J3.4

DATA ASSIMILATION WITH A MULTISCALE CHEMISTRY-TRANSPORT FORECAST MODEL

Blond N.*, Bel L.**, Vautard R.
Laboratoire de météorologie dynamique, France, Palaiseau

1. INTRODUCTION

To study and understand phenomena responsible for high pollution episodes, observations cannot be use as only source of information because they are most of the time disperse over the studied area. That makes difficult the diagnosis, precisely where there are no stations. Moreover, this present study evaluates a simple interpolation of ozone observations, collected during summer 1999 over Ile de France area. Results show that such a method cannot find out a urban plume if there is no station to locate it (see section 6).

A large variety of numerical models designed to study tropospheric air pollution episode exists and can help to estimate air pollution state between stations. They cover various spatial scales and show reasonable simulation skills. However a lot of phenomena are still badly understood and then, some physic or chemistry parameterisations need to be constantly improved to avoid unacceptable errors. Then, in order to produce an accurate image of the true state of system, observational information needs to be introduced in the model, for instance, by interpolating differences between observations and simulations at station locations (i.e. innovations). The result of the process, called analysis, can be used in itself as a comprehensive and self-consistent diagnosis of the state of the atmospheric pollution. Moreover it can be useful as the initial state for a numerical forecast models.

This study focuses on the use of observational data of ozone to analyse ozone field coming from the multi-scale chemistry transport model CHIMERE over summer 1999, and to re-initialise the forecast model with analyses. Indeed as CHIMERE is ultimately intended for operational forecasting and long-term simulations, the computational effort has to be kept low. Then a series of analysis experiments have been performed at regional scale over Ile de France area using low computational analysis methods: Statistical Interpolation (Daley, 1991) and Kriging (Cressie, 1993). At this time, a more elaborate assimilation process, for example 4D-Var (Elbern, 2001) can not be used in real time. The first method is borrowed from meteorology. It consists in correcting model simulation at every grid point using a linear combination of innovations. A new anisotropic and inhomogeneous estimation of background error covariances is presented and evaluated with a leave–one-out method in Section 4. Moreover the method is compared to a classic homogeneous and isotropic process. Surface ozone observations and aircraft measurements (see section 3) collected during ESQUIF experiment (Menut L. and al., 2000) are presented.

The second relates to the Kriging technique widely used in mining science. The method is close to the previous one, but uses only observations. However, Kriging technique is also adapted in this study to interpolate innovations. Then, main differences between the Statistical interpolation method and the Kriging technique on innovations are that the estimation of background error covariance is calculated for a summer with a function of the ozone concentration covariance itself for the first one, and every day with a function of the distance between two grid points for the last one. The Kriging technique is described in section 5, and a comparison with the Statistical interpolation is given in Section 6.

First, a short description of the model is given.

2. THE MULTI-SCALE MODEL : CHIMERE

CHIMERE is a simplified 3D eulerian chemistry-transport model. It is extended to a multi-scale version, with a continental domain (see Figure 6) at low horizontal resolution (0.5°, so about 50km) and presently two sub domains : The Ile-de-France (1.3°W ; 48.1°N ; 3.3°E ; 49.4°N) and the Alsace (6.6°W ; 47.3°N ; 8.3°E ; 49.2°N) areas. The extension of the model centred on Paris is presented on Figure 1. These two regional versions of the model have a better horizontal resolution of few kilometers (6km for Ile-de-France and 4km for Alsace). They are forced at boundaries by the continental model with a nesting one way. The model has five vertical layers, going from surface to about three thousand meters above ground, so it encloses the boundary layer in anticyclonic conditions. The emissions are derived from the EMEP annual totals (1997), modulated in time and VOC speciation by GENEMIS profiles (1994). For the sub domains, emissions data were provided to us by the two air quality networks. The land-use data are derived from the RIVM data basis.

The chemical mechanism is adapted from the original scheme MELCHIOR (Beekmann and Lattuati, 2000). It describes 116 reactions of 44 gaseous species. Advection is performed by the PPM (Piecewise Parabolic Method) 3D order scheme. Vertical mixing is

* *Corresponding author address:* Blond N., Ecole Polytechnique, Laboratoire de météorologie dynamique, Palaiseau, France; e-mail: blond@lmdx.polytechnique.fr.

** *Corresponding author address:* Bel l., Univ. de Paris-Sud, Probabilités, Statistique et Modélisation, Orsay, France; e-mail: Liliane.Bel@math.u-psud.fr .

parameterised by a diffusion depending only on the height of the boundary layer, which is calculated from Richardson number profiles. Photolytic rates are attenuated as a function of cloudiness. Boundary concentrations are prescribed for fourteen species relevant for photoxidant formation and with longer lifetime, using a climatology monthly mean data from the global MOZART CTM (Hauglustaine et al. 1998) at continental scale, and using CHIMERE continental outputs at regional scale. The numerical solver is the TWOSTEP method. Meteorological first guess data come from the European Center for Medium Range Weather Forecast (ECMWF). The study focuses on the simulation of summer 1999. The temporal resolution of the meteorological data used, is of six hour.

The model is simplified enough to allow long simulations or real-time forecast on a simple workstation, but realistic enough to allow quantitative simulations of a various scale ozone distribution.

3. OBSERVATIONAL DATA BASE

3.1 *Surface observations*

At regional scale, ozone surface observations come from AIRPARIF, see web page: http://www.airparif.asso.fr/english/reseau/default.htm.
The distribution of measurements sites is depicted in Figure 1, by symbolic discrimination in terms of prevalent local chemical regime, either rural, urban, or urban background.

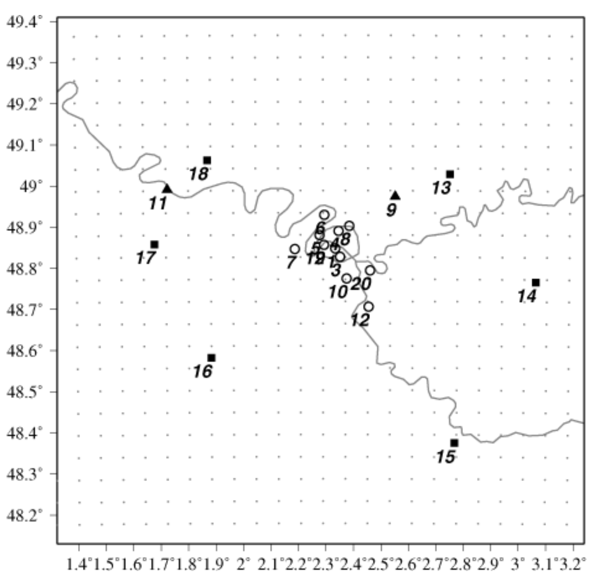

Fig. 1. *Model domain of 25x25 grids. Paris is in the center of the map (represented by a circle line). Open circles, shaded squares and triangle denote urban, rural, and urban background surface observation sites, respectively. The names of these stations : 1. Paris-6, 2. Paris-7, 3. Paris-13, 4. Paris-18, 5. Neuilly-sur-Seine, 6. Gennevilliers, 7. Garches, 8. Aubervilliers, 9. Tremblay-en-France, 10. Vitry-sur-Seine, 11. Mantes-la-Jolie, 12. Montgeron, 13. Montge-en-Goele, 14. Saints, 15. Fontainebleau, 16. Rambouillet, 17. Prunay, 18. Frémainville, 19. Tour-Eiffel1, 20. Créteil-église.*

At continental scale, surface observations used, come from several air quality networks over Western Europe. A list of these organisation is given on web page:
http://euler.polytechnique.fr/pioneer/analyses/index.html.

Most of surface ozone measurement are introduced in the first level of the model.

3.2 *Aircraft measurements*

In the framework of ESQUIF project (Etude et simulation de la Qualité de l'air en Ile de France), which was designed to improve the understanding of photochemical pollution events in the Paris region, a large number of observation flights was performed in order to study the chemical composition around the agglomeration. For more information, see page web: http://moebius.polytechnique.fr/~esquif .
These measurements are not use in analyses processes, but permit to evaluate their skill in altitude.

4. STATISTICAL ANALYSIS METHOD

4.1 *Description*

The analysis algorithm enters within the general framework of Daley 1991. The aim is to obtain a realistic estimation of a system state using different information sources. The first guess is represented by the state vector $\mathbf{x}^f$. In the present study, this first guess is the surface ozone field issued from the model CHIMERE. The second information comes from the observed values, gathered into an observation vector $\mathbf{y}^o$. The estimator is denoted $\mathbf{x}^a$.

The Statistical Interpolation method consists in estimating the true state by correcting the first guess value by a linear combination of the innovations. The estimator is written :

$$\mathbf{x^a} = \mathbf{x^f} + \mathbf{PH}^{\mathbf{T}}(\mathbf{HPH}^{\mathbf{T}} + \mathbf{R})^{-1}(\mathbf{y^o} - \mathbf{y^f})$$

where **P** and **R** are the background and the observation error covariance matrix, respectively. **H** is a linear operator from model state space of dimension N (i=1,N) to observation space of dimension K (k=1,K).
Background and the observational biases are retrieved to obtain an unbiased estimation. Hence distinct measurements are affected by physically independent errors, observational errors are assumed to be not mutually correlated. The causes of errors in the background and in the observations are supposed to be completely independent, then observational

errors are also assumed to be not correlated with first guess errors.
Thus the key point to data assimilation is the estimation of biases, and determination of the matrix **R** and **P**. The study of the vector of departures $\mathbf{y}^{o}$ - $\mathbf{x}^{f}$ at observations points provides an important information about them.

The average of innovations gives a global information about biases at stations location. Then it is necessary to decide which part of the bias is from the observation and which part is from the model, in order to estimate model bias all over the analysis domain. Instrumental bias are assumed to be zero.
At regional scale, observational bias of representativness is assumed to be zero because an observation is a hourly mean data, then such an observation is rather representative of about 4-7km around for a wind flow of about 1-2m.s^{-1}. The resolution of the regional version of the model is 6km. This is why the global bias is assumed to be only the model bias which is taken as a linear function of the mean ozone concentration given by the model. The model tends to overestimate ozone concentration on rural area and tends to underestimate the pollutant on Paris area. At continental scale, only rural station are used to avoid biases due to representativness.

The background error covariances are used to be determined by first plotting the correlation of each station pair (l, k) on scatter diagram as a function of the absolute distance between stations l and k. Such an estimation assumes space isotropy, which is not appropriate to study pollution atmospheric phenomena, controlled by emissions. This is why a second method is performed. It consists in estimating the background error correlation according to the correlation of observational ozone concentration themselves. The error variances are modelled as a function of observational concentration variances. This assimilation process avoids homogeneous and isotropic hypothesis. The process is also used to correct the model in altitude, but every day, it is stopped at atmospheric boundary layer height, because the statistic study of error correlation, made over summer 1999, do not take into account dynamic evolution from a day to another.

4.2 *Validation at regional scale*

Using ground-based measurements. In order to objectively estimate the representation of pollution state improvement and test the different analysis processes quality, a leave-one-out method is used. It means that sequentially, observations of a single station are omitted in the assimilation process, but are compared to ozone analysis, to estimate its skill. It turns out a reduction of the standard deviation of about 40% after an analysis process on each station of Ile de France area, in comparison with a forecast without assimilation (see section 6, Figure 7). Figure 2, presents standard deviation versus time, for a rural station (Prunay) and an urban station (Vitry). Some differences appear from 16UTC to the night until 12UTC between the homogeneous and isotropic analyses process and the inhomogeneous and anisotropic method. During these times, the later process permits a RMS reduction in comparison with the homogeneous and isotropic case. The difference could reach more than 15%. At 15UTC there is no significant difference.

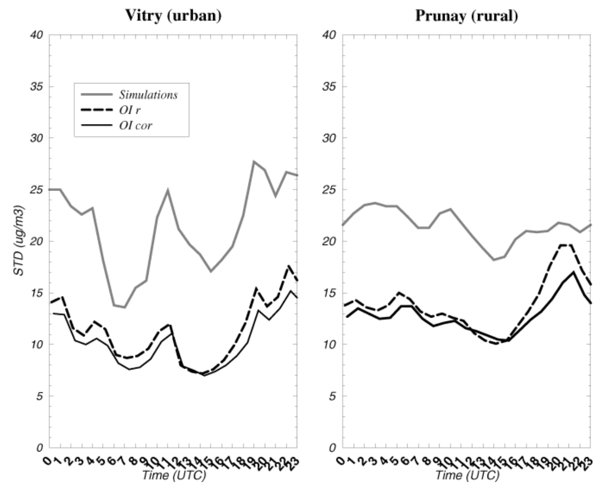

Fig. 2. *STandard Deviation (STD - differences between simulation and observation in gray line, between analysis and observation in black line : dashed line for the isotropic OI, and solid line for the anisotropic OI) versus time, for two stations (Vitry, an urban station and Prunay a rural station), calculated for the period of May 1999 to September 1999.*

Two phenomena can explain this sort of results. The first is that during night time, the atmosphere is stratified because of temperature inversion. Some pollutants could be blocked in small structures, so that the atmosphere is inhomogeneous. The second one relates to emissions. Indeed, emissions break space homogeneity. Emission hours correspond to the intervals 5-11h and 16-19h. This is why, an estimation of **P** with background field correlations is more appropriate during the night and emission hours.

Using aircraft measurements. These observations are not introduced in the analysis process but only used to evaluate its performance in upper level. Figure 3 and Figure 4 present the comparison between measurements of DIMONA aircraft and analysis of ozone field at 14UTC calculated by CHIMERE, for the 16th of July 1999. While the observations show the urban plume eastward, CHIMERE simulates it south-eastward. Analysis, using only surface measurements does not change this spatial distribution in altitude (Figure 3).
Nevertheless, it decreases underestimations of about 7 ppb out of the urban plume (Figure 4).

Figure 5 shows results of a comparison between aircraft measurements of ozone and simulations or analyses for 12 flight over Paris area (8 DIMONA and 4 ARAT aircraft). A reduction of the root mean square is observed, after an analysis, for most of them.

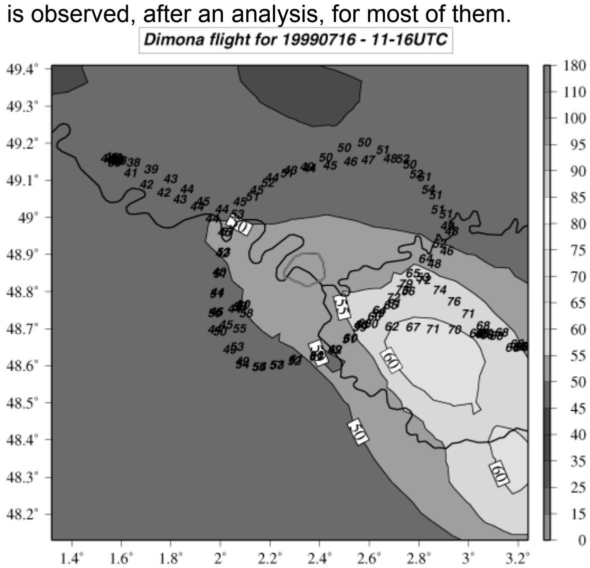

Fig. 3. Measurements of DIMONA aircraft (11.35-15.63 UTC) and analysis of ozone field at 14UTC calculated by CHIMERE, for the 16th of July 1999.

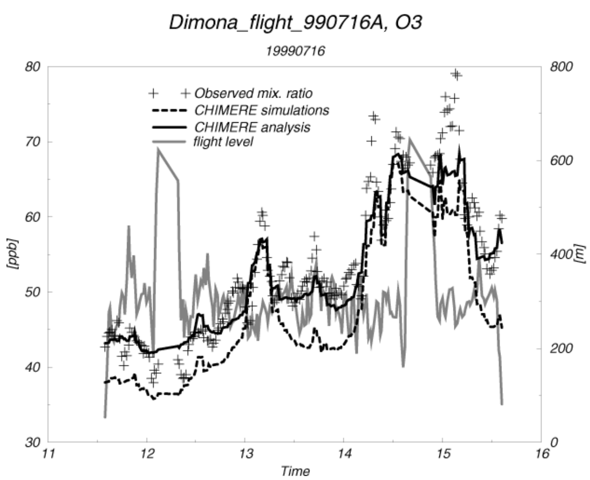

Fig. 4. Dimona aircraft measurements (black crosses), compared to simulated (black long-dashed line) and analysis (black solid line) of ozone concentrations (in ppb), calculated by CHIMERE for the 16th of July 1999. The gray line indicates aircraft altitude.

4.3 ***Validation at continental scale***

A leave-one-out method is also applied at continental scale. Figure 6 represents root mean square of difference between surface measurements of ozone and simulated or analysis field calculated by CHIMERE for the period of May 1999 to September 1999. It turns out that the best reduction (about 40%) of the RMS after an analysis is bigger where there is a lot of station.

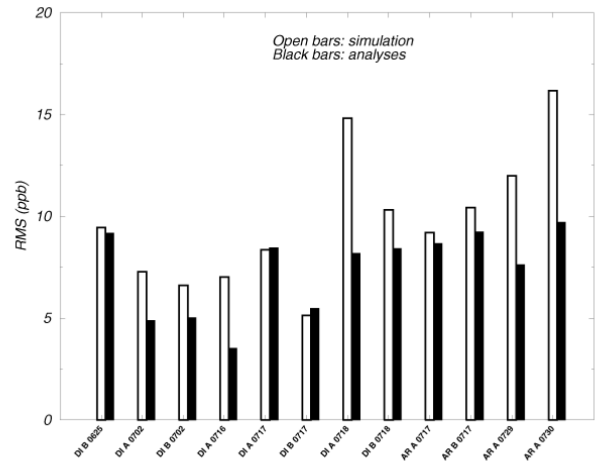

Fig. 5. Root mean square of differences between aircraft measurements of ozone and simulated (open bars) or analysis (black bars) calculated by CHIMERE for 12 flight over Paris area. DI correspond to DIMONA. AR to ARAT. A for morning flight and B for flight in afternoon. DI A 0716 correspond to the DIMONA aircraft of the afternoon of the 16th of July.

5. KRIGING TECHNIQUE

Let the space-time model be :

$$y_t^o(s_k) = x_t^f(s_k) + W_t(s_k) + \varepsilon_t(s_k)$$

where s is the location, t is the time, $x_t^f(s_k) = x_{k,t}^f$ is the model value at location s_k and time t, W is a random process representing the dependant scale variation and ε is a zero-mean white noise process independent of W, representing the measurement error. It is assumed that there is no correlation on time. For each day, an ordinary kriging is performed on the model :

$$\tilde{Z}_{k,t} = y_{k,t}^o - x_{k,t}^f = W_{k,t} + \varepsilon_{k,t}$$

where $(Y^o{}_{k,t})_{k=1,K}$ are the K observed values, and $(\tilde{Z}_{k,t})_{k=1,K}$ are the innovations. The estimation at location i is given by:

$$x_{i,t}^a = x_{i,t}^f + \sum_{k=1}^{K} \lambda_{i,t}^k \tilde{Z}_{k,t}$$

$\lambda_{i,t}^k$ are the kriging weights for time t. This last equation is the same equation, from which equation written in section 4.1, comes. The difference lies in the dependence of weights with the time. $\tilde{Z}_t$ is assumed to be an intrinsically stationary process with constant mean and isotropic. Hence

$$\gamma_t(h) = \frac{1}{2}E(\tilde{Z}_t(s+h) - \tilde{Z}_t(s))^2$$

is a function of the distance h.

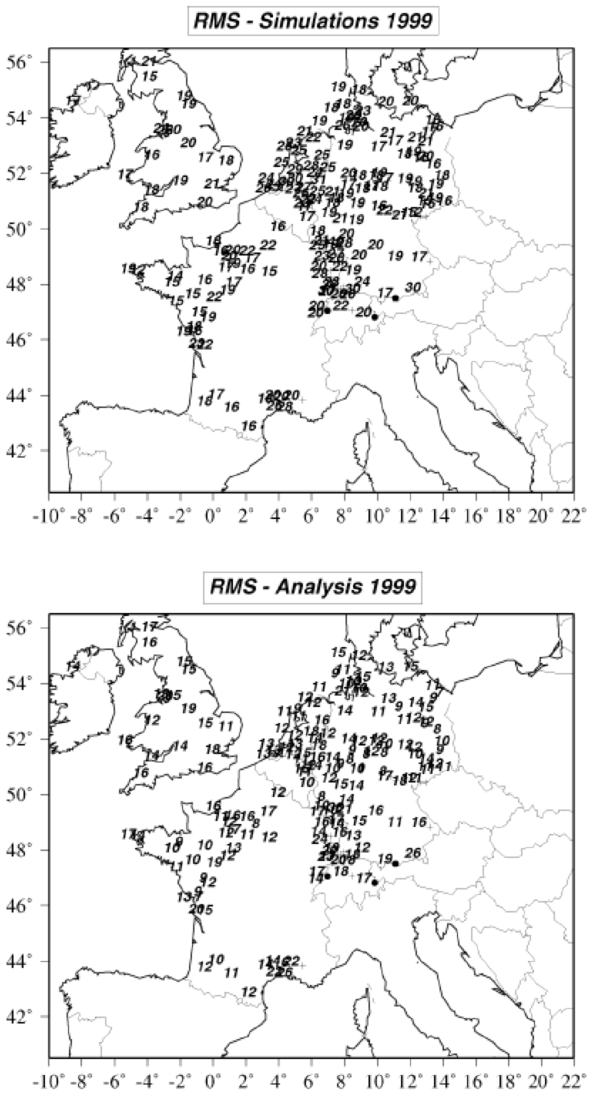

Fig. 6. Root mean square (μg/m³) of differences between surface measurements of ozone and simulated or analysis field calculated by CHIMERE for the period of May 1999 to September 1999. Cross correspond to observations introduced in the second level of the model, black dots are in the third one, and no symbol in the thirst level.

The kriging weights $\lambda_{i,t}^k$ are solution of the linear system (Cressie, 1993) :

$$\Gamma_t \lambda_{i,t} = \gamma_{i,t}$$

with

$$\Gamma_t(k,l) = \begin{cases} \gamma_t(s_k - s_l) & if\ k = 1, K\ and\ l = 1, K \\ 1 & if\ k = K+1\ and\ l = 1, K \\ 1 & if\ k = 1, K\ and\ l = K+1 \\ 0 & if\ k = K+1\ and\ l = K+1 \end{cases}$$

$$\lambda_{i,t} = \begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda^K \\ \mu \end{pmatrix} \qquad \gamma_{i,t} = \begin{pmatrix} \gamma_t(s_i - s_1) \\ \vdots \\ \gamma_t(s_i - s_K) \\ 1 \end{pmatrix}$$

μ is a Lagrange coefficient to obtain an unbiased estimation. Weights $\lambda_{i,t}$ have to be calculated every day, and consequently the variogram $\gamma_t(h)$ has to be estimated every day. The estimator of γ_t is given by

$$2\gamma(h_j) = \frac{1}{|N_j|}\sum_{N_j}(\tilde{Z}_t(s_k) - \tilde{Z}_t(s_l))^2 \qquad j = 1, \cdots, J$$

with

$$N_j = \left\{(s_k, s_l); \| s_k - s_l| - (2j-1)\frac{L}{2}| < \frac{L}{2} + tol\right\}$$

$$h_j = \frac{1}{|N_j|}\sum_{N_j}|s_k - s_l|$$

The value of the length interval L is L=15 km, the value of the tolerance tol is tol=10 km. J=7 is the number of intervals. The model for the variogram is the exponential one :

$$\gamma_t(h) = c_t(1 - \exp(-\frac{h}{a_t})) + nug$$

Parameters c_t and a_t are estimated each day. The error measurement is assumed to be about 10 $\mu g.m^{-3}$, then the tce nugget effect equals:

$$nug = \frac{\sigma_\varepsilon^2}{2} = 50$$

	Mean	Per25	Med.	Per75	min	max
a	1393	4	15	45	0.0056	17353
c	6437	59	141	445	0	233860

Table 1. *Statistics on parameters a and c of variogram model, calculated over summer 1999.*

Table 1 gives statistics on parameters a and c, calculated over summer 1999. It turns out that they vary on several order of magnitude. When a is large, the semivariance varies as a linear function with h. When a is low, the semivariance varies rapidly near h=0 and then is a constant with h.

6. COMPARISONS

Figure 7 shows the comparison between processes described above. It turns out that in comparison with a Kriging on innovation, a Kriging on only observations is more efficient on urban stations (first picture). Indeed urban stations are nearest from each other and are in largest number than rural stations. Thus the model is not useful to interpolate their observations, and can only bring additional errors.

Moreover the optimal interpolation gives result as good as a Kriging on observations over urban area (second picture). Then differences between a kriging on innovations and OI (third figure) are probably due to the statistic way used in the case of OI. This process learnt systematic error correlation (apart from biases because they have been retrieved) over a long period. Assumption of no correlations in time, made in the case of a Kriging on innovations is probably not adapted. Differences are not due to the isotropic or anisotropic approaches because at 15UTC they give same results (see section 4.2).

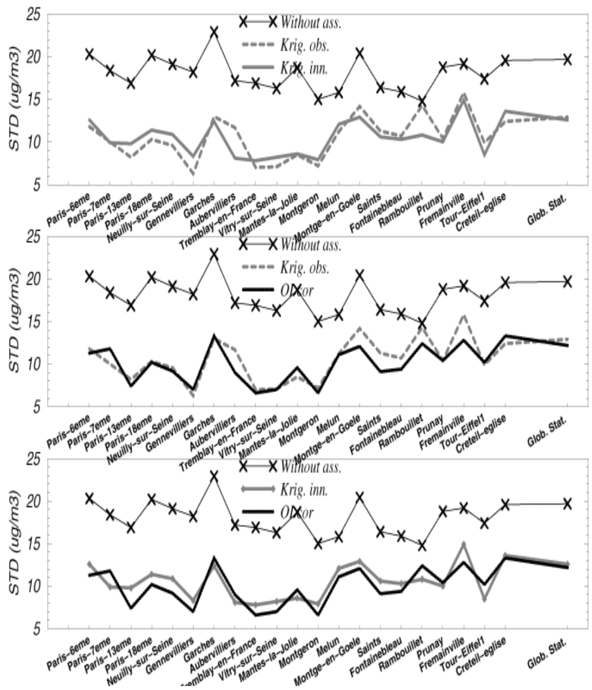

Fig. 7. *Standard deviation (differences between simulation and observation in black line, between analysis and observations in colored lines : black solid line with crosses for Kriging on observations, a long-dashed gray line for a Kriging on innovation, and black dot line for an anisotropic interpolation) at 15UTC, for all stations of Paris area, calculated for the period of May 1999 to September 1999.*

On the other hand, the model is useful to catch the urban plume structure. If observations on Rambouillet station (only one in South-West of Ile de France area – see Figure 1) is not available, how can the analysis find out urban plumes if the wind comes from North-East? As shows Figure 7, it is a difficult exercise. OI improves the standard deviation on rural and isolated stations.

7. REAL FORECAST EXPERIMENTS

Since summer 2001, a web site (http://euler.polytechnique.fr/pioneer/) displays real time ozone forecasts using the multi-scale chemistry-transport models, described in section 2. Statistical interpolation method, described also above, is used to initialise the model with analyses at 15UTC the day before of the forecast, and this, every day. This experiment enters within the general project PIONEER (Prévisibilité et Incertitude de l'Ozone a l'Echelle Européenne et Régionale).

Analyses do not improved forecast more than 2 days after and the improvement is too little to be significant. A reduction of the RMS of about 1-3 μg/m for the next day of the forecast day is observed.

8. CONCLUSIONS AND PERPECTIVES

A series of analyses is performed with the multi-scale chemistry-transport model CHIMERE. Each developed process allows a reduction of about 40% in terms of RMS of differences between simulations and observations. A leave-one-out method is used. A Kriging on observations cannot find out a urban plume if there is no stations to locate it. A statistical interpolation is more efficient than a kriging on innovations, probably because the hypothesis of no correlation in time, made for the second method is not appropriated. The real time experiment performed during summer 2001 showed few improvement of the analysis on the forecast. Taking into account precursors of ozone in the analysis process should improve results.

Daley, R., 1991: Atmospheric data analysis. Cambridge Univ. Press, New York.

Beekmann M., Lattuati M., 2000: Development of the detailed gas phase mechanism MELCHIOR for photooxidant modelling and comparison with EMEP and RACM mechanism, *submitted to Atmospheric Environment.*

Cressie, N. A. C., 1993. Statistics for Spatial Data (revised edition), Wiley, New York, NY.

Elbern, H., Schmidt H., 2001: Ozone episode analysis by four-dimensional variational chemistry data assimilation. J*. of Geophysical Research*, **106**, 3569-3590.

Hauglustaine D. A., Brasseur G. P., Walters S., Rasch P. J., Müller J.-F., Emmons L. K., Carroll M. A., 1998: MOZART: A global chemical transport model for ozone and related chemical tracers, 2. Model results and evaluation. *J. of Geophysical Research* **103**, 28291-28336.

Menut, L. and al., 2000: Radar attenuation by wet ice spheres. *Ann. Geophysicae*, **18**, 1467-1481.

GENEMIS project, 1994, *EUROTRAC annual report 1993*, part 5, EUROTRAC international scientific secreteriat, Garmisch-Partenkirchen, Germany.