|---|---|---|---|
| 2554 | 0.839 | 1.185 | 19.610 | 0.0194 | 0.320 | 23.860 | 0.512 | 0.130 | 0.565 |

In Table 1 we evaluate the model statistic for the concentration of the ground ozone $O_3$ with the experimental data from the four EMEP rural background stations. Let us discuss the symbols in first row of the Table 1. The symbol $n$ represents the number of data which we compared with the model. The *FAC2* is the *Fraction of the predictions within a factor of two*. Symbols *MB*, *MGE*, *NMB* and *NMGE* stands for the *Mean bias*, *Mean Gross Error*, *Normalized Mean bias*, and *Normalized Mean Gross Error*, respectively. The *RMSE* is the *Root mean square error*. *Correlation coefficient* is denoted as $r$. The symbol *COE* is the *Coefficient of Efficiency*, which for the perfect model has the value $COE = 1$. The value of $COE = 0$ implies that the model is no more able to predict the observed values than does the observed mean. Therefore, since the model can explain no more of the variation in the observed values than can the observed mean, such a model can have no predictive advantage. For negative values of *COE*, the model is less effective than the observed mean in predicting the variation in the observations. The *Index of Agreement IOA* spans between 1 and +1 with values approaching +1 representing better model performance. More information and the exact definition of the specific statistical functions can be found in [19].

In Table 1 we can see that with Correlation coefficient $r = 0.512$, modeled values are not in very good agreement with the observations. However, the model has slightly predictive advantage since the $COE = 0.13 > 0$. The discrepancies between model and observations could be caused by a) inaccurate emission input b) low grid resolution ($14 \times 14$ km).

We used the conditional quantiles for graphical representation of the model performance against observations for continuous measurements. These plots are included in the OpenAir package [18]. In Fig. 4 there is conditional quantiles plot applied to the CMAQ model for 4 rural EMEP $O_3$ monitoring sites in Slovakia in January 2017. In Fig. 5there are conditional quantiles plots for the specific EMEP $O_3$ monitoring sites. We can see that for the CHOPOK

EMEP site the model underestimates the observed values. This is caused by the low grid resolution, which does not account for the elevation of Chopok site properly ($O_3$ concentrations increase with altitude). On the other hand, CMAQ overestimates the concentrations for the Topoľníky site . This could be caused by the constant emissions which we used for the simulation.

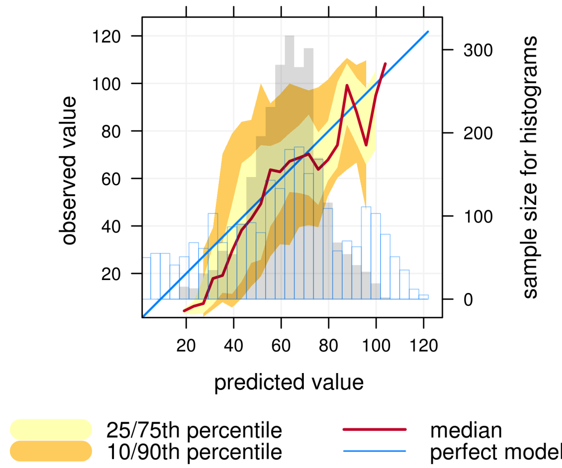


**FIGURE 4.** Conditional quantiles plot applied to the CMAQ model for 4 rural EMEP $O_3$ monitoring sites in Slovakia in January 2017, for hourly data. The blue line shows the results for a perfect model. The red line shows the median value of the predictions. The shading shows the predicted quantile intervals i.e. the 25/75th and the 10/90th. A perfect model would lie on the blue line and have a very narrow spread. The shadow and blue histogram shows the counts of predicted values and measured values, respectively.

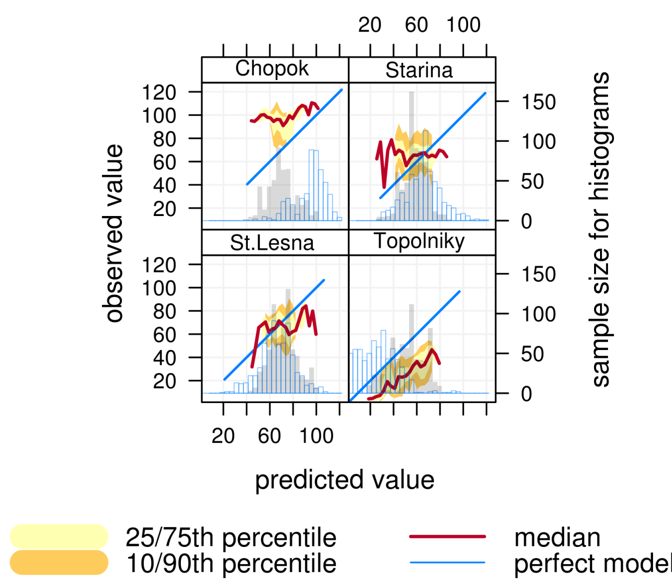


**FIGURE 5.** Conditional quantiles plots for the specific EMEP $O_3$ monitoring sites. For the description of the plot see FIG. 5.

## Results for the $NO_2$ concentration

Figure 6 shows plotted $NO_2$ hourly and daily concentration series for the modeled and observed values at station Topoľníky. We can see that the model more less reproduces the observed trends.

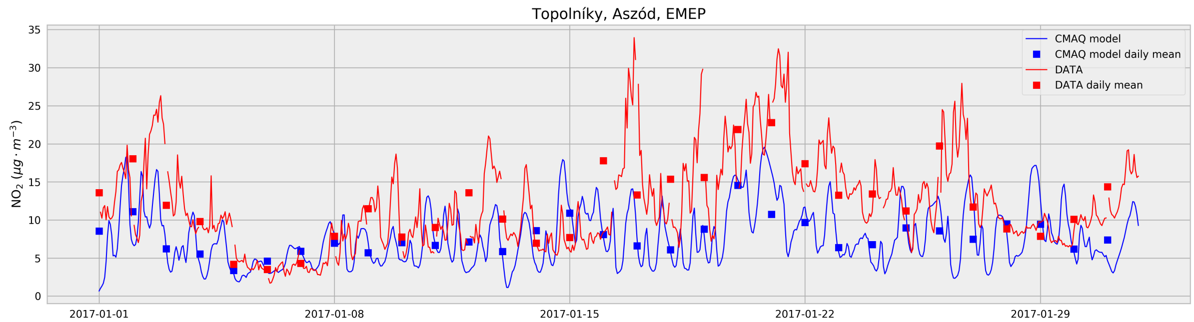


**FIGURE 6.** Comparison of the model output with measurement at EMEP background station Topoľníky for $NO_2$ concentration

For the evaluation of statistics comparing modeled $NO_2$ concentration against measurements we used the modStats function from the OpenAir package [18] as we did for the $O_3$ concentrations. From the Table 2 we can see that we obtain similar Correlation coefficient $r = 0.442$ and $COE = 0.121$ as in the case of the comparison of the $O_3$ concentrations.

**TABLE 2.** The modStats evaluation of the model comparison $NO_2$ concentrations with the experimental data from the four EMEP rural background stations in Slovakia. The statistical functions are explained in text.

| n | FAC2 | MB | MGE | NMB | NMGE | RMSE | r | COE | IOA |
|---|---|---|---|---|---|---|---|---|---|
| 2807 | 0.487 | -3.039 | 4.369 | -0.412 | 0.592 | 6.849 | 0.442 | 0.121 | 0.560 |

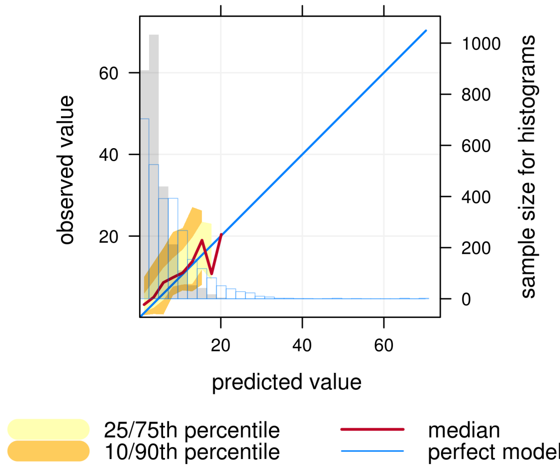


**FIGURE 7.** Conditional quantiles plot applied to the CMAQ model for 4 rural EMEP $NO_2$ monitoring sites in Slovakia in January 2017, for hourly data. For the description of the plot see FIG. 5.

In Fig. 7 there is conditional quantiles plot applied to the CMAQ model for 4 rural EMEP $NO_2$ monitoring sites in Slovakia in January 2017. In Fig. 8 there are conditional quantiles plots for the specific EMEP $NO_2$ monitoring sites. We can see that for the CHOPOK EMEP site the model overestimates the observed values, which is caused by low grid resolution underestimating the station elevation (generally $NO_2$ decrease with altitude).

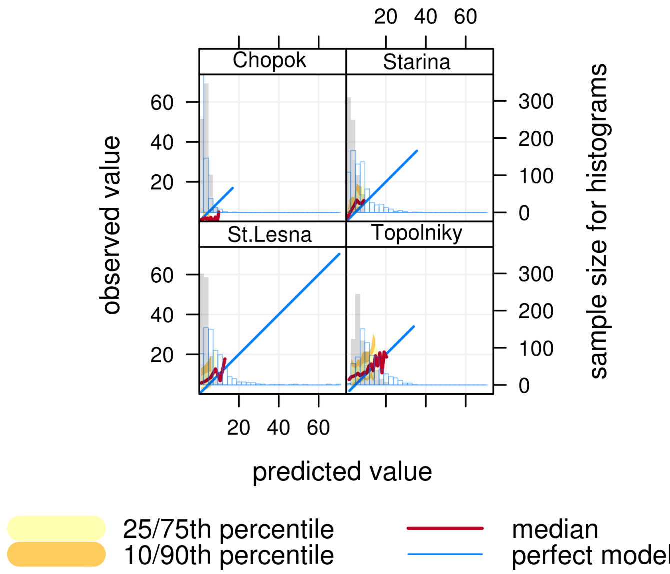


**FIGURE 8.** Conditional quantiles plots for the specific EMEP $NO_2$ monitoring sites. For the description of the plot see FIG. 5.

### Results for the Particule Matter (PM) concentration

Fig. 9 shows hourly and daily concentrations of $PM_{10}$ for the modeled and observed values at Topoľníky station. We can see that the model strongly underestimates the $PM_{10}$ values. This is caused by fact that experimental input emissions did not include domestic heating, which is the major source of $PM_{10}$ and $PM_{2.5}$ concentrations. Therefore, PM values are strongly underestimated even when compared with the measurements at the rural background station.

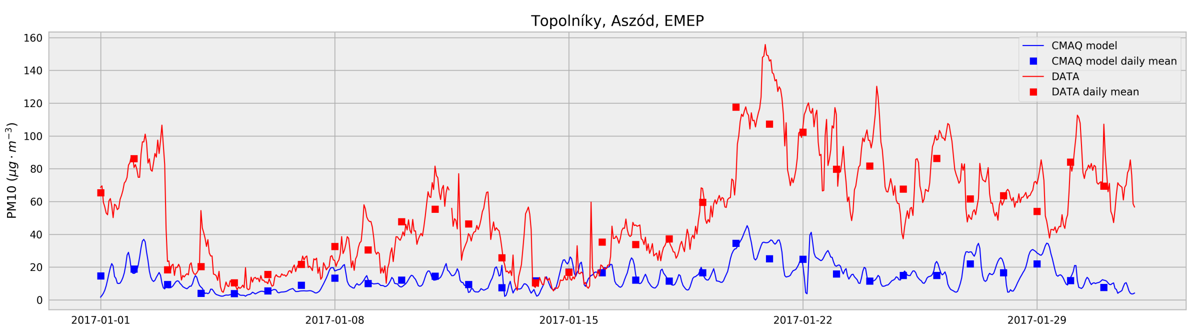


**FIGURE 9.** Comparison of the model output with measurement at Topoľníky EMEP background station for $PM_{10}$ concentration

## CONCLUSIONS AND FUTURE PLANS

CMAQ air-quality simulation was performed on the large domain covering the substantial part of the Europe with the grid resolution (14 × 14 km). 32 cores from the SHMU high-power computation system was used for the simulation taking around 42 minutes per day. For the evaluation of model against measurements, modStats function from the

OpenAir package was used. We found out that our model has small predictive advantage for the concentrations of $NO_2$ and $O_3$. As experimental input emissions did not contain domestic heating, which is the major source of $PM_{10}$ and $PM_{2.5}$ concentration of these pollutants is strongly underestimated even when compared with the measurements on the rural background stations. This fact seem to agree with the hypothesis that small combustion sources are responsible for large part even of the background concentrations of atmospheric aerosol, especially during adverse dispersion situations. In case of gaseous species the model performance is reasonably better and shows the role of adverse dispersion conditions on the concentration of the pollutants.

In the next steps we plan to run the simulation with complete emissions, including the SNAP 2 sector with domestic heating sources obtained using improved emission model for Slovakia. We also plan to perform nested simulation on the domain covering the Czech Repulic and Slovakia and south regions of Poland with the better grid resolution ($4 \times 4$ km). There is also plan to create preprocessor which enables the use of the Aladin meteorological output as an input to the CMAQ instead of the WRF meteorological input. In future there are plans to use the CMAQ model for short-term air quality forecasts.

## ACKNOWLEDGMENTS

I would like to thank my collaborators Jana Matejovicova and Jana Krajcovicova for many advices they gave me and also for their support. I also thank to Johannes Bieser from the Institute for Coastal Research, Helmholtz Centre Geesthacht, for preparing emission input data.

## REFERENCES

[1] U.S. EPA. EPA's Report on the Environment (Roe) (2008 Final Report). U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-07/045F (NTIS PB2008-112484), 2008

[2] CMAQv5.2 Operational Guidance Document https://www.cmascenter.org/cmaq/, 2017

[3] Milbrandt, J. A. and Yau, M. K.: A multimoment bulk microphysics parameterization, Part I: Analysis of the role of the spectral shape parameter, J. Atmos. Sci., 62, 30513064, doi:10.1175/JAS3535.1, 2005a.

[4] Milbrandt, J. A. and Yau, M. K.: A multimoment bulk microphysics parameterization, Part II: A proposed three-moment closure and scheme description, J. Atmos. Sci., 62, 30653081, doi:10.1175/JAS3535.1, 2005b.

[5] Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, J. Geophys. Res-Atmos., 113, D13103, doi:10.1029/2008JD009944, 2008.

[6] Pleim, J. E.: A combined local and nonlocal closure model for the atmospheric boundary layer, Part I: Model description and testing, J. Appl. Meteor. Climatol., 46, 13831395, doi:10.1175/JAM2539.1, 2007.

[7] Pleim, J. E. and Xiu, A. J.: Development of a land surface model, Part II: Data assimilation, J. Appl. Meteor., 42, 18111822, doi:10.1175/1520-0450(2003)

[8] Grell, G. A.: Prognostic Evaluation of Assumptions Used by Cumulus Parameterizations,Mon. Weather Rev., 121, 764787, doi:10.1175/1520-0493(1993)1212.0.CO;2, 1993.

[9] Grell, G. A. and Devenyi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophys. Res. Lett., 29, 1693, doi:10.1029/2002GL015311, 2002

[10] Otte, T. L. and Pleim, J. E.: The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: updates through MCIPv3.4.1, Geosci. Model Dev., 3, 243-256, https://doi.org/10.5194/gmd-3-243-2010, 2010

[11] Snyder, J. P.: Map Projections - A Working Manual, U.S. GEOLOGICAL SURVEY PROFESSIONAL PAPER 1395, WASHINGTON, 1987

[12] Byun, D. and Schere, K. L.: Review of the governing equations,computational algorithms, and other components of theModels-3 CommunityMultiscale Air Quality (CMAQ) modeling system,Appl.Mech. Rev., 59, 5177, 2006.

[13] Emery, C., Jung, J., Koo, B., Yarwood, G., 2015. Improvements to CAMx Snow Cover Treatments and Carbon Bond Chemical Mechanism for Winter Ozone. Final report for Utah DAQ, project UDAQ PO 480 52000000001.

[14] Byun, D.W.: Dynamically consistent formulations in meteorological and air quality models for multiscale atmospheric studies, Part II: Mass conservation issues, J. Atmos. Sci., 56, 37893807,doi:10.1175/1520-0469(1999)

[15] Byun, D.W.: Dynamically consistent formulations in meteorological and air quality models for multiscale atmospheric studies, Part II: Mass conservation issues, J. Atmos. Sci., 56, 38083820, doi:10.1175/1520-0469(1999)

[16] Fountoukis C. and Nenes, A., 2007. ISORROPIA II: a computationally efficient thermodynamic equilibrium model for K+-Ca2+-Mg2+-NH4+-Na+-SO42- -NO3- -Cl–H2O aerosols. Atmospheric Chemistry and Physics, 7, 4639-4659. Grell,

[17] Bieser, J. et al., 2011. SMOKE for Europe  adaptation, modification and evaluation of a comprehensive emission model for Europe. Geoscientific Model Development, 4(1), pp.4768. g System. Applied Mechanics Reviews, 59(2), p.51 Byun,

[18] Carslaw, D.C. and K. Ropkins, (2012). openair  an R package for air quality data analysis. Environmental Modelling & Software. Volume 27-28, pp. 5261.

[19] Carslaw, D.C. (2015). The openair manual  open-source tools for analysing air pollution data. Manual for version 1.1-4, Kings College London.