Lessons from a low-order coupled chemistry meteorology model and applications to a high-dimensional chemical transport model

Jean-Matthieu Haussaire, Marc Bocquet
CEREA, Joint Laboratory École des Ponts ParisTech/EDF R&D, Université Paris-Est, Champs-sur-Marne, France
jean-matthieu.haussaire@enpc.fr

Context

Atmospheric chemistry models are becoming increasingly complex, with multiphasic chemistry, size-resolved particulate matter, and possibly coupled to numerical weather prediction models. In the meantime, data assimilation methods have also become more sophisticated. Hence, it will become increasingly difficult to disentangle the merits of data assimilation schemes, of models, and of their numerical implementation in a successful high-dimensional data assimilation study. That is why we believe that the increasing variety of problems encountered in the field of atmospheric chemistry data assimilation puts forward the need for simple low-order models, albeit complex enough to capture the relevant dynamics, physics and chemistry that could impact the performance of data assimilation schemes.

Low-order model

The photochemistry module is based on the GRS, which consists of seven chemical species meant to represent the atmospheric chemistry of ozone formation from VOC and NO_x emissions.

$$\begin{aligned}
\mathrm{ROC} + \mathrm{h}\nu &\xrightarrow{k_1} \mathrm{RP} + \mathrm{ROC} \\
\mathrm{RP} + \mathrm{NO} &\xrightarrow{k_2} \mathrm{NO_2} \\
\mathrm{NO_2} + \mathrm{h}\nu &\xrightarrow{k_3} \mathrm{NO} + \mathrm{O_3} \\
\mathrm{NO} + \mathrm{O_3} &\xrightarrow{k_4} \mathrm{NO_2} \\
\mathrm{RP} + \mathrm{RP} &\xrightarrow{k_5} \mathrm{RP} \\
\mathrm{RP} + \mathrm{NO_2} &\xrightarrow{2k_6} \mathrm{S(N)GN}
\end{aligned}$$

where ROC = reactive organic compounds;
RP = radical pool;
S(N)GN = stable (non-)gaseous nitrogen product.

We use the quasi-steady-state approximation (QSSA) for the radical pool species RP, which is highly reactive and has the shortest lifetime among all the GRS species.

$$[\mathrm{RP}] = \frac{k_2[\mathrm{NO}] + 2k_6[\mathrm{NO_2}]}{2k_5}\left(\sqrt{1 + \frac{4k_1k_5[\mathrm{ROC}]}{(k_2[\mathrm{NO}] + 2k_6[\mathrm{NO_2}])^2}} - 1\right)$$

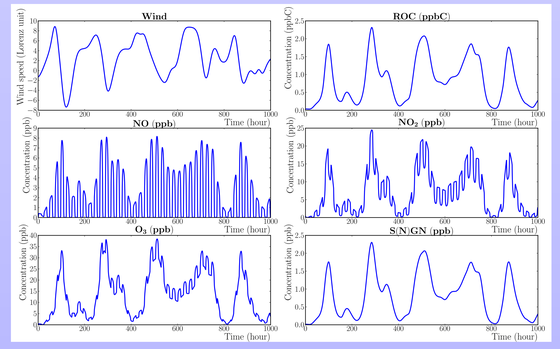
GRS is coupled to the Lorenz-95 model whose 40 state variables, which extend over a mid-latitude circle, are seen as wind speeds and direction that advect the GRS chemical species.

$$\frac{\mathrm{d}x_m}{\mathrm{d}t} = (x_{m+1} - x_{m-2})x_{m-1} - x_m + F, \text{ with } F = 8$$

There is therefore a total of 40 wind variables and 200 concentration variables, defined on the circle using an Arakawa C-grid. The equation for the ozone concentration at a grid point $[\mathrm{O_3}]_{m+\frac{1}{2}}$ is for example

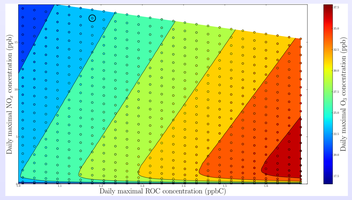
$$\frac{\mathrm{d}[\mathrm{O_3}]_{m+\frac{1}{2}}}{\mathrm{d}t} = \psi_m - \psi_{m+1} - \lambda[\mathrm{O_3}]_{m+\frac{1}{2}} + k_3[\mathrm{NO_2}]_{m+\frac{1}{2}} - k_4[\mathrm{NO}]_{m+\frac{1}{2}}[\mathrm{O_3}]_{m+\frac{1}{2}}$$

with $\psi_m = x_m[\mathrm{O_3}]_{m-\frac{1}{2}}$ if $x_m \geq 0$, $= x_m[\mathrm{O_3}]_{m+\frac{1}{2}}$ if $x_m < 0$. and where λ is a scavenging ratio.



Time evolution of the L95-GRS variables at one grid point. The L95 variables, flagged "Wind", are shown with the original Lorenz unit, while the concentration unit is ppb (ppbC for ROC).

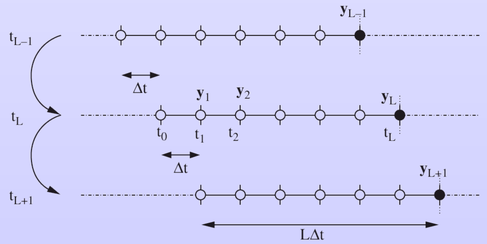
Ozone isopleths



This model, even if not chaotic, is highly nonlinear, exhibiting distinct chemical regimes. The figure opposite represents ozone isopleths for different mean ROC and NO_x concentrations. This feature is typical of lower troposphere ozone chemistry.

Iterative ensemble Kalman smoother

This algorithm estimates the state at the time t_0, L time-steps in the past from present time t_L. Each observation is assimilated a single time.



This smoother minimizes at each cycle

$$\mathcal{J}(\mathbf{w}) = \frac{1}{2}(N-1)\mathbf{w}^{\mathrm{T}}\mathbf{w} + \frac{1}{2}||\mathbf{y}_L - H_L \circ \mathcal{M}_{L\leftarrow 0}\left(\bar{\mathbf{x}}_0 + \mathbf{A}_0\mathbf{w}\right)||^2_{\mathbf{R}_L^{-1}}$$

with
- $\mathbf{w}$ state vector in ensemble space ;
- N the ensemble size ;
- $\mathbf{y}_L$ the observation at time t_L;
- H_L the observation operator at time t_L ;
- $\mathcal{M}_{L\leftarrow 0}$ the model propagating the state from t_0 to t_L ;
- $\bar{\mathbf{x}}_0$ the average of the prior ensemble ;
- $\mathbf{A}_0$ the anomaly matrix ;
- $\mathbf{R}_L$ the observation error covariance matrix at time t_L;

To avoid optimizing the inflation coefficient, we use the "finite-size" scheme by changing the cost function as such

$$\frac{1}{2}(N-1)\mathbf{w}^T\mathbf{w} \rightarrow \frac{N}{2}\ln(\varepsilon_N + \mathbf{w}^T\mathbf{w})$$

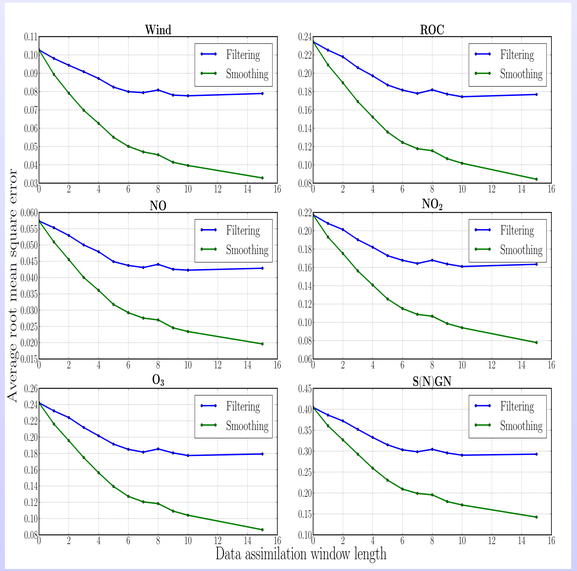
- Unlike 4D-Var, no need for the full adjoint model; ☺
- Like the EnKF, propagation of uncertainties; ☺
- Like 4D-Var, nonlinear variational analysis; ☺
- Like the EnKF, sampling errors of the ensemble and need to use inflation and/or localization. ☹

References

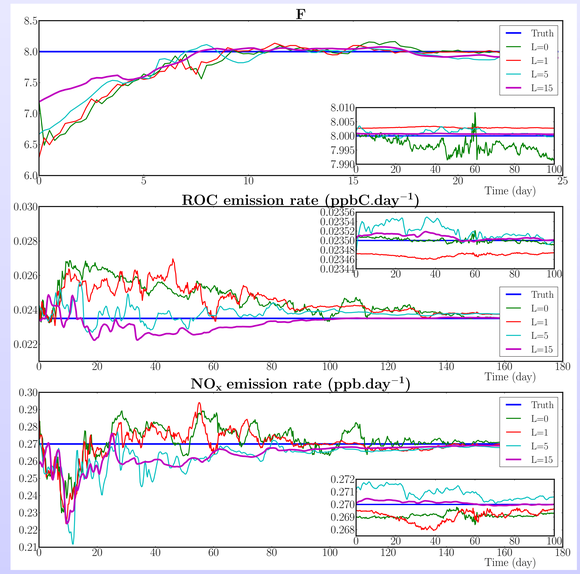
Haussaire, J.-M. and Bocquet, M.: A low-order coupled chemistry meteorology model for testing online and offline data assimilation schemes: L95-GRS (v1.0), Geosci. Model Dev., 9, 393-412, 2016.
Bocquet, M. and Sakov, P.: Joint state and parameter estimation with an iterative ensemble Kalman smoother, Nonlin. Processes Geophys., 20, 803–818, 2013.

Results

A twin experiment was conducted on this model with the SDA IEnKS-N. One grid point out of five is observed every 6 hours during 100 000 data assimilation cycles. The forcing parameter of the wind field and the two emission fluxes were also estimated.



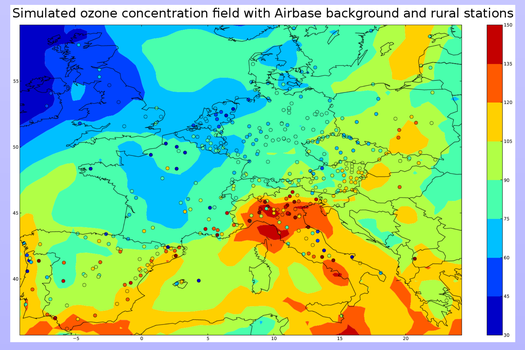
Average filtering and smoothing analysis RMSEs of the L95-GRS variables, as a function of the DAW length (in units of Δt = 6h). The RMSEs are normalized by the standard deviations of the observation error.

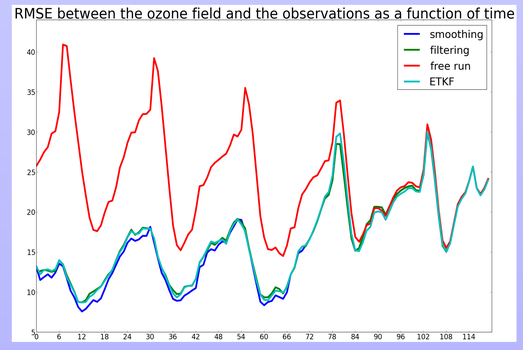


Time evolution (days) of the parameter variables for several DAW lengths without spin up (main) or after a long time (inset). F is shown with the original Lorenz unit, while the emission rate unit is ppbC.day^{-1} or ppb.day^{-1}

Preliminary results with a CTM

Building on the results of this low-order model, we are testing the localized IEnKS with $L = 1$ on a state-of-the-art chemical transport model Polyphemus. We assimilate hourly ozone concentrations from background and rural stations of the Airbase network over Europe and estimate the ozone concentration field. The assimilation period spans 72h, followed by 48h of forecast.





The IEnKS performs as well as the ETKF when it comes to filtering and the smoothing improves noticeably the score. This first result is promising even though the localization of the IEnKS can be troublesome and an (offline) CTM implies a lot of model error. Moreover, the benefit for forecasting seems arguable. However, we hope to get a stronger impact of the IEnKS when it will come to parameter estimation.