Identification of chemical species emission and sensitivity analysis with variational assimilation of restricted artificial observations

S. Massart¹, J.P. Argaud², A. Pia centini³, O. Thual³

¹ CNRS, France.
² EDF R&D, France.
³ CERFACS, France.

Abstract

Recent developments in the field of chemical data assimilation have led to a variational analysis where the results are generally used as the initial condition for forecasting. Such analysis method is able to adjust other model parameters like emissions which are not available with conventional techniques. Using a simplified one-dimensional urban chemistry transport model, this paper presents the identification of the ground flux pollutants with data assimilation twin experiments. It also describes the initial effort in the understanding of the influence of each observed species on the quality of a two-dimensional variational analysis system.

1 Introduction

Since emission credits is one solution to reduce greenhouse gas emissions, there is a need to understand the distribution of pollutant sources in each country. But it remains difficult to measure ground flux pollutants directly. Nevertheless, the urban networks of pollutant concentration measurements have been developed since the growing interest in air quality. In this context, numerical modeling appears as a unique potential tool to rebuild the distribution of emissions from a set of measured concentrations. Given a numerical model, the approach consist on researching the set of emissions which, once integrated, best fit the measurements. This method requires to know the mechanism of the generation of pollutants due to ground fluxes. Thus Trevino and Méndez [5] have tried to understand and explain the process of ozone molecule production in the lower atmosphere with a reduced
set of chemical reaction, including the presence of hydrocarbon and NO\textsubscript{x} sources.

When the chemistry transport is physically acceptable, the computation of the pollutant fluxes can be done by discretizing the space domain and giving the type of source emission for each grid point (like road, factory, countryside, ...). Some authors then propose to associate each type of source with time evolution values. For example, Salles [2] worked on a road traffic emission source inventory model to estimate as precisely as possible precursor species fluxes which are used in ozone chemical mechanism.

Our strategy to assess pollutant fluxes is to use variational assimilation which can be described as the process by which all the available information is used in order to estimate as accurately as possible the state of the chemical atmospheric flow. Given a set of observations over a time interval and a numerical model, the principle of variational assimilation is first to define a scalar function $\mathcal{J}$ which, for any solution of the model over every time will measures the distance between that solution and the observations. This objective function depends on the model parameters such as initial conditions, boundary conditions, and extra terms. In our case, one can view it as a function of spatio-temporal emission distributions. Talagrand [3] then suggested to add a term which measures the distance to the background (or prior estimate) which can come from source inventory. Thus, the task of variational analysis is to adjust the parameters in order to minimize $\mathcal{J}$. The determination of the gradient of $\mathcal{J}$ in respect with all of the adjusting parameters is required in many descent algorithms. As explained in detail in Talagrand and Courtier [4], the most economical way of determining the gradient in the case of variational assimilation is to integrate the adjoint equation of the numerical model.

This last step being expensive in terms of implementation time and computational cost, we have chosen a chemical scheme which can simply describe the urban ozone chemical mechanism on a one-dimensional domain, ozone being one of the main common indicators for air pollution. Numerical results obtained with this model are presented in the first part of this paper.

The first purpose of this study is to show what can be expected on improving emissions while measuring concentrations. Then, the role of each observed species played in the analysis step is clarified.

## 2 Modeling a pollution episode

Using a three-dimensional meteorological model coupled with a complex chemistry scheme is so expensive in terms of computational cost that it seems difficult to realize some assimilations with a such model. Our purpose being to realize experiments in various configurations, we have had to reduce the complexity of both the meteorological model and the chemistry scheme. The first reduction was on the resolution in space with a one-dimensional domain in the spatial coordinate $x$, on which the following system equations are integrated over the time $t$:

\[
\frac{\partial C_i}{\partial t} + U \cdot \frac{\partial C_i}{\partial x} = \mathcal{F}_i (C_0, \ldots, C_j, \ldots, C_p) + S_i - V_{i,dep} \cdot C_i \tag{1}
\]
where $C_i(x,t)$ is the concentration of pollutants $i$; $p$ is the number of species; $U(x,t)$ the velocity field; $F_i$ a temperature dependent chemical formation or depletion rate of species $i$; $S_i(x,t)$ the source flux and $V_{i}^{\text{dep}}(x,t)$ the deposition velocity.

### 2.1 The reduced urban chemistry scheme

The chemistry scheme which describes the function $F$ of the system (1) comes from Tulet [6] where it was developed under the constraint of optimal simulation of the urban ozone concentration. The introduction of an hybrid class of hydrocarbons (Non-Methane Hydrocarbon or NMHC) is mainly at the origin of its reduction to 10 species and 17 reactions which are detailed in table 1. This scheme, which not depend on the space dimension $x$, is solved with a Two Step algorithm [7] in order to calculate concentration levels in each grid points before being advected.

#### Table 1: Reduce chemical scheme

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Kinetic constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $\text{NO} + \text{O}_3 \rightarrow \text{NO}_2 + \text{O}_2$</td>
<td>$k_1 = 1.82 \times 10^{-14}$</td>
</tr>
<tr>
<td>2. $\text{CO} + \text{OH} \rightarrow \text{HO}_2$</td>
<td>$k_2 = 2.41 \times 10^{-13}$</td>
</tr>
<tr>
<td>3. $\text{HO}_2 + \text{NO} \rightarrow \text{OH} + \text{NO}_2$</td>
<td>$k_3 = 8.56 \times 10^{-12}$</td>
</tr>
<tr>
<td>4. $\text{NO}_2 + \text{OH} \rightarrow \text{HNO}_3$</td>
<td>$k_4 = 1.14 \times 10^{-11}$</td>
</tr>
<tr>
<td>5. $\text{CH}_4 + \text{OH} \rightarrow \text{RO}_2$</td>
<td>$k_5 = 6.45 \times 10^{-15}$</td>
</tr>
<tr>
<td>6. $\text{NMHC} + \text{OH} \rightarrow \text{RO}_2$</td>
<td>$k_6 = 3.32 \times 10^{-11}$</td>
</tr>
<tr>
<td>7. $\text{R}_2 \text{O}_2 + \text{NO} \rightarrow \text{NO}_2 + \text{HO}_2 + \text{RCHO}$</td>
<td>$k_7 = 9.00 \times 10^{-12}$</td>
</tr>
<tr>
<td>8. $\text{OH} + \text{HO}_2 \rightarrow \text{loss}$</td>
<td>$k_8 = 1.11 \times 10^{-10}$</td>
</tr>
<tr>
<td>9. $\text{HO}_2 + \text{HO}_2 \rightarrow \text{loss}$</td>
<td>$k_9 = 5.45 \times 10^{-12}$</td>
</tr>
<tr>
<td>10. $\text{RCHO} + \text{OH} \rightarrow \text{RCO}_3 + \text{H}_2\text{O}$</td>
<td>$k_{10} = 1.76 \times 10^{-11}$</td>
</tr>
<tr>
<td>11. $\text{NO}_2 + \text{hv} \rightarrow \text{O}_3 + \text{NO}$</td>
<td>$J_{\text{NO}_2} = 1.1 \times 10^{-2}$</td>
</tr>
<tr>
<td>12. $\text{O}_3 + \text{hv} \rightarrow 2.\text{OH}$</td>
<td>$J_{\text{O}_3} = 5.1 \times 10^{-6}$</td>
</tr>
<tr>
<td>13. $\text{RCHO} + \text{hv} \rightarrow \text{CO} + 2.\text{HO}_2$</td>
<td>$J_{\text{CHO}} = 2.4 \times 10^{-5}$</td>
</tr>
<tr>
<td>14. $\text{RCHO} + \text{hv} \rightarrow \text{CO}$</td>
<td>$J_{\text{COH}_2} = 3.6 \times 10^{-5}$</td>
</tr>
<tr>
<td>15. $\text{RCHO} + \text{hv} \rightarrow \text{RO}_2 + \text{CO} + \text{HO}_2$</td>
<td>$J_{\text{CH}_2\text{CHO}} = 5.1 \times 10^{-6}$</td>
</tr>
<tr>
<td>16. $\text{RCHO} + \text{hv} \rightarrow \text{RO}_2 + \text{CO} + \text{HO}_2$</td>
<td>$J_{\text{CHO}_2\text{CHO}} = 4.2 \times 10^{-6}$</td>
</tr>
<tr>
<td>17. $\text{RCHO} + \text{hv} \rightarrow \text{RO}_2 + \text{CO} + \text{HO}_2$</td>
<td>$J_{\text{F}_2\text{CHO}} = 4.2 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

### 2.2 Numerical characteristics

The one-dimensional domain, discretized into 25 grid points, is 86.4 km long in order that, advected with a constant wind of one meter per second, any incoming pollutant crosses the whole domain in a day. It is divided into two parts which are assumed to be a simplified representation of a region with a town from grid points two to ten (about 28 kilometers) surrounded by countryside. We also suppose that this countryside is long enough so that there is no space evolution for concentrations far from the town. This consideration leads us to impose Neumann boundary...
conditions where there is no horizontal flux on the edge of the domain.

Among the differences between a town and a countryside parameterization, we choose to distinguish them only by the rate of ground fluxes of NO, NO$_2$, CO and NMHC. The characterization of the countryside is a homogeneous spatial pollutant emissions with a rate of an half. On the contrary, in the city, fluxes increase on approaching the heart of the town, with a rate of one for the heart ($x = 20.74$km) and for $x \in [6.9km, 34.6km]$:

$$\text{Emission rate} = 0.75 + 0.25 \cdot \cos \left( \frac{\pi \cdot 20.74 - x}{20.74 - 6.9} \right)$$

(2)

Regarding the time evolution of pollutant emission, it is simplified to one value for the night and one value for the day (see table 2).

Table 2: Day and night emission values in molecules.cm$^{-2}$.s$^{-1}$ for the countryside and the heart of the city. The evolution between both is described by equation (2).

<table>
<thead>
<tr>
<th>Species</th>
<th>Heart of the Town</th>
<th>Countryside</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day</td>
<td>Night</td>
</tr>
<tr>
<td>NO</td>
<td>$1.199 \cdot 10^5$</td>
<td>$1.004 \cdot 10^6$</td>
</tr>
<tr>
<td>NO$_2$</td>
<td>$2.995 \cdot 10^4$</td>
<td>$2.460 \cdot 10^5$</td>
</tr>
<tr>
<td>CO</td>
<td>$6.150 \cdot 10^6$</td>
<td>$8.200 \cdot 10^7$</td>
</tr>
<tr>
<td>NMHC</td>
<td>$2.496 \cdot 10^5$</td>
<td>$4.510 \cdot 10^6$</td>
</tr>
</tbody>
</table>

The simulation characteristics and all chemical parameterization is the same each day. Moreover the dynamical field is homogeneous all over the grid and is constant in time. With a given initial condition, the numerical model will reach an equilibrium, which means that the time evolution of each pollutant concentration is the same each day. It also means that the evolution is controlled by the flux pollutants and not by the initial condition. The concentration space-time evolutions of a few species obtained after four days of simulation, are displayed in figure 1.

2.3 Direct simulation interpretation

In spite of the relative simplicity of the model used for the concentration level calculation, few remarks can be drawn from figure 1. The main source of O$_3$ and NO is a photolysis of NO$_2$. So, during the day, NO$_2$ concentration decreases while producing O$_3$ and NO. The last being a fast species, it promptly reacts and does not increase. In contrast, ozone concentration continues to increase all over the day and is slowly transported to the countryside. Close to sunset, the photochemical equilibrium between NO$_2$ and the couple NO, O$_3$ is reversed so that the quantity of NO$_2$ grows and the two others species are destroyed.
In order to perform data assimilation experiments, having the numerical model described in section 2, we need both observations and backgrounds. The first comes from the integration of the numerical model during one day, from reference emissions. Measurements are then constructed by discretizing the time evolution of each species and adding random perturbations which are an image of measurement equipment noises. All the experiments have been realized with an observation time step of one hour.

The urban network of pollutant measurements is growing for some years but it remains difficult to have one measure per grid point if the mesh is small enough to well represent the urban chemistry transport. Nevertheless, we assumed that each grid point have a sensor. So, we defined the reference set of observations when all the species are observed each hour during 24 hours in each grid point.

Emissions prior estimates are constructed by divided by a factor two the reference values, values which were used for the construction of the synthetic measurements. These emissions backgrounds have been integrated and the result which is explained in section 3.2, justify the assimilation with the fact that the reconstructed concentration fields do not fit with observations.

Having a set of 24 hours observations, backgrounds, a numerical model and its adjoint, we have realized a variational assimilation using PALM software [1] which is developed at CERFACS.

The control space, in which the analysis is perform, is a 200 dimensional space (four emitted species with a day and a night value over a twenty five point mesh). The size of the reference observations is 6250 (ten species with a time step of one hour during one day, over the same mesh).

In order to evaluate the distance between background and iterative temporary solutions, it is supplied with a norm which requires the covariance matrix. This last was specified as constant and diagonal with elements proportional to the square of the reference emissions for the background matrix and to the square of the computed concentrations from reference emissions for the observation matrices.
3.2 Sensitivity to emissions

Initial concentrations are obtained by a stabilized simulation using the same emission rates each day, which means that the concentration evolutions of each species are nearly similar from one day to the next. So, when emission ratios are divided by two, with the same initial conditions, in order to compute the evolution from backgrounds, the balance is broken. Two extreme behaviors were then noticed and an example of each one is displayed in figure 2. The first, illustrated by NO, is a fast reaction of concentrations which quickly reach another equilibrium state. This also means that the concentration level of these species is mostly controlled by the emission values. These emissions will thus be easier to identify.

![Figure 2: Relative differences in percentage plotted as a function of space and time between reference evolution and evolution computed from background emissions, for NO and NO₂ concentrations.](image)

The second behavior, illustrated by NO₂, is a slow evolution of the error, all the more for the night evolution. This certainly means that day and especially night emissions of this species will be identified with more difficulty. Nevertheless, the concentration level of one primary species is determined by its emissions but also by other ground fluxes, chemistry scheme, initials and boundaries conditions and so on. The next essential step is thus to assimilate the reference measurements to best understand their impact on both emissions and evolutions.

3.3 Emissions identification

The quality of the source identification using variational assimilation can be first measured by the consistency between the analyzed emissions and the reference ground fluxes. Figure 3 represents the space evolution of the day and night relative differences in percentage for each of the emitted species for the reference experiment.

Let’s first remark that the results of the analysis step depend on the emitted species so it is difficult to draw a common behavior. Nevertheless, except for NO, the night sources are worst identified than day ones. This means that the chemical concentrations are more sensitive to day sources which have more important
values than night. Because of a relative error on analysis of around 40% near the initial error of 50%, NO\textsubscript{2} fluxes are not so well identified, as predicted by the previous section. But assimilation permits to access reference day emissions for NO, CO and hydrocarbons with an error less or about noise measurements. Moreover, being more important, city emissions might have more influence on the concentration levels and would have been better identified. This is not visible except maybe for night NMHC. At last, let's notice that a large error on the analysis of a species could mean that the concentrations levels are not sensitive to this emission which thus could be suppress from control emissions.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Relative differences in percentage plotted as a function of space between reference and analyzed emissions for the reference experiment. The two black vertical lines delimit the town.}
\end{figure}

In order to have a better understanding of the assimilation impact on the spatio-temporal evolution of pollutant concentrations, O\textsubscript{3} observed fields, numerical integrations of the prior estimate and analysis are displayed on figure 4.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Spatio-temporal evolution of O\textsubscript{3} concentration: (a) observed and numerical fields from integration of prior estimate (b) and analyzed emissions (c).}
\end{figure}

Given this reference experiment, the analysis integration fits better with measurements than the background integration which is qualitatively closed to the observations (even if the ozone plume is too spread out in the countryside) but quantitatively false. Moreover, in spite of the added noise on observations, the analyzed evolutions are a very good representation of reference evolutions (see figure 1).
4 Sensitivity analysis to observations

4.1 Description of experiments

In order to understand how each observed species can affect the analysis, the aim of this part is to assimilate new observations sets constructed by keeping successively synthetic measurements of only one of the ten species from the reference observations. The size of the ten sets is now 625, which has to be compared to the 200 from the control space.

Instead of plotting the four day and night emissions from the analysis of each of the ten species observations, the next graphs are obtained for each emitted species, by drawing simultaneous the relative error on the analysis of the ten sets. So, this kind of representation can show which observed species improve best analysis on a specific emission and on another way, which species do not have to be measure whether they deteriorate or have no impact on this analysis.

4.2 Sensitivity for night emissions

The contribution of the assimilation on night emissions identification is similar to the reference case with no improvement on NO and CO. Nevertheless, these experiments have shown that analysis on hydrocarbons emissions is only the results of the observations of its own concentrations, because the assimilation of measurements of all the others species lets emissions to their prior estimates.

NO is quite different because two observed species allowed to reach low error: NO itself and HO2 too (see figure 5). Then three others have a positive impact: in inverse order, ROZ, HO2 and NOZ.

![Figure 5: Relative differences in percentage plotted as a function of space between reference and analyzed night emissions of NO when analysis come from the assimilation of each species observations.](image)

This study shows that there are two night emissions which can be corrected (NO and NMHC) while measuring only six species. There is no need to observe HNO3, O3, CO and RCHO to adjust night emissions.
4.3 Sensitivity for day emissions

For each observed species, the relative error rarely go over the initial error of 50% and in this case, they are closed. Moreover, each observed species improve one or another analysis on day emissions. Generally speaking, all the species are good to be observed in the sense of the global behavior of the assimilation.

Nevertheless, whatever species is observed, the maximum gain in relative error of NO₂ emissions is around 10%. This means that the weight of observations is not sufficient to remove from prior estimates which have then to be chosen cautiously.

![Figure 6: Relative differences in percentage plotted as a function of space between reference and analyzed day emissions of NO when analysis come from the assimilation of each species observations.](image)

A good representation of reference emissions can be obtained for the three last species provided that observations are well chosen. The observed species which give a reasonable analysis with a error less than 25% are four for NO (figure 6) and hydrocarbons and only one the carbon monoxyde with itself.

5 Conclusion

The identification of the pollutant sources has been studied by the means of the assimilation of chemical atmospheric observations. The assimilation technique was a variational method which offer the advantage to use the temporal evolution of the studied phenomenon through a numerical model. This last one was a reduced urban chemistry transport model so that the size of the control space was reasonable and results can be interpreted.

This paper intended to give understanding of the impact of restricted observations on the analyzed emissions. Numerical results, although preliminary, have shown that, even with a great number of observations, the assimilation does not well catch all reference emissions. In particular, NO₂ analyzed emissions and CO night analyzed emissions are closed to their prior estimate values.

The assimilation of one primary species measurements gives the best analysis on emissions. But, even if this is the only way to improve night emissions of hydrocarbons, each observed species acts on at least one emission. Thus, one
can think that the more species are observed, the better analysis are. And the fact that analysed emissions from one observed species is always similar or worst than the analysis from reference observations, should confirm this conclusion. But the generalization remains difficult because of interactions while assimilating simultaneous few species.

Finally, if a sensitivity study tend to prove that concentrations fields mainly depend on specific emissions, this study can help to choose which species can be added to observations of the assimilation process in order to improve these emissions and then simulations.

References


