Post-processing of a CTM with observed data: downscaling, unbiasing and estimation of the subgrid scale pollution variability

Giovanni Bonafè, Michele Stortini, Enrico Minguzzi and Marco Deserti
ARPA Emilia Romagna - Servizio IdroMeteoClima

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.

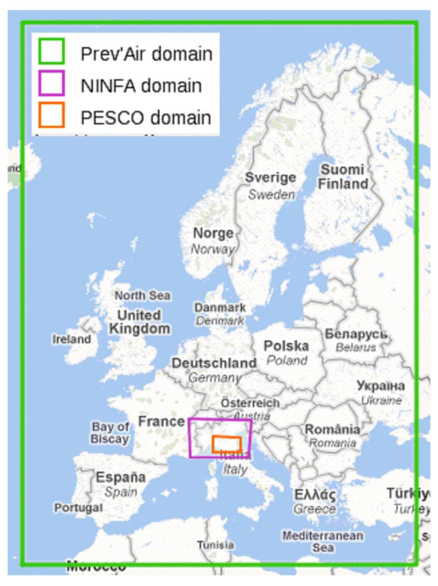

Figure 1: Domains of the continental CTM (Prev'Air), the regional CTM (NINFA) and the geostatistical module (PESCO)

NINFA: chemistry-transport model

The implementation of the CTM is called NINFA:

- chemistry transport model Chimere (Bessagnet et al., 2004), implemented over the Northern Italy (horizontal resolution 10km); runs operationally at ARPA-SIMC;
- daily analysis and forecast of PM10, PM2.5, ozone, nitrogen dioxide and other pollutants.
- meteorological input is provided by COSMO-I7, 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 mixing height, friction velocity and cloud water content;
- emission input comes from the regional inventory of 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.

PESCO: geostatistical module for unbiasing and downscaling

analysis 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. It 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. For every timestep (every day for PM, every hour for O_3 and NO_2):

1. differences between CTM fields and observed concentrations are calculated;
2. 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
   - coefficients of the linear combination are fitted with minimum square method, but
   - they are constrained to stay in the range between the 5th and the 95th percentile of the coefficients estimated through a first guess, unconstrained fit
   - several external parameters were tested; the most useful turned out to be elevation and total annual emissions
3. the interpolated differences are added to the CTM output, to obtain the final fields.

forecast The final PESCO fields (analysis) are also used to post-process operational air quality forecasts produced by NINFA system, in order to unbias and downscale surface concentration fields:

1. the ratios of PESCO final fields to NINFA analysis are evaluated for every time and then
2. averaged over the standard three-months seasons (DJF, MAM, JJA, SON);
3. these seasonally-averaged fields are then used as correction factors for daily operational forecasts.

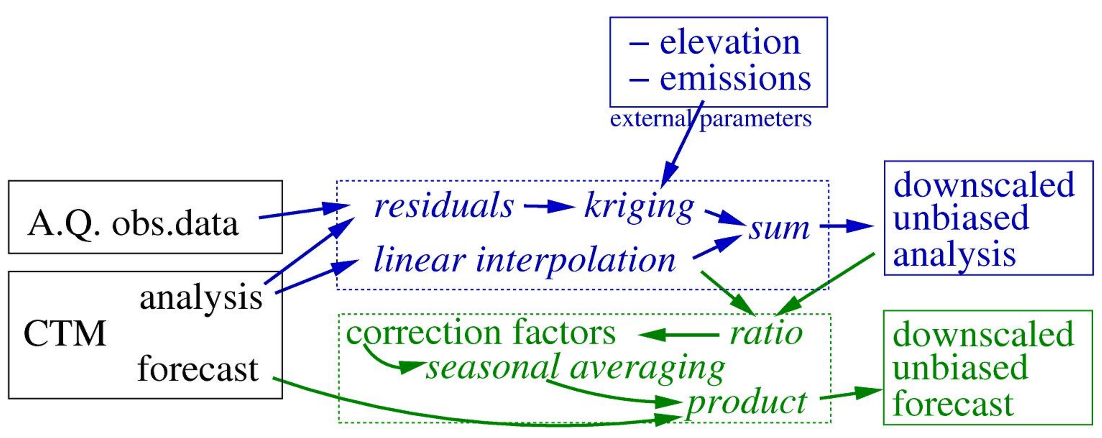

Figure 2: Diagram of the NINFA+PESCO modelling system

Air quality assessment

PESCO geostatistical module was applied to the evaluation of the compliance with EU legislation requirements for PM10, PM2.5, O_3 and NO_2. 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.

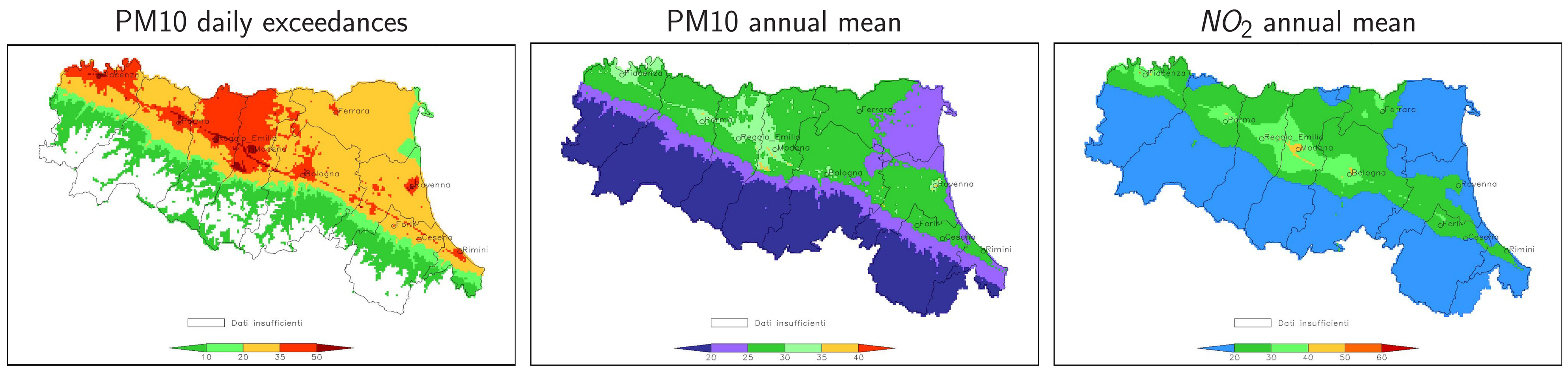

Figure 3: Year 2010: evaluation of the compliance with EU legislation requirements

Cross-validation

Figure 4 shows the results of a leave-one-out cross-validation fo the PESCO module, performed on the 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;
- performances is critically dependent on the availability of reliable background measurements.

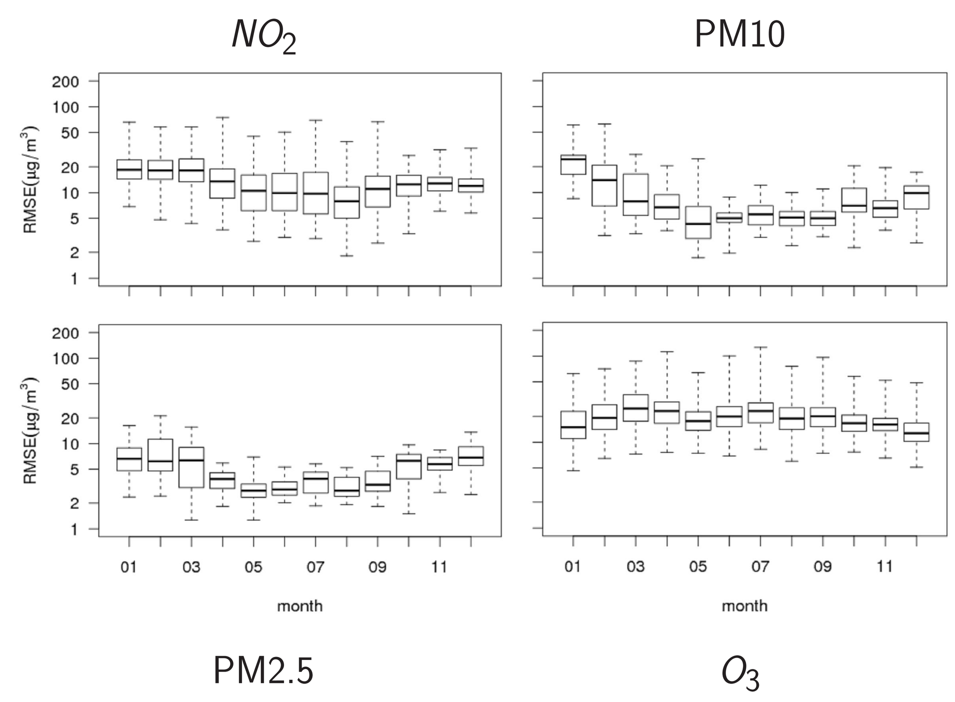

Figure 4: Cross-validation of the geostatistical module PESCO: monthly minimum, maximum, median, 25^{th} and 75^{th} percentile of the root mean square error, represented as box-and-whyskers plots

UltraPESCO: evaluation of subgrid-scale variability

Inside each 1km·1km cell of PESCO grid, pollutant concentration - even if yearly averaged - cannot be considered uniform, especially if population exposure is investigated. A simplified methodology to estimate the subgrid-scale variability of NO_2 is described and tested.

1. ADMS-Urban dispersion model (Carrhuters et al., 1994; CERC, 2006) is used to produce annual simulations of NO_x concentrations, over 5 sample domains 5
   - included in the PESCO domain
   - containing part of a city and the adjacent sub-urban and rural areas
2. after removing a buffer near the boundaries of the sample domains, a total of 248 1km·1km cells are selected for the following processing:
   - interpolation from the variable resolution working grid of the model to regularly spaced grids with 50m step
   - the resulting fields are then compared with annual NO_2 concentrations measured by the available monitoring stations included in the sample domains (10 kerbside stations)
   - a linear regression is evaluated, and ADMS-Urban fields are corrected accordingly
3. 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
4. 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
   - a regression tree calibrated on the sample domains 6 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
5. σ values are corrected to ensure that concentrations at the monitoring sites are within the 95^{th} percentile of the distribution of the corresponding PESCO cell

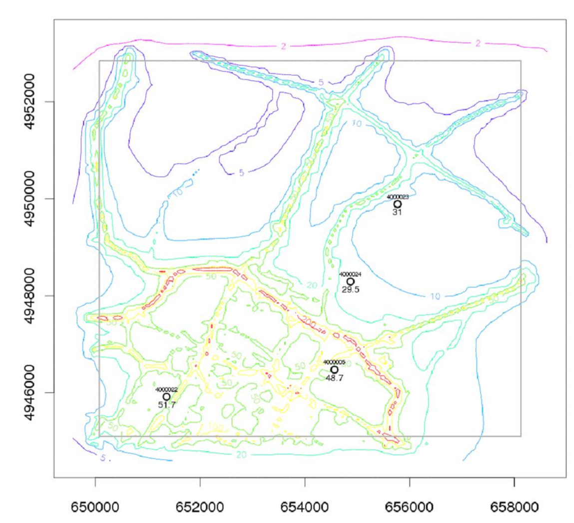
Figure 5: NO_x annual mean (colored isolines) simulated by ADMS-Urban over one of the sampling domains and NO_2 annual mean observed by the monitoring stations (black circles)

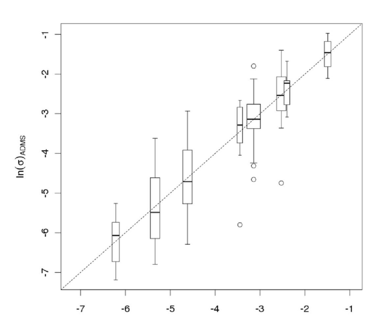
Figure 6: $ln(\sigma)$ forecasted by the regression tree versus $ln(\sigma)$ fitted on the high resolution ADMS-Urban simulations

Population exposure

Figure 7 shows the first application of the UltraPESCO module (evaluation of the subgrid scale varibility)

- the distribution frequency evaluated for each 1km cell has been used to estimate the fraction of each cell in which NO_2 annual mean concentrations are expected to exceed 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
- results are promising:
  - PESCO (no subgrid scale evaluation) → 172000 exposed inhabitants
  - UltraPESCO (with subgrid scale evaluation) → 325000 exposed inhabitants

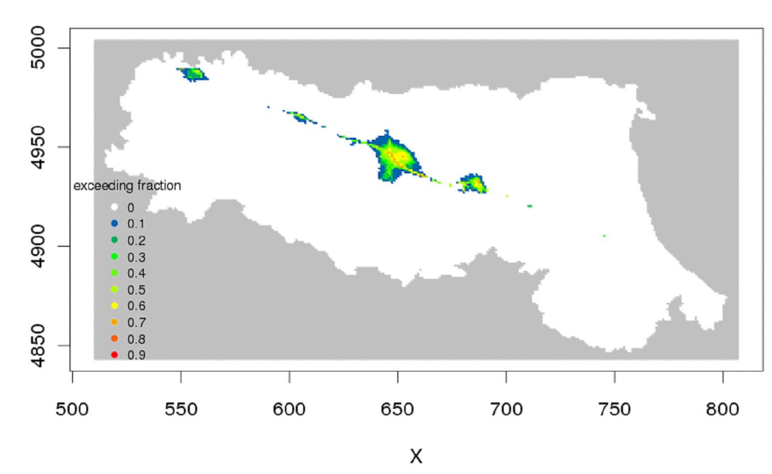

Figure 7: Fraction of land where NO_2 annual mean exceeded the $40\mu g \cdot m^{-3}$ threshold, as estimated by the UltraPESCO module.

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. Cóndor, 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.

gbonafe@arpa.emr.it