

# Testing stratospheric chemical-dynamical ensemble Kalman filtering

Thomas Milewski, Michel S. Bourqui  
Department of Atmospheric and Oceanic Sciences  
McGill University  
thomas.milewski@mail.mcgill.ca

## Abstract

Data assimilation is an essential component of the numerical prediction of the atmosphere, but its use in the stratosphere is still at an early stage. This layer of the atmosphere poses a lot of scientific challenges, including its strong nonlinearity in wave-breaking regions and the coupling between ozone chemistry, radiation and dynamics. The Ensemble Kalman Filter (EnKF) is chosen here since it is ensemble-based and thus, unlike variational data assimilation schemes, it does not require the tangent linear model and its adjoint to propagate covariances and therefore retains the nonlinear properties of the model used. The EnKF is here coupled with a chemistry-climate model with a fast interactive ozone chemistry scheme, the IGCM-FASTOC (Taylor and Bourqui, 2005).

This study presents first results towards exploring the application of the EnKF to stratospheric chemical-dynamical data assimilation. Initial stability tests have been performed for two analysis schemes of the EnKF, one with explicit formulation of the error-covariance matrices and observations localization (Houtekamer and Mitchell, 2001, henceforth HM), the other with singular-value decomposition (SVD) of the error fields (Evensen, 2003). Assimilation is performed with synthetic satellite observations, mimicking MIPAS (nadir) and ACE-FTS (solar-occultation) temperature retrievals. Assimilation test scores for these different types of observations are here presented.

## Data Assimilation Experiment

### Model forecasts :

Climatological ensemble from a T21L26 (lid at 0.1hPa) time-slice run of 129 years on  $\sigma$  levels. State vector:

$$\psi = (u, v, T, q, O_x, N_2O_5, NO_x, HNO_3, P_s)^T$$

### Twin experiment :

128 members + 1 true state chosen so as to minimize RMS error over the stratosphere. 24-hour assimilation window, forecast and analyses at 00 UTC.

### HM assimilation :

EnKF with Schur-product localization of forecast error-covariance matrix using a compactly-supported 5<sup>th</sup>-order piecewise rational function (Gaspari and Cohn, 1999) with 1400-km horizontal decorrelation length and 1 unit in  $\ln(P)$  scale for vertical decorrelation length. Data assimilated sequentially in 100-observation batches.

### SVD assimilation :

EnKF with SVD decomposition of sum of the observation perturbations and mapped-to-observation-space forecast departures from ensemble average. Number of singular vectors defined by the smallest of the number of ensemble members or the number of observations.

### Synthetic observations:

Random perturbations (Burgers et al., 1998) from true temperature interpolated on pressure levels. Perturbations are normally-distributed, unbiased, with a standard deviation equal to 10% of the true temperature.

Two types of observations (Fig.1), MIPAS and ACE-FTS, both span a vertical region ranging from 10 to 300 hPa. MIPAS orbits moves longitudinally with time, with a recurrence interval of 36 days. Each ACE-FTS occultation latitude cycle has a period of 60 days.

## Results

### Climatological ensemble (Fig.2, dashed and solid red lines)

The evolution of the climatological ensemble give us ground for comparison with assimilation products (blue lines). Here, the most "central" member is initially chosen as truth, but it will eventually diverge away from ensemble mean : the error grows and surpasses the climatological spread after day 70.

### MIPAS HM assimilation (Fig.2, upper panel)

The spread of the ensemble sharply decreases in the first few assimilation cycles to settle to a constant value. The error wiggles about the spread, but does not diverge, indicating good filter stability. Notably at the end of the sequence, the climatological natural tendency towards divergence is corrected by the assimilation.

### ACE-FTS HM assimilation (Fig.2, lower left panel)

The spread displays two regimes for such data assimilation. The first, distinguishable at the beginning and end of the sequence, shows little correction from the forecast stage to the analysis (thin envelope), while the second regime has a smaller spread and larger envelope. The two regimes are linked to the occultation latitudes, the first regime happens when observations are taken near the equator (e.g. Fig. 1, right panel), the second when the observations are taken near both poles. Thus, considering the latitudinal restraint and sparseness of the observations, the analysis can not correct the error in the end of the sequence and we observe divergence between the error and the spread.

### MIPAS SVD assimilation (Fig.2, lower right panel)

The error shows improvement in comparison with the climatological error, however the spread displays unexpected behavior. Opposite to typical assimilation cycles, the analysis stage actually increases the spread, while the forecast tries to bring the spread back towards climatology. The net result is a constant increase in the spread until model failure. The limited number of singular vectors (128) onto which the state vector (of much higher dimension) can project is the most likely candidate for such behavior.

## Conclusions

Examples of twin experiment assimilations of synthetic observations in climatological-ensemble Kalman filters have highlighted possibilities and failures of the system. A larger window of experiment is necessary to draw more accurate conclusions but a few observations can be stated.

Satisfactory stability tests of the HM EnKF with a latitudinally-dense observation network (MIPAS), will allow to test ozone assimilation and its effects on dynamical properties of the analysis ensemble.

Unexplained divergence of the spread in SVD assimilation will be investigated by removing chemical tracers from the state vector, increasing the size of the ensemble or doing local analysis, to see if dynamical balance can be retained.

Future improvements to the HM data assimilation filter will include implementing space and variable-dependent covariance localization and possibly covariance inflation.

## References

- Burgers, G., P. J. Van Leeuwen, and G. Evensen, 1998: Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, **126**, 1719–1724.
- Evensen, G., 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.
- Gaspari, G. and S. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125**, 723–757.
- Houtekamer, P. L. and H. L. Mitchell, 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, **129**, 123–137.
- Taylor, C. P. and M. S. Bourqui, 2005: A new fast stratospheric ozone chemistry scheme in an intermediate general-circulation model. i: Description and evaluation. *The Quarterly Journal of the Royal Meteorological Society*, **131**, 2225–2242.

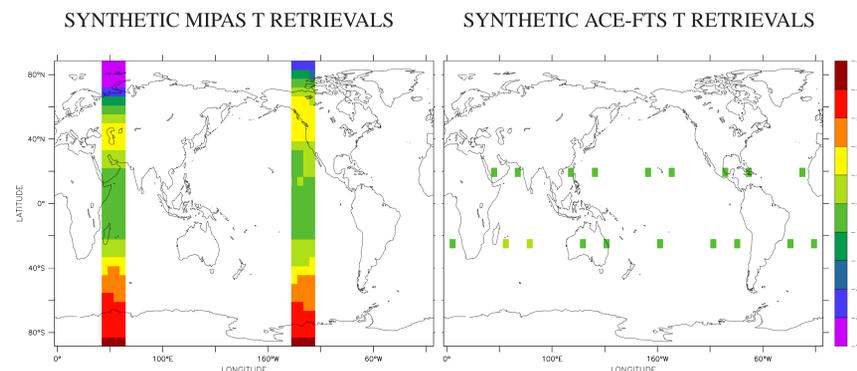


Figure 1: Example of horizontal location for the two types of temperature observations used.

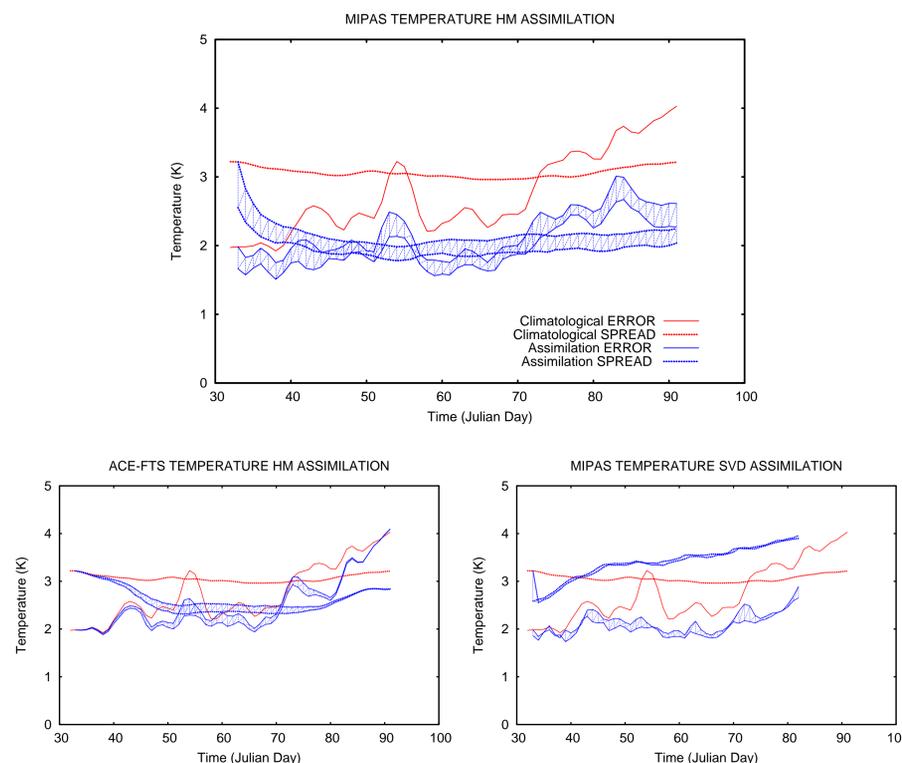


Figure 2: Global mid-stratospheric forecast and analysis temperature error and spread for climatological ensemble and three different assimilation scenarios.

## Diagnostics

Diagnostics for state ensemble matrix  $A(N, M) = (\psi_1, \psi_2, \dots, \psi_M)$ , N representing the model state space and M the ensemble members :

Diagnostics are performed for a given variable and a spatial subset of the model state space spanning the whole globe but restricted to the mid-stratosphere (between pressure levels 30 to 90 hPa).

RMS difference between ensemble mean and true state :

$$\text{ERROR} = \sqrt{\frac{1}{N} \sum_{i=1}^N [\bar{A}(i) - A^t(i)]^2}$$

Mean RMS difference between ensemble members and the ensemble mean :

$$\text{SPREAD} = \frac{1}{N} \sum_{i=1}^N \sqrt{\frac{1}{M-1} \sum_{m=1}^M [A(i, m) - \bar{A}(i)]^2}$$