Chapter 82
Application of Performance Indicators Based on Observation Uncertainty to Evaluate a Europe-Wide Model Simulation at Urban Scale

Philippe Thunis, Bertrand Bessagnet, Etienne Terrenoire, and Augustin Colette

Abstract In the frame of the European Consortium for Modeling of Air Pollution and Climate Strategies (EC4MACS) the CHIMERE chemistry transport model has been run over Europe for the entire year 2009 with a spatial resolution of 7 km with the aim of assessing the urban impact on daily exceedances of PM and NO2 in European cities. In order to better capture these urban impacts, improvements on urban scale meteorology, vertical resolution and emissions have been implemented. In the current work an evaluation of the model results against the AIRBASE European monitoring network measurements is done using model performance indicators (MPC) based on observation uncertainty.

The MPC used in this approach, constructed on the hypothesis that model results are allowed the same margin of uncertainty as measurements, are developed for four statistical indicators (Root Mean Square Error, Normalized Mean Bias, Normalized Mean Standard Deviation and temporal correlation) to summarize the model-observation errors in terms of phase, amplitude and bias. The utility of this approach is to provide a performance scale to inform the user on the expected value an indicator should reach for a particular modeling application. These indicators are then used to identify the strengths and weaknesses of the model application in terms of geographical areas, cities, pollutants and/or period of the year.

P. Thunis (✉)
European Commission, JRC, Institute for Environment and Sustainability,
Air and Climate Unit, Via E. Fermi 2749, 21027 Ispra, VA, Italy
e-mail: philippe.thunis@jrc.ec.europa.eu

B. Bessagnet • E. Terrenoire • A. Colette
Institut National de l'Environnement Industriel et des Risques, INERIS,
60550 Verneuil en Halatte, France
e-mail: bertrand.bessagnet@ineris.fr; etienne.terrenoire@ineris.fr; augustin.colette@ineris.fr

82.1   Introduction

As Chemistry-Transport Models (CTMs) were initially designed to simulate ozone concentrations within the lower troposphere, a coarse horizontal resolution was sufficient to reach this objective. But during the last decade, air quality legislation has focussed more and more on particulate matter (PM) and CTMs were required to refine their resolution to capture the urban signals as high PM concentrations usually occur in urban areas. In this work a fine resolution ($0.0625 \times 0.125°$ i.e. about 7 km) simulation is performed over Europe for the meteorological year 2009 with the CHIMERE CTM [1]. For the evaluation model performance indicators based on the observation uncertainty [2] are used to identify the main strengths and weaknesses of the CHIMERE application in terms of geographical area, pollutant and period of the year. Results are presented here for NO2 and PM10 based on a comparison with the AIRBASE monitoring network measurements.

82.2   Methodology

The offline Chemistry Transport Model CHIMERE model is fed with ECMWF meteorological fields. Despite their relatively coarse horizontal resolution (16 km) this meteorological input dataset has been preferred to the higher resolution WRF fields since the latter tends to overestimate significantly the magnitude of the wind fields [3]. Anthropogenic emission fields were derived using a top-down approach over the entire domain extending from 10°W to 30°E in longitude and from 36°N to 62°N in latitude. Boundary conditions were obtained from the monthly mean LMDz-INCA climatology for gaseous species and from the GOCART model for aerosols. Biogenic species are calculated using the MEGAN model while wildfire emissions are issued from GFED3. Some important modifications were made to the code itself and to the input dataset:

- The relatively coarse resolution of the ECMWF meteorological fields prevents a good representation of the urban effects. Based on literature overview wind fields intensities in urban centres have arbitrarily been decreased by a factor two. Similarly the turbulent diffusion coefficients have been decreased within the urban canopy by a factor 2.
- Based on a comparison between bottom-up and top-down approaches on one hand and on expert judgments on the other, anthropogenic emissions in some Eastern country regions have been increased significantly to reflect the larger effective residential heating emissions in these regions.
- In addition new temporal profiles for residential emissions based on a "degree-days" concept have been generated to account for the fact that emissions related to heating would increase during the colder winter days (same total emissions distributed according to temperature).

The evaluation of the model performances is based on a comparison with the AIRBASE monitoring stations which are classified in terms of urban, suburban and rural background. The evaluation is performed by grouping stations around a series of cities (30 city areas). For each of these 30 cities the monitoring stations within a circle of radius 200 km are used for the evaluation. In total about 650 stations are used both for PM10 and NO2.

The evaluation itself is based on performance indicators normalised by the observation uncertainty [2]. The main performance indicator is constructed as the ratio between the model to observed root mean square error and a function of the observation uncertainty. As discussed in Thunis et al. [2], values of this ratio between 0 and 0.5 indicate that, on average, differences between model results and measurements are within the range of their associated uncertainties. Conversely, values larger than 1 indicate statistically significant differences between model and measured values. Based on measurement inter-comparison exercises measurement uncertainties values for NO2 and PM10 are provided [4]. As observation uncertainties generally exhibit much larger relative uncertainties at low concentration levels the required on model performance within the range of concentrations becomes less stringent. In this way the less certain the measurements are the less stringent the model performances requested should be.

For visualization of this indicator the target diagram proposed by [5] has been modified by normalizing all quantities by the observation uncertainty (Fig. 82.1, top and bottom rows). The X and Y axis of the diagram represent the observation normalized centered root mean square error (CRMSE) and bias (BIAS), respectively. The distance from the origin then represents the observation uncertainty normalized root mean square error (RMSE). The green area circle identifies the fulfillment of the performance criteria while the dashed green area represents the zone for which model results are within the observation uncertainty range. The negative and positive sides of the X axis are used to identify observed-model differences which are dominated by amplitude (standard deviation) or phase (correlation), respectively. The negative and positive sides of the Y axis identify negative and positive biases, respectively.

82.3 Results

Figure 82.1 provides an overview of the results for PM10 (right) and NO2 (left) where results have been averaged by city areas (each circle represents a city area) and where all station types (urban, suburban and rural) have been included in the analysis. The PM10 target diagram points out an underestimation of the observed levels, especially in Eastern country cities (Prague, Warsaw, Sofia and Krakow) with an error dominated by a lack of amplitude. Despite the increase of domestic emissions in these countries and the new temporal profiles based on degree days, the amplitude of the signal is yet underestimated. For many of the Mediterranean
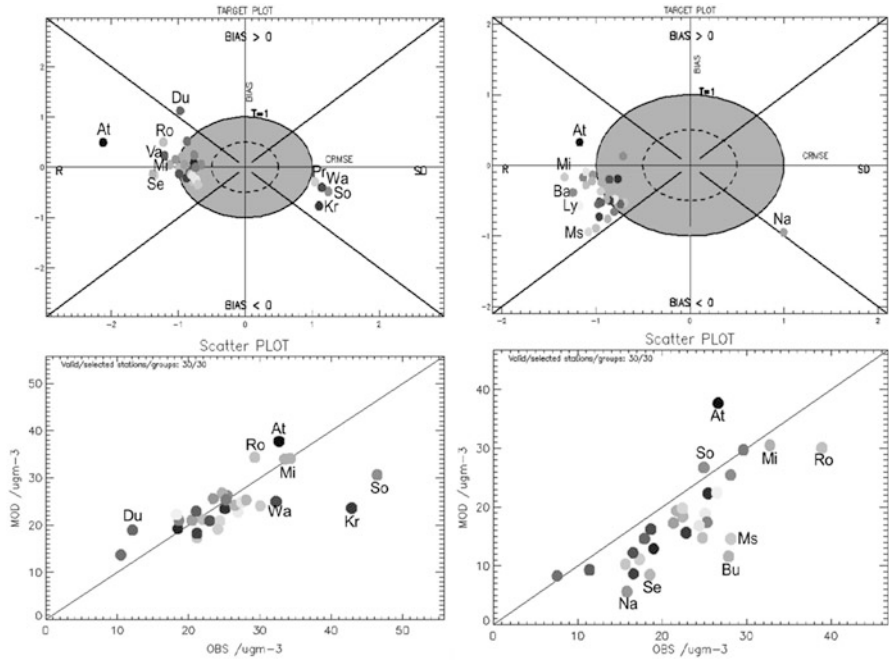
Fig. 82.1 PM10 – all station types (*left*) and NO2 – urban background stations (*right*) model-observed evaluation. Target diagrams (*top*) and scatter plots (*bottom*) provide information for statistics averaged by city areas. For the Target diagrams, stations statistics are sorted and the 10 % worst ones are eliminated. The *circle* on the plot represents the worst remaining statistic. Acronyms are: Du (Dublin), Ro (Rome), Va (Valencia), At (Athens), Mi (Milan), Wa (Warsaw), Kr (Krakow), So (Sofia), Na (Napoli), Ms (Marseille), Se (Sevilla), Pr (Prague). For more information on target diagrams, refer to text

cities (e.g. Rome, Valencia, Milan, Sevilla, Athens) the model results show a lack of temporal correlation which might be caused by to the difficulty to adequately capture the local scale effects (e.g. sea-breezes) with the 16 km meteorology. The scatter plot (bottom) provides some information on the absolute concentrations and confirms the Eastern Europe underestimation while other city areas do not show significant biases.

For NO2 model concentrations exhibit a significant underestimation at almost all cities. According to the Target diagram, the error is dominated by a lack of temporal correlation. The general underestimation could be explained by the fact that NOx emissions generally occur at street level, scale which cannot be sufficiently captured with the current spatial resolution. The lack of temporal correlation might be due to a lack of sufficiently accurate temporal emission profiles for the NOx traffic emissions, as these are emissions that are directly linked to the observed NO2 concentrations.

82.4 Conclusion

CTMs are currently able to simulate air quality over large domains with a refined resolution. This allows assessing model performances with respect to PM10 and NO2 in urban areas in different geographical areas in the frame of a single simulation. In this work the CHIMERE CTM has been run over the entire European territory with a spatial resolution of about 7 km. To better capture urban scale effects some improvements have been made to the model itself (e.g. urban sub-scale paramerisations) but also to some input datasets (e.g. emissions). Model performances have been evaluated around 30 European city areas against the AIRBASE measurements. To perform this evaluation model performance criteria based on observation uncertainty have been used.

Despite corrections made to the PM anthropogenic emissions in some eastern country areas (increase of overall emitted totals, degree days corrections) CHIMERE still underestimates the observed concentration levels. Although the timing of the peaks is quite well reproduced the peaks amplitude is underestimated. Problems also arise from some Mediterranean areas where the model faces difficulties in reproducing the temporal variations of the concentrations. For NO2 the concentration levels are generally underestimated and show a lack of temporal correlation probably due to a lack of sufficiently accurate temporal emission profiles for the NOx traffic emissions.

The proposed evaluation methodology has proved its usefulness to screen model performances. The same methodology can then be used to check the model behaviour in more details (seasonal variations, station details). Further work will also be done to relate model performances in terms of meteorological and air quality fields.

References

1. Bessagnet B, Menut L, Curci G, Hodzic A, Guillaume B, Liousse C, Moukhtar S, Pun B, Seigneur C, Schulz M (2009) Regional modeling of carbonaceous aerosols over Europe – focus on secondary organic aerosols. J Atmos Chem 61:175–202
2. Thunis P, Pederzoli A, Pernigotti D (2012) Performance criteria to evaluate air quality modeling applications. Atmos Environ 59:476–482
3. Miglietta M, Thunis P, Georgieva E, Pederzoli A, Bessagnet B, TErrenoire E, Colette A (2012) Evaluation of WRF model performance in different European regions with the DELTA-FAIRMODE evaluation tool. Int J Environ Pollut 50(1/2/3/4):83–97
4. Pernigotti D, Thunis P, Gerboles M, Belis C (2013) Model quality objectives based on measurement uncertainty. Part II: NO2 and PM10. Atmos Environ 79:869–878
5. Jolliff JK, Kindle JC, Shulman I, Penta B, Friedrichs MAM, Helber R, Arnone RA (2009) Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. J Mar Syst 76:64e82

Questions and Answers

Q: What is the reason for selecting a 200 km radius around the cities?

A: Two different radiuses around cities have been considered in this work: the first one of 30 km corresponding approximately to the area covered by a 50×50 km regional scale model grid cell. But to capture rural background stations away from the city centers, a second radius of 200 km, selected arbitrarily has been used.

Q: Do the newly introduced performance indicators imply that we assume Gaussian distributions of measurements errors, model errors, differences?

A: Yes. From an experimental point of view this assumption is quite reasonable. From a modeling point of view, this is more difficult to assess but the central limit theorem can be used to justify this assumption. In fact, any variable Y that is a combination of Xi (even without normal distributions) can be approximated by a normal distribution because of the Central Limit Theorem provided that the Xi are independent and that the variance of Y is much larger than the variance of any single component from a non-normally distributed Xi. Therefore, it can be assumed that the distribution of individual observations and modeled values is approximately normal.

Q: Your measurement uncertainty is based on measurement error, estimated by the variability of measurements from co-located instruments. Have you considered the issue of spatial representativeness of your monitoring sites. In many cases we would expect the point measurements to be more reliable than a volume average which is what the model predicts. In this case you may have created a performance criteria that is unrealistically strict in terms of identifying "good" or "acceptable" model performance.

A: It is indeed important to consider spatial representativeness and the presented approach could include these impacts whenever realistic estimates for this spatial representativeness are available (work in progress). The performance criteria are not however unrealistically stringent as they are based on an estimate based on the maximum observation uncertainty. The resulting performance criteria in terms of bias and/or correlation are also very close to those proposed in other studies. One possible evolution will be to substitute the maximum by an average measurement uncertainty together with an estimate of the spatial representativeness so that the performance criteria become more station specific.