



Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction

M. Buehner¹, J. Morneau², and C. Charette¹

¹Data Assimilation and Satellite Meteorology Research Section, Environment Canada, Dorval, Quebec, Canada

²Data Assimilation and Quality Control Development Section, Environment Canada, Dorval, Quebec, Canada

Correspondence to: M. Buehner (mark.buehner@ec.gc.ca)

Received: 27 May 2013 – Revised: 16 July 2013 – Accepted: 5 August 2013 – Published: 24 September 2013

Abstract. The goal of this study is to evaluate a version of the ensemble-variational data assimilation approach (EnVar) for possible replacement of 4D-Var at Environment Canada for global deterministic weather prediction. This implementation of EnVar relies on 4-D ensemble covariances, obtained from an ensemble Kalman filter, that are combined in a vertically dependent weighted average with simple static covariances. Verification results are presented from a set of data assimilation experiments over two separate 6-week periods that used assimilated observations and model configuration very similar to the currently operational system. To help interpret the comparison of EnVar versus 4D-Var, additional experiments using 3D-Var and a version of EnVar with only 3-D ensemble covariances are also evaluated. To improve the rate of convergence for all approaches evaluated (including EnVar), an estimate of the cost function Hessian generated by the quasi-Newton minimization algorithm is cycled from one analysis to the next.

Analyses from EnVar (with 4-D ensemble covariances) nearly always produce improved, and never degraded, forecasts when compared with 3D-Var. Comparisons with 4D-Var show that forecasts from EnVar analyses have either similar or better scores in the troposphere of the tropics and the winter extra-tropical region. However, in the summer extra-tropical region the medium-range forecasts from EnVar have either similar or worse scores than 4D-Var in the troposphere. In contrast, the 6 h forecasts from EnVar are significantly better than 4D-Var relative to radiosonde observations for both periods and in all regions. The use of 4-D versus 3-D ensemble covariances only results in small improvements in forecast quality. By contrast, the improvements from using 4D-Var versus 3D-Var are much larger. Measurement of the fit of the background and analyzed states to the observations

suggests that EnVar and 4D-Var can both make better use of observations distributed over time than 3D-Var. In summary, the results from this study suggest that the EnVar approach is a viable alternative to 4D-Var, especially when the simplicity and computational efficiency of EnVar are considered. Additional research is required to understand the seasonal dependence of the difference in forecast quality between EnVar and 4D-Var in the extra-tropics.

1 Introduction

For more than a decade, numerical weather prediction (NWP) centers have been increasingly adopting the four-dimensional variational data assimilation (4D-Var) approach for global (Rabier et al., 2000; Rabier, 2005; Rawlins et al., 2007; Gauthier et al., 2007) and regional (Honda et al., 2005; Tanguay et al., 2012) deterministic prediction. This has contributed to significant improvements in analysis and forecast quality. During the same period, ensemble data assimilation approaches, including the Ensemble Kalman Filter (EnKF; Houtekamer and Mitchell, 1998; Burgers et al., 1998), have become increasingly used for initializing ensemble forecasts (Charron et al., 2010) and for providing flow-dependent background-error statistics used to produce deterministic analyses (Clayton et al., 2012). Both 4D-Var and ensemble data assimilation approaches rely on output from forecast models, though in different ways, within the data assimilation procedure when using observations to compute a correction to a short-term forecast (i.e., the background state). Several past studies have compared these approaches from both theoretical (Lorenc, 2003; Kalnay et al., 2007; Gustafsson, 2007) and empirical (Caya et al., 2005;

Whitaker et al., 2009; Buehner et al., 2010a, b; Miyoshi et al., 2010; Zhang et al., 2011) perspectives. The present study focuses on a comparison between 4D-Var and another variational approach, called ensemble-variational data assimilation (EnVar), which relies heavily on the ensembles produced by an ensemble data assimilation approach, the EnKF (Houtekamer and Mitchell, 2005; Houtekamer et al., 2009) in our case. The comparison is made in a context very close to that of the systems currently operational at Environment Canada.

To make more efficient use of the limited resources available and accelerate the development of future NWP systems, there is currently an effort at Environment Canada to move towards a more unified approach for data assimilation. The current situation of developing separate state-of-the-art assimilation approaches and software libraries for the deterministic and ensemble prediction systems will gradually be replaced by systems that all make effective use of ensembles and share large amounts of computer code for common tasks. Consequently, any improvements made to the quality of the ensembles should benefit both the deterministic and ensemble prediction systems. The EnVar approach uses 4-D ensemble covariances in a way that essentially replaces the use of tangent-linear and adjoint versions of the forecast model in 4D-Var. The use of 4-D ensemble covariances in EnVar is also similar to how they are used within the EnKF itself (Hunt et al., 2004; Houtekamer and Mitchell, 2005; Buehner et al., 2010a). Since a significant effort is required to develop and maintain computationally efficient tangent-linear and adjoint models, replacing 4D-Var with an EnVar approach would significantly reduce the effort required to further develop the data assimilation component of the deterministic prediction systems.

The goal of this study is to evaluate a version of EnVar for possible replacement of 4D-Var in the operational global deterministic prediction system (GDPS) at Environment Canada. The configurations of the systems included in this study were chosen specifically with this goal in mind. For example, the horizontal resolution of the analysis increment in EnVar is chosen to match the resolution of the operational EnKF, even though this resolution is higher than the analysis increment in 4D-Var. In the context of comparing approaches for potential operational use, such a difference in resolution is appropriate because of the significantly lower computational cost of EnVar as compared with 4D-Var. Similarly, no experiments were performed using 4D-Var in combination with an ensemble-based covariance matrix (as in the 4D-Var-Benkf experiment by Buehner et al. (2010a, b) and the system currently operational at the United Kingdom Meteorological Office, described by Clayton et al., 2012) instead of the simple static background-error covariance matrix. While such an approach provides improved background-error covariances for 4D-Var, it does not result in any reduction in the effort required to maintain and develop the system. Finally, the direct use of EnKF analyses for initializing

deterministic forecasts was also not tested, since it is not currently being considered as a possible replacement of 4D-Var. This is mostly because the horizontal spatial resolution, vertical extent, and volume of assimilated observations are all significantly lower in the EnKF than in the deterministic system. In addition, an approach was evaluated for accelerating the minimization of the EnVar cost function by cycling an estimate of the Hessian generated by the quasi-Newton minimization algorithm.

The previous study of Buehner et al. (2010a, b) does include comparisons with the approaches just mentioned in a context where all experiments use the same spatial resolution (for the analysis increment), model configuration and set of assimilated observations. In that study, using ensemble covariances in 4D-Var resulted in improved forecast scores when compared with either 4D-Var using simple static covariances or the EnKF ensemble mean analysis. This is also consistent with the results obtained in an idealized context with a low-dimensional toy model by Fairbairn et al. (2013). A similar comparison of 4D-Var using either simple static background-error covariances or ensemble covariances by Kuhl et al. (2013) also showed significant forecast improvements from using the ensemble covariances.

In the next section the configurations of the variational data assimilation approaches evaluated in this study are described. Verification results from a set of data assimilation experiments with the full set of operationally assimilated observations are presented in Sect. 3. In Sect. 4, some simple diagnostic results are presented that demonstrate the ability of EnVar to represent the temporal dimension within the assimilation window in comparison with other approaches. Finally, some conclusions are given in Sect. 5.

2 Data assimilation approaches evaluated

Several data assimilation approaches were chosen for evaluation with the goal of understanding how EnVar compares with 4D-Var in the context of operational global deterministic weather prediction. Each approach was tested in 6 week data assimilation experiments for each of two seasons with the same configuration of the forecast model and the same set of assimilated observations, both being very similar to the system implemented operationally at Environment Canada on 13 February 2013. Briefly, the Global Environmental Multiscale (GEM) forecast model is configured with a uniform horizontal latitude-longitude grid with 1024 by 800 grids points (resulting in a grid spacing of about 25 km at 50° latitude) and 80 vertical levels with the top level at 0.1 hPa (CMC, 2013). The operationally assimilated observations include those from radiosondes; aircraft; wind profilers; land stations, ships and buoys (near-surface observations); scatterometers; atmospheric motion vectors; satellite-based radio occultation; and microwave and infrared satellite sounders and imagers. All experiments use an incremental

approach to generate an analysis at the spatial resolution of the forecast model from an analysis increment computed on a lower resolution horizontal grid and a slightly different set of vertical levels. The 4D-Var experiment uses 2 iterations of an outer-loop during which the high-resolution forecast model is integrated to obtain an updated trajectory for the tangent-linear and adjoint models and an updated measure of the fit to the observations. The first inner-loop minimization uses 35 iterations and the second uses 30 iterations. The other assimilation approaches all use only a single inner-loop minimization with 70 iterations.

EnVar is tested using 4-D background-error covariances obtained from 192 EnKF background ensemble members stored every hour over the 6 h assimilation window, as they are also used in the EnKF version that was also implemented operationally on 13 February 2013. To evaluate the impact of using 4-D covariances, another experiment using only the 3-D covariances valid at the centre of the assimilation window was performed. When discussing these experiments the terms 4D-EnVar and 3D-EnVar will be used to distinguish between the two, however, elsewhere the simpler name of EnVar is used to refer to the experiment with 4-D covariances. Note that 3D-EnVar is similar to the approach called 3D-Var-Benkf and 4D-EnVar is similar to the approach called En-4D-Var by Buehner et al. (2010a, b).

In addition, a 3D-Var experiment is also included in the comparisons. The same static and highly parameterized background-error covariances are used in both 3D-Var and 4D-Var. These are generated using lagged forecast differences (48 h forecasts minus 24 h forecasts valid at the same time) following the so-called “NMC method” (Parrish and Derber, 1992; see Charron et al., 2012 for details). The horizontal resolution for the analysis increment in the 3D-Var and both EnVar experiments is chosen to match the resolution of the EnKF, which uses a Gaussian grid with 600 by 300 grid points (grid spacing of about 66 km at the equator). The 4D-Var experiment computes an analysis increment at lower resolution on a Gaussian grid with 400 by 200 grid points (grid spacing of about 100 km at the equator), as in the system that became operational on 13 February 2013.

The EnVar experiments in this study use hybrid background-error covariances (Hamill and Snyder, 2000) that are a weighted average of the flow-dependent 4-D (or 3-D) ensemble covariances (\mathbf{B}_{enkf}) and the same static covariances used in 3D-Var and 4D-Var (\mathbf{B}_{nmc}). The approach for incorporating spatially localized ensemble covariances within a preconditioned cost function, including the spatial localization parameters used, are the same as described by Buehner et al. (2010a). Additional details related to the approach are given by Bishop et al. (2011). Unlike the study of Buehner et al. (2010a), the ensemble covariances are combined with the static covariances by computing the analysis increment

$$\Delta \mathbf{x} = \beta_{\text{nmc}}^{1/2}(\text{level}_k) \mathbf{B}_{\text{nmc}}^{1/2} \boldsymbol{\xi}_{\text{nmc}} + \beta_{\text{enkf}}^{1/2}(\text{level}_k) \mathbf{B}_{\text{enkf}}^{1/2} \boldsymbol{\xi}_{\text{enkf}}, \quad (1)$$

where the two β factors control the contributions of \mathbf{B}_{nmc} and \mathbf{B}_{enkf} as a function of the vertical level, and the vectors $\boldsymbol{\xi}_{\text{nmc}}$ and $\boldsymbol{\xi}_{\text{enkf}}$ are the portions of the control vector associated with each covariance matrix. The complete control vector

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_{\text{nmc}} \\ \boldsymbol{\xi}_{\text{enkf}} \end{bmatrix} \quad (2)$$

is used by the minimization algorithm to find the minimum of the preconditioned cost function

$$J(\boldsymbol{\xi}) = \frac{1}{2} \boldsymbol{\xi}^T \boldsymbol{\xi} + \frac{1}{2} \left[H(\mathbf{x}^b) + \mathbf{H} \Delta \mathbf{x}(\boldsymbol{\xi}) - \mathbf{y} \right]^T \mathbf{R}^{-1} \left[H(\mathbf{x}^b) + \mathbf{H} \Delta \mathbf{x}(\boldsymbol{\xi}) - \mathbf{y} \right], \quad (3)$$

where $H(\mathbf{x}^b)$ is the nonlinear observation operator applied to the background state trajectory, \mathbf{H} is the tangent linear version of $H(\bullet)$, \mathbf{y} is the vector containing all observations being assimilated and \mathbf{R} is the observation-error covariance matrix. The analysis increment, $\Delta \mathbf{x}$, in Eq. (3) is obtained from the control vector using Eqs. (1) and (2). The analysis, \mathbf{x}_a , is then obtained by summing the analysis increment and the background state, all valid at the middle of the 6 h assimilation window.

Because the top model level in the operational EnKF (2 hPa) is lower than the top model level of the GDPS (0.1 hPa), the ensemble covariances are not available for the upper portion of the GDPS levels. To overcome this, the weighting between the two matrices (controlled by the β factors in Eq. 1) depends on the vertical level and gradually changes from being equal ($\beta_{\text{nmc}} = \beta_{\text{enkf}} = 0.5$) from the surface up to about 40 hPa to become fully weighted towards the \mathbf{B}_{nmc} matrix ($\beta_{\text{nmc}} = 1.0$, $\beta_{\text{enkf}} = 0.0$) above 10 hPa and up to the top level of the GDPS (see Fig. 1 in which the vertical model co-ordinate very approximately equals the pressure divided by the surface pressure, more details on this relationship in CMC, 2013). Consequently, above 10 hPa the EnVar analysis is nearly equivalent to 3D-Var and therefore cannot be expected to be as good as 4D-Var. A brief summary of the four different data assimilation approaches is given in Table 1. An alternative weighting between the two covariances matrices was tested in preliminary experiments with more weight given to \mathbf{B}_{enkf} in the troposphere ($\beta_{\text{nmc}} = 0.25$, $\beta_{\text{enkf}} = 0.75$). The resulting forecast scores were similar or slightly worse than when using equal weighting and are not included in this study. It should be noted that some additional vertical covariance localization is imposed by vertically varying the weighting between the two covariance matrices. This results from the assumption that the increments from the two matrices are independent of each other. However, the use of a gradual variation in the weightings, spread out over many model levels, minimizes this effect.

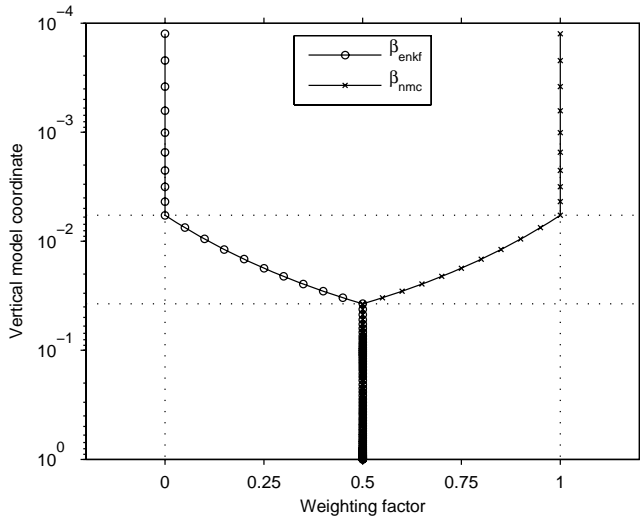


Fig. 1. The hybrid covariance weights used in computing the average of the \mathbf{B}_{Enkf} (line with circles) and \mathbf{B}_{nmc} (line with crosses) covariance matrices for the EnVar experiments.

Like 4D-Var, the EnVar approach could potentially also use an outer-loop to provide an updated measure of the fit to observations for computing the cost function. This can be accomplished by producing an analysis at the beginning of the assimilation window for initializing a model integration over the window. However, since the increment is temporally constant above 10 hPa and below this level the increment still has a significant contribution from the static background-error covariances that are temporally constant, the analysis increment computed at the beginning of the window is not fully appropriate for that time. Consequently, significant errors would result from starting a model integration from the EnVar analysis at the beginning of the assimilation window. Nevertheless, such an approach was attempted with the result of significantly degraded forecast scores that are also not included in this study.

It should be noted that the computational cost of the EnVar approach is currently significantly lower than the cost of 4D-Var at Environment Canada. Since the EnKF is already an operational system, the cost of producing the ensemble of background states used in both the EnKF and EnVar is not considered as being associated with EnVar. Given this fact, the cost of only producing an EnVar analysis requires less than one fifth of the time and less than half of the number of processors as compared with 4D-Var, even though the analysis increment horizontal resolution is significantly higher in EnVar than in 4D-Var.

The minimization algorithm M1QN3 (Gilbert and Lemaréchal, 1989) is used in the operational system and in the experiments performed for this study. This algorithm is a limited-memory quasi-Newton approach that generates an approximate estimate of the cost function Hessian during the minimization to accelerate convergence. The operational

4D-Var and all of the experiments included in this study use the cost function Hessian estimated during the previous analysis cycle to initialize the Hessian for the current analysis. While being very efficient and simple to implement, this approach results in a significantly improved convergence of the minimization problem when using a fixed number of iterations. The full Hessian depends on the specified background-error covariances, observation error covariances and the observation operator. Therefore, if these three quantities remain relatively similar over time, the strategy of cycling the Hessian from one analysis time to the next can be beneficial, with negligible additional computational cost. Figure 2 shows the impact of cycling the Hessian in the context of EnVar. For a single analysis time well after the beginning of the data assimilation experiment, the minimization is performed both with (dashed line) and without (solid line) the estimate from the previous analysis used to initialize the Hessian. Even though the background-error covariances change each analysis time in EnVar, the impact on the rate of convergence of using the Hessian from the previous analysis is clearly positive. The total cost function (Fig. 2a) is reduced much more rapidly with Hessian preconditioning, but the value is nearly the same as without Hessian preconditioning after 70 iterations. The impact on the total observation cost function (Fig. 2b) and the cost function component associated with only the satellite radiance observations (Fig. 2c) and only the aircraft observations (Fig. 2d) are also shown. This shows that the fit of the analysis to the observations, most noticeably for the aircraft observations, is improved when using Hessian preconditioning. Consequently, even in EnVar in which all three quantities (background- and observation-error covariances and the observation operator) change from one analysis time to the next, these changes are small enough such that the strategy of cycling the Hessian estimated by the minimization algorithm is still effective.

3 Forecast verification results

In this section, forecast verification scores are presented from using analyses produced by 4D-Var, 3D-Var and versions of EnVar that use either 3-D or 4-D ensemble covariances. For each approach, the data assimilation experiments span the periods 1 February to 14 March 2011 and 1 July to 14 August 2011. First the impact of using EnVar analyses instead of either 4D-Var or 3D-Var analyses is shown. Then the impact of using 4-D versus 3-D ensemble covariances within EnVar is examined. Though not entirely analogous with this last comparison, it is also interesting to evaluate the impact of including the time dimension in 4D-Var versus 3D-Var.

3.1 EnVar versus 4D-Var and 3D-Var

Figure 3 shows the standard deviation (solid curves) and bias (dashed curves) relative to radiosonde observations of

Table 1. Summary of experiments.

Experiment name	Increment/analysis grid spacing	Minimization iterations	B matrix
EnVar or 4D-EnVar	66 km/25 km	70	Hybrid: 4-D \mathbf{B}_{enkf} + 3-D \mathbf{B}_{nmc}
3D-EnVar	66 km/25 km	70	Hybrid: 3-D \mathbf{B}_{enkf} + 3-D \mathbf{B}_{nmc}
4D-Var	100 km/25 km	35 + 30 (2 outer-loop iterations)	3-D \mathbf{B}_{nmc}
3D-Var	66 km/25 km	70	3-D \mathbf{B}_{nmc}

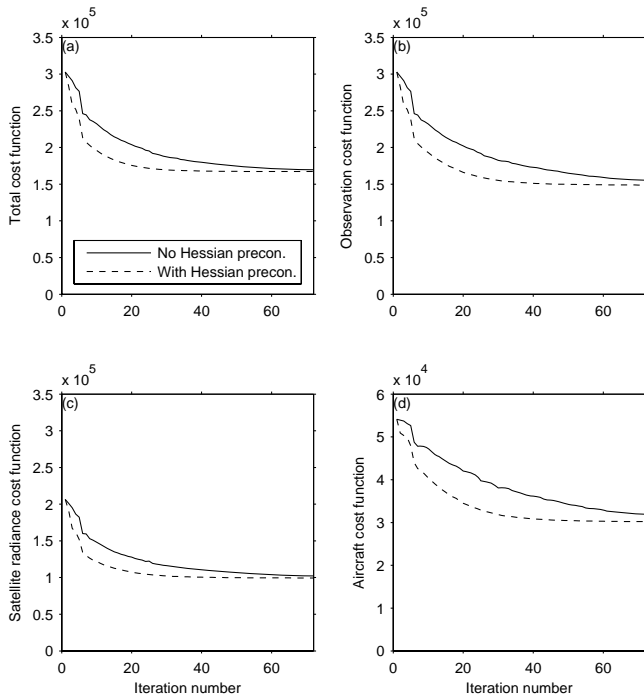


Fig. 2. Evolution of (a) the total cost function, (b) the total observation cost function, (c) the cost function for satellite radiance observation, and (d) the cost function for aircraft observations. These are shown for a single EnVar minimization without Hessian preconditioning (solid line) and with Hessian preconditioning from the previous analysis cycle (dashed line).

the 6 h forecasts (forecast-minus-observed values) from the EnVar (red) and 4D-Var (blue) experiments over the February/March period. The results are shown for temperature (upper panels) and zonal wind (lower panels) on pressure levels between 100 hPa and 1000 hPa and for three regions: the northern extra-tropics (left panels), tropics (middle panels) and southern extra-tropics (right panels). The small colored boxes indicate the level of statistical significance with which the scores for the two experiments can be considered distinct from each other and are only shown when the significance level is 90 % or higher. The color of the boxes indicates which experiment has the lower magnitude for standard deviation or bias. The predominantly red shading of most of the boxes on the right of each panel indicates that the EnVar experiment produces 6 h forecasts with significantly smaller

standard deviation than 4D-Var (however note the small, but significant degradation for temperature near 250 hPa in the northern extra-tropics with EnVar relative to 4D-Var). The boxes on the left of each panel indicate that the bias is not consistently larger or smaller in magnitude for EnVar than 4D-Var, except possibly for zonal wind in the tropics, which is significantly smaller for EnVar than 4D-Var for almost all levels. Figure 4 shows similar results for the EnVar and 4D-Var experiments, except for the July/August period. The results for this period are quite similar to the February/March period.

The next set of verification scores are computed relative to the ERA-Interim reanalyses (Dee et al., 2011). The standard deviation of the difference between forecasts and the reanalyses was computed for all experiments after interpolating the forecasts with spatial averaging onto a coarse resolution global 1.5° latitude-longitude grid. Figure 5 shows contours of the differences in these standard deviations computed from the EnVar and 3D-Var experiments for the February/March period as a function of pressure from 100 hPa to 1000 hPa and lead time every 24 h from 0 h to 120 h. Negative values correspond with a lower standard deviation for EnVar than for 3D-Var. These are shown for geopotential height for both the northern extra-tropics (Fig. 5a) and the southern extra-tropics (Fig. 5c) and for zonal wind in the tropics (Fig. 5b). The forecasts initialized with EnVar analyses have better verification scores than with 3D-Var analyses (i.e., smaller standard deviations relative to the reanalyses) for all regions, levels and lead times, except for some later lead times in the tropics near the surface for which the scores are nearly equal. Interestingly, the impact in the tropics is largest at the shortest lead time, whereas in the extra-tropics the impact is largest at 120 h, most noticeably near the tropopause. Figure 6 shows similar results as the previous figure, except for the July/August period. These results are similar as for the February/March period, except that the difference in standard deviation is smaller in the northern extra-tropics (summer season) and larger in the southern extra-tropics (winter season).

Similar to the previous two figures, Figs. 7 and 8 show the difference in standard deviation relative to the ERA-Interim reanalyses between the EnVar and 4D-Var experiments. For the February/March period, shown in Fig. 7, the forecast scores are only very slightly better for EnVar as compared with 4D-Var in the northern extra-tropics and worse in

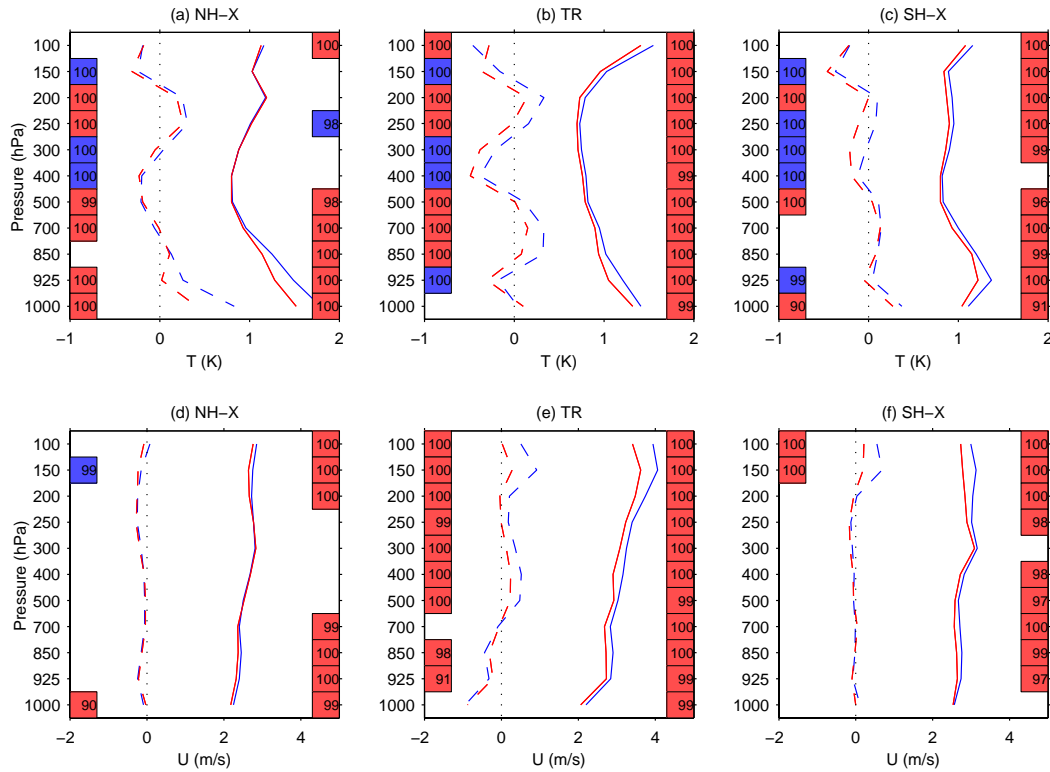


Fig. 3. The standard deviation (solid curves) and bias (dashed curves) for the 6 h forecasts produced by the EnVar (red) and 4D-Var (blue) experiments for the February/March 2011 period relative to radiosonde observations. The results are shown for temperature (upper panels) and zonal wind (lower panels) for the northern extra-tropics (left panels), tropics (middle panels) and southern extra-tropics (right panels). The small shaded squares on each side of the panels indicate the significance level of the differences between the statistics (bias for left side, standard deviation for right side) and the color indicates which experiment has the better score.

the southern extra-tropics. In the tropics, the scores are improved for EnVar relative to 4D-Var, similar to the comparison between EnVar and 3D-Var. Similarly, Fig. 8 shows the same comparison for EnVar versus 4D-Var, except for the July/August period. The difference in the forecast scores in the northern extra-tropics is again quite small, but for this season the 4D-Var has slightly better forecasts. In the southern extra-tropics, the scores also have the opposite sign as in the February/March period, with a small improvement in the scores for EnVar relative to 4D-Var, especially above 850 hPa and at lead times beyond 48 h. In the tropical region, the scores are again better for EnVar than 4D-Var and similar to those seen in the comparison between EnVar and 3D-Var.

The hybrid background-error covariances used in the EnVar experiments gradually transition from being a simple average of the \mathbf{B}_{nmc} and \mathbf{B}_{enkf} covariances in the troposphere and lower stratosphere to being equivalent with the \mathbf{B}_{nmc} covariances above 10 hPa. To evaluate the impact of this on the stratospheric analyses and forecasts, Fig. 9 shows similar results as in the previous figures, except for the layer of the atmosphere between 1 hPa and 100 hPa for the entire global domain. The difference in the standard deviation of temperature

forecasts relative to the reanalyses are shown comparing the EnVar experiment with either 3D-Var (Fig. 9a) or 4D-Var (Fig. 9b). As expected, the comparison with 3D-Var shows a small consistent improvement for EnVar below about 10 hPa and very similar forecast scores above. When compared with 4D-Var, the forecasts are of similar quality below 10 hPa and significantly degraded above. Consequently, it appears that the gradual transition with vertical level of the weighting between the two covariance matrices in EnVar has the predictable effect of producing similar quality forecasts as 3D-Var for the levels where the covariances are equivalent with 3D-Var and forecasts of improved quality for the levels where \mathbf{B}_{enkf} makes a significant contribution to the hybrid covariances.

In summary, the comparison between EnVar with 3D-Var shows that EnVar nearly always produces improved forecasts when compared with 3D-Var. When compared with 4D-Var, the forecasts from EnVar analyses always have either similar or better scores than 4D-Var in the troposphere of the tropics and the winter extra-tropical region (i.e., northern extra-tropics in February/March and southern extra-tropics in July/August). Conversely, in the summer extra-tropical

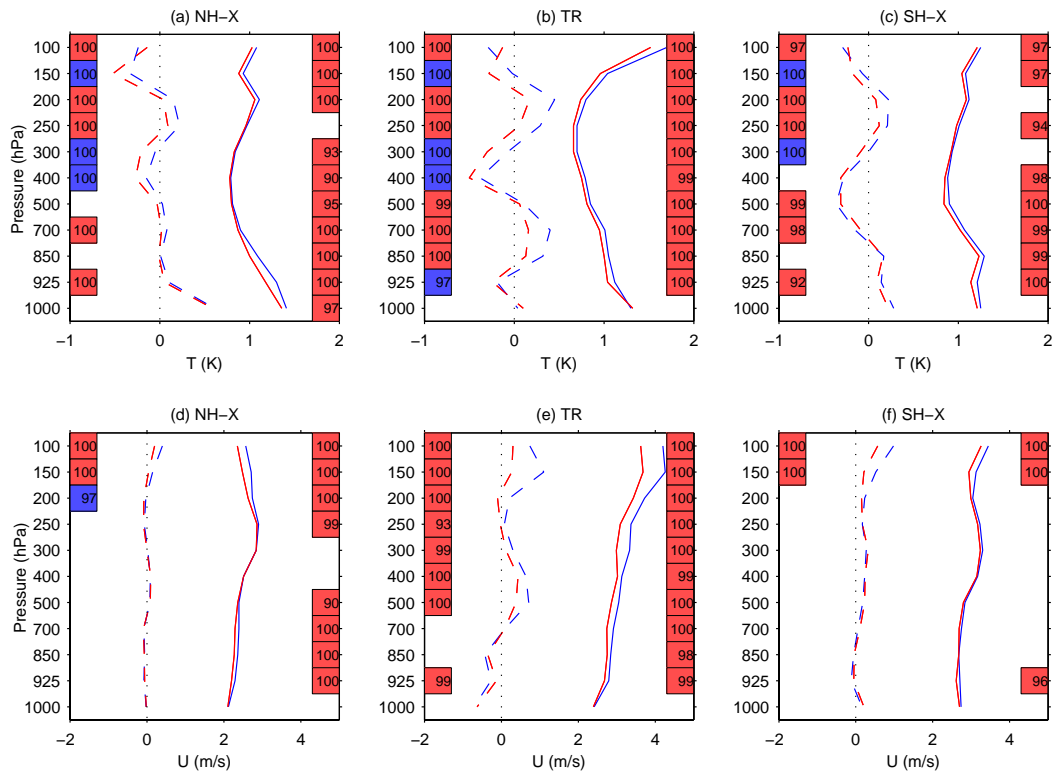


Fig. 4. Same as the previous figure, except for the July/August 2011 period.

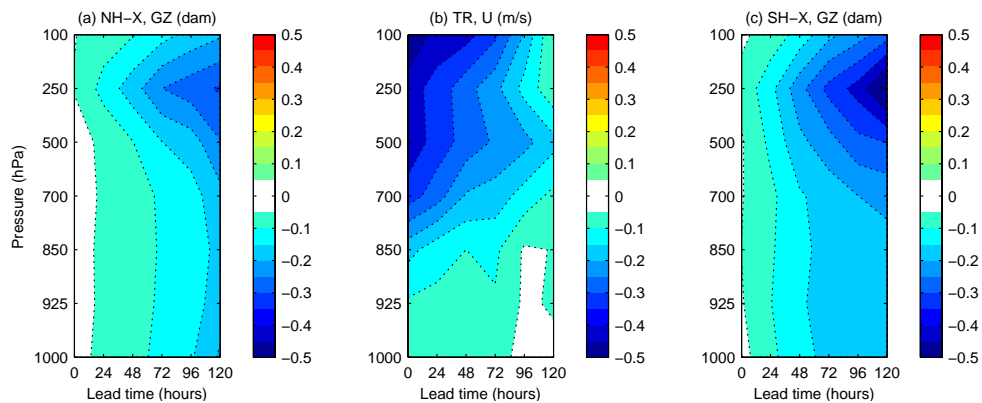


Fig. 5. The difference in standard deviation relative to ERA-Interim reanalyses for the EnVar and 3D-Var forecasts for the February/March 2011 period for (a) geopotential height in the northern extra-tropics, (b) zonal wind in the tropics, and (c) geopotential height in the southern extra-tropics. Note that negative values corresponds with a smaller standard deviation relative to ERA-Interim reanalyses (i.e., better quality) for the forecasts from EnVar analyses than for the forecasts from 3D-Var analyses.

region, the medium-range forecasts from EnVar have either similar or worse scores than 4D-Var in the troposphere. These seasonal differences in the extra-tropics are largest in the southern extra-tropics, where the EnVar is significantly worse in February/March and better in July/August than 4D-Var. In contrast, the short-range forecasts are consistently improved for EnVar when compared with 4D-Var. In the

stratosphere above 10 hPa, the forecasts from EnVar are of similar quality as 3D-Var and significantly worse than 4D-Var.

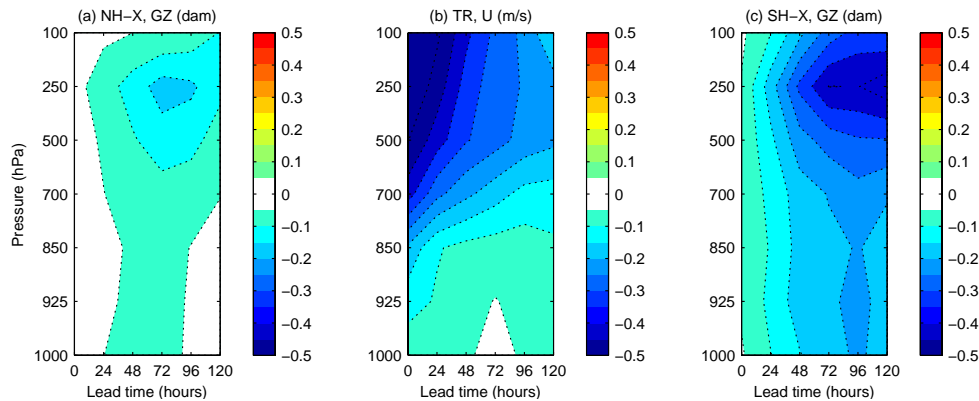


Fig. 6. Same as the previous figure, except for the July/August 2011 period.

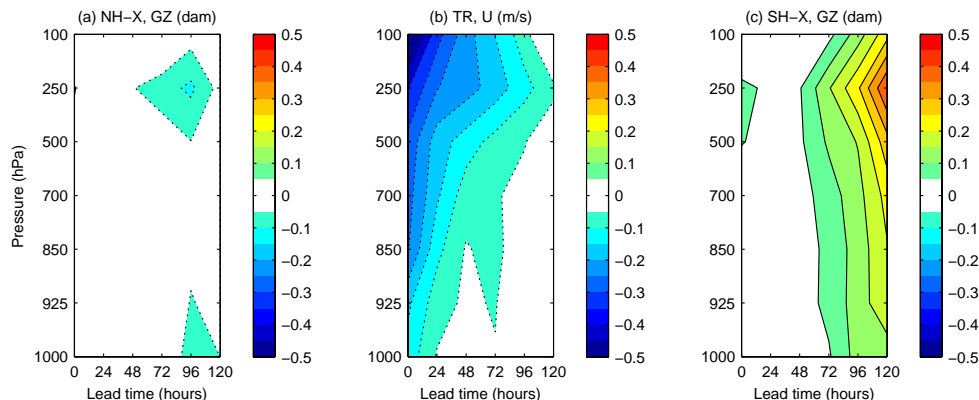


Fig. 7. Same as Fig. 5, except showing the difference in standard deviation relative the ERA-Interim reanalyses for the EnVar and 4D-Var experiments for the February/March 2011 period. Note that negative values correspond with improved forecasts from EnVar analyses as compared with forecasts from 4D-Var analyses.

3.2 4D-EnVar versus 3D-EnVar and 4D-Var versus 3D-Var

The 4-D ensemble covariances in EnVar act to propagate information from observations distributed throughout the 6 h assimilation window to the middle of the window where the computed analysis increment is used to produce the final analysis. This ability is likely to be somewhat limited because the ensemble covariances are used in combination with the 3-D climatological covariances. To help determine the impact of using the 4-D ensemble covariances in this context, an additional EnVar experiment was performed using the 3-D ensemble covariances valid at the middle time. The experiments are referred to as 3D-EnVar and 4D-EnVar to indicate the type of ensemble covariances used in each. Figure 10 is similar to those shown in the previous subsection, but comparing the forecasts from the 4D-EnVar and 3D-EnVar experiments for the February/March period. These results demonstrate that the use of 4-D ensemble covariances gives a generally small improvement (note that the contour intervals in Fig. 10 are 5 or 10 times smaller than in Figs. 5–8)

as compared with using the 3-D ensemble covariances. The impact in the southern extra-tropics is larger than the northern extra-tropics. Since the background-error covariances are identical at the middle time in the 4D-EnVar and 3D-EnVar experiments, observations near the middle time will have the same influence on the analysis increment in the two experiments. In contrast, observations near either the beginning or end of the time window should be more accurately assimilated when using the 4-D covariances than the purely 3-D covariances. Similar results were obtained for the July/August period (not shown), except that the impact is slightly larger in the southern extra-tropics and slightly smaller in the northern extra-tropics.

In 4D-Var, the tangent linear and adjoint versions of the forecast model are used to propagate information between the observations distributed throughout the 6 h assimilation window and the analysis increment computed at the beginning of the window. The analysis increment computed at the beginning of the window is then propagated with the non-linear forecast model to the middle time to produce the final

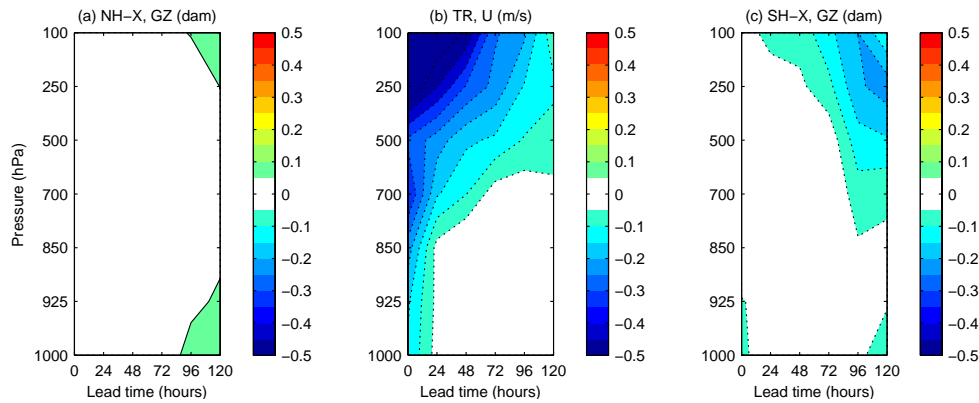


Fig. 8. Same as the previous figure, except for the July/August 2011 period.

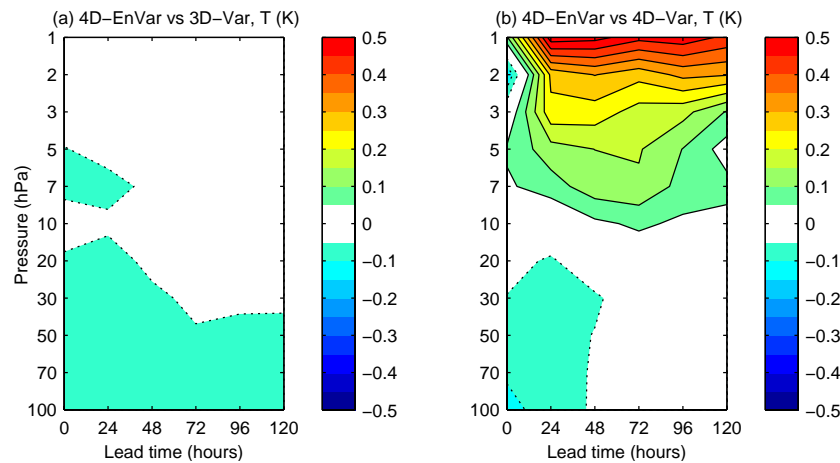


Fig. 9. The difference in standard deviation relative to ERA-Interim reanalyses in the stratosphere for (a) the EnVar-minus-3D-Var forecasts and (b) the EnVar-minus-4D-Var forecasts for the February/March 2011 period for temperature over the entire global domain.

“analysis” used to initialize the medium-range forecasts. In 3D-Var, the analysis increment is computed directly at the middle of the assimilation window, like in EnVar, based on observations distributed throughout the window. However, in 3D-Var the information from the observations is propagated assuming it is unchanged through time. Therefore, all of the observations are treated differently in 4D-Var versus 3D-Var and such a comparison is not directly analogous with the comparison between 4D-EnVar and 3D-EnVar. Nonetheless, Fig. 11 shows the results comparing 4D-Var with 3D-Var. The verification scores from 4D-Var are generally better than 3D-Var in all three regions. Note that the differences in the southern extra-tropics are as much as an order of magnitude larger than the difference between 4D-EnVar and 3D-EnVar. Similar results were obtained for the July/August period (not shown), except that the impact is slightly smaller in the southern extra-tropics and slightly larger in the northern extra-tropics.

4 Temporal fit to observations over the assimilation window

The ability of the EnVar analyses to fit observations distributed through time over the assimilation window provides evidence of how well the 4-D ensemble covariances capture the spatial-temporal structure of the errors. In 4D-Var, the temporal covariances are implicitly modeled by the tangent-linear and adjoint versions of the forecast model, whereas in 3D-Var the errors are assumed to be constant through time. In this section, the ability of all three assimilation approaches to fit observations distributed over the assimilation window is examined to evaluate the accuracy of the spatial-temporal background-error covariances used by each.

Figure 12 shows the fit to temperature (upper panels) and zonal wind (lower panels) observations near 250 hPa (between 225 hPa and 275 hPa) from aircraft computed over the entire February/March period for the EnVar (red curves), 4D-Var (blue curves) and 3D-Var (green curves) experiments. For 4D-Var, the fit to the observations is measured with

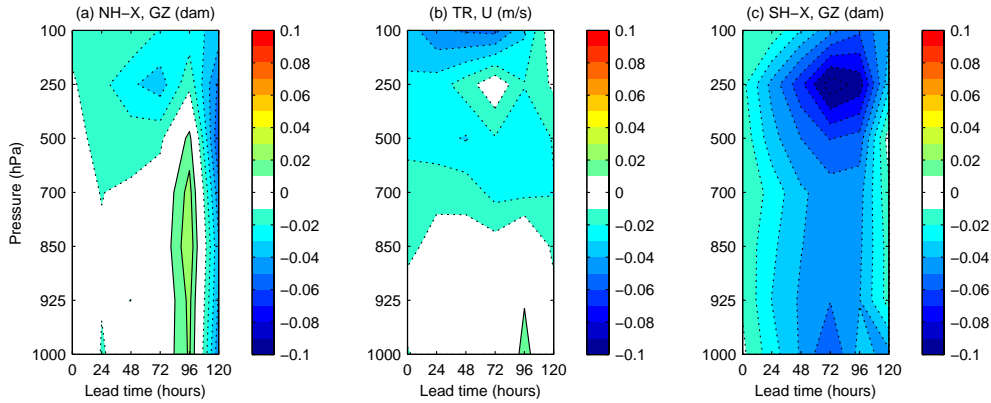


Fig. 10. Same as Fig. 5, except showing the difference in standard deviation relative the ERA-Interim reanalyses for the 4D-EnVar and 3D-EnVar experiments for the February/March 2011 period. Note that negative values correspond with improved forecasts from 4D-EnVar analyses as compared with forecasts from 3D-EnVar analyses.

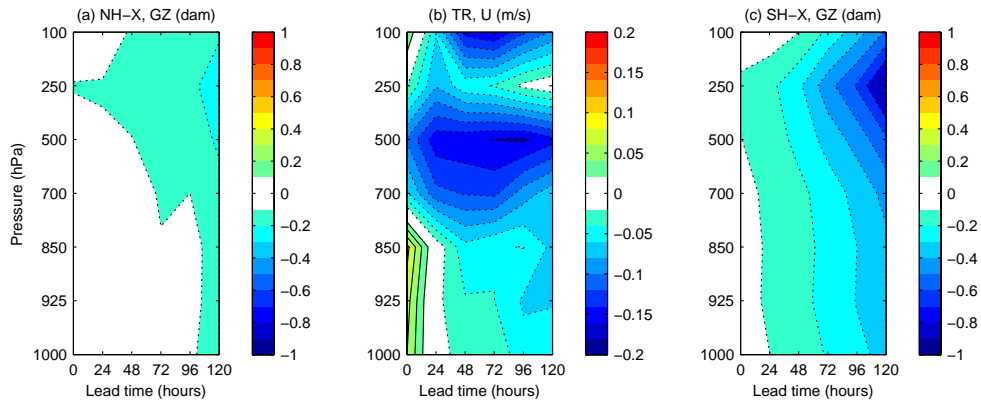


Fig. 11. Same as the previous figure, except showing the difference in standard deviation relative the ERA-Interim reanalyses for the 4D-Var and 3D-Var experiments for the February/March 2011 period.

respect to either the final integration of the tangent-linear version of the forecast model (blue), or an integration of the high-resolution nonlinear forecast model starting from the analysis at the beginning of the assimilation window (not shown). The standard deviation of the observation-minus-background (solid curves) and observation-minus-analysis (dashed curves) is shown in Fig. 12a and c as a function of the relative time within the assimilation window. For all experiments the background state fits the observations more closely at the beginning of the window than at the end. While the analysis fits the observations more closely than the background state over the entire window, it appears to vary less over the assimilation window than for the background state. Note that the background state from 4D-Var consistently agrees more with the aircraft temperature observations than for EnVar, consistent with the similar measure at 250 hPa shown for radiosonde observations in Fig. 3a. To give a more direct measure of how the analysis from each assimilation approach fits the observations over time, Fig. 12b and d show the proportional reduction in the variance of the

fit to the observations,

$$D(t) = \frac{\text{var}[y(t) - H(x_b(t))] - \text{var}[y(t) - H(x_a(t))]}{\text{var}[y(t) - H(x_b(t))]}, \quad (4)$$

after being normalized by its value at the middle of the assimilation window,

$$\hat{D}(t) = \frac{D(t)}{D(t_{\text{middle}})}, \quad (5)$$

where the variable t denotes the relative time within each 6 h assimilation window, that is, from -3 h to 3 h, and t_{middle} is 0 h. Since this quantity is normalized to have a value of one at the middle of the window, it provides a relative average measure of how each assimilation approach improves the fit to observations at times away from the middle of the window. Values less than one for a particular time in the assimilation window indicate that the analysis is drawn to these observations less than it is drawn to the observations at the middle of the assimilation window. Figure 12b and d show that all of

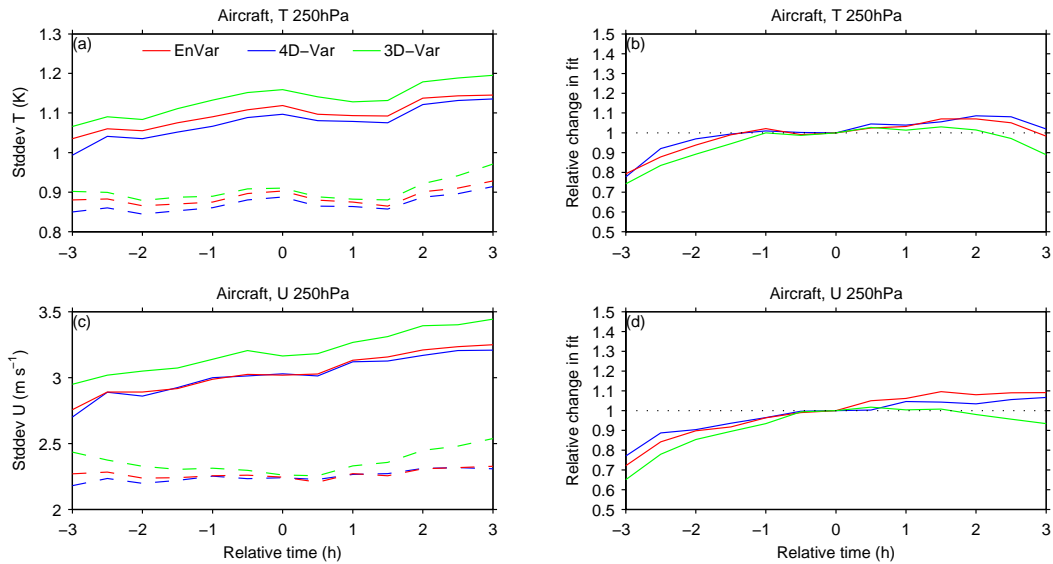


Fig. 12. Results showing the mean fit of the analyses from different assimilation approaches over the 6 h assimilation window. Left panels: the standard deviation of the observation-minus-background (solid curves) and observation-minus-analysis (dashed curves) for aircraft observations. Right panels: the relative improvement in the fit to the observations made by the assimilation procedure, normalized by this measure at the middle of the assimilation window (see Eqs. 4 and 5). These are shown for all observations of 250 hPa temperature (upper panels) and 250 hPa zonal wind (lower panels) from aircraft over the February/March 2011 period for the EnVar (red), 4D-Var (blue), and 3D-Var (green) experiments. The 4D-Var results are from propagating the analysis increment with the tangent linear model.

the assimilation approaches generally move the background state a greater “distance” towards the aircraft observations in the second half of the assimilation window than for the observations in the first half of the assimilation window, consistent with the results seen in the panels on the left. Also, the 3D-Var approach (green) appears less able to fit observations before and after the middle of the assimilation window than both EnVar (red) and 4D-Var (blue). In general, the EnVar and 4D-Var approaches fit the observations before and after the middle of the assimilation window similarly, except that EnVar fits the zonal wind observations in the second half consistently more closely than 4D-Var. Only small differences are seen in the fit to the observations with 4D-Var when using either the tangent-linear or nonlinear model to propagate the analysis increment from the beginning of the assimilation window to later times (not shown).

Similarly, Fig. 13 shows the same type of information as the previous figure, but for the brightness temperature observations from channels 6, 10 and 14 of the AMSU-A instruments. For channel 14, which is most sensitive to temperature around 2 hPa, the 3D-Var and EnVar approaches give very similar results, consistent with the equivalence of the two approaches above 10 hPa, as mentioned previously. For both 3D-Var and EnVar, the relative fit to observations by the analysis is slightly less near the beginning and end of the assimilation window as compared with the middle of the window. In contrast, with 4D-Var the analysis fits the observation much less closely near the beginning of the window

as compared with 3D-Var and EnVar, and significantly more closely in the second half of the window relative to the middle of the window. This can possibly be explained by a large growth rate of the temperature perturbations around 2 hPa during the 6 h forecast model integration. This would make it easier in 4D-Var to create large analysis increments near the end of the assimilation window as compared with EnVar and 3D-Var which both have analysis increments at this level that are constant in time due to the use of 3-D background-error covariances (since EnVar fully uses \mathbf{B}_{nmc} at this level). The impact of partially using the 4-D ensemble covariances in EnVar at lower levels is seen in the results for channels 10 and 6. In both cases, the EnVar analyses are drawn more strongly towards the observations away from the middle of the assimilation window than with 3D-Var. For channel 10, which is most sensitive to temperature around 40 hPa, the EnVar analyses generally provide the closest relative fit to the observations, especially near the beginning of the assimilation window as compared with 3D-Var and 4D-Var. For channel 6, which is most sensitive to temperature around 300 hPa, both EnVar and 3D-Var provide a larger relative fit to the observations near the beginning of the assimilation window, whereas both EnVar and 4D-Var give a similarly larger fit to the observations than for 3D-Var near the end of the assimilation window.

The same results were computed also for the 3D-EnVar experiment, though they are not shown in Figs. 12 and 13 for the sake of clarity. Consistent with the forecast

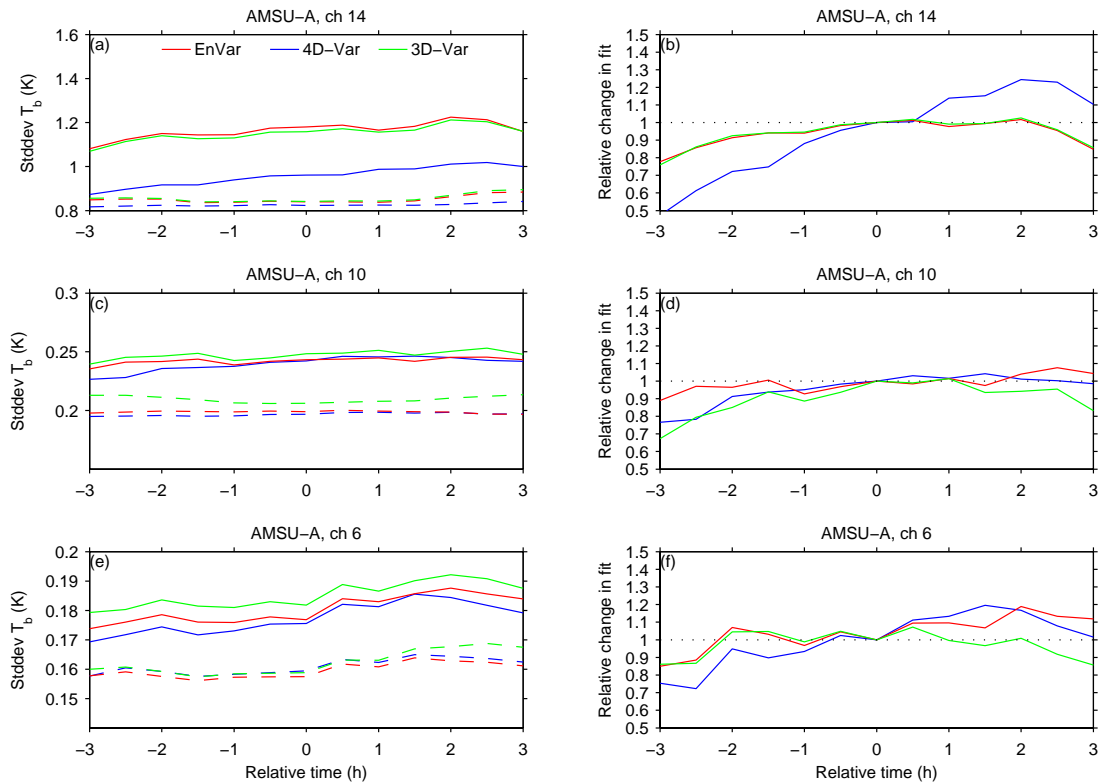


Fig. 13. Same as the previous figure, except for brightness temperature observations from AMSU-A channel 14 (upper panels), 10 (middle panels), and 6 (lower panels).

verification scores shown in the previous section, the fit of the background state to the observations is similar to, but slightly worse than for the 4D-EnVar experiment for both aircraft and AMSU-A observations. However, for the relative measure given by Eqs. (4) and (5) the results from 3D-EnVar are very similar to those for 3D-Var. This confirms that the improved relative fit of 4D-EnVar analyses to observations near the beginning and end of the assimilation window (as compared with 3D-Var in Figs. 12 and 13) is due to the use of the 4-D ensemble covariances.

The results shown in this section suggest that the use of 4-D ensemble covariances below about 10 hPa enables EnVar to produce a 4-D analysis increment that is reasonably consistent with the misfit between the observations and the background state throughout the 6 h assimilation window. This allows the EnVar approach, like 4D-Var, to better make use of the observations distributed throughout the assimilation window than assimilation approaches that assume the error in the background state is constant for all times in the assimilation window, like 3D-Var or 3D-EnVar. However, this aspect of EnVar, as implemented for this study, is certainly limited by the weighted average between the 4-D ensemble covariances and the 3-D static covariances used to specify the background-error covariances. From the comparison of 4D-EnVar with 3D-EnVar in the previous section, it can be

concluded that the inclusion of this aspect in EnVar only results in a relatively small improvement to forecast quality.

5 Conclusions

The goal of this study was to evaluate a version of the ensemble-variational data assimilation approach (EnVar) for possible replacement of 4D-Var for operational global deterministic weather prediction at Environment Canada. This implementation of EnVar relies on the 4-D ensemble covariances obtained from the Canadian ensemble Kalman filter, currently used for initializing ensemble forecasts. Verification against ERA-Interim reanalyses were generated from a set of data assimilation experiments over two separate 6 week periods for EnVar, 4D-Var and also for 3D-Var and a version of EnVar that uses 3-D ensemble covariances. In these experiments, EnVar analyses nearly always result in improved, and never degraded, forecasts when compared with 3D-Var. Compared with 4D-Var, the forecasts from EnVar analyses have either similar or better scores in the troposphere of the tropics and the winter extra-tropical region. In the summer extra-tropical region the medium-range forecasts from EnVar have either similar or worse scores than 4D-Var in the troposphere. The seasonal differences in the extra-tropics of medium-range forecast quality are largest in

the southern extra-tropics, where the EnVar is significantly worse in February/March and better in July/August than 4D-Var. In contrast, the 6 h forecasts from the EnVar experiments are significantly better than those from 4D-Var relative to radiosonde observations for both periods and in all regions. In the stratosphere above 10 hPa, the forecasts from EnVar analyses are of similar quality with those from 3D-Var, consistent with the fact that both use the same static background-error covariances here, and significantly worse than those from 4D-Var. A possible approach for improving the EnVar analyses and resulting forecasts in the stratosphere is to raise the top model level in the EnKF (currently 2 hPa) to the same level as in the deterministic system (0.1 hPa). The use of 4-D ensemble covariances as compared with 3-D ensemble covariances only results in small improvements in forecast quality. By contrast, the improvements from using 4D-Var as compared with 3D-Var are much larger.

In conclusion, the results from this study suggest that the EnVar approach is a viable alternative to 4D-Var, especially when the simplicity and computational efficiency of EnVar are considered. It should be emphasized that, due to practical reasons, a version of 4D-Var that uses 3-D ensemble background-error covariances in place of the static covariances was not considered. In preparation for a possible operational implementation, research is underway with the goal of understanding the causes of the lower quality medium-range forecasts in the summer extra-tropical regions as compared with 4D-Var. If this problem is eventually resolved by making improvements to the EnKF itself, then both the deterministic and ensemble forecasting systems will automatically benefit from this work. EnVar is also currently being tested in combination with other planned upgrades to the global deterministic system, including a version of the forecast model that uses a global Yin-Yang grid with 15 km horizontal grid-spacing (Qaddouri and Lee, 2011), additional AIRS and IASI channels assimilated, and improved use of radiosonde observations (Laroche and Sarrazin, 2013). Encouraging preliminary results have also been obtained from tests to evaluate the impact of using EnVar instead of the currently operational 4D-Var approach for regional deterministic analyses over North America (Tanguay et al., 2012).

Acknowledgements. The authors thank Peter Houtekamer and Xingxiu Deng for providing the large quantity of EnKF 4-D background ensembles used in the EnVar experiments. Stéphane Laroche and Ron McTaggart-Cowan are thanked for assistance with the verification against ERA-Interim. Stéphane Laroche is also thanked for comments that helped improve an earlier version of the paper. The official reviews of Andrew Lorenc and an anonymous reviewer were also helpful.

Edited by: G. Desroziers

Reviewed by: A. Lorenc and one anonymous referee

References

- Bishop, C. H., Hodyss, D., Steinle, P., Simms, H., Clayton, A. M., Lorenc, A. C., Barker, D. M., and Buehner, M.: Efficient ensemble covariance localization in variational data assimilation, *Mon. Weather Rev.*, 139, 573–580, 2011.
- Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B.: Intercomparison of variational data assimilation and the Ensemble Kalman Filter for global deterministic NWP, Part I: Description and Single-Observation experiments, *Mon. Weather Rev.*, 138, 1550–1566, 2010a.
- Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B.: Intercomparison of variational data assimilation and the Ensemble Kalman Filter for global deterministic NWP, Part II: One-Month Experiments with Real Observations, *Mon. Weather Rev.*, 138, 1567–1586, 2010b.
- Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis Scheme in the Ensemble Kalman Filter, *Mon. Weather Rev.*, 126, 1719–1724, 1998.
- Canadian Meteorological Centre (CMC): Improvements to the Global Deterministic Prediction System (from version 2.2.2 to 3.0.0), and related changes to the Regional Deterministic Prediction System (from version 3.0.0 to 3.1.0), Canadian Meteorological Centre technical note, available at: http://collaboration.cmc.ec.gc.ca/cmc/CMOI/product_guide/docs/lib/op_systems/doc_opchanges/technote_gdps300_20130213_e.pdf, 2013.
- Caya, A., Sun, J., and Snyder, C.: A comparison between the 4dvar and the ensemble Kalman filter techniques for radar data assimilation, *Mon. Weather Rev.*, 133, 3081–3094, 2005.
- Charron, M., Pellerin, G., Spacek, L., Houtekamer, P. L., Gagnon, N., Mitchell, H. L., and Michelin, L.: Toward Random Sampling of Model Error in the Canadian Ensemble Prediction System, *Mon. Weather Rev.*, 138, 1877–1901, 2010.
- Charron M., Polavarapu, S., Buehner, M., Vaillancourt, P. A., Charette, C., Roch, M., Morneau, J., Garand, L., Aparicio, J. M., MacPherson, S., Pellerin, S., St-James, J., and Heilliette, S.: The stratospheric extension of the Canadian global deterministic medium range weather forecasting system and its impact on tropospheric forecasts, *Mon. Weather Rev.*, 140, 1924–1944, 2012.
- Clayton, A. M., Lorenc, A. C., and Barker, D. M.: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office, *Q. J. Roy. Meteor. Soc.*, 139, 1445–1461, doi:10.1002/qj.2054, 2012.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, doi:10.1002/qj.828, 2011.
- Fairbairn, D., Pring, S. R., Lorenc, A. C., and Roulstone, I.: A comparison of 4DVar with ensemble data assimilation methods, *Q. J. Roy. Meteor. Soc.*, doi:10.1002/qj.2135, online first, 2013.
- Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S., and Morneau, J.: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at

- the Meteorological Service of Canada, *Mon. Weather Rev.*, 135, 2339–2354, 2007.
- Gilbert, J.-Ch., and Lemaréchal, C.: Some numerical experiments with variable-storage quasi-Newton algorithms, *Mathematical Programming*, 45, 407–435, 1989.
- Gustafsson, N.: Discussion on “4D-Var or EnKF?”, *Tellus A*, 59, 774–777, doi:10.1111/j.1600-0870.2007.00262.x, 2007.
- Hamill, T. M. and Snyder, C.: A hybrid ensemble kalman filter–3d variational analysis scheme, *Mon. Weather Rev.*, 128, 2905–2919, 2000.
- Honda, Y., Nishijima, M., Kopizumi, K., Ohta, Y., Tamiya, K., Kawabata, T., and Tsuyuki, T.: A pre-operational variational data assimilation system for a non-hydrostatic model at the Japan Meteorological Agency: Formulation and preliminary results, *Q. J. Roy. Meteor. Soc.*, 131, 3465–3475, 2005.
- Houtekamer, P. L. and Mitchell, H. L.: Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Weather Rev.*, 126, 796–811, 1998.
- Houtekamer, P. L. and Mitchell, H. L.: Ensemble Kalman filtering, *Q. J. Roy. Meteor. Soc.*, 131, 3269–3289, 2005.
- Houtekamer, P. L., Mitchell, H. L., and Deng, X.: Model Error Representation in an Operational Ensemble Kalman Filter, *Mon. Weather Rev.*, 137, 2126–2143, 2009.
- Hunt, B., Kalnay, E., Kostelich, E., Ott, E., Patil, D., Sauer, T., Szunyogh, I., Yorke, J., and Zimin, A.: Four-dimensional ensemble Kalman filtering, *Tellus A*, 56, 273–277, 2004.
- Kalnay, E., Li, H., Miyoshi, T., Yang, S. C., and Ballabrera-Poy, J.: 4-D-Var or ensemble Kalman filter?, *Tellus A*, 59, 758–773, doi:10.1111/j.1600-0870.2007.00261.x, 2007.
- Kuhl, D., Rosmond, T., Bishop, C., McLay, J., and Baker, N.: Comparison of hybrid ensemble/4D-Var and 4D-Var within the NAVDAS-AR data assimilation framework, *Mon. Weather Rev.*, 141, 2740–2758, doi:10.1175/MWR-D-12-00182.1, 2013.
- Laroche, S. and Sarrazin, R.: Impact of Radiosonde Balloon Drift on Numerical Weather Prediction and Verification, *Weather Forecast.*, 28, 772–782, doi:10.1175/WAF-D-12-00114.1, 2013.
- Lorenc, A. C.: The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var, *Q. J. Roy. Meteor. Soc.*, 129, 3183–3203, doi:10.1256/qj.02.132, 2003.
- Miyoshi, T., Sato, Y., and Kadowaki, T.: Ensemble Kalman filter and 4D-var intercomparison with the Japanese operational global analysis and prediction system, *Mon. Weather Rev.*, 138, 2846–2866, 2010.
- Parrish, D. and Derber, J.: The National Meteorological Center’s spectral statistical interpolation analysis scheme, *Mon. Weather Rev.*, 120, 1747–1763, 1992.
- Qaddouri, A. and Lee, V.: The Canadian Global Environmental Multiscale model on the Yin-Yang grid system, *Quart. J. Roy. Meteor. Soc.*, 137, 1913–1926, doi:10.1002/qj.873, 2011.
- Rabier, F.: Overview of global data assimilation developments in numerical weather-prediction centres, *Q. J. Roy. Meteor. Soc.*, 131, 3215–3233, doi:10.1256/qj.05.129, 2005.
- Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A.: The ECMWF operational implementation of four-dimensional variational assimilation, I: Experimental results with simplified physics, *Q. J. Roy. Meteor. Soc.*, 126, 1143–1170, 2000.
- Rawlins, F., Ballard, S., Bovis, K., Clayton, A., Li, D., Inverarity, G., Lorenc, A., and Payne, T.: The Met Office four-dimensional variational data assimilation scheme, *Q. J. Roy. Meteor. Soc.*, 133, 347–362, 2007.
- Tanguay, M., Fillion, L., Lapalme, E., and Lajoie, M.: Four-Dimensional Variational Data Assimilation for the Canadian Regional Deterministic Prediction System, *Mon. Weather Rev.*, 140, 1517–1538, 2012.
- Whitaker, J. S., Compo, G. P., and Thépaut, J.-N.: A comparison of variational and ensemble-based data assimilation systems for re-analysis of sparse observations, *Mon. Weather Rev.*, 137, 1991–1999, 2009.
- Zhang, M., Zhang, F., Huang, X.-Y., and Zhang, X.: Intercomparison of an ensemble Kalman filter with three- and four-dimensional variational data assimilation methods in a limited-area model over the month of June 2003, *Mon. Weather Rev.*, 139, 566–572, 2011.