Addition of time interpolation in the ensemble Kalman filter

Peter Houtekamer and Herschel Mitchell
Environment Canada

Since 12 January 2005, an ensemble Kalman filter (EnKF) is being used in Canada for global atmospheric data assimilation. It provides the initial conditions for a medium-range ensemble prediction system.

Initial EnKF analyses were not of the same quality as the higher resolution analyses that were obtained with a 4D-Var algorithm for the high-resolution deterministic forecast at our center.

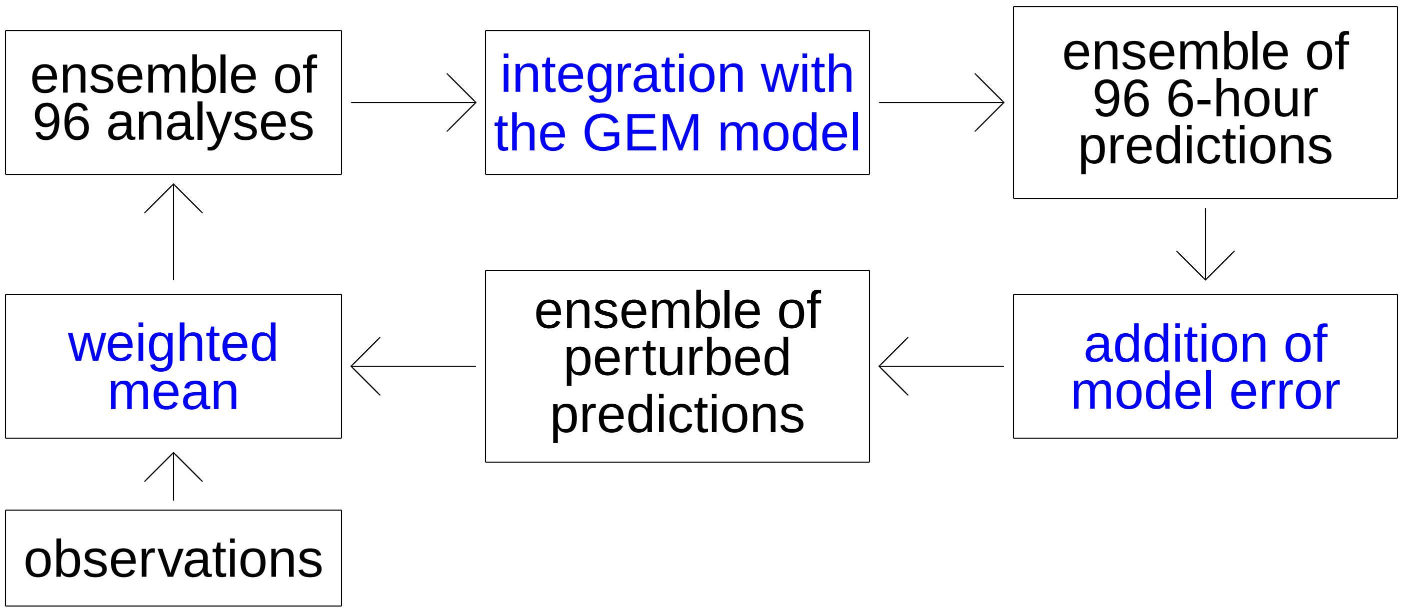
In part this was due to the lower resolution - and correspondingly lower computational cost - of the EnKF. In part this was due to weaknesses of the initial EnKF algorithm.

We have modified the EnKF algorithm and are working towards a clean comparison with 4D-Var.

Overview

- introduction
- the January 2005 operational EnKF implementation
- adding time interpolation
  - expansion of the control variable
  - digital filter finalization to obtain a smooth trajectory
  - addition of model error at the beginning of the trajectory
- preliminary comparison with 4D-Var
- future work

The January 2005 Operational EnKF Implementation



We use the EnKF and the GEM forecast model to generate 96 analyses every 6 hours.

The fourth dimension of the (Canadian) ensemble Kalman filter

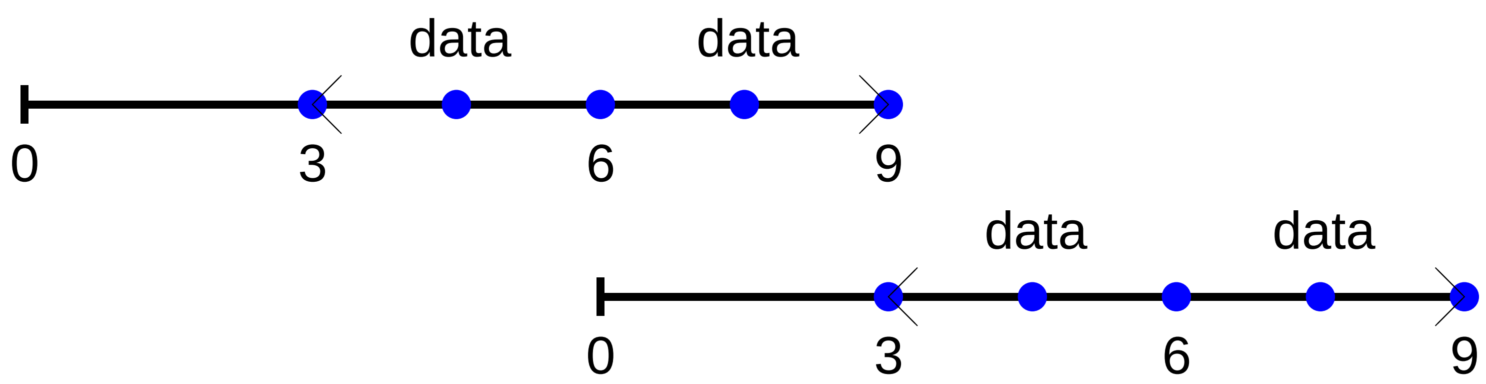
With the EnKF, we transport covariance information from one assimilation cycle to the next. In this regard, our EnKF is a truly 4D algorithm.

However, to interpolate the model state to the observations, we use a unique model state that is valid at the middle of the 6-hour assimilation window. Consequently, in the forward operator H, we neglect the model evolution during the 6-hour window. Thus, for the forward interpolation, the Canadian EnKF is a 3D algorithm.

It follows that, unlike 4D-Var, we are unable to benefit from all available observations in the 6-hour window and have to limit ourselves to observations in the central 3-hour portion of the assimilation interval, as was done in the 3D-Var algorithm at our center.

6-h assimilation window

An EnKF can assimilate all data in a 6-h window at the appropriate time (Hunt et al. 2004, Tellus A), as is currently done in 4D variational algorithms:



For the time interpolation, we need the model state at $t = 3\text{h}, t = 4\text{h}30, t = 6\text{h}, t = 7\text{h}30$ and $t = 9\text{h}$. Only the analysis at the central time $t = 6\text{h}$ is used to start the subsequent integration.

Length of the state vector

The cost of the data assimilation step is dominated by operations involving the matrix PH^T. These have a cost proportional to the number of model coordinates $N_{model} * N_{timelevels}$, to the number of observations N_{obs} and to the number of ensemble members N_{ens}:

$$cost = O(N_{model} * N_{timelevels} * N_{obs} * N_{ens}).$$

Using $N_{timelevels} = 5$ leads to a problematic five-fold increase of the cost of the analysis.

In our sequential algorithm, we assimilate the observations one batch at a time. The ensemble of forecast trajectories gradually evolves into an ensemble of analyzed trajectories as more and more batches of observations are assimilated. We need to keep track of the trajectories because the forward operator H first interpolates the model state to the time of the observation.

Although we obtain an analysis trajectory valid at $N_{timelevels}$ times, we only need the central model state (valid at $t = 6$h) to start the subsequent integration.

Expansion of the state vector

The vector Hx of interpolated observations can be added to the state vector (Tarantola 1987; Anderson, MWR, 2001; Gauthier 2005), which then becomes (x, Hx). The classical advantage is that H can be a complex operator - like a parameterization of convection - that exists in the forecast model but not normally in the assimilation code. Combined with a sequential algorithm, we obtain an evolving state vector (x, Hx). During the sequential algorithm we do not have to re-evaluate H.

With this algorithm, we precompute H and consequently we no longer need to keep track of the entire evolving trajectory. It is sufficient to have as state vector $(x(t = 6\text{h}), Hx(t = t_{obs}))$. All relevant information about the temporal evolution is in the correlations between $x(t = 6\text{h})$ and $Hx(t = t_{obs})$.

This revised algorithm, with time interpolation and an expanded state vector, has a cost:

$$cost = O((N_{model} + N_{obs}) * N_{obs} * N_{ens}) \approx O(N_{model} * N_{obs} * N_{ens})$$

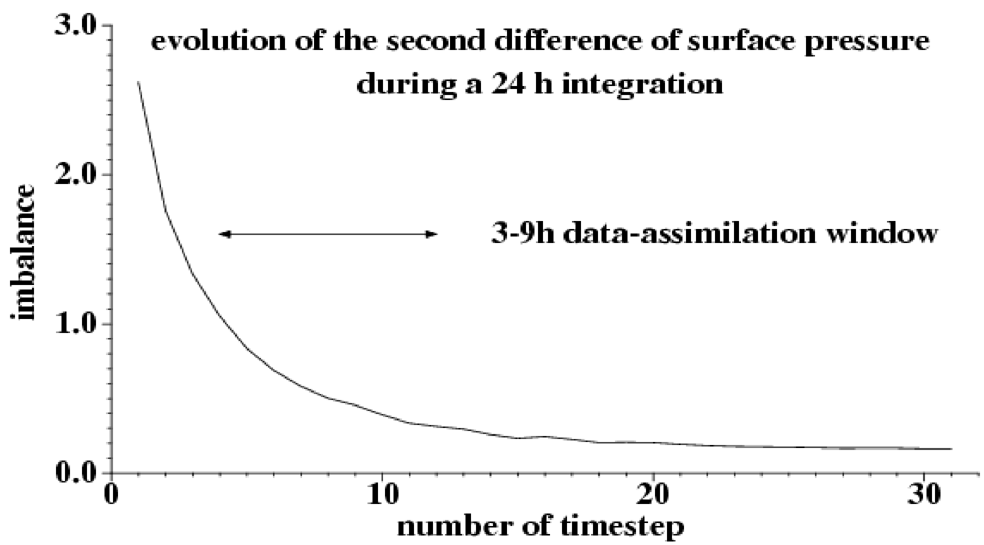
Imbalance in the initial conditions

To quantify the imbalance in a given analysis, we estimate the second difference $D^2(p_s)$ for surface pressure, p_s, from the time series provided by a 24-h integration starting from that analysis.

$$D^2(p_s) = p_s(t + 45\text{min}) - 2p_s(t) + p_s(t - 45\text{min})$$

The global r.m.s. value $\parallel D^2(p_s) \parallel$ is subsequently obtained using:

$$\parallel D^2(p_s) \parallel^2 = (D^2(p_s), D^2(p_s)) = \frac{1}{S}\int_S D^2(p_s) D^2(p_s) dS$$



Imbalance and localization

$r_h(km)$	Initial imbalance $\parallel D^2(p_s(\text{45min})) \parallel$				
	$r_z = 2$	$r_z = 4$	$r_z = 6$	$r_z = 8$	$r_z = 100$
2800	2.618	2.913	3.100	3.236	3.622
5600	2.600	2.725	2.783	2.815	2.925
8400	2.744	2.520	2.381	2.345	2.423
11200	2.883	2.420	2.316	2.133	2.145
280000	1.996	1.558	1.223	1.153	0.454

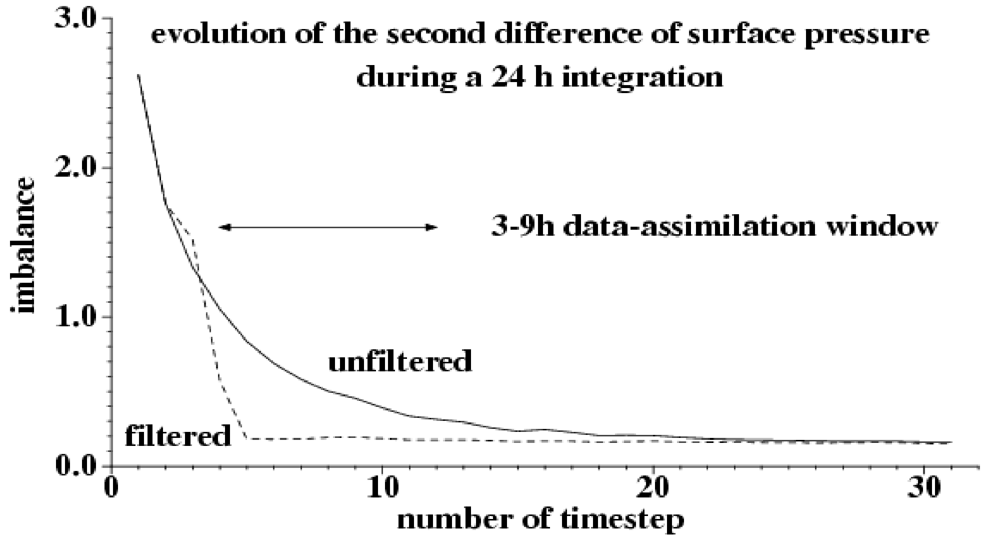
The second difference of surface pressure estimated at t=45 min for different values of the localization parameters r_h and r_z. Horizontal correlations are forced to zero at a distance of r_h km and vertical correlations at r_z units of ln p.

Relaxing the localization has only a small impact on the imbalance observed in the surface pressure component. In the limit of no localization, we obtain balanced initial conditions.

Impact of a digital filter

A digital filter finalization technique (Fillion et al., 1995, Tellus A) can be used to obtain an integration that is balanced from time step 4 (at 3 hours) onwards.

Applying a digital filter in the operational EnKF system, which assumes all data are valid at 6 h, leads to a small positive impact on our verification scores.



With the digital filter, we can consider time interpolation in the 6-h observation window (much like a 4D-Var).

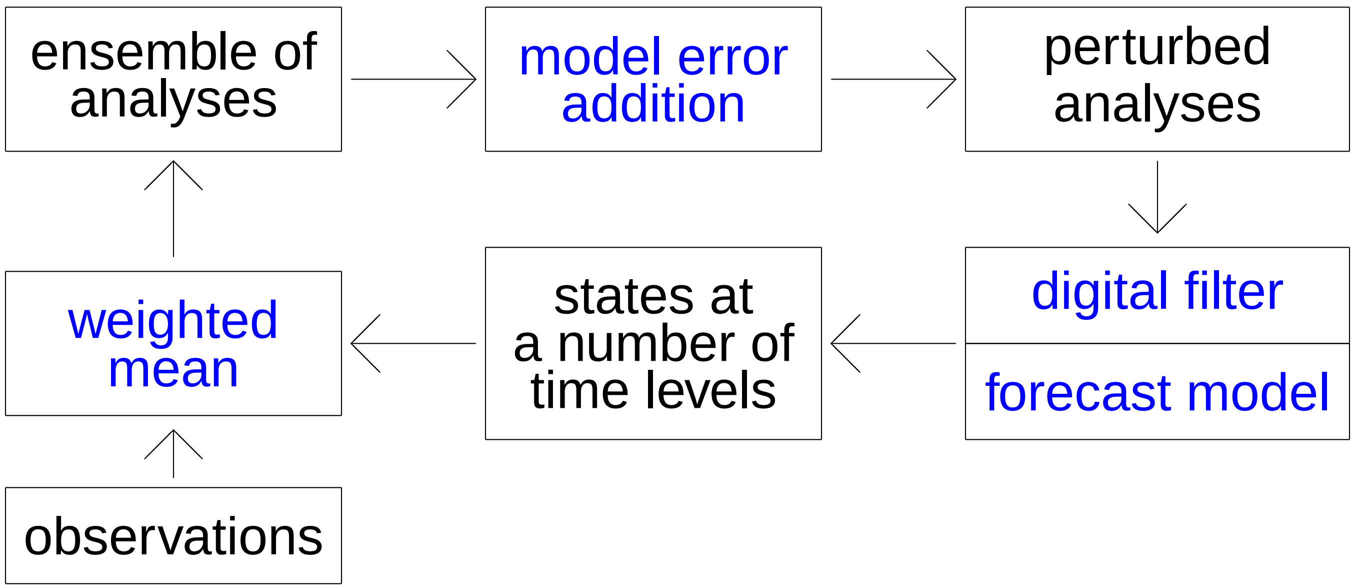
Interpretation of model error

The parameterized model error has a length of 1.39 m/s as measured with an energy norm. This is almost as large as the total impact of the model physics (1.52 m/s). It is hard to believe that the model physics can be 90% in error. However, the regular addition of a model error that is this large enables us to maintain sufficient spread in the ensemble.

Alternatively, we can take the position that every imperfection in the data-assimilation cycle should be accounted for by the model-error term. If, for instance, the H operator does not include time interpolation, we will need to compensate for this by using a larger model error, Q.

In view of this, we can argue that the parameterized model-error perturbations could actually be added to the ensemble of analyses!

An EnKF configuration with time interpolation



We obtain incremental improvements by relocating the model-error addition, by using a digital filter and by adding time interpolation. The revised configuration has been used for a preliminary comparison with 4D-Var.

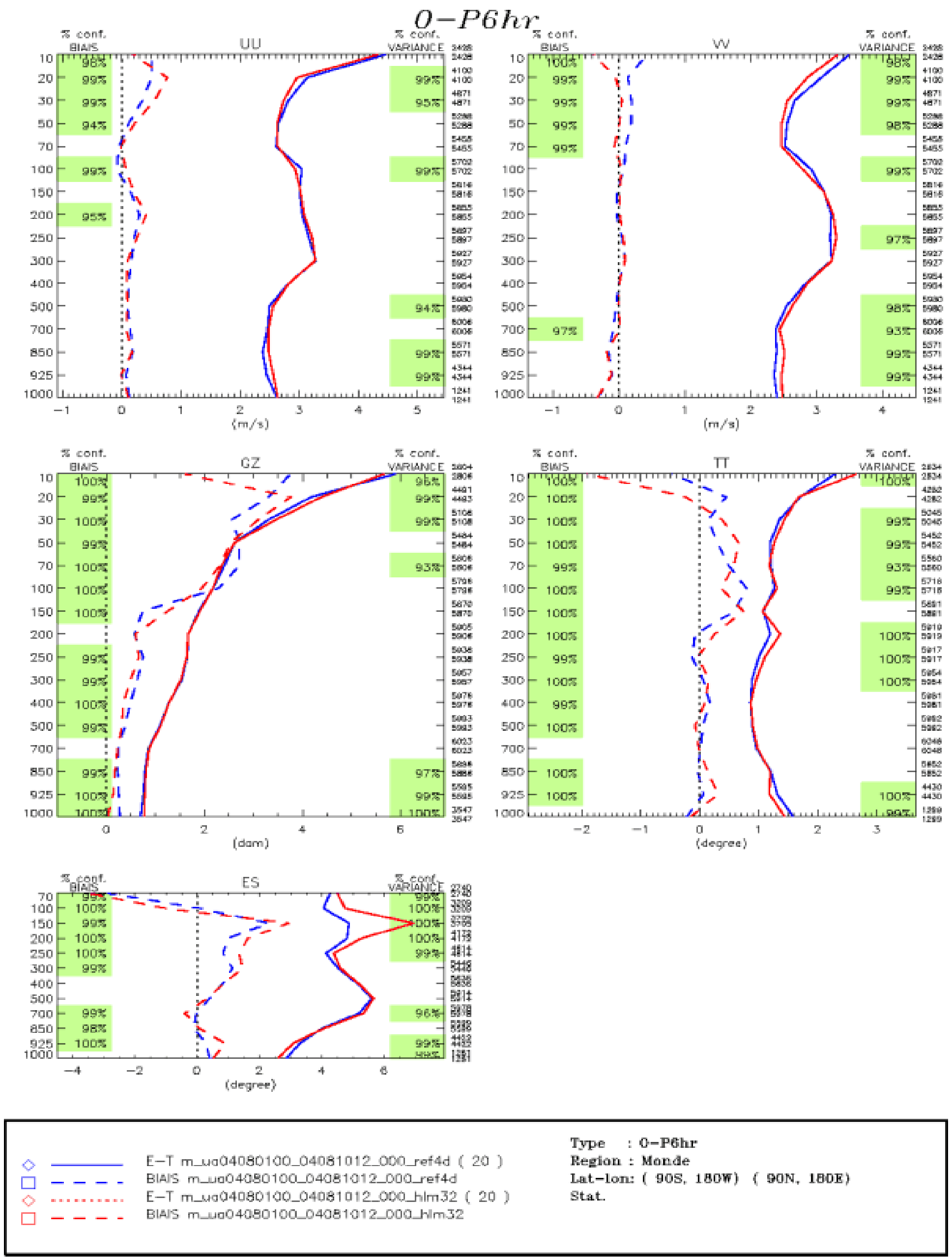
A first comparison of the EnKF with 4D-Var

For now, the comparison suffers from a series of important shortcomings:

1. the GEM forecast model was not run in the same configuration,
2. the 4D-Var assimilates more observation types (notable GOES humidity data and dewpoint depression above 200 hPa),
3. the observation operators H are not by construction the same (some differences may therefore exist),
4. the 4D-Var has been able to reject some of the observations that would otherwise have been used for the verification,
5. the quality of the ensemble mean 6h forecast from the EnKF system is compared with the unique 6h forecast from the 4D-Var system,
6. only one short experiment was performed in summer.

The preliminary results are as follows:

Verifying the EnKF and 4D-Var against radiosonde data



panels for u,v, height, T and T-Td

Period: 1-10 August 2004.

Blue curves: the pre-implementation 4D-Var

Red curves: The EnKF on a 400 $\times$ 200 grid with time interpolation for H

The EnKF is inferior for winds below 400 hPa, for temperature at 200 hPa and humidity above 300 hPa.

The EnKF is superior for winds above 100 hPa, and for temperature and humidity at the surface.

Same ballpark results

A comparison between the EnKF and 4D-Var is being performed at Environment Canada. Preliminary results would appear to be in the same ballpark.

Unfortunately, due mostly to different development paths, some avoidable differences exist between the experimental environments in which the two algorithms are run. At this point, these may have a bigger impact than the differences found. It is, therefore, too early to arrive at any definitive statements.

We are working towards having more similar experimental environments.

This comparison may lead to a revision and improvement of both algorithms.

Future work

We will reduce the amplitude of the homogeneous and isotropic model-error parameterization.

To maintain an appropriately large ensemble spread, we will test the use of different physical parameterizations for different members of the EnKF ensemble. In this, we would follow and expand upon what is currently being done for the Canadian medium-range forecast ensemble.

Thank you for your attention