H14-273

POSTPROCESSING OF A CTM WITH OBSERVED DATA: DOWNSCALING, UNBIASING AND ESTIMATION OF THE SUBGRID SCALE POLLUTION VARIABILITY

Giovanni Bonafè[1], Michele Stortini[1], Enrico Minguzzi[1] and Marco Deserti[1]

[1]Arpa Emilia Romagna, Servizio IdroMeteoClima, Italy

Abstract: In order to integrate CTM outputs with measured data, a post-processing downscaling and unbiasing procedure has been implemented: the procedure is based on a kriging algorithm with external variables, and it provides long-term evaluation of PM10, PM2.5, ozone and nitrogen dioxide at 1 km horizontal resolution. A similar downscaling and unbiasing procedure was also applied to operational air quality forecast.
Moreover, a statistical method was implemented to estimate the subgrid scale variability of pollutants concentrations, based on some deterministic simulations run over selected sample areas and on the data collected by monitoring stations of diverse typologies. The method was tested over the Emilia Romagna region, and applied to estimate population exposure to nitrogen dioxide. Results are presented and discussed.

Key words: *chemistry-transport model, downscaling, Po Valley, subgrid-scale variability, kriging, air quality forecast, air quality assessment, population exposure assessment.*

INTRODUCTION

Even if provided by a well-designed monitoring network, observed data of concentration of pollutants are not enough to accurately evaluate air quality; this is particularly true in an area like Emilia Romagna region, which is geographically heterogeneous and complex and includes cities with more than 100000 inhabitants, many smaller towns, industrial areas, hills, mountains, valleys, coastal areas and a relevant part of the Po river plain. Chemistry-transport models can provide a satisfactory representation of the large scale air pollution, but even if they are run with high horizontal resolution, so far they are not able to accurately reproduce the concentration observed at urban background sites, at least in the closed basin of the Po Valley. Following a simple but effective conceptual representation, the concentrations of many pollutants can be considered as resulting from the superposition of a large scale (regional) background, an urban background and a local scale signal, characterized by peaks or "hot spot". Depending on the considered pollutant, the characteristics of the area and the purpose of the analysis, some of these three components may be negligible; nevertheless, to evaluate population exposure to particulate matter and nitrogen dioxide in the Po Valley, all the three must be taken into account.

In this work, CTM outputs are first integrated with measured data, by means of a post-processing downscaling and unbiasing procedure: this procedure is based on a kriging algorithm with external variables, and provides long-term evaluation of PM10, PM2.5, ozone and nitrogen dioxide with 1 km resolution. A similar downscaling and unbiasing procedure, based on the results of this long term analysis, was also applied to the air pollution daily forecast.

The output of a CTM, post-processed by the downscaling and unbiasing module, reproduces satisfactory the first two components of air pollution: the regional and the urban background. In order to provide the needed additional information about the local scale contribution over a large region, the implementation of a very high scale deterministic model would be unaffordable. Therefore, a statistical method is implemented to estimate the subgrid scale variability of pollutant concentrations, based on a set of deterministic simulations run over selected sample areas and on the data collected by monitoring stations of diverse typologies. The method is tested over the Emilia Romagna region, and applied to estimate population exposure to nitrogen dioxide.

METHODOLOGY

Chemistry-transport model

The core of the modelling system is the chemistry transport model Chimere (Bessagnet *et al.*, 2004). It has been implemented over the Northern Italy and runs operationally at ARPA-SIMC, providing daily analysis and forecast of PM10, PM2.5, ozone, nitrogen dioxide and other pollutants. This implementation (called NINFA) has an horizontal resolution of 10km*10km.

Meteorological input is provided by COSMO-I7, the operational implementation over Italy of the non-hydrostatic meteorological model COSMO (Steppeler *et al.*, 2003). Some post-processing of the COSMO output is performed, in order to provide Chimere with the parameters which are not directly calculated by the model, namely mixing height, friction velocity and cloud water content. Emission input comes from the regional inventory fo Emilia Romagna, from the national inventory of ISPRA and from a large scale inventory provided by project MACC. Boundary conditions are provided by PREV'AIR, the continental implementation of Chimere.

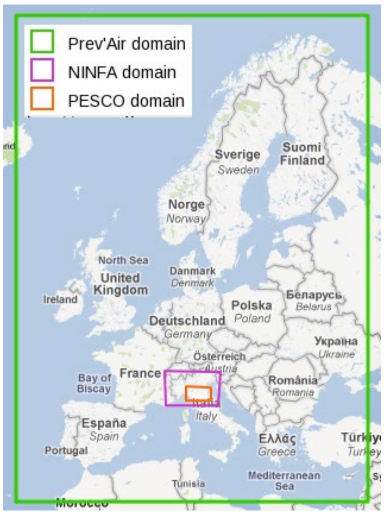

Figure 14. Domains of the continental CTM (Prev'Air), the regional CTM (NINFA) and the geostatistical module (PESCO)

Geostatistical module for unbiasing and downscaling
A geostatistical module (PESCO) is implemented, in order to post-process the output of NINFA with the aim to remove bias and increase horizontal resolution; in this module, CTM ouputs are integrated with surface measurments to produces fields of PM10, PM2.5, ozone and nitrogen dioxide surface concentrations over a regular grid covering the region Emilia-Romagna with a resolution of 1km*1km.

First the differences between CTM fields and observed concentrations are evaluated. This procedure is repeated for every time in which observations are available (daily for aerosols, hourly for gases), and uses all background stations (urban, suburban and rural) of the regional air quality network of Emilia Romagna. These differences are then interpolated on the 1km*1km grid, by means of a kriging algorithm (Honoré and Malherbe, 2003; Cressie, 1993), assuming that they can be expressed as the sum of a linear combination of some external parameters plus a “small” residual term. The coefficient of the linear combination are fitted with the minimum square method, but they are constrained to stay in the range between the 5^{th} and the 95^{th} percentile of the coefficients estimated through a first guess, unconstrained fit. To choose the appropriate external parametres, several candidates were tested, and the most useful turned out to be elevation and total annual emissions (NO_2 emissions are used to process ozone and NO_2, PM10 emissions for PM10 and PM2.5). Finally, the interpolated differences are added to the CTM output, to obtain the final fields.

The final PESCO fields are also used to post-process operational air quality forecasts produced by NINFA system, in order to unbias and downscale surface concentration fields. The ratios of PESCO final fields to NINFA analysis (i.e. d-1 forecasts) are evaluated for every time and then averaged over the standard three-months seasons (DJF, MAM, JJA, SON); these seasonally-averaged fields are then used as correction factors for daily operational forecasts.

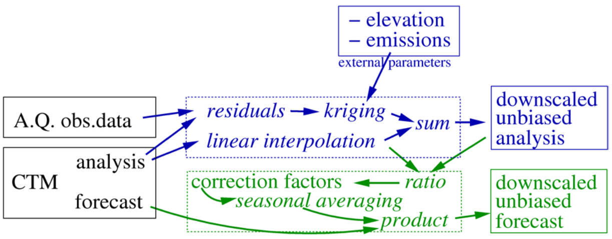

Figure 15. Diagram of the NINFA+PESCO modelling system

Evaluation of subgrid-scale variability
Inside each 1*1 km^2 cell of PESCO grid, pollutant concentration – even if yearly averaged - cannot be considered uniform, especially if population exposure is investigated. In this work, a simplified methodology to estimate the subgrid-scale variability of nitrogen dioxide is described and tested. To do this, the ADMS-Urban dispersion model is used to produce annual simulations of NO_X concentrations, over 5 sample domains each of whom contains part of a city and the adjacent sub-urban and rural areas (figure 3, left). ADMS-Urban (Carrhuters *et al.*, 1994; CERC, 2006) is an advanced gaussian model, specifically designed to deal with the complex emission pattern typical of urban areas. After removing a buffer near the boundaries of the sample domains, a total of 248 1*1 km^2 PESCO grid cells are selected for the following processing.

ADMS-Urban outputs are used to evaluate NO_X annual average concentrations, which are then interpolated from the variable resolution working grid of the model to regularly spaced grids with 50 m step. The resulting fields are then compared with annual NO_2 concentrations measured by the available monitoring stations included in the sample domains; a total of 10 kerbside stations are used in this study. A linear regression is evaluated (figure 3, right), and ADMS-Urban fields are corrected accordingly: the intercept (36 $\mu g \cdot m^{-3}$) corresponds to the large-scale background concentrations, which are neglected by ADMS-Urban simulatons and are assumed to be constant in Emilia Romagna; the slope is a correction factor which accounts for non-linear mechanism not described by the gaussian model, such as the tranformation of NO_x in NO2, the chemical reactions involving ozone and some small-scale sources of turbulence in urban areas.

Subsequently, the distribution frequency of corrected ADMS-Urban fields is evaluated for each of the PESCO grid cells included in ADMS-Urban domains: according to the Akaike Information Criterion (Burnham and Anderson, 2002), the log-normal distribution is found to provide the best fit in almost all of the cells. The two parameters of the lognormal distibution, μ and σ, are then estimated in the entire PESCO domain: the mode of the distribution is assumed to be equal to the background concentration evaluated by PESCO for each cell. Moreover a regression tree calibrated on the sample domains (figure 4) is used to evaluate $\ln(\sigma)$ as a function of quantities whose values are available on the whole domain: three predictors were selected, namely the mode of the lognormal distribution, the number of inhabitants and the total NO_x emissions in each cell. Finally σ values are corrected to ensure that concentrations at the monitoring sites are within the 95th percentile of the distribution of the corresponding PESCO cell.

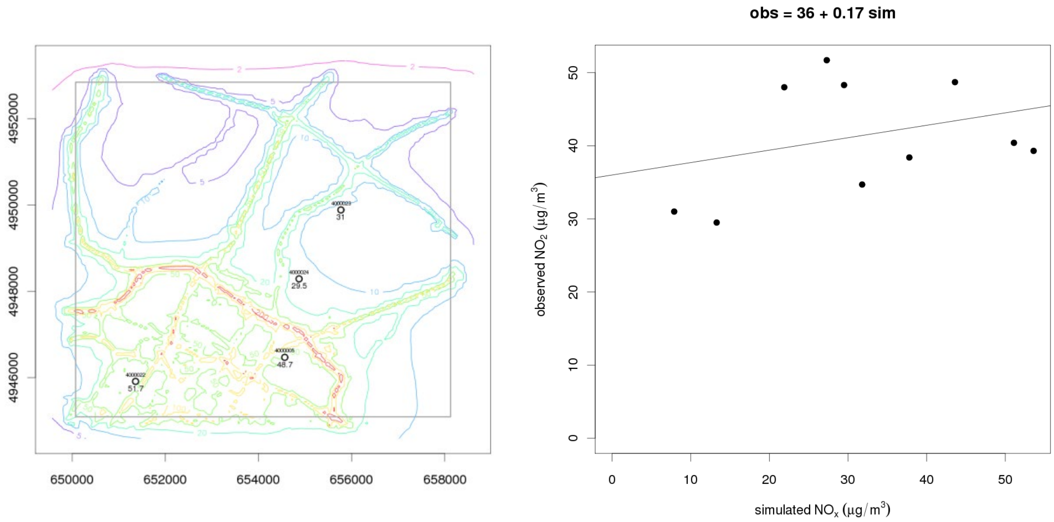

Figure 16. Left: concentrations of NO_x (annual mean, colored isolines) simulated by ADMS-Urban over one of the sampling domains and NO_2 annual mean observed by the monitoring stations (black circles). Right: linear fitting of observed NO_2 versus simulated NO_x, over the monitoring stations included in the sampling domains where high resolution simulations were performed with ADMS-Urban.

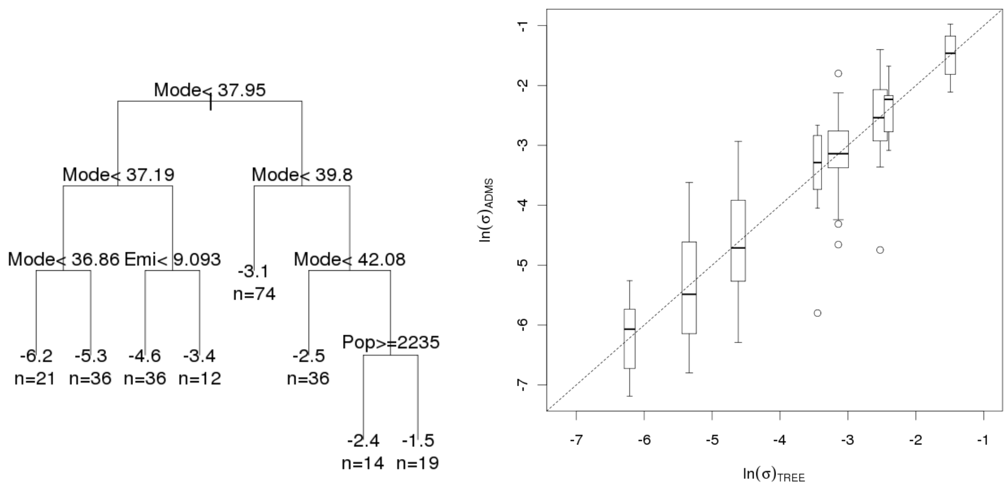

Figure 17. Regression tree to forecast $\ln(\sigma)$, natural logarithm of the scale parameter of the log-normal distribution, using as predictors the mode of the distribution, NO_x emissions and population density. Left: diagram of the tree. Right: $\ln(\sigma)$ forecasted by the tree versus $\ln(\sigma)$ fitted on the high resolution ADMS-Urban simulations, used for the calibration of the tree.

RESULTS AND CONCLUSIONS

Figure 5 shows the results of PESCO geostatistical module applied to the compliance with EU legislation requirements. As expected, the most critical requirement is the number of days in which PM10 concentrations exceed the 50 $\mu g \cdot m^{-3}$ threshold, which is not fulfilled in most urban and suburban areas, along the whole extent of "via Emilia" (a road running on the foothills of Apennines across the whole region and connecting its main cities), and also in large portions of rural areas in western Emilia Romagna. Small but significant areas in the heavily industrialised central Emilia Romagna do not respect the standards also for PM2.5 and NO_2.

Figure 6 shows the results of a leave-one-out cross-validation, performed on the same 2010 data. For all pollutants except ozone, errors are larger in winter, when concentrations are higher and their spatial gradients sharper; this is particularly true for PM10, for which PESCO performs very well in summer but has large errors in January and February. Errors seems to be generally smaller for aerosols than for gases, but this could be partly due to the fact that for gases the scores refers to hourly values, while for aerosols they refer to daily averages. Finally, it should be noted that the performance of this geostatistical approach is also critically dependent on the availability of reliable background measurements: in this study the availability of the few rural measurements has proven critical.

Figure 7 shows the first application of the evaluation of the subgrid scale varibility. The distribution frequency evaluated for each PESCO cell has been used to estimate the fraction of each cell in which NO_2 annual mean concentrations are expected to excced the 40 $\mu g \cdot m^{-3}$ threshold. These values has been multiplied for the number of inhabitants, to obtain an estimate of the total number of people exposed to concentrations above the legislation limits. Although results are preliminary, and an independent validation has not yet been performed, results are promising: the approach leads to "reasonable" results, and the total population exposed (325000 inhabitants) resulted to be significantly larger than estimated with the 1 km resolution PESCO maps (172000 inhabitants).

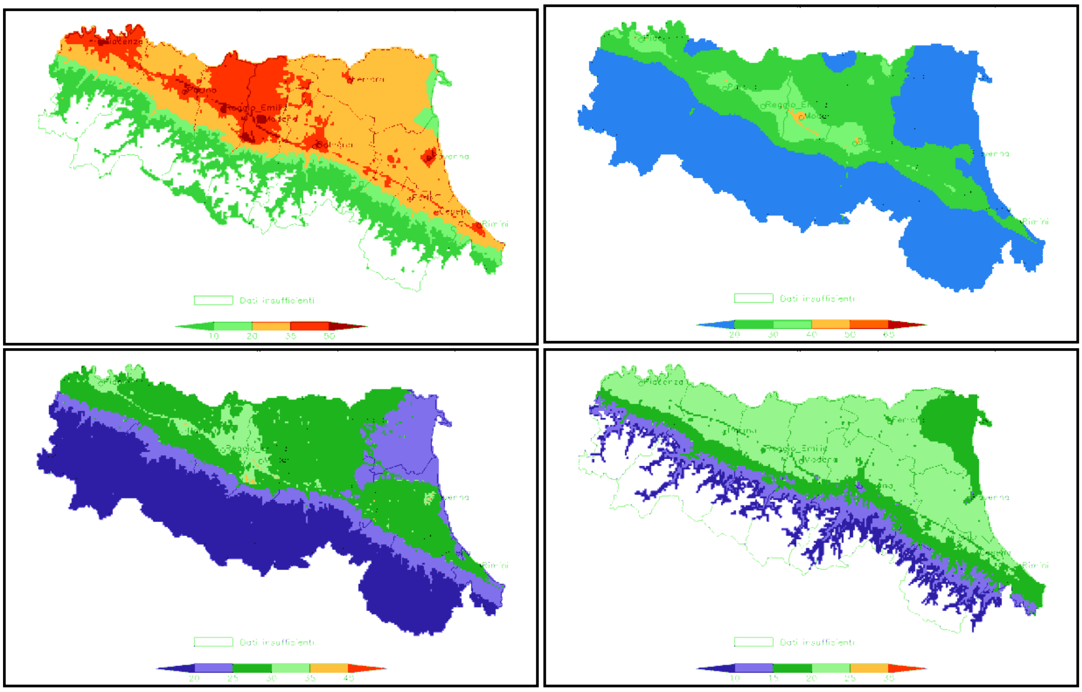

Figure 18. Background concentrations estimated for year 2010. Top left: PM10, number of exceedances of the daily threshold. Top right: NO_2, annual mean. Bottom left: PM10, annual mean. Bottom right: PM2.5, annual mean.

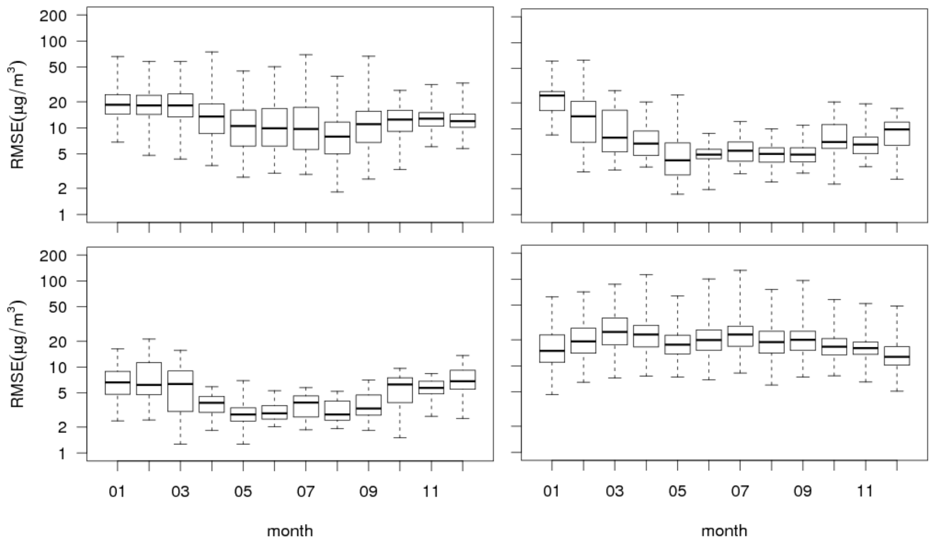

Figure 19. Cross-validation of the geostatistical module PESCO: monthly minimum, maximum, median, 25th and 75th percentile of the root mean square error, represented as box-and-whyskers plots. Top left: nitrogen dioxide; top right: PM10; bottom left: PM2.5; bottom right: ozone.

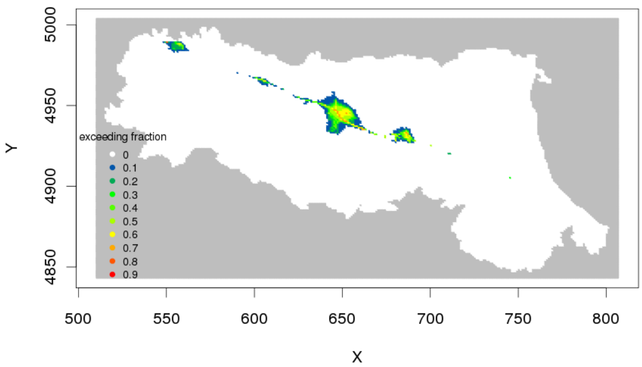

Figure 20. Fraction of land where NO_2 annual mean exceeded the 40µg·m^{-3} threshold, as estimated by the UltraPESCO module. Total exposed population was estimated in 325000

REFERENCES

Bessagnet, B., A. Hodzic, R. Vautard, M. Beekmann, S. Cheinet, C. Honoré, C. Liousse and L. Rouil, 2004: Aerosol modeling with CHIMERE: preliminary evaluation at the continental scale. *Atmos. Environ.*, **38**, 2803-2817.

Burnham, K. P., and Anderson, D.R., 2002: Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. Springer-Verlag. ISBN 0-387-95364-7.

Carruthers, D., Holroyd, R., Hunt J., Weng, W.-S., Robins, A., Apsley, D., Thomson, D., Smith, F., 1994: UK-ADMS: a new approach to modelling dispersion in the earths atmospheric boundary layer. Journal of Wind Engineering and Industrial Aerodynamics 52, 139-153

CERC, 2006 : ADMS-Urban, User Guide. Available from Cambridge Environmental Research Consultant, Cambridge, UK

Cressie, N.A.C., 1993: Statistics for Spatial Data. Wiley Series in Probability and Statistics. *Wiley-Interscience*.

Cullen A.C. and Frey H.C., 1999: Probabilistic techniques in exposure assessment. Plenum Press, USA, pp. 81-159.

De Lauretis, R., A. Caputo, R.D. Còndor, E. Di Cristofaro, A.Gagna, B. Gonella, F. Lena, R. Liburdi, D. Romano, E. Taurino, and M. Vitullo, 2009: La disaggregazione a livello provinciale dell'inventario nazionale delle emissioni. *Technical Report 92, Istituto Superiore per la protezione e la Ricerca Ambientale*

Dosio, A., S. Galmarini, and G. Graziani, 2002: Simulation of the circulation and related photochemical ozone dispersion in the Po plains (northern Italy): comparison with the observations of a measuring campaign. *Journal of Geophysical Research-Atmospheres*, **107**(**D22**), 8189

Evans M., Hastings N. and Peacock B., 2000: Statistical distributions. John Wiley and Sons Inc.

Honoré, C. and L. Malherbe, 2003: Application de modèles grande échelle à la problématique régionale: cas de l'ozone. *Technical report, INERIS-LCSQA*, December 2003.

Steppeler, J., G. Doms, U. Schättler, H.W. Bitzer, A. Gassmann, U. Damrath, G. Gregoric, 2003: Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteorology and Atmospheric Physics*, **82**, 75-96

Tugnoli, S. and V. Rumberti, 2010: Inventario delle emissioni in atmosfera. Technical report, ARPA Emilia Romagna. www.arpa.emr.it/cms3/documenti/_cerca_doc/meteo/ambiente/inventario_emissioni_2007.pdf