A Comparison Study of Data Assimilation Algorithms for Ozone Forecasts

Lin Wu[1,2], Vivien Mallet[2,1], Marc Bocquet[1,2] and Bruno Sportisse[2,1]

[1] CEREA (joint laboratory ENPC - EDF R&D), Université Paris-Est, France
[2] INRIA, CLIME project-team (joint with ENPC), France

Lin.Wu@cerea.enpc.fr

There is a recent diffusion of data assimilation expertise from numerical weather prediction (NWP) to air quality community. However, the atmospheric chemistry-transport models (CTM) are stiff but stable systems; the perturbations on initial conditions tend to be smoothed out rather than amplified. Therefore the conclusions from meteorological experiences cannot be applied directly. We perform a comparison study of assimilation algorithms [Wu et al., 2008]. Hopefully this could serve as a base point for the design of assimilation algorithms suitable for one-day ozone forecasts in realistic applications.

Algorithms

Chemistry-Transport Equation for Air Quality Model

$$\frac{\partial c_i}{\partial t} = \underbrace{-\mathrm{div}(Vc_i)}_{\text{advection}} + \underbrace{\mathrm{div}\left(\rho K \nabla \frac{c_i}{\rho}\right)}_{\text{diffusion}} + \underbrace{\chi_i(\mathbf{c})}_{\text{chemistry}} + \underbrace{S_i - L_i}_{\text{sources and losses}}$$

Facts

1. Nonlinear due to chemical reaction term $\chi_i(\mathbf{c})$;
2. High dimension, typically $10^6 \sim 10^7$;
3. Strong uncertainties mainly due to uncertain parameters [Mallet and Sportisse, 2006]; initial conditions tend to be forgotten.

Assimilation

Estimate the uncertainties for a better prediction from diverse resources, i.e. model simulations, observations and statistics information. Variational vs. sequential algorithms.

1. Model and observations at time step k:
$$\begin{cases} \mathbf{c}_k = M[\mathbf{c}_{k-1}] + \eta_k & \text{Model} \quad M \\ \mathbf{y}_k = H[\mathbf{c}_k] + \epsilon_k & \text{Observation} \quad \mathbf{y}_k \end{cases}$$
2. Minimization of a cost function $J(\mathbf{c})$ that deals with obs.:
$$\tfrac{1}{2}(\mathbf{c}-\mathbf{c}_k)^T \mathbf{P}_k^{-1}(\mathbf{c}-\mathbf{c}_k) + \tfrac{1}{2}(\mathbf{y}_k - H[\mathbf{c}])^T \mathbf{R}^{-1}(\mathbf{y}_k - H[\mathbf{c}])$$

where $\mathbf{P}_k = \mathbf{E}(\eta_k\eta_k^T), \mathbf{R} = \mathbf{E}(\epsilon_k\epsilon_k^T)$.

Implemented Algorithms

1. Optimal interpolation (OI);
2. Ensemble Kalman filter (EnKF);
3. Reduced-rank square root Kalman filter (RRSQRT);
4. Four dimensional variational assimilation (4DVar): time interval $k = 0, \dots, N$.

The model uncertainties are either parameterized with homogeneous error correlations (Balgovind parameterization for OI and 4DVar) or estimated by perturbing some (mostly lognormal) uncertain parameter $\mathbf{p}$ (RRSQRT and EnKF) as:

$$\forall k, i \quad \hat{\mathbf{p}}_k^i = \mathbf{p}_k^i \times \sqrt{\alpha}^{\,\gamma},$$

where γ is sampled according to a standard normal distribution, independent of the time index k and of the space index i. The same sample of γ is used to perturb all values of the field $\hat{\mathbf{p}}$. The uncertain parameters sets and their magnitudes are listed as follows

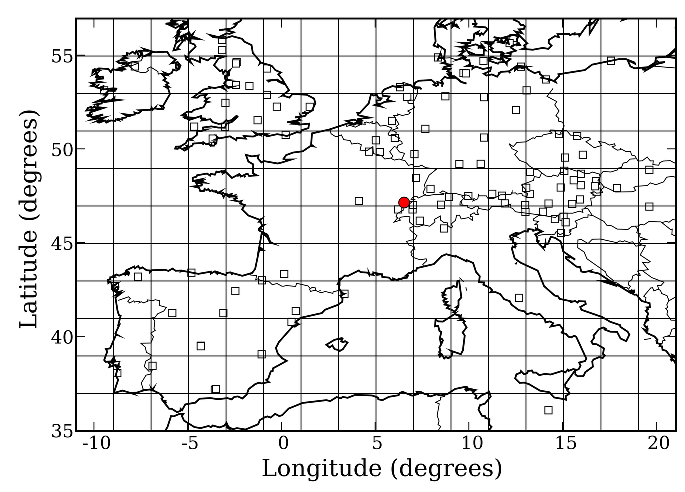
	Parameter name	α_0	α_1	α_2
Ω_0	Boundary condition	3.	3.	3.
	Deposition velocity	1.5	2.	3.
	Photolysis rate	1.3	1.5	2.
	Surface emission	1.5	2.	3.
	Attenuation	1.3	1.5	2.
	Vertical diff. coef.	1.3	1.5	2.
Ω'	Cloud height	1.3	1.5	2.
	Vertical wind	1.3	1.5	2.
	Meridional wind	1.3	1.5	2.
	Zonal wind	1.3	1.5	2.
	Specific humidity	1.3	1.5	2.
Ω''	Pressure	1.3	1.5	2.
	Air density	1.3	1.5	2.
	Merid. diff. coef.	1.3	1.5	2.
	Zonal diff. coef.	1.3	1.5	2.
	Temperature	0.005	0.01	0.015

In this table, uncertain parameters sets and magnitudes are detailed. We denote Ω_1 as $\{\Omega, \Omega'\}$, and Ω_2 as $\{\Omega, \Omega', \Omega''\}$. The distribution of temperature is supposed to be normal, and its magnitude should be interpreted as relative standard derivation.

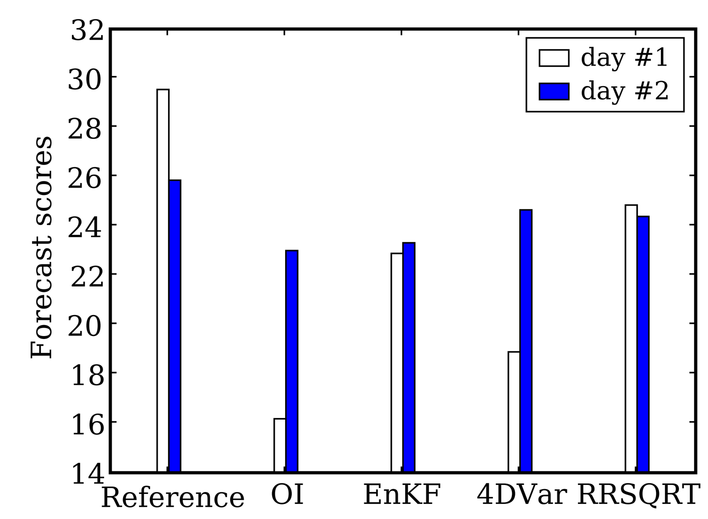
Comparison Results

The four assimilation algorithms are compared under the same experimental settings.

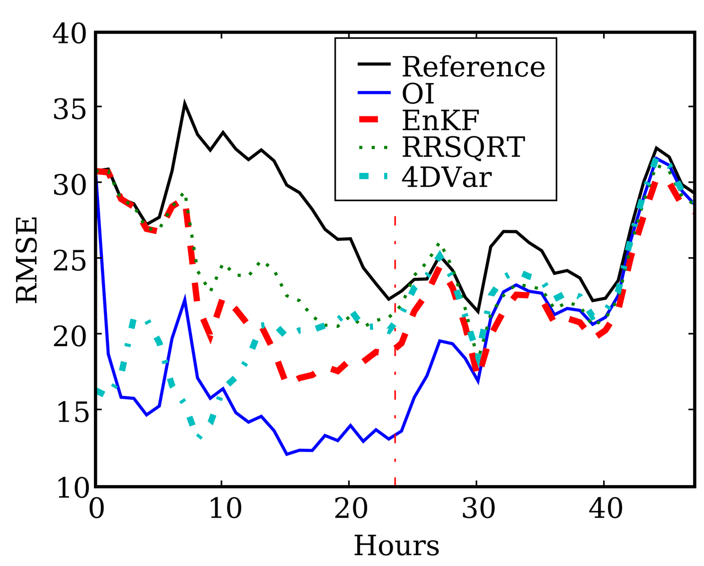
2 Map of model domain.



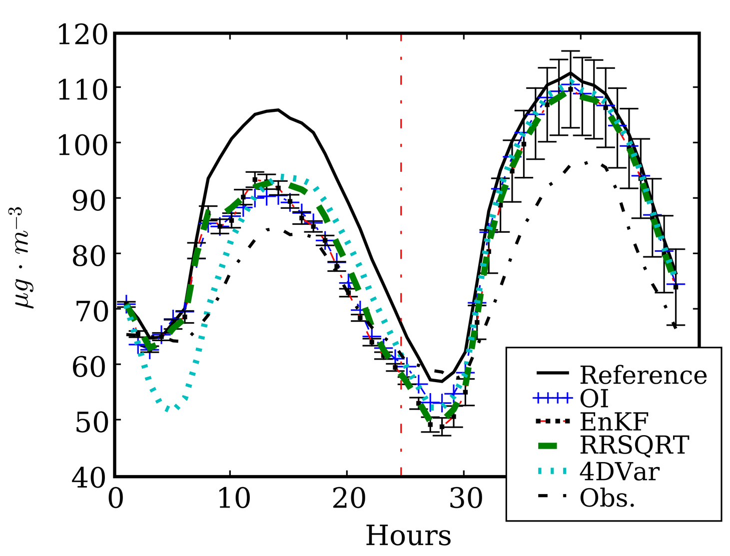
The coarse discretization of model domain and the observation network EMEP stations are shown in FIG. 2. The squares show the locations of EMEP monitoring stations, and the disc shows the location of the monitoring station Montandon. For model details, we refer to Mallet et al. [2007].



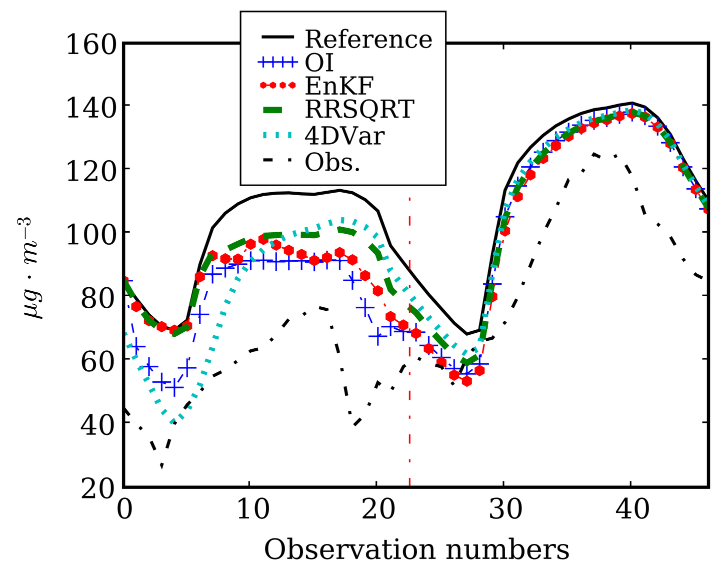
3 Forecast scores of ozone concentrations during the assimilation period (day #1) and the prediction period (day #2)..



4 Time evolution of the RMSE for the ozone forecasts. The vertical lines delimits the assimilation period from the prediction period.



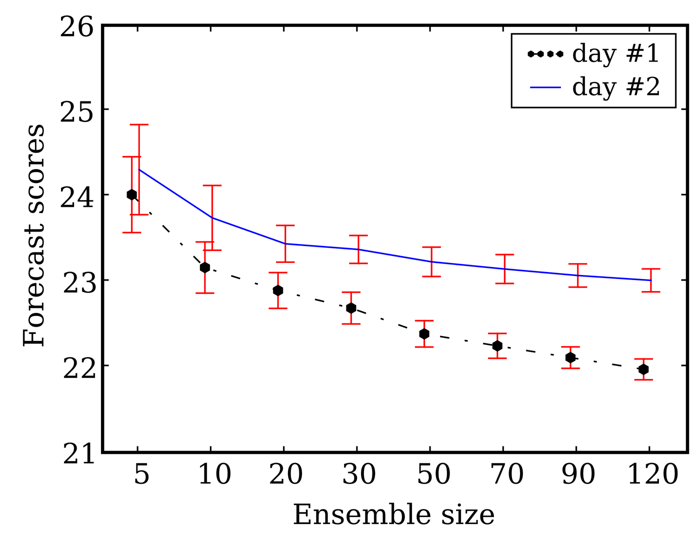
5 Time evolution of average ozone forecasts over all available stations. The error bar shows the average spread of the EnKF forecast ensemble calculated over these stations.



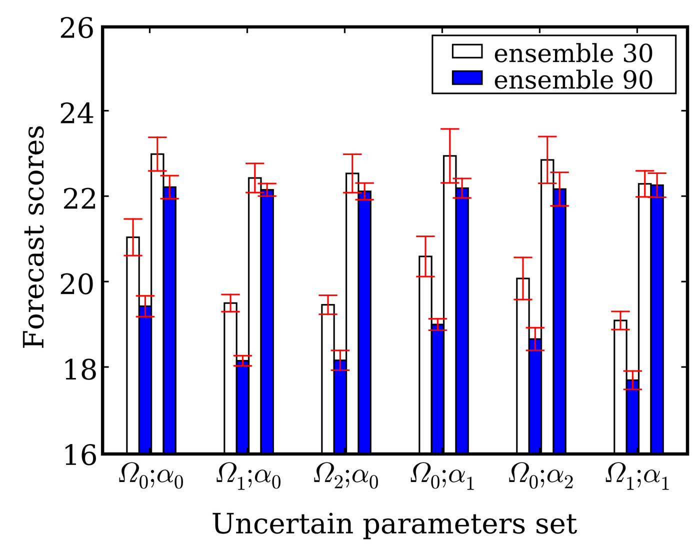
6 Time evolution of ozone forecasts against available observations over two days at EMEP station Montandon.

Sensitivities to Assimilation Algorithms

Modifications of configurations on each component of the data assimilation systems, i.e. model, observation and algorithm, may influence the assimilation performance. In FIG. 7 - 10 we show the results on sensitivities to algorithms.

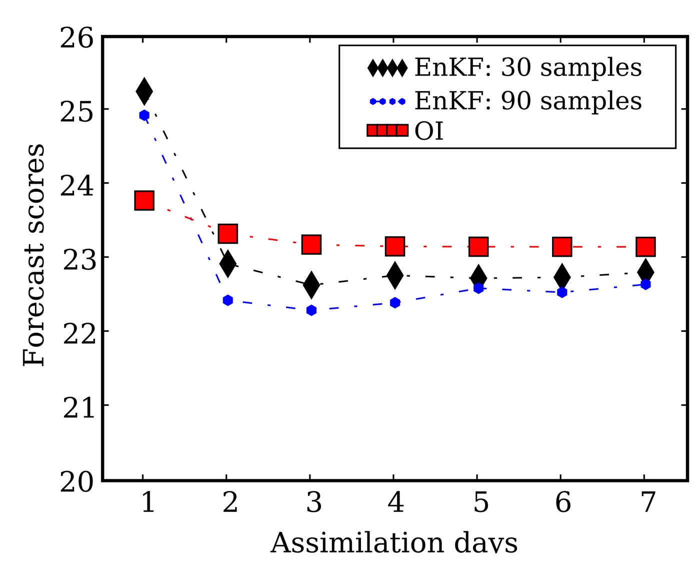


7 Forecast scores of EnKF against the ensemble size. The curve shows mean scores and the errorbar shows the standard derivations over 10 random seed numbers.

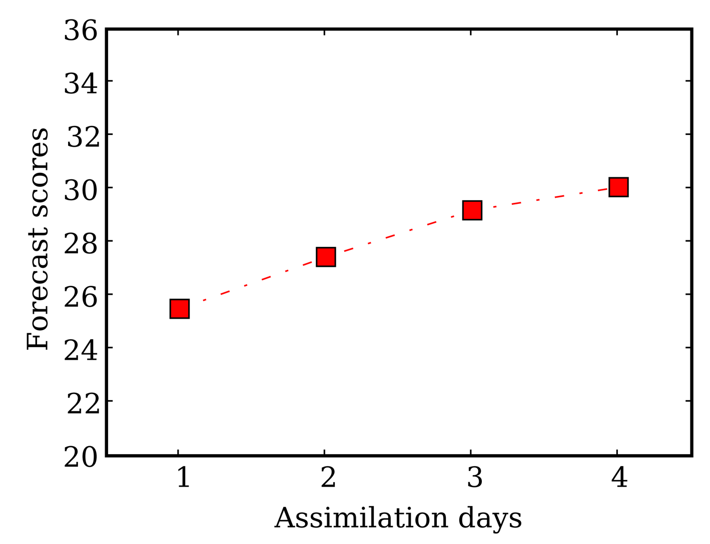


8 Forecast scores for EnKF against different uncertain parameter definitions.

The parameter sets and perturbation magnitudes are defined in TAB. 1. The EnKF sample number is chosen to be 30 (white columns) and 90 (dark columns) respectively. The two columns of scores for each case show the forecast scores during the assimilation and prediction periods. The bar values are mean scores, and the errorbar shows the standard derivations over 10 random seed numbers.

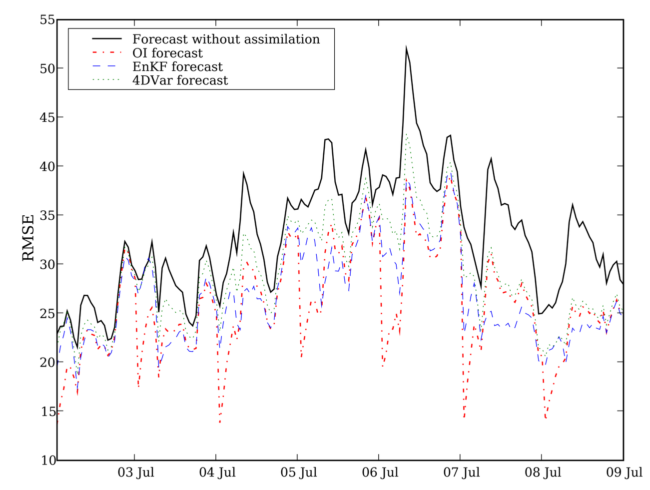


9 Forecast scores of OI and EnKF (with 30 and 90 members) against the number of assimilation days.

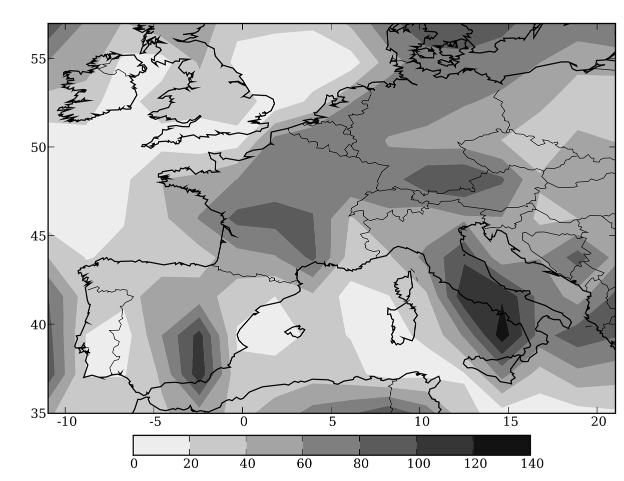


10 Forecast scores against the number of assimilation days for the two experiments using 4DVar.

Cycling and Model Error



11 The one-day forecast performances based on model simulations with/without assimilations in the context of cycling assimilation/predictions.



12 The covariance approximated by EnKF forecast ensemble between the error at the station Montandon and the error in all ground cells at 13:00 UT, 2nd July 2001.

Conclusions

It is found that the assimilations significantly improve the ozone forecasts. The comparison results reveal the limitations and the potentials of each assimilation algorithm. In the four-dimensional variational method, it is shown that the model error has to be accounted for to further improve the forecasts. In the sequential methods, the ensemble approach demonstrates great potential for the forecasts during the end of the prediction periods.

NOTE 1 - RMSE: *the root mean square error.*

NOTE 2 - SCORE: RMSE *over given time length.*

References

Mallet, V., Quélo, D., Sportisse, B., Ahmed de Biasi, M., Debry, É., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M., and Foudhil, H. (2007). Technical Note: The air quality modeling system Polyphemus. *Atmos. Chem. Phys.*, 7(20):5,479–5,487.

Mallet, V. and Sportisse, B. (2006). Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling. *J. Geophys. Res.*, 111:D01302.

Wu, L., Mallet, V., Bocquet, M., and Sportisse, B. (2008). A comparison study of data assimilation algorithms for ozone forecasts. *J. Geophys. Res.*, 113:D20310.

WWRP/THORPEX WORKSHOP on 4D-VAR and ENSEMBLE KALMAN FILTER INTER-COMPARISONS, Buenos Aires, Argentina, November 10–13, 2008