Coupling 4D-Var with 4D-En-Var

David Fairbairn and Stephen Pring

University of Surrey/Met Office

13^{th} June 2012

David's Supervisors: Ian Roulstone (Surrey) and Andrew Lorenc (Met Office)

Thanks also to Neill Bowler and Adam Clayton for their help.

Funded by EPSRC, Met Office and University of Surrey

Introduction

- Data assimilation in NWP combines a prior forecast (background state) with the latest observations of the atmosphere to provide the best estimate of the state of the atmosephere.
- The prior forecast and the observations are weighted by their respective covariance matrices, which are a measure of their expected errors.
- We aim to compare four data assimilation methods using toy models, for their ability to produce a deterministic analysis:
  1. 4D-Var (Four-dimensional variational data assimilation with a climatological background error covariance matrix);
  2. 4D-En-Var (Four-dimensional-Ensemble variational data assimilation);
  3. 4D-Var-Ben (4D-Var with a flow-dependent background error covariance matrix coming from 4D-En-Var);
  4. DEnkf (Deterministic Ensemble Kalman filter).

Incremental 4D-Var

- 4D-Var (Le-Dimet and Talagrand, 1986) Provides a least squares fit between observations and a prior forecast in the assimilation window;
- 4D-Var is an initial value problem;
- 4D background state ($\underline{\mathbf{x}}^b$) propagated through the window using the forecast model:

$$\underline{\mathbf{x}}^b = \underline{M}(\mathbf{x}^b(t_0)). \tag{1}$$

- Incremental formulation designed to improve efficiency (Courtier et al., 1994):

$$\delta\mathbf{x}(t_0) = \mathbf{x}^b(t_0) - \mathbf{x}(t_0), \tag{2}$$

  where $\mathbf{x}$ is the current best estimate (the first guess equal to the background).
- Cost function conditioning improved by control variable transform from $\delta\mathbf{x}$ to $\mathbf{v}$:

$$\delta\mathbf{x}(t_0) = \mathbf{U}\mathbf{v}, \tag{3}$$

  where $\mathbf{U}$ is designed such that $\mathbf{U}\mathbf{U}^T = \mathbf{B}$.
- The increment is propagated using the Tangent Linear model:

$$\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\delta\mathbf{x}(t_0). \tag{4}$$

Incremental 4D-Var

- The Cost function is defined by:

$$\begin{aligned} J[\mathbf{v}] &= \frac{1}{2}\mathbf{v}^T\mathbf{v} \\ &+ [\underline{\mathbf{H}}\delta\underline{\mathbf{x}} - \underline{\mathbf{d}}]^T\underline{\mathbf{R}}^{-1}[\underline{\mathbf{H}}\delta\underline{\mathbf{x}} - \underline{\mathbf{d}}], \end{aligned} \tag{5}$$

  where

$$\underline{\mathbf{d}} = \underline{H}(\underline{\mathbf{x}}) - \underline{\mathbf{y}}^o \tag{6}$$

  is the difference between the observations ($\mathbf{y}^o$) and the model predicted values of the observations $H(\underline{\mathbf{x}})$.
- The Observation operator H transforms $\mathbf{x}$ from model space to observation space. $\mathbf{H}$ is the linearised observation operator.
- Background and Observation error covariance matrices ($\mathbf{B}$ and $\mathbf{R}$) are assumed to be Gaussian and unbiassed.
- The cost function gradient is defined by:

$$\left[\frac{\partial J}{\partial \mathbf{v}}\right] = \mathbf{v} + \mathbf{U}^T\underline{\mathbf{M}}^T\underline{\mathbf{H}}^T\underline{\mathbf{R}}^{-1}[\underline{\mathbf{H}}\delta\underline{\mathbf{x}} - \underline{\mathbf{d}}], \tag{7}$$

  where $\mathbf{M}^T$ is the adjoint model.

Advantages of 4D-Var

- Tangent linear model provides effective time correlation of the observations;
- Full rank climatological background error covariance matrix;
- 4D-Var provides accurate time-correlations of background errors (within the linear approximation of $\mathbf{M}$), which allow observations distributed in time to be effectively used.

Disadvantages of 4D-Var

- Requires complex tangent linear and adjoint models;
- No flow-dependent uncertainty information;

4D-En-Var (notation by Andrew Lorenc)

- 4D-En-Var is effectively an EnKF in a variational framework;
- 4D-En-Var [Liu et al., 2008a,b, Buehner et al., 2010] weights an ensemble of model trajectories according to how well they fit the observations in an assimilation window;
- The design adopts as much as possible from 4D-Var;
- Fundamental difference is that 4D-Var uses a propagated climatological background error covariance matrix ($\underline{\mathbf{M}}\mathbf{B}\underline{\mathbf{M}}^T$), which is replaced in 4D-En-Var by a localised 4-dimensional ensemble covariance matrix.
- Also, unlike 4D-Var, 4D-En-Var is not an intitial value problem;
- 4D-En-Var requires a background ensemble of model trajectories

$$\underline{\mathbf{x}}_j^b = \underline{M}(\mathbf{x}_j^b(t_0)), \text{ for } j = 1, \ldots, m. \tag{8}$$

from which we we calculate the mean trajectory $\overline{\underline{\mathbf{x}}}^b$ and hence the perturbations from the mean ($\delta\underline{\mathbf{x}}_j^b$).

4D-En-Var

- The tangent linear model equation in 4D-Var is replaced by a locally weighted linear combination of these perturbation trajectories:

$$\delta\underline{\mathbf{x}} = \sum_{j=1}^{m} \frac{1}{\sqrt{m-1}} \delta\underline{\mathbf{x}}_j^b \circ \underline{\boldsymbol{\alpha}}_j, \tag{9}$$

  where $\underline{\boldsymbol{\alpha}}_j$ is the smooth 4-dimensional field of weights given to the jth perturbation trajectory and $\circ$ is the Schur (element by element) product.
- Control variable transform to condition the J_α (like 4D-Var):

$$\underline{\boldsymbol{\alpha}}_j = \underline{\mathbf{U}}^\alpha \mathbf{v}_j^\alpha \text{ for } j = 1, \ldots, m, \tag{10}$$

  where

$$(\mathbf{U}^\alpha)^T(t_i)\mathbf{U}^\alpha(t_i) = \mathbf{C}, \tag{11}$$

  where $\mathbf{C}$ is the correlation matrix defined by the localization function.
- Sequence of control vectors $\mathbf{v}_j$ concatenated to make $\mathbf{v}$;
- We can define a new operator $\underline{\mathbf{U}}^{\alpha s}$ to represent (9) and (10):

$$\delta\underline{\mathbf{x}} = \underline{\mathbf{U}}^{\alpha s}\mathbf{v}. \tag{12}$$

4D-En-Var

- The cost function:

$$J[\mathbf{v}] = \frac{1}{2}\mathbf{v}^T\mathbf{v} + [\underline{\mathbf{H}}\delta\underline{\mathbf{x}} - \underline{\mathbf{d}}]^T\underline{\mathbf{R}}^{-1}[\underline{\mathbf{H}}\delta\underline{\mathbf{x}} - \underline{\mathbf{d}}] \quad (13)$$

- The cost function gradient:

$$\left[\frac{\partial J}{\partial \mathbf{v}}\right] = \mathbf{v} + \underline{\mathbf{U}}^{\alpha s\,T}\underline{\mathbf{H}}^T\underline{\mathbf{R}}^{-1}[\underline{\mathbf{H}}\delta\underline{\mathbf{x}} - \underline{\mathbf{d}}], \quad (14)$$

- For our experiments $\underline{\boldsymbol{\alpha}}_j$ is constant in time, which removes the time dimension from $\mathbf{v}_j$;
- The analysis can be taken at any point in the assimilation window;
- In our experiments we take the analysis at timestep 1 (if all the obs are located at timestep 1) or at timestep 4 (if the obs are evenly distributed between timesteps 2-6).

Generating the ensemble in 4D-En-Var

Either, we can

1. Use an ensemble of background states from a separate DA method to calculate the perturbations. The analysis comes from minimizing (13);
2. Or, 4D-En-Var can be run for each member to generate its own ensemble i.e. (13) is minimized for each ensemble member. The cost function for each ensemble member uses the same background error covariance matrix, but uses the ensemble member to calculate the innovations in (13). This is how 4D-En-Var is formulated in these experiments; The analysis comes from the ensemble mean.

Maintaining the spread of 4D-En-Var

Like EnKF methods, the spread of 4D-En-Var must also be maintained. Two methods are compared in the experiments:

1. Perturbed observations (as in the EnKF of Evensen [1994]). This will be referred to as 4D-En-Var;
2. Halving the influence of the analysis perturbations (equivalent to the DEnkf but applied as post-processing step):

$$\begin{aligned} \mathbf{x}_j^a(t_0) \leftarrow \overline{\mathbf{x}}^a(t_0) \quad &+ \quad 0.5(\mathbf{x}_j^b(t_0) - \overline{\mathbf{x}}^b(t_0)) \\ &+ \quad 0.5(\mathbf{x}_j^a(t_0) - \overline{\mathbf{x}}^a(t_0)). \end{aligned} \tag{15}$$

This will be referred to as 4D-En-VarD:

Advantages of 4D-En-Var

- Flow-dependent background error covariance matrix;
- Avoids complex tangent linear and adjoint models of 4D-Var;
- It shares many of the features of 4D-Var (e.g. minimization algorithm).

Disadvantages of 4D-En-Var

- Added expense over an EnKF/EnKS;
- In NWP it requires localization to make the covariance matrix full rank;
- Localization function and the nonlinear model do not commute $\longrightarrow$ 4D structure of 4D-En-Var degraded by severe localization (Bishop and Hodyss, 2011);

4D-Var-Ben

- 4D-Var-Ben is identical to 4D-Var except it uses flow-dependent background error covariance matrix ($\mathbf{P}_e^b(t_0)$) from 4D-En-Var instead of using the climatological **B**.
- The 4D-Var-Ben increment at the beginning of the window has the same equation as the 4D-En-Var increment (9):

$$\delta\mathbf{x}(t_0) = \sum_{j=1}^{m} \frac{1}{\sqrt{m-1}} \delta\mathbf{x}_j^b(t_0) \circ \boldsymbol{\alpha}_j, \tag{16}$$

- It differs to 4D-En-Var in that the increment in (16) is propagated forwards in time using the tangent linear model.

Advantages of 4D-Var-Ben

- Flow-dependent background error covariance matrix;
- Takes advantage of the tangent linear model in 4D-Var, which calculates implicit 4D background error covariance $\underline{\mathbf{M}}\mathbf{P}_e^b\underline{\mathbf{M}}^T$;
- 4D-Var-Ben may be more accurate than 4D-En-Var in NWP, since it provides accurate time-correlations of background errors (within the linear approximation of $\mathbf{M}$), which allow observations distributed in time to be effectively used.

Disadvantages of 4D-Var-Ben

- Requires complex tangent linear and adjoint models;
- In NWP it requires localization to make the covariance matrix full rank;

DEnkf

- The DEnkf of Sakov and Oke [2008] is an approximation of the ensemble square root filter (EnSRF, Whitaker and Hamill, 2002).
- We use it as a control to test 4D-En-Var with a DEnkf analysis perturbation update step;
- The deterministic analysis $\mathbf{x}^a$ comes from the ensemble mean:

$$\overline{\underline{\mathbf{x}}}^a = \overline{\underline{\mathbf{x}}}^b + \underline{\mathbf{K}}(\underline{\mathbf{y}}^o - \overline{\underline{H}(\underline{\mathbf{x}}^b)}). \tag{17}$$

- The Kalman gain determines the weight to give to the background and the observations:

$$\mathbf{K}_i = (\mathbf{P}^b(t_i))^T \mathbf{H}_i^T (\mathbf{H}_i \mathbf{P}^b(t_i) \mathbf{H}_i^T + \mathbf{R}_i)^{-1}, \tag{18}$$

where

$$\mathbf{P}^b = \frac{1}{\sqrt{m-1}} (\mathbf{X}^b)^T \mathbf{X}^b \tag{19}$$

and

$$\mathbf{X}^b = \delta \mathbf{x}_j^b, j = 1, ..., m. \tag{20}$$

DEnkf

- In these experiments, the DEnkf is formulated by serially assimilating the observations;
- Formulated as a fixed-lag Kalman smoother within the assimilation window;
- The localization function **C** is applied as a Schur product with **K**;
- The DEnkf maintains the ensemble spread by halving the Kalman gain matrix in the analysis perturbation update step:

$$\delta \mathbf{x}_j^a(t_i) = \delta \mathbf{x}_j^b(t_i) - \frac{1}{2}\mathbf{K}_i \mathbf{H}_i(\delta \mathbf{x}_j^b(t_i))). \tag{21}$$

This is equivalent to halving the influence of the analysis perturbations (as used by 4D-En-VarD).

Key differences between the methods

Key differences between the methods:

1. 4D-Var uses a climatological background error covariance matrix. The other methods use flow-dependent background error covariance matrices, which are generated from an ensemble of model trajectories;
2. 4D-Var/4D-Var-Ben covariances evolved implicitly with Tangent linear/adjoint models, whilst DEnKF/4D-En-Var covariances evolved explicitly with nonlinear model in subspace of background ensemble;

Motivation

In 2009, a group of DA scientists at Environment Canada compared variational and EnKF approaches for producing a global deterministic analysis in NWP (Buehner et al., 2010):

1. 4D-Var;
2. 4D-En-Var;
3. 4D-Var-Benkf (4D-Var with a flow-dependent background error covariance matrix coming from EnKF);
4. EnKF;

Their results

They judged the methods by the accuracy of short and medium range deterministic forecasts. They found that:

1. 4D-Var performed marginally better (worse) than the EnKF for short-range (medium-range) forecasts in the extratropics;
2. 4D-En-Var peformed as well as the EnKF;
3. 4D-Var-BEnKF showed large (modest) improvements compared with 4D-Var in Southern extratropics (tropics);

Possible explanations for their results

- Observation density: Greater conventional obs density in Northern extratropics than Southern extratropics $\rightarrow$ Do sparser obs favour flow-dependent covariance matrices?
- Observation type: Much lower volume of surface obs in Southern extratropics, thus much higher reliance on satellite radiances. Satellite radiances only indirectly related to analysis variables temperature and humidity; $\rightarrow$ Can flow-dependent covariance matrices capture better indirect observations?;
- Differences in temporal/spatial localization of covariances.

The aim of our toy model experiments

- We aim to replicate some of their experiments using toy models, to gain better understanding of their results;
- The methods we wish to compare are similar to their methods;
- Toy model experiments allow the truth to be caluculated exactly (no ambiguity);
- Spatial and temporal obs density can be varied to simulate differences in conventional obs coverage between Northern and Southern Extratropics;

Key results from our experiments

- The flow-dependent data assimilation methods perform better than 4D-Var when the ensemble size is sufficent;
- When severe localization is required, 4D-Var-Ben performs better than 4D-En-Var and the DEnkf;
- 4D-En-VarD (deterministic) performs better than 4D-En-Var (stochastic);
- 4D-En-VarD and the DEnkf perform similarly for a large ensemble size, but 4D-En-VarD performs better for a small ensemble size.

The Lorenz 2005 Model 2 (Lorenz, 2004)

$$\frac{dX_n}{dt} = [X, X]_{K,n} - X_n + F, \tag{22}$$

where

$$[X, Y]_{K,n} = \sum_{j=-J}^{J} \sum_{i=-J}^{J} (-X_{n-2K-i} X_{n-K-j} + X_{n-K+j-i} X_{n+K+j})/K^2. \tag{23}$$

- Advection of waves across a latitude circle;
- Quadratic advection term, linear damping term and forcing term F;
- Increasing F increases the nonlinearity of the system;
- Increasing the summation K increases the average wave length and allows for smooth spatial correlations;

Perfect/Imperfect models

1. Perfect model experiments (identical twin): $n = 40, K = 1, J = 0, F = 10$. Error doubling time $\equiv$ 0.3 time units (about 1.5 days)
2. Imperfect model experiments: Model error from truth model at higher resolution than DA model:
   - Truth: n=240, K=8, J=4, F=15;
   - DA: n = 180, K=6, J=3, F=15.

   Error doubling time $\equiv$ 0.4 time units (about 2 days)

Experimental parameters

- Timestep 0.01 units with 6 timesteps in the window;
- Integration with 4th order Rungekutta method;
- Window length 0.05 units (about 6 hours in the atmosphere);
- A range of observation densities for two scenarios:
  1. 25,50,75,100,125,150 random obs at timestep 1 in the window;
  2. 5,10,15,20,25,30 random obs at timesteps 2-6 in the assimilation window.

  The first scenario means the time correlation of the observations is irrelevant;
  The second scenario is a test of the time correlation;
- Observation error uncorrelated and drawn from a Gaussian distribution with standard deviation 0.1 ($\mathbf{R} = 0.01\mathbf{I}$);
- Range of ensemble sizes: 3,4,5,9,13,17 (to explore a range of different sampling errors).

Experimental parameters

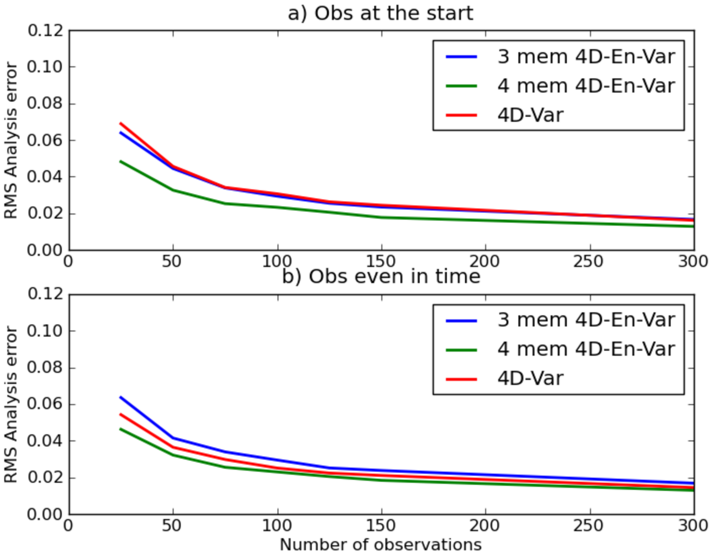
- Sampling error alleviated by tuned localization and tuned fixed covariance inflation;
- Gaspari and Cohn [1999] Localization function;
- Perfect climatological background error covariance matrix tuned for 4D-Var.

Tuned localization/inflation

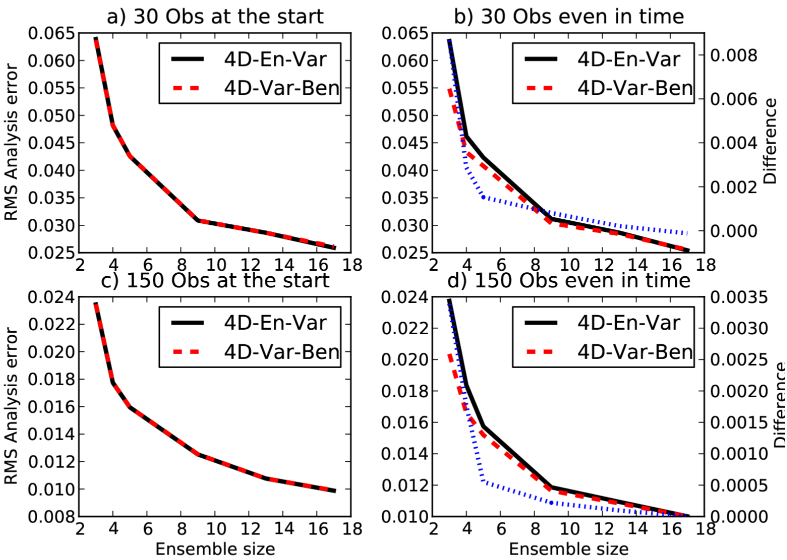
Ensemble size	Gaspari-Cohn half width	Inflation
3	0.125	1.11
4	0.2	1.08
5	0.275	1.09
9	0.5	1.03
13	0.75	1.03
17	2.5	1.02

Table: Tuned localizations/fixed inflations for various ensemble sizes in 4D-En-Var (perfect model).

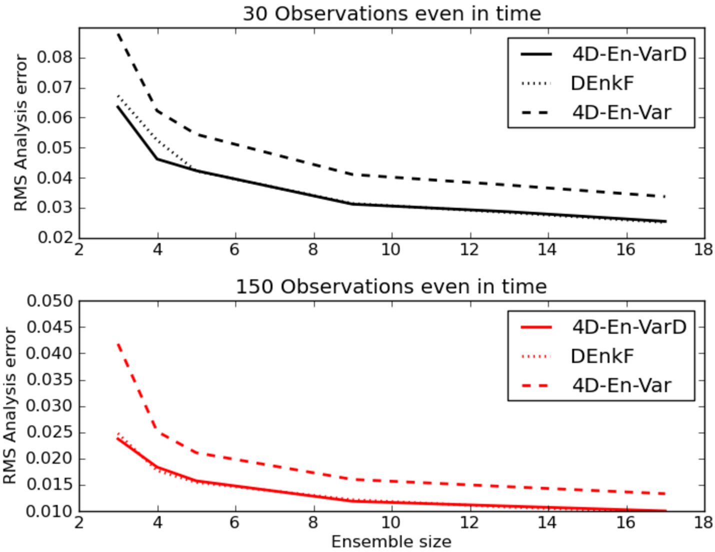
Perfect model: 4D-En-VarD vs 4D-Var



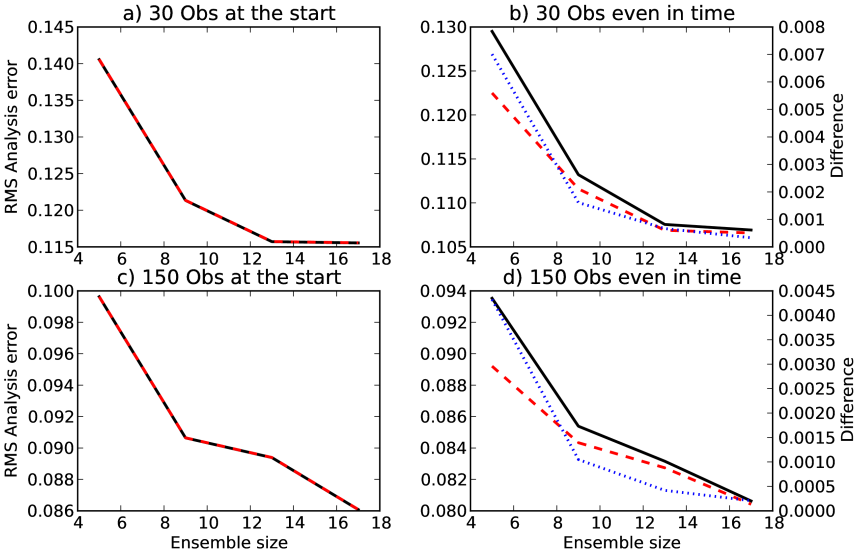
Perfect model: 4D-En-VarD vs 4D-Var-Ben



Perfect model: 4D-En-VarD vs DEnkf



ImPerfect model: 4D-En-VarD vs 4D-Var-Ben



Conclusions

The toy model experiments showed that:

- The flow-dependent data assimilation methods perform better than 4D-Var (when the ensemble size is sufficent), due to the importance of measuring the 'errors of the day' in the background error covariance matrix (agrees with Kalnay et al. [2007], Zhang et al. [2009]);
- No clear evidence here that 4D-Var is more effective at handling dense observations than the flow-dependent methods;
- When severe localization is required, 4D-Var-Ben performs better than 4D-En-Var and the DEnkf, since the tangent linear model is more effective at evolving the covariance matrix than the background ensemble; (similar to results by Zhang et al. [2009] and Bishop and Hodyss [2011]);
- 4D-En-VarD performs better than 4D-En-Var due to sampling error from perturbed observations (similar to results by Whitaker and Hamill [2002]);
- The serial assimilation of observations makes the DEnkf perform worse than 4D-En-VarD for a small ensemble. Otherwise they are similar.

- The results show some similarities with the operational results in Buehner et al. [2010];
- In NWP, severe localization is required since the ensemble dimension $\ll$ model dimension;
- However, the toy model is vastly less complicated than the NWP model e.g. unable to capture important balance assumptions;
- DA less complicated than operational systems e.g. no outer loop in 4D-Var;
- We did not simulate different observation types i.e. Conventional/radiance obs;

Future work

- Repeat the experiments for the Lorenz Model 3, which has two different scales;
- What is the effect of model error on the performance of all the methods?

References

C. H. Bishop and D. Hodyss. Adaptive ensemble covariance localization in ensemble 4d-var state estimation. *mon. Wea. Rev*, 139:1241–1255, 2011.

M. Buehner, P. Houtekamer, C. Charette, H. L. Mitchell, and B. He. Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part 1: Description and Single-Observation Experiments. Part 2: One-Month Experiments with Real Observations. *Mon. Wea. Rev*, 138:1550–1586, 2010.

P. Courtier, J. Thepaut, and A. Hollingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q.J.R. Meteorol. Soc*, 120:1389–1408, 1994.

G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res*, 99:10143–10162, 1994.

G. Gaspari and S. Cohn. Construnction of correlation functions in two and three dimensions. *Q.J.R. Meteorol. Soc*, 125:723–757, 1999.

E. Kalnay, H. Li, T. Miyoshi, S. Yang, and J. Ballabrera-Poy. 4D-Var or ensemble Kalman filter. *Tellus*, 59A:758–773, 2007.

F. X. Le-Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, 38A: 97–110, 1986.

C. Liu, Q. Xiao, and B. Wang. An ensemble-based four-dimensional variational data assimilation scheme. part i: Technical formulation and preliminary test. *Mon. Weather Rev*, 136:3363–3373, 2008a.

C. Liu, Q. Xiao, and B. Wang. An ensemble-based four-dimensional variational data assimilation scheme. part ii: Observing System Simulation Experiments with advanced research WRF (ARW). *Mon. Weather Rev*, 136:3363–3373, 2008b.

E. N. Lorenz. Designing Chaotic Models. *J. Atmos. Sci*, 62:1574–1587, 2004.

P. Sakov and P. R. Oke. A deterministic formulation of the ensemble kalman filter: an alternative to ensemble square root filters. *Tellus*, 60A:3035–3049, 2008.

J. S. Whitaker and T. M. Hamill. Ensemble Data Assimilation without Perturbed Observations. *Mon. Weather Rev*, 130:1913–1924, 2002.

F. Zhang, M. Zhang, and J. A. Hansen. Coupling ensemble Kalman filter with four- dimensional variational data assimilation. *Adv. Atmos. Sci*, 26(1):1–8, 2009. doi: 10.1007/s00376-009-0001-8.