H13-137

DATA ASSIMILATION IN AIR QUALITY MODELLING OVER PO VALLEY REGION

Gabriele Candiani[1], Claudio Carnevale[1], Giovanna Finzi[1], Enrico Pisoni[1] and Marialuisa Volta[1]

[1]Department of Information Engineering (DII), University of Brescia, Brescia, Italy

Abstract: An accurate description of the air quality state is a very challenging task due to the complexity and non-linearity of the mechanisms taking place in atmosphere. This is especially true when dealing with secondary pollutants such as ozone (O_3). Deterministic chemical transport models are valuable tools to understand the spatial distribution of an air pollutant over a certain domain. These models can provide spatially consistent air quality data as they consider the main physical and chemical processes governing the atmosphere. However, uncertainties in formalization and input data (emission fields, initial and boundary conditions, meteorological patterns) can heavily affect simulation results. This issue can be addressed through the use of Data Assimilation (DA), which combines concentration fields of secondary atmospheric pollutants simulated by a chemical transport model (background) with measured data (observations).
In this work, O_3 concentration fields over the Po Valley region (Northern Italy) have been simulated by the chemical transport model TCAM. For the period from 15th May to 15th July 2007, the Optimal Interpolation (OI) algorithm has been used to assimilate O_3 and NO_2 data measured by both ground stations and satellite data, into the TCAM simulations. The study area (640 × 410 km^2) is characterized by complex topography, high anthropogenic emissions and frequent stagnant meteorological conditions leading to high concentration levels of ozone in summer. Preliminary results show that the estimate of O_3 concentrations improves significantly assimilating O_3 data, while no impact or even a negative impact can be seen when performing the assimilation of NO2 for both ground-based and satellite data. The research has been developed in the framework of the Pilot Project QUITSAT (QUalità dell'aria mediante l'Integrazione di misure da Terra, da SAtellite e di modellistica chimica multifase e di Trasporto - contract I/035/06/0 – http://www.quitsat.it), sponsored and funded by the Italian Space Agency (ASI).

Key words: *Air Quality, Data Assimilation, Optimal Interpolation*

INTRODUCTION

An accurate description of the air quality state is a challenging task due to the complexity of the atmosphere mechanisms. In fact, the dynamic of a pollutant concentration is usually non-linear and partially unknown. This is especially true when dealing with secondary pollutants such as ozone (O_3), which constitutes one of the major concerns in Europe, due to its harmful effects on human health and on natural ecosystems. Ozone formation and accumulation are non-linear processes, which depends on a large number of reactions taking place in the atmosphere. These phenomena are favoured by meteorological stagnating conditions and by primary emissions of precursors (namely nitrogen oxides and volatile organic compounds). Due to these reasons, modeling tools are required to provide a good estimate of the air quality state. In fact, Chemical transport model (CTM) are able to reproduce in a 3D-gridded domain the complex chemical and physical processes governing the atmosphere. However, uncertainties in formalization and input data (emission fields, initial and boundary conditions, meteorological patterns) can heavily affect simulation results. Thus, a Data Assimilation (DA) technique such as Optimal Interpolation (OI; Kalnay, E., 2003), can be used to assimilate measurements in the simulations of a deterministic chemical transport model, in order to obtain a better estimate of the atmospheric state.

The paper is organized as follows: Section 2 describes the formalization of the implemented scheme; Section 3 illustrates a case study in which O_3 and NO_2 data, measured by ground stations and a satellite sensor, are assimilated into a chemical transport model to estimate the ozone concentration; in Section 4 conclusions are given.

METHODOLOGY

Aim of the DA scheme is to compute the best possible estimate (analysis state $x_a(t)$) of the unknown pollutant concentration (true state $x(t)$) starting from a model simulation (background $x_b(t)$) and from a series of measured data (observations $y_o(t)$). The background is modeled as

$$\dot{x}_b(t) = f(x_b(t), v(t)) \tag{1}$$

where $x_b(t) \in \Re^n$, f represents the CTM and $v(t)$ represents the uncertainties of the model. The observations are related to the true state $x(t)$ by

$$y_o(t) = H(x(t), t) + w(t) \tag{2}$$

where $y_o(t) \in \Re^p$, the matrix $H(t) \in \Re^{p \times n}$, called observation operator, maps the model true state $x(t)$ into the observations true state $y(t)$ and $w(t)$ is the observation error. The model error $v(t)$ and the observation error $w(t)$ are assumed to be white, gaussian and mutually uncorrelated:

$$v(t) \approx N(0, B) \tag{3}$$

$$w(t) \approx N(0, R) \tag{4}$$

$$\mathrm{E}[v(\cdot)] = \mathrm{E}[w(\cdot)] = \mathrm{E}[v(\cdot)w(\cdot)'] = 0 \tag{5}$$

$B \in \Re^{n \times n}$ and $R \in \Re^{p \times p}$ are the error covariance matrices of the background and the observations respectively. The specification of the error covariance matrices B and R is usually a difficult task.

The matrix B is of utmost importance as it describes the model error spread over the domain. In this work, B is estimated using the Gaussian exponential function

$$B = s^2 \cdot exp\left(-\frac{d_h^2}{2L_h^2}\right) \cdot exp\left(-\frac{d_v^2}{2L_v^2}\right) = s^2 \cdot \tilde{B} \tag{6}$$

where:

- d_h and d_v are the distances between two grid cells of the domain in the horizontal and vertical direction;
- L_h and L_v are two parameters defining the decay of covariance in the horizontal and vertical direction;
- s^2 is the model error variance computed on the basis of previous simulations.

This approximation states that the model error variance is constant for each cell of the domain, while the error covariance between two grid points is a function of the horizontal and vertical distance between them.

The R matrix can be considered diagonal when the measurements performed by the monitoring stations are independent. Moreover, if the same type of instruments are used for the measurements, it could be possible to assume that all the monitoring stations have the same error variance r^2, thus R could be rewritten as

$$R = r^2 \cdot I \tag{7}$$

The assimilation scheme follows a two steps approach: 1) at time t, the model is solved to obtain the background field $x_b(t)$; 2) the analysis $x_a(t)$ is computed through the OI algorithm merging the information of the background with the information coming from the observations $y_o(t)$. The analysis represents the initial state for the step 1 at time $t+1$.

Step 1

The computation of $x_b(t)$ is performed using a CTM, able to reproduce the complex and non-linear phenomena involving pollutants in atmosphere. Such model describes the concentration of each pollutant over a certain domain, through a mass conservation equation, considering transport, emission, deposition and chemical transformation phenomena. To solve the resulting partial differential equation system it is necessary to provide the initial and boundary conditions for each involved species:

- Initial conditions, the values for the concentrations of transported species over all the domain ($x_a(t-1)$).
- Boundary conditions, the values for the concentrations of transported species on the boundaries of the application domain during the simulation time.

Other fundamental inputs are the emission and meteorological fields, that must be supplied to CTMs with detailed spatial and temporal resolution. In general, the CTM equation system is solved on a regular 3D grid through a numerical approach. The simulations performed using the equation (1) over the simulation period, provide the background $x_b(t)$ needed for the assimilation step.

Step 2

At each time step t, the analysis state $x_a(t)$ is computed assimilating observations $y_o(t)$ into the background state $x_b(t)$ estimated at step 1, applying the OI algorithm

$$x_a(t) = x_b(t) + K\left(y_o(t) - \tilde{H} \cdot x_b(t)\right), \tag{8}$$

where $\tilde{H}$ is a linear operator which performs the interpolation (based on the Inverse Distance Weighted algorithm) of the background state field $x_b(t)$ at the observation points.

The matrix $K \in \Re^{n \times p}$ is computed by:

$$K = B\tilde{H}' \cdot \left(\tilde{H}B\tilde{H}' + R\right)^{-1}. \tag{9}$$

It should be noted that, under these assumptions about B and R matrices, the Kalman gain can be written as:

$$K = s^2 \cdot \tilde{B}\tilde{H}' \cdot \left(s^2 \cdot \tilde{H}\tilde{B}\tilde{H}' + \frac{r^2}{s^2} \cdot I \right)^{-1} = \tilde{B}\tilde{H}' \cdot \left(\tilde{H}\tilde{B}\tilde{H}' + \sigma \cdot I \right)^{-1}. \tag{10}$$

where the only degree of freedom $\sigma = \frac{r^2}{s^2}$ is the ratio between the observations and model error variances.

CASE STUDY

In this study, OI algorithm is used to assimilate observed and simulated hourly ozone concentration over a 640 × 410 km^2 domain placed in Northern Italy (Figure 1). This region is characterized by densely inhabited and industrialized area, and by high anthropogenic emissions, frequent stagnating meteorological conditions and Mediterranean solar radiation regularly causing high ozone levels in particular during summer months. The period selected for the analysis ranges from 15th May to 15th July 2007.

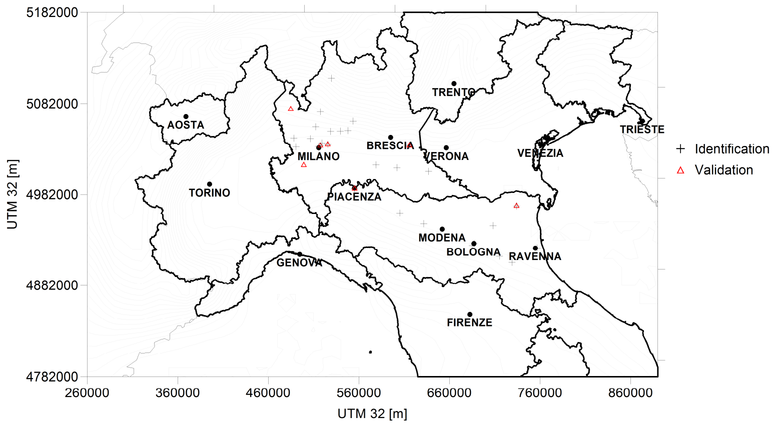

Figure 1. Study area with the ground stations used for the assimilation (black crosses) and for the validation (red triangles).

Background fields

The background field is computed by means of the Transport Chemical Aerosol Model (TCAM; Carnevale, C. et al., 2008). TCAM is a part of the Gas Aerosol Modeling Evaluation System (GAMES; Volta, M. and G. Finzi, 2006) shown in Figure 2, which also includes: the meteorological pre-processor PROMETEO, that provides TCAM all the meteorological input fields in the correct spatial-temporal resolution, starting from the output of continental scale models; the emission processor POEM-PM (Carnevale, C. et al., 2006); a boundary condition pre-processor, that computes the boundary conditions for TCAM model in the application domain starting from the simulation of continental scale models.

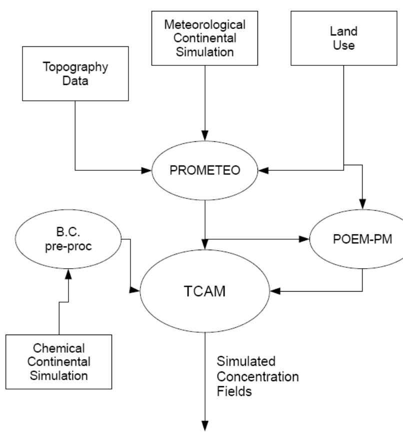

Figure 2. The GAMES modelling system.

The emission fields are estimated by POEM-PM pre-processor starting from the 2004 CTN emission inventory (Deserti, M. et al, 2009). Meteorological fields are computed by means of MM5 model (Grell, G. et al., 1994). The boundary conditions are computed starting from continental scale simulation of CHIMERE model (Schmidt, H. et al., 2001).

Observations data

The observations consist of data measured by ground stations and OMI sensor during the same period of the TCAM simulations. Data from 26 and 32 stations where used for the assimilation of O_3 and NO_2 respectively. These measurements were collected from the regional monitoring networks for Italy, and from AirBase database for France and Switzerland. The 75% of the stations (black crosses in Figure 1) are used in the assimilation scheme, while the remaining stations (red triangles in Figure 1) are used for the validation. Measurements of the tropospheric NO_2 column provided by the OMI sensor are available for 20 days during the simulation period.

DA parameters

The parameters used in the OI algorithm have been chosen on the base of sensitivity analysis and literature values. In the case of ground stations, the horizontal and vertical influence lengths L_h and L_v are equal to 80 km and 80 m respectively. In the case of OMI data, L_h is set to 10 km. The σ is set to 0.1 as suggested in Kalnay, E., 2003.

Results

The validation is carried out comparing the results of the model simulations, before a new assimilation step is performed, with the values measured by the stations chosen for the validation and therefore not used in the assimilation process. Figure 3 and Figure 4 show the results of the ozone statistics in terms of box plots, for the assimilation of O_3 and NO_2 respectively. The box plots present the comparison between the case in which the assimilation is performed (OI) or not (TCAM).

Figure 3 shows the results for the ozone obtained assimilating O_3 measurements. It can be noticed that all the box plots confirm an improvement due to the assimilation of O_3 data. In particular, all the error box plots show a reduction of the error value when the assimilation is performed. The Correlation coefficient instead presents an improvement, with an increment in the statistic from 0.4 to 0.6.

Figure 4 and 5 show the impact on the ozone due to the assimilation of NO_2 measured by ground stations and OMI sensor. From the box plots it can be seen that ozone estimate has no benefit from the NO_2 assimilation, neither using sparse ground stations observations frequent in time (hourly) nor with dense OMI observations with a poor temporal resolution (daily). In particular, all the box plots present a very similar statistics for both the case with (OI) and without (TCAM) the assimilation.

CONCLUSION

In this work, the impact on the estimate of ozone fields is evaluated assimilating two different species (O3 and NO2) from both ground stations and a satellite sensor. The data assimilation is performed using the OI algorithm, a sequential process which performs the assimilation in two steps. In the first step, the model is integrated forward in time to obtain the background field. In the second step, the observations coming from measurement stations are assimilated into the background field to produce the analysis; this new field is the input for the next model integration. The comparison of the assimilation of the two species shows different results. The statistics show that the assimilation of O_3 significantly improves the estimation of ozone concentrations. On the contrary, the NO_2 assimilation does not seem to give any benefit to the estimate of ozone fields, independently on the use of ground-based data or the OMI sensor.

AKNOWLEDGEMENTS

This work has been developed in the frame of Pilot Project QUITSAT (QUalità dell'aria mediante l'Integrazione di misure da Terra, da SAtellite e di modellistica chimica multifase e di Trasporto - contract I/035/06/0 - http://www.quitsat.it), sponsored by the Italian Space Agency (ASI). We also acknowledge the Italian Ministry of University and Research (MIUR), the COST728 action (Enhancing mesoscale meteorological modeling capabilities for air pollution and dispersion applications) and the EU Network of Excellence ACCENT (T&TP).

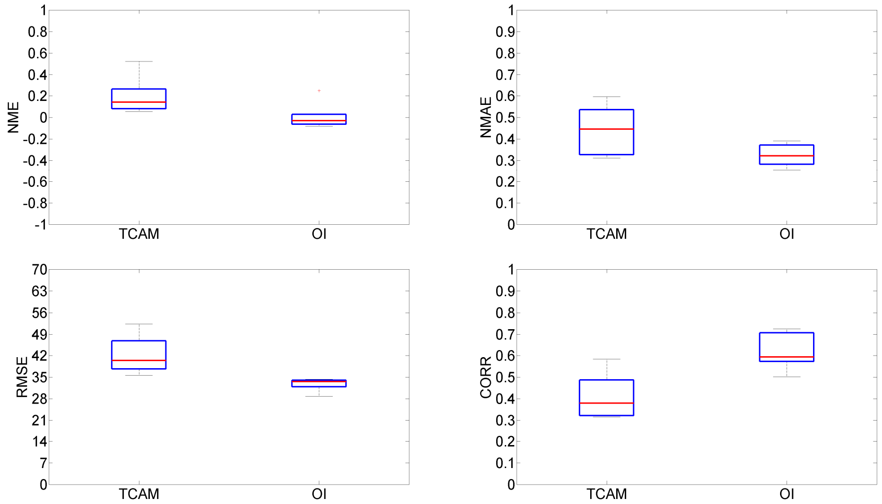

Figure 3. Validation of the impact of O3 data assimilation over the O3 concentrations. Box plots show: NME; NMAE; RMSE; Correlation.

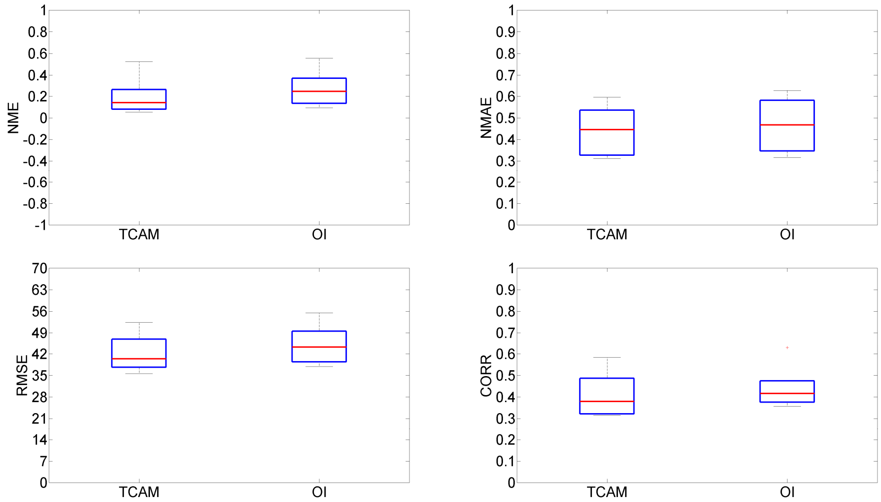

Figure 4. Validation of the impact on the O_3 concentrations due to the assimilation of NO_2 data measured by ground stations. Box plots show: NME; NMAE; RMSE; Correlation.

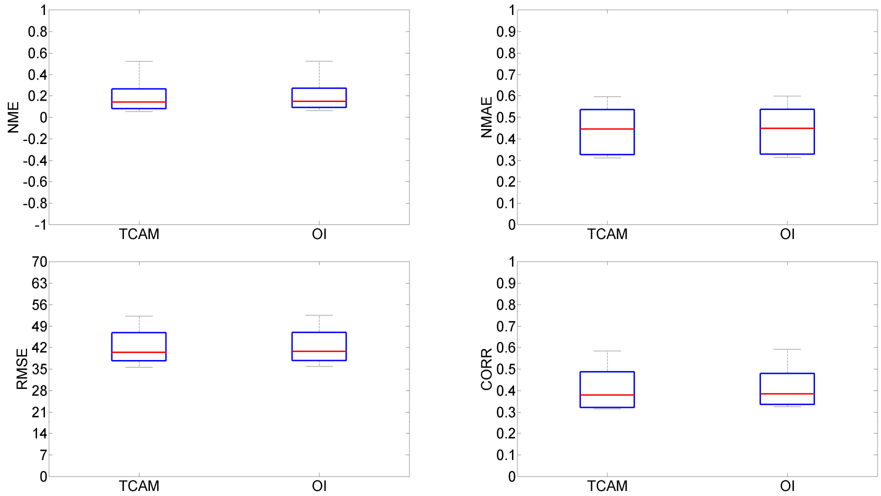

Figure 5. Validation of the impact on the O_3 concentrations due to the assimilation of NO_2 data measured by OMI sensor. Box plots show: NME; NMAE; RMSE; Correlation.

REFERENCES

Carnevale, C., E. Decanini, M. Volta, 2008: Design and validation of a multiphase 3D model to simulate tropospheric pollution. *Science of the Total Environment*, vol. 390, p. 166-176, ISSN: 0048-9697, doi: 10.1016/j.scitotenv.2007.09.017.

Carnevale, C., V. Gabusi, and M. Volta, 2006: POEMPM: an emission model for secondary pollution control scenarios, *Environmental Modelling and Software*, vol. 21, pp. 320–329.

Denby, B., Horlek, J., S. E. Walker, K.E., Fiala, J.: Interpolation and assimilation methods for Europe scale air quality assessment and mapping. Part I: review and recommendations. Technical report, ETC/ACC Technical Paper 2005/7 (2005)

Deserti, M., E. Minguzzi, M. Stortini, S. Bande, E. Angelino, M. Costa, G. Fossati, E. Peroni, G. Pession, F. Dalan, S. Pillon, C. Carnevale, G. Finzi, E. Pisoni, G. Pirovano, and M. Bedogni, 2009: A performance evaluation of chemical transport models in the Po' Valley, Italy, in *Proceedings of 7th International Conference in Air Quality*.

Grell, G., J. Dudhia, and D. Stauffer, 1994: A description of the Fifth generation Penn State/NCAR Mesoscale Model (MM5), tech. rep., NCAR Tech Note TN-398 + STR. 122 pp.

Kalnay, E., 2003: Atmospheric modelling, data assimilation and predictability, Cambridge University press.

Schmidt, H., C. Derognat, R. Vautard, and M. Beekmann, 2001: A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe, *Atmospheric Environment*.

Volta, M. and G. Finzi, GAMES, 2006: a comprehensive Gas Aerosol Modeling Evaluation System, *Environmental Modelling and Software*, vol. 21, pp. 587–594.