Environment Canada   Environnement Canada

Canada

Status of EnKF

Workshop on 4D-Var EnKF inter-comparisons

Buenos Aires Argentina

Peter Houtekamer

Data Assimilation and Satellite Meteorology Research Section

10-13 November 2008

Overview

1) Monte Carlo methods,
2) Kalman filter equations
   a) nonlinear evolution of error statistics,
   b) many ways to do the analysis step,
   c) imbalance due to localization,
3) many options for the simulation of model error,
4) issues and options for the future.

Blue fonts indicate issues that cannot easily be addressed with a single run 4D-Var.

Red fonts are used for issues that pose a problem in an EnKF environment.

Monte-Carlo methods: a simple view

Original system:

Short-hand notation:

Monte-Carlo method: generate many sets of randomly perturbed observations to obtain a random sample of forecasts.

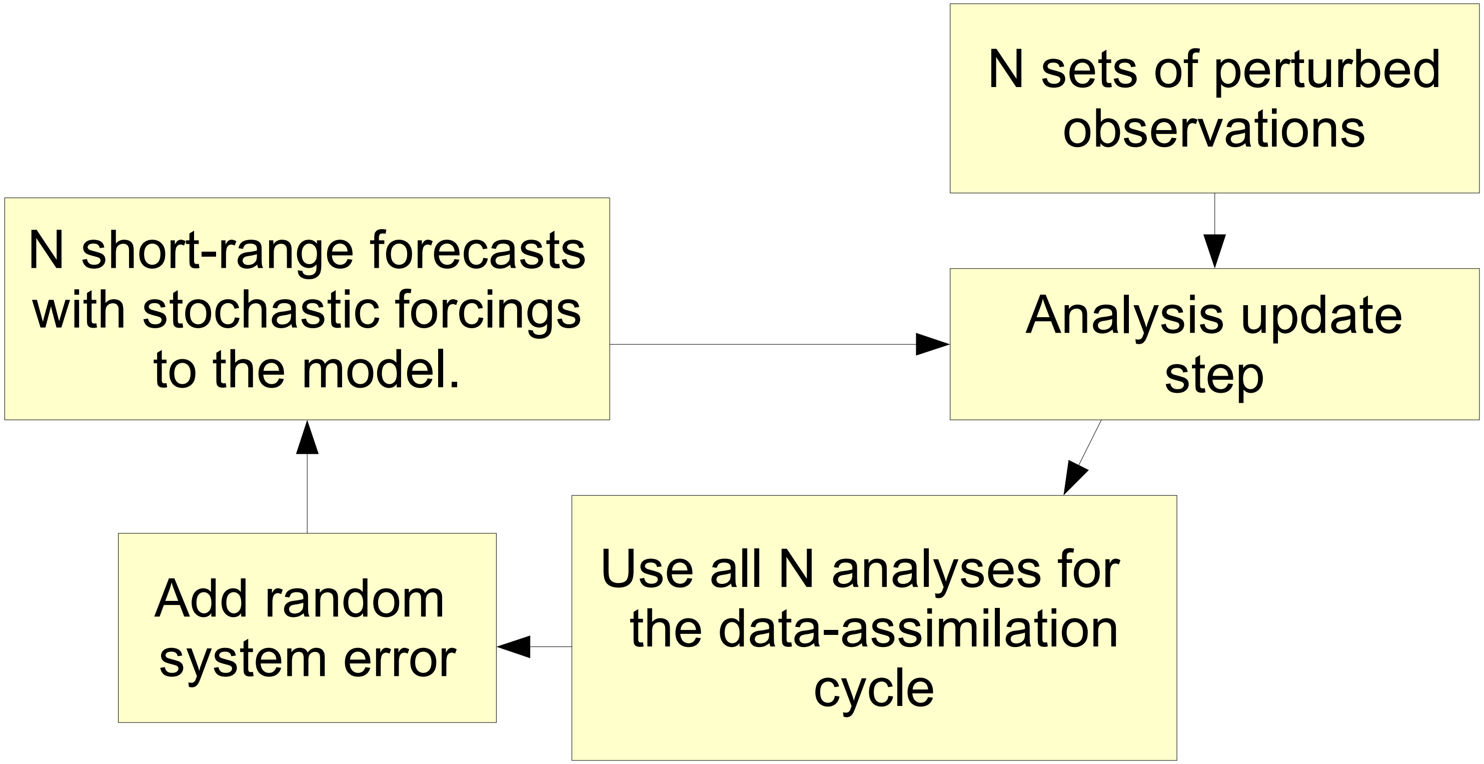
Result: statistics on forecast error

Error sources in the ensemble Kalman filter

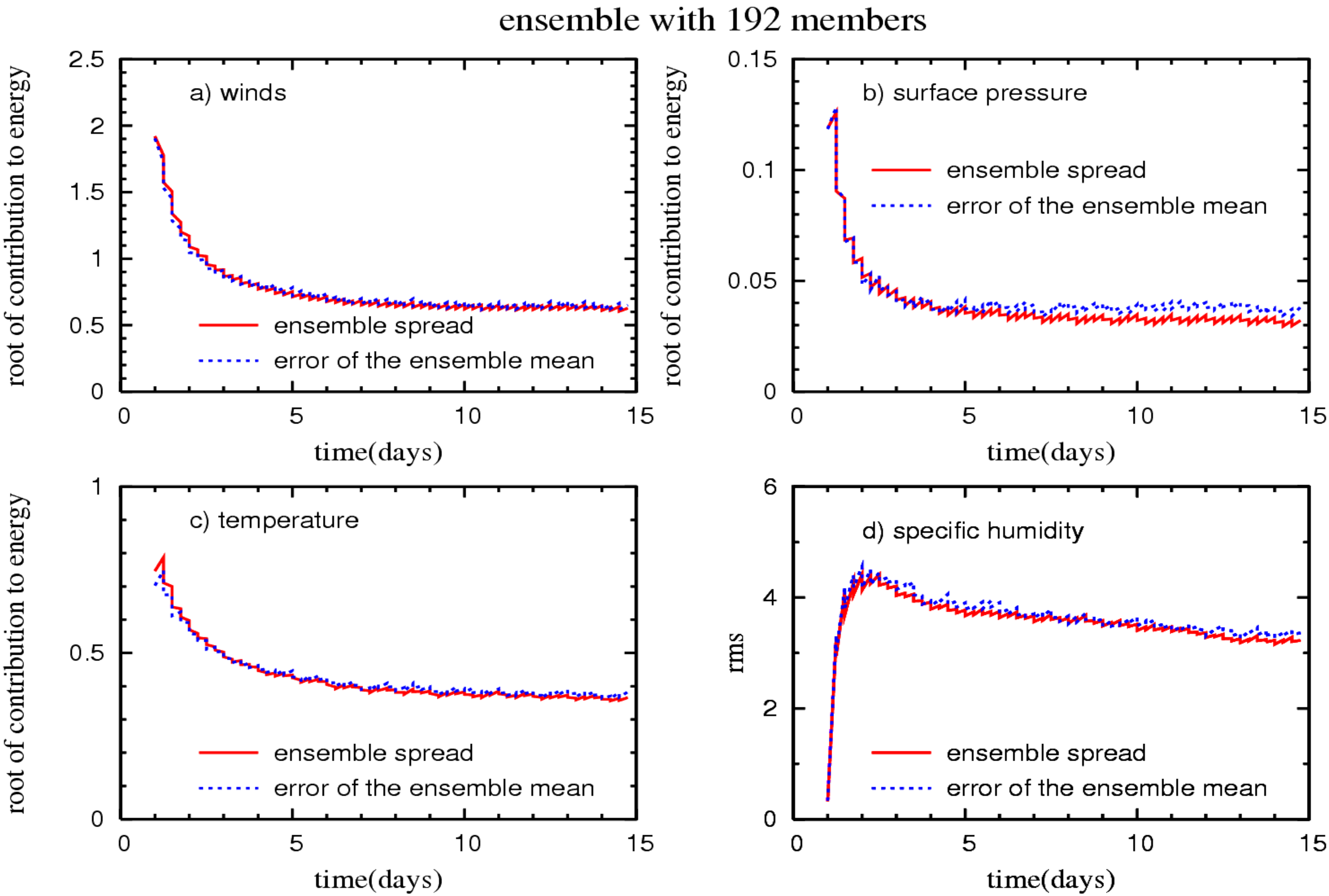
Errors in the data-assimilation cycle have several origins:

1) uncertain observations,
2) uncertain error statistics (like assuming that errors are independent and have no bias),
3) differences between the forecast model and the atmosphere.

All significant sources of error will have to be sampled.



A perfect model experiment with the Canadian EnKF



In a controlled environment, without tuning any parameters, the EnKF is able to maintain ensemble statistics that are representative of the ensemble mean error.

A scientific work environment

The EnKF provides a closed coherent framework to deal with error statistics:

1) estimated error statistics are specified for uncertain basic inputs (like the observations),
2) the EnKF will maintain a representative ensemble if the specified statistics are realistic,
3) consequently a comparison of innovation statistics with the ensemble spread and observational uncertainty provides information about the quality of the specified error statistics.

Note: current EnKF implementations converge with O(100) members. Beyond this, other sources of error become more important than errors due to the sample size.

The ensemble of analyses can be used to initialize an ensemble prediction system (EPS).

Kalman filter equations

The Kalman filter (Kalman 1960; Kalman and Bucy 1961) provides ``*an elegant and comprehensive mathematical description of the data assimilation problem*'' (Daley 1991).

$$x^a(t) = x^f(t) + K(o(t) - H x^f(t))$$ (1) analysis increment

$$K = P^f(t) H^T (H P^f(t) H^T + R)^{-1}$$ (2) Kalman Gain matrix

$$P^a(t) = (I - K H) P^f(t)$$ (3) analysis error covariance

$$P^f(t+1) = M P^a(t) M^T + Q$$ (4) evolve error covariances with model dynamics

$$x^f(t+1) = M(x^a(t))$$ (5) evolve the best estimate

Assumption*: the model and observations have independent errors with* ***no bias***.

Practical problems (Daley 1991, Ghil and Malanotte-Rizzoli 1991):

i) the computational cost of the matrix equations (2), (3) and (4),

ii) the model error Q is not well known (Dee, 1995).

(note that in 4D-Var the model error can be neglected due to a regular re-initialization of covariance information).

Evolution of error covariances with model dynamics

The Extended Kalman Filter EKF (Evensen 1992, Miller et al. 1994, Gauthier et al., Bouttier 1994) uses linearized dynamics in the covariance evolution equation. The neglect of higher moments leads to unbounded error covariance growth (due to the absence of saturation).

$$P^f(t+1) = M P^a(t) M^T + Q \qquad (4)$$

The EnKF (Evensen 1994) uses the nonlinear forecast model to evolve error covariances. It uses a Markov Chain Monte Carlo (MCMC) method to solve the fundamental Fokker-Planck equations (book on the ensemble Kalman filter by Evensen 2006). Because the nonlinear forecast model is used error covariances do saturate.

Options:

$$x_i^f(t+1) = M(x_i^a(t)) + q_i, i = 1, \cdots, N$$

(6) Like the EKF, but non-linear

$$x_i^f(t+1) = M(x_i^a(t) + q_i), i = 1, \cdots, N$$

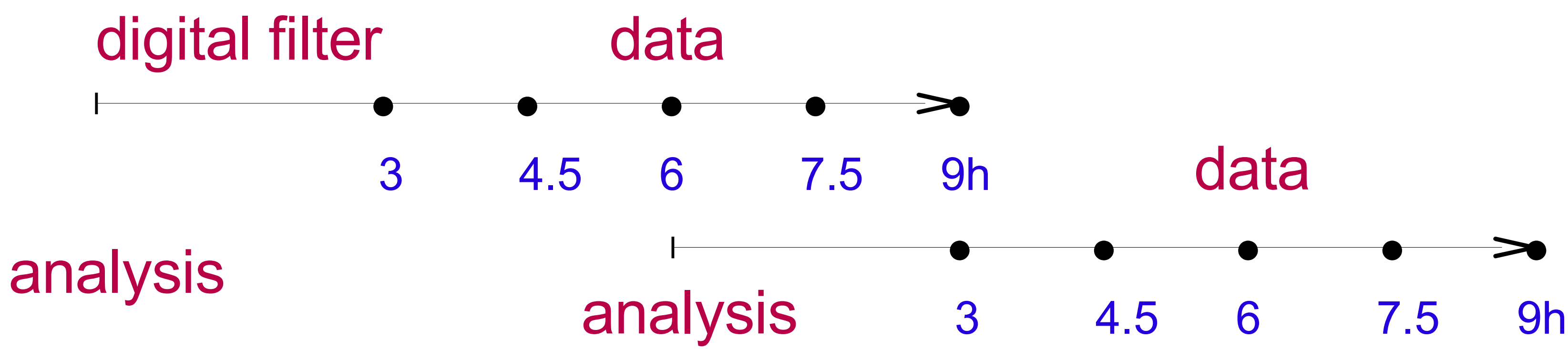
(7) Like 4D-Var, but non-linear, permits time-interpolation

$$x_i^f(t+1) = M_i(x_i^a(t) + q_i), i = 1, \cdots, N$$

(8) sampling model error using perturbed models.

Time interpolation

With the EnKF we can assimilate all data in a 6-h window, as is currently done in 4D variational algorithms (Evensen 2006; Hunt et al. 2004).



To permit time interpolation, the state vector consists of the five dotted points:

$$(x(t=3\mathrm{h}), x(t=4.5\mathrm{h}), x(t=6\mathrm{h}), x(t=7.5\mathrm{h}), x(t=9\mathrm{h}))$$

Only the analysis at the central time is used to start the subsequent integration. Therefore, it is more economical to use an ***extended state-vector*** consisting (Anderson 2001) of:

$$(x^f(t=6\mathrm{h}), Hx)$$

Many ways to do the analysis step
1: stochastic or deterministic

1) ***Stochastic***: Early EnKF implementations (Houtekamer and Mitchell 1998) used sets of randomly perturbed observations as suggested by the Monte Carlo methodology.

2) ***Deterministic***: Whitaker and Hamill (2002) propose an Ensemble Square Root Filter (EnSRF) in which observations are not perturbed. Instead, ensemble background perturbations are combined with the observations using a modified gain matrix. The modification is such that the appropriate analysis error covariances are obtained.

Tippett et al (2003) compare different proposed EnSRF algorithms.

Lawson and Hansen (2004) show that, in the presence of nonlinear dynamics, the deterministic filter can lead to pathological non-Gaussian ensemble distributions (with one or a few outlying members).

In his book, Evensen proposes the use of random rotations in the EnSRF to distribute the ensemble variance among members.

The issue is unresolved and depends also on other choices for the analysis step.

Many ways to do the analysis step
2: with or without cross-validation

In the EnKF, when using a ***single ensemble*** of N members, the ensemble members are used both

1) to determine the gain matrix K and
2) to determine the quality of the K (from the spread of the ensemble of analyses).

Doing (1) prior to (2) violates Monte Carlo principles (since the analysis system has been modified it is no longer a black box). The result is an underestimate of the analysis error (Houtekamer and Mitchell 1998).

As in a ***cross-validation approach***, it is possible to use certain ensemble members to compute the gain matrix K and other members to test the quality of that gain. In the operational Canadian EnKF, 4 sub-ensembles are used. To assimilate each group of 24 members, the remaining 72 members are used.

It is not evident how to combine determinstic and cross-validation algorithms. All other groups use a single ensemble approach. The resulting underestimate of the ensemble spread is treated as a model error.

Many ways to do the analysis step 3: localization

Due to the small ensemble size O(100), it is necessary to localize the impact of observations.

The impact of an observation can be smoothly forced to zero at large separation using a ***Schur product*** of the ensemble based covariances and a covariance function with compact support (Hamill et al. 2001 and Houtekamer and Mitchell 2001). Sometimes, the vertical location of an observation (like a radiance) or a model coordinate (like surface pressure) is ill defined.

Alternatively one can use a ***box analysis method*** as in the LETKF (Local ensemble transform Kalman filter, Ott et al. 2004).

The presentation by Tom Hamill gives a complete overview of the various proposed localization strategies.

Localization is a deviation from the original Kalman filter equations and causes imbalance.

Imbalance

In practice, permitting imbalance (deviations from the model attractor) in the analysis often leads to a higher quality analysis (because the model and the atmosphere have different attractors).

Examples:

1) the use of different physical parameterizations for different members of an ensemble (e.g: Meng and Zhang with regional models and Houtekamer and Mitchell with global models).
2) the use of an unbalanced temperature component in the background or model error description (Canadian EnKF, 3D/4D-variational methods).

In the EnKF, the use of localization is the dominant source of imbalance. The imbalance is an ***undesirable side effect*** of localization. The resulting rapid oscillations in the surface pressure can be removed using a balancing method such as a digital filter finalization.

Imbalance in 4D systems is the subject of the afternoon session.

Model error

It has long been known (Daley 1991, Dee 1995) that a Kalman filter implementation needs to account for model error in an appropriate manner to be successful.

Neglecting model error in the EnKF leads to an rms ensemble spread that is too small by about a factor of two (Houtekamer et al., in print).

Performing an ensemble of 4D-Var assimilation cycles with sets of randomly perturbed observations and using a stochastic backscatter algorithm similarly leads to an ensemble spread that is too small by about a factor of two (Isaksen et al., ECMWF, 2007).

Thus, if our assumptions about the data-assimilation system were correct, errors would be smaller by about a factor of two.

This raises the question: what is wrong in our current systems and what can we do about it?

Strategies to account for model error

Within the Monte Carlo framework, there are two main ways to account for model error:

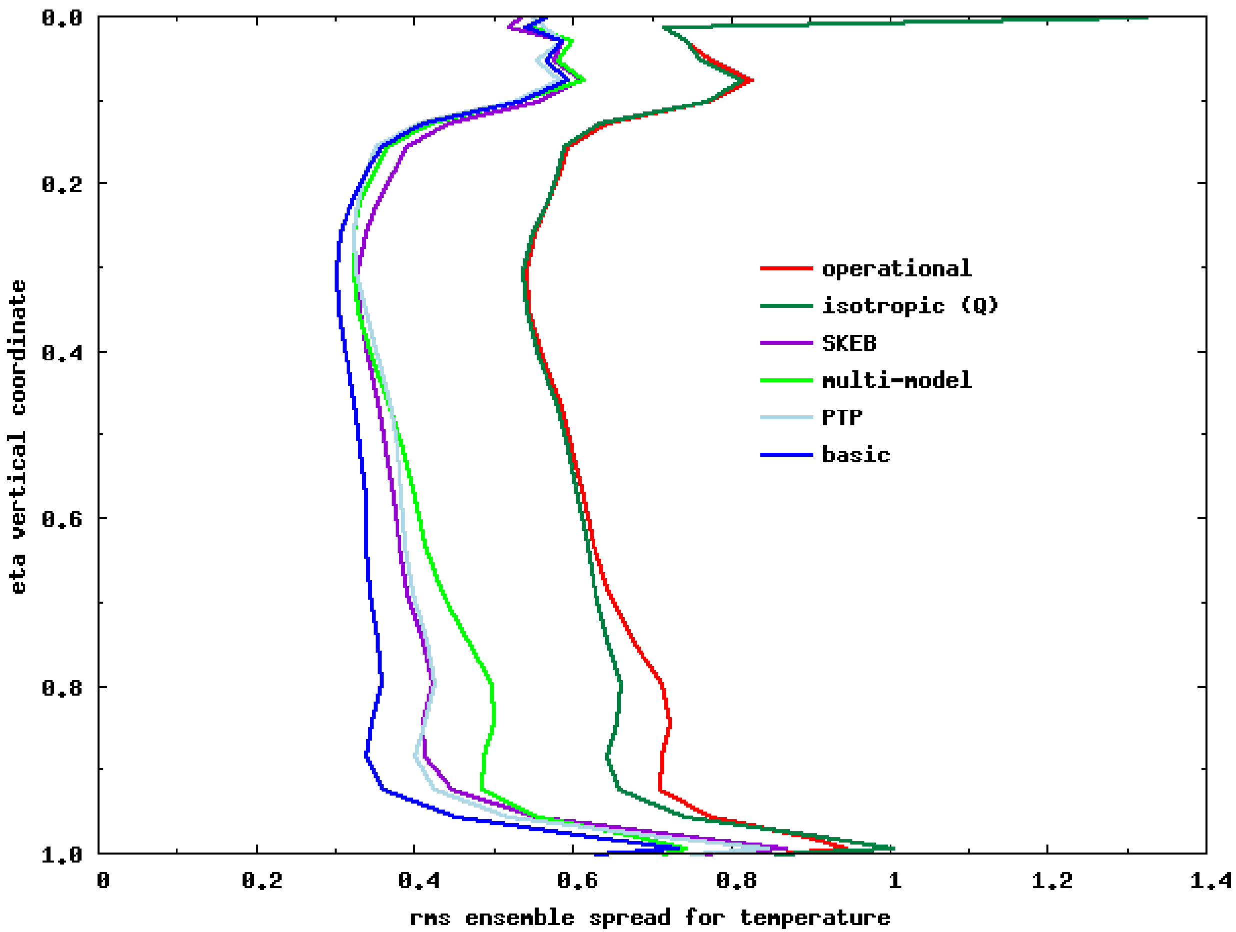
1) ***addition of random perturbation fields***,
2) ***using a perturbed model for the model integration***.

$$x_i^f(t+1) = M_i(x_i^a(t) + q_i),\, i = 1, \cdots, N$$

1) To generate random perturbation fields q, we need knowledge about the statistical properties of the model error. Examples are: (a) sample differences from runs with different resolution (Hamill and Whitaker 2005), (b) import (and adjust) a background matrix from 3D-Var (Canadian ensemble).

2) To perturb the model, we need some reasonable hypotheses about the weakest components of the model. Examples are: (a) stochastic physics (Buizza et al. 1999), (b) Stochastic Kinetic Energy Backscatter (Shutts 2005), (c) using different physical parameterizations (Meng and Zhang 2007).

Model error that is (not) clearly due to the model



Uncertainty in the model physics explains a sizable fraction of the model error in the lower atmosphere only.

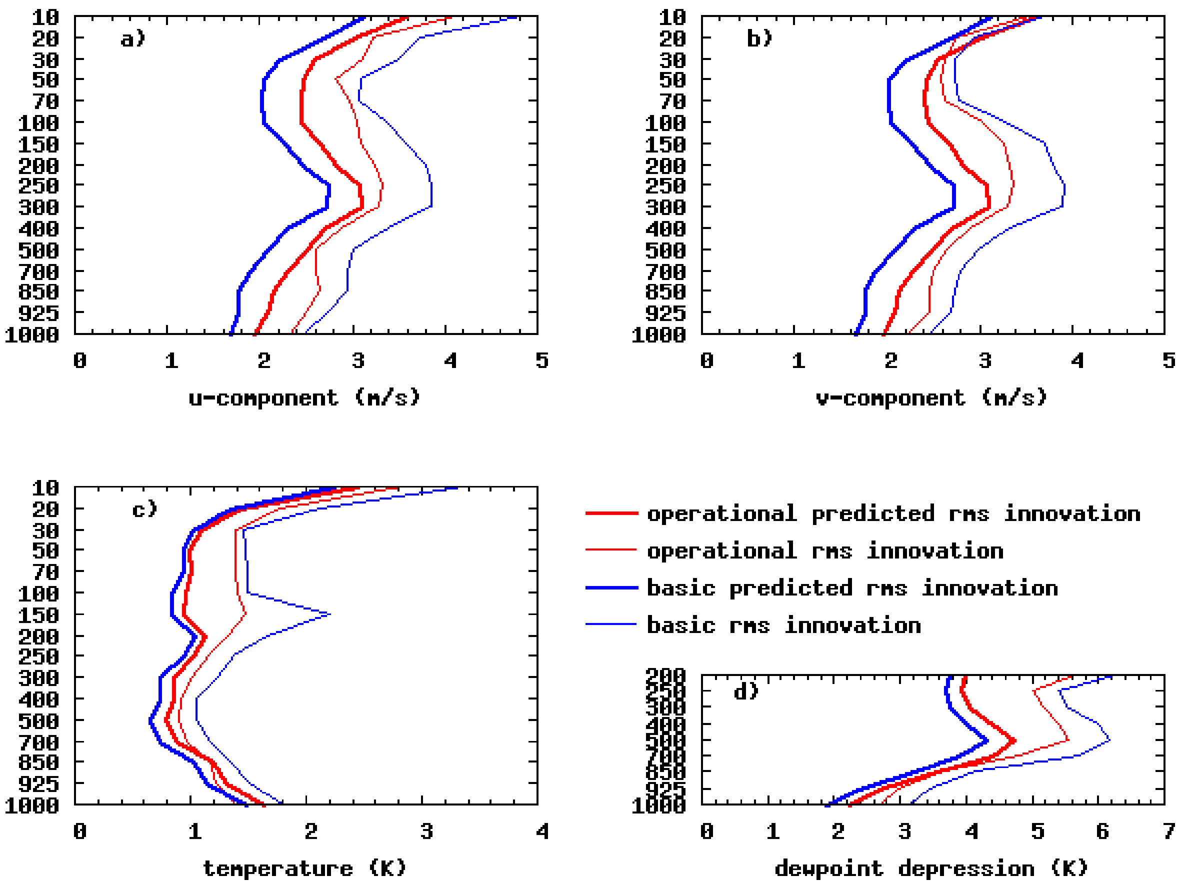
In the upper atmosphere, the isotropic component dominates. It has been obtained from 3D-Var and subsequent tuning.

i.e: ***we don't know the origin of most of the model error***.

The impact of model error simulation

For a representative ensemble: $$\overline{(O-\overline{P^f})^2} = H\,P^f\,H^T + R$$

Radiosonde observations have been used to verify this equation.



Basic: EnKF with real observations and assuming that the model is perfect.

Operational: as above, but adding random perturbations (from Q) and using different physical parameterizations for different members.

Simulation of model error dramatically improves results.

Summary of the simulation of model error

To maintain reasonable ensemble statistics, we need to add a model error term of significant amplitude.

We do not know the cause of the model error. Uncertainty in the model physics can only explain a certain fraction.

Reduction of the unexplained model error appears most important for the upper atmosphere (above about 300 hPa).

See the talk by Eugenia Kalnay for an overview of methods to account for model error in the ensemble Kalman filter.

Computational aspects.

Most EnKF studies use N_ensemble ~ 100. Amazingly, it has not been necessary to increase the ensemble size as more observations and higher resolution models started to be used (both N_obs and N_model increased by orders of magnitude between 1998 and 2008). Horizontal and vertical localization have been sufficient so far.

Can we continue along the same lines or will we need to significantly increase the ensemble size as we want to have small-scale details in global analyses?

Most proposed EnKF algorithms scale as

O(N_obs x N_model x N_ensemble).

Will we need to move to variational analysis algorithms when N_obs becomes very large or can we use more severe localization?

examples: (1) Zupanski, 2004, Maximum Likelihood Ensemble Filter,

(2) Hamill and Snyder, 2000, Hybrid EnKF-3D-Var.

Towards smaller scales

1) A dual-resolution approach (Gao and Xue, 2008, talk by Mark Buehner) can be used in which a low-resolution ensemble supports a high-resolution analysis (much like in an incremental 4D-Var).

2) As the resolved scales become smaller, it is possible to reduce the length of the assimilation window in agreement with the shorter predictability limits for these scales (with no loss of information).

3) With the current parameters for the digital filter finalization the window of the Canadian global EnKF cannot be made shorter than 3 hours. Do we still want to initialize as we move to higher resolution and if so how should it best be done?

Complexity

Data assimilation with any method is a complex procedure.

Developing the EnKF towards operational use (1997-2005) in Montreal would not have been made possible without:

1) continuous strong support from our managers,
2) the help of our many colleagues in the Meteorological Research Division (data assimilation, modeling and computing),
3) many ideas developed at other centers (like DAO and NCAR),
4) about 200 experiments with an evolving configuration.

It is hard to predict what the external parameters (like the number and type of observations, the quality of the forecast model and the type of computational platform) of future data assimilation systems will be and which algorithm will be most suitable in such a context.

It is important to proceed in such a way that no critical knowledge or experience is lost.

Thank you!

This talk did focus on global atmospheric data assimilation. References to other application areas can be found in the book by Evensen (2006).

E.g:

1) ocean data assimilation (Kepenne and Rienecker 2003),
2) hydrologic data assimilation (Reichle et al. 2002),
3) convective scale assimilation (Dowell et al. 2004).