Sensitivity of Global Ensemble Forecasts to the Initial Ensemble Mean and Perturbations: Comparison of EnKF, Singular Vector, and 4D-Var Approaches

MARK BUEHNER AND AHMED MAHIDJIBA

Meteorological Research Division, Environment Canada, Dorval, Quebec, Canada

(Manuscript received 30 November 2009, in final form 27 April 2010)

ABSTRACT

This study examines the sensitivity of global ensemble forecasts to the use of different approaches for specifying both the initial ensemble mean and perturbations. The current operational ensemble prediction system of the Meteorological Service of Canada uses the ensemble Kalman filter (EnKF) to define both the ensemble mean and perturbations. To evaluate the impact of different approaches for obtaining the initial ensemble perturbations, the operational EnKF approach is compared with using either no initial perturbations or perturbations obtained using singular vectors (SVs). The SVs are computed using the (dry) total-energy norm with a 48-h optimization time interval. Random linear combinations of 60 SVs are computed for each of three regions. Next, the impact of replacing the initial ensemble mean, currently the EnKF ensemble mean analysis, with the higher-resolution operational four-dimensional variational data assimilation (4D-Var) analysis is evaluated. For this comparison, perturbations are provided by the EnKF. All experiments are performed over two-month periods during both the boreal summer and winter using a system very similar to the global ensemble prediction system that became operational on 10 July 2007. Relative to the operational configuration that relies on the EnKF, the use of SVs to compute initial perturbations produces small, but statistically significant differences in probabilistic forecast scores in favor of the EnKF both in the tropics and, for a limited set of forecast lead times, in the summer hemisphere extratropics, whereas the results are very similar in the winter hemisphere extratropics. Both approaches lead to significantly better ensemble forecasts than with no initial perturbations, though results are quite similar in the tropics when using SVs and no perturbations. The use of an initial-time norm that does not include information on analysis uncertainty and the lack of linearized moist processes in the calculation of the SVs are two factors that limit the quality of the resulting SV-based ensemble forecasts. Relative to the operational configuration, use of the 4D-Var analysis to specify the initial ensemble mean results in improved probabilistic forecast scores during the boreal summer period in the southern extratropics and tropics, but a near-neutral impact otherwise.

1. Introduction

As is common at many NWP centers, both deterministic and ensemble forecasts are produced operationally by the Meteorological Service of Canada (MSC). The systems used to produce the initial conditions for these two forecast systems operate almost completely independently. Since 2005, the global deterministic analysis is produced using a four-dimensional variational data assimilation (4D-Var) system (Gauthier et al. 2007). An ensemble Kalman filter (EnKF) approach (Houtekamer and Mitchell 2005) has been used operationally, also since 2005, to supply the initial conditions for the global ensemble prediction system (EPS).

The operational 4D-Var and EnKF data assimilation systems both use the Global Environmental Multiscale (GEM) model (Côté et al. 1998), but with different configurations. The configuration for the EnKF that became operational on 10 July 2007 uses a lower horizontal and vertical resolution than in the deterministic system to compensate for the larger number of forecasts produced. Both systems assimilate a similar set of observations using nearly identical observation operators and observation-error statistics, though several additional types of remotely sensed data are assimilated in the 4D-Var that are not used in the EnKF. A study was recently conducted to intercompare the two systems in the context

Corresponding author address: Mark Buehner, Meteorological Research Division, Environment Canada, 2121 TransCanada Hwy., Dorval QC H9P 1J3, Canada.
E-mail: mark.buehner@ec.gc.ca

DOI: 10.1175/2010MWR3296.1
of initializing global deterministic forecasts (Buehner et al. 2010a,b). Specifically, the ensemble mean of the EnKF was used to initialize deterministic forecasts and also the EnKF background ensembles were used to specify flow-dependent background-error covariances within the variational data assimilation system. These alternative approaches were evaluated against the currently operational approach of using 4D-Var with a static and relatively simple estimate of the background-error covariances. Results from that study demonstrate the potential of a significant positive impact on analysis and forecast accuracy in the deterministic forecast system from combining elements of the two existing systems.

The present study examines alternative strategies for initializing the ensemble forecasts. The availability of the tangent-linear and adjoint versions of the GEM forecast model, originally developed for 4D-Var, make it possible to compute singular vectors (SVs; Zadra et al. 2004; Buehner and Zadra 2006; Li et al. 2008). The use of SVs to generate initial perturbations for an EPS is operational at several NWP centers, including the European Centre for Medium-Range Weather Forecasts (ECMWF; Buizza and Palmer 1995; Molteni et al. 1996). Another alternative strategy is to replace the ensemble mean, currently defined by the 96-member EnKF ensemble mean analysis, with a spatially interpolated version of the high-resolution 4D-Var deterministic analysis. Since more satellite observations are assimilated and the spatial resolution of the background state and analysis is higher (though the analysis increment is actually at a lower resolution), the 4D-Var may produce a more accurate analysis than the EnKF. The goal of this study is to evaluate these two alternative strategies for computing the initial ensembles in the Canadian EPS relative to the currently operational approach. The EnKF/EPS configuration that became operational on 10 July 2007 is used for the comparisons in this study. As far as we are aware, this represents the first comparison of EnKF, SV, and 4D-Var approaches in the context of an operational global EPS and using the same forecast model for each. As mentioned below, however, other studies have been performed to compare other ensemble data assimilation techniques (e.g., ensemble of 4D-Var analyses) to the SV approach.

Several previous studies have compared the use of different strategies for computing the initial conditions for ensemble forecasts. Hamill et al. (2000) used a quasigeostrophic channel model without model error to compare SVs with a system simulation approach (Houtekamer et al. 1996) applied to a 3D-Var data assimilation cycle. The latter approach, which can be considered an approximation to the standard EnKF, provided the best ensemble forecasts due to a more accurate ensemble mean and perturbations that were more representative of analysis uncertainty. Bowler (2006) compared several techniques, including the EnKF and SVs, using highly idealized low-order models and found that the EnKF consistently provided the best ensemble forecasts. Descamps and Talagrand (2007) compared a similar set of approaches in a slightly more realistic context using a three-level quasigeostrophic model. For several probabilistic forecast scores, they also showed that the EnKF provides better ensemble forecasts than SVs over the first portion of the forecasts. A comparison of the ECMWF, MSC, and National Centers for Environmental Prediction (NCEP) operational ensemble prediction systems was conducted by Buizza et al. (2005). At that time MSC used a system simulation approach applied to an optimal interpolation data assimilation scheme (Houtekamer et al. 1996), NCEP used bred vectors (Toth and Kalnay 1993), and ECMWF used SVs. The results generally favored the ECMWF system. However, because of differences in forecast model, data assimilation system, and simulation of model error, the authors point out that these differences cannot be used to evaluate the different strategies for obtaining the initial ensemble perturbations. It was thought that the improved performance of the ECMWF was mostly due to a more accurate ensemble mean state and a better forecast model. A more recent comparison of these and other ensemble prediction systems is provided by Park et al. (2008). In that study the MSC ensemble system was already based on the EnKF and was very similar to the system evaluated in the present study. To isolate the influence of the initial ensemble perturbations, Magnusson et al. (2008) compared using SV and bred vectors within the ECMWF EPS. This showed a much more similar performance from the two methods than from the comparison of the full ECMWF and NCEP systems, mentioned above. Wei et al. (2008) compared the use of bred vectors with several approaches based on the ensemble transform approach using the NCEP operational system. They found that the ensemble transform technique with rescaling provided improved ensemble forecasts relative to the operational approach at that time, which was based on bred vectors. Buizza et al. (2008) compared the use of SVs (both initial-time and evolved SVs) with perturbations computed from an ensemble of perturbed 4D-Var analyses for initializing the ECMWF ensemble prediction system. To simulate various sources of uncertainty in the 4D-Var analysis cycles, the observations and sea surface temperatures were perturbed and the forecast model included a stochastic backscatter scheme (Shutts 2005). This approach is similar to the EnKF except that, whereas the EnKF uses background-error covariances computed from the ensemble itself, all analyses...
in the perturbed 4D-Var assimilation cycles are performed using the same background-error covariances that do not depend on the ensembles. They found that the perturbations obtained from the ensemble 4D-Var led to underdispersive and less skillful ensemble forecasts in the northern extratropics than when using SVs. However, relative to the use of SVs alone, an overall improvement in the ensemble forecasts was obtained, especially for the tropics, when the perturbations from initial-time SVs were combined with those from the ensemble 4D-Var. The present paper complements past studies by comparing the EnKF and SV approaches within the Canadian operational EPS. The impact of recentering the EnKF ensemble members on the high-resolution 4D-Var analysis, instead of the EnKF ensemble mean analysis, is also examined.

In the next section, a description of the approaches for generating initial ensemble members is given. Section 3 provides details about the numerical experiments performed and also the verification tools used. Results from using initial perturbations obtained with SVs are given in section 4. In section 5, results from using the 4D-Var deterministic analysis to specify the initial ensemble mean are shown. Finally, some conclusions are given in section 6.

2. Approaches for generating initial ensemble members

During the second half of 2007 and until June 2009, the Canadian operational EnKF and EPS both used the GEM model with a uniform 400 × 200 latitude–longitude grid (~100-km grid spacing at the equator), 28 vertical levels, and the top at 10 hPa. The EPS consists of a 20-member medium-range ensemble forecast that is initialized by always selecting the same set of 20 members out of the 96 members in the EnKF analysis ensemble (for more details, see Houtekamer et al. 2007). These initial states are adjusted so that their ensemble mean is equal to the ensemble mean of the 96 EnKF analyses. Next, balanced and spatially smooth random perturbations are added to the members at the initial time to account for all sources of error (not only error related to the forecast model) that are not otherwise simulated and to obtain realistic ensemble spread in the early portion of the ensemble forecast. The average amplitude of these perturbations is almost as large as the spread in the original EnKF analysis ensemble (e.g., standard deviation of 0.7 K for the random perturbations and 0.8 K for the EnKF ensemble spread for temperature near 850-hPa averaged globally over the period 1–14 December 2007). These relatively large perturbations are required at the beginning of the ensemble forecasts due to the insufficient initial growth in the EPS ensemble spread (Houtekamer et al. 2007). During the forecasts, model errors are accounted for by using a different set of physical parameterizations for each member, applying stochastic perturbations to the tendencies output from the physical parameterizations, and by using a stochastic kinetic energy backscatter scheme (Shutts 2005). The EPS and EnKF configurations considered in this study are those that were operational after 10 July 2007. More details concerning the simulation of model error in the EPS are reported by Charron et al. (2010).

In the present study, only changes to the procedure for obtaining the initial conditions of the EPS are evaluated. Therefore, the procedures for accounting for model error during the ensemble forecasts mentioned above are used in all experiments, but the addition of random perturbations to the initial states are only used for the experiments employing EnKF-derived ensemble perturbations. To help distinguish the impact on the ensemble forecasts from the initial condition perturbations and the procedures for simulating model errors, results from a simple experiment with no initial perturbations are also included in the comparisons.

a. Operational approach: EnKF

As already stated, the model states used to initialize the operational Canadian global EPS are obtained from an ensemble of analyses produced by the EnKF. The EnKF is based on the application of a Monte Carlo procedure to the standard data assimilation cycle such that important sources of uncertainty are represented by relatively small ensembles of random samples (Evensen 1994; Burgers et al. 1998; Houtekamer and Mitchell 1998). Specifically, during the EnKF analysis step, each of the ensemble members is updated by assimilating the observations after they have been independently perturbed for each member. These perturbations simulate the effect of random observation error and are computed such that the ensemble mean of the perturbations for each observation is zero. For the purpose of assimilating the observations, flow-dependent background-error covariances are obtained from the ensemble of short-term forecasts using the four-subensemble configuration described by Mitchell and Houtekamer (2009). These background-error covariances include the temporal cross covariances between five time steps, separated by 90 min, over the 6-h assimilation window. The separation of the 96-member EnKF ensemble into 4 subensembles for the analysis step may cause each subensemble to have slightly different statistics and therefore it was decided to select 5 analyses from each for initializing the 20-member EPS. During the EnKF forecast
step used to produce the set of background fields for the subsequent analysis time, perturbation fields (with an ensemble mean equal to zero at each grid point) are added to the initial conditions and different configurations of the model physical parameterizations are used to simulate the effect of model error (Houtekamer et al. 2009). The perturbation fields added to the 96-member analysis ensemble within the EnKF are similar to those added to the 20 members used to initialize the medium-range ensemble forecasts, except the amplitude is about half as large. The overall amplitude of these perturbations was empirically tuned to maintain sufficient ensemble spread (Houtekamer et al. 2009). The net effect of simulating the numerous sources of uncertainty is to produce and maintain a spread in the ensemble of background and analysis states that is representative of the error in the corresponding ensemble mean. The goal is to provide an appropriate means of generating both the ensemble mean and perturbations of the initial states for an EPS.

b. Ensemble perturbations obtained from singular vectors

For a given optimization time interval, SVs are defined as the orthogonal set of initial-time perturbations that maximize the growth (in the tangent-linear sense) with respect to specified norms. It has previously been shown that, under certain conditions, the SVs are the optimal set of perturbations with respect to representing the forecast-error covariance matrix (i.e., the leading eigenvectors) at optimization time (Ehrendorfer and Tribbia 1997). In contrast, the EnKF approach is designed to produce a random sample of model states from the analysis and forecast distributions that, in theory, would require a much larger sample to explain the same fraction of forecast-error variance than with the SVs. This potential ability of SVs to efficiently explain forecast-error uncertainty with small ensembles makes the approach attractive when considering the large computational expense of producing ensemble forecasts. Consequently, a major goal of this study is to evaluate the impact of using SVs to generate the initial ensemble perturbations relative to the current EnKF approach. The specific approach for using SVs in this regard is largely based on the approach used at ECMWF (Leutbecher and Palmer 2008). However unlike their approach, no evolved SVs are used when computing the initial perturbations. In addition, SVs for the tropics are computed following the same procedure as for the extratropics and not only in the region of tropical cyclones as they are at ECMWF (Puri et al. 2001). This can be expected to lead to initial perturbations that do not efficiently capture the uncertainty in the tropics of the ensemble mean forecast (Barkmeijer et al. 2001).

By assuming that the growth of perturbations over the optimization time interval is governed by the linearized dynamics, the orthogonal set of most rapidly growing perturbations can be efficiently computed with an iterative algorithm employing the tangent-linear and adjoint versions of the forecast model. For the present study, the Arnoldi Package (ARPACK) implementation of the implicitly restarted Lanczos method was used. The optimization time interval is 48 h and the total energy norm (Zadra et al. 2004) with the moisture term neglected is used to define the norms at both the initial and optimization times. The SVs were computed on a global grid with a horizontal grid spacing of 3° in latitude and longitude and 28 vertical levels. In this study, the only physical parameterization included in the tangent-linear and adjoint models is the simplified boundary layer scheme. To ensure a reasonable distribution of the initial perturbations globally, a local projection operator (Buizza 1994) is applied at the optimization time to compute a separate set of SVs for each of three regions: northern extratropics (30°N ≤ lat ≤ 75°N), tropics (30°S ≤ lat ≤ 30°N), and southern extratropics (30°S ≤ lat ≤ 75°S). To avoid capturing spurious perturbation growth near the model vertical boundaries, the norm at optimization time is only evaluated over the vertical model levels between approximately 100 and 800 hPa. In total, 60 SVs are obtained for each region.

The 20 perturbations used to initialize EPS are obtained by computing random linear combinations of all the SVs (Leutbecher and Palmer 2008). To ensure the perturbations do not change the specified ensemble mean state, the mean perturbation must be zero. At ECMWF this is obtained by constructing the perturbations in positive–negative pairs. However, because the use of positive–negative perturbation pairs tends to reduce the number of dimensions spanned by the perturbations, relative to the case in which all perturbations are independent, this approach may be particularly inefficient when the number of perturbations is relatively small (i.e., 20 in the Canadian EPS versus 50 at ECMWF). Consequently, the ensemble mean of the coefficients applied to each SV is simply removed in the present study. Because the SV perturbations already require the application of an empirical scaling (described below) and the coefficients used to perform the linear combination of SVs are a random sample, the impact on the ensemble spread of subtracting the ensemble mean of the coefficient is not as important as for approaches such as the ensemble transform Kalman filter (Wang et al. 2004). Therefore, the coefficient $a_j$ used to compute the contribution of SV $j$ to
perturbation \( i \) is computed separately for each region according to
\[
\tilde{a}_{ij} = a_{ij} - \frac{1}{20} \sum_{r=1}^{20} a_{ij},
\]
where \( a_{ij} \) are random values obtained by sampling a Gaussian distribution with a mean of 0 and a variance of 1. To eliminate excessively large amplitude perturbations that could cause numerical instabilities, the Gaussian distribution is truncated at \( \pm 3 \). Using these coefficients, the initial perturbations are then given by
\[
x_i' = \bar{x} + \beta_{NH} \sum_{j=1}^{60} (\tilde{a}_{ij} u_{ij}^{NH}) + \beta_{TR} \sum_{j=1}^{60} (\tilde{a}_{ij} u_{ij}^{TR})
\]
\[
+ \beta_{SH} \sum_{j=1}^{60} (\tilde{a}_{ij} u_{ij}^{SH}),
\]
where \( x_i' \) are the perturbed analyses \((i = 1, \ldots, 20)\); \( \bar{x} \) is the EnKF ensemble mean analysis; and \( u_{ij}^{NH} \), \( u_{ij}^{TR} \), and \( u_{ij}^{SH} \) are initial-time SVs over the northern extratropics, tropics, and southern extratropics, respectively. (\( j = 1, \ldots, 60 \)). The constants \( \beta_{NH}, \beta_{TR}, \) and \( \beta_{SH} \) are determined empirically to yield an ensemble spread for the 48-h forecasts that is comparable to that obtained with the operational EnKF approach when averaged over a 2-month period. Therefore, compared with the currently operationally approach described previously, the SV approach uses the same initial ensemble mean and has ensemble perturbations tuned to produce a similar ensemble spread after 48 h.

c. Ensemble mean obtained from 4D-Var analysis

Another alternative strategy for initializing the EPS is to use the operational 4D-Var analysis to specify the ensemble mean. This approach is motivated by the fact that, compared to the EnKF system, the 4D-Var analysis employs a higher-resolution version of the forecast model and assimilates more types of observations. Consequently, the analysis produced by the 4D-Var may be of higher quality than the mean of the 96-member EnKF analysis ensemble used to specify the initial ensemble mean state.

Both the EnKF and 4D-Var systems, operational during the second half of 2007 and early 2008, use the GEM model with a uniform latitude–longitude grid and the top level at 10 hPa. The EnKF uses a 400 \( \times \) 200 horizontal grid (\( \sim 100 \)-km grid spacing) with 28 vertical levels and the 4D-Var produces analyses on an 800 \( \times \) 600 horizontal grid (\( \sim 33–50 \)-km grid spacing) and 58 vertical levels (Bélair et al. 2009). During the same period, the types of observations assimilated operationally in the EnKF system were wind, temperature, and humidity from radiosondes; wind and temperature from aircraft; wind, temperature, pressure, and humidity from in situ surface observations; atmospheric motion wind from geostationary and polar-orbiting satellites; and radiances from Advanced Microwave Sounding Unit-A/B (AMSU-A/B). In addition to these observation types, the 4D-Var also assimilated wind from profilers over the United States and radiances from geostationary satellites.

Because of the difference in horizontal and vertical resolution, the 4D-Var analyses were interpolated to the lower-resolution grid used by the EPS. To account for differences in the surface topography fields between the two grids, the surface pressure field was adjusted after interpolation using the change in elevation and the hydrostatic relation. The same EnKF-derived ensemble perturbations as used in the operational approach were then added to the interpolated 4D-Var analysis.

3. Description of experiments and verification tools

a. Experiment configurations

A set of ensemble forecast experiments were performed over 42 cases separated by 36 h during the periods of 1 December 2007–31 January 2008 (boreal winter) and 11 July–10 September 2007 (boreal summer). The experimental approaches used to initialize the ensemble forecasts are as described in section 2 and summarized in Table 1. For this study, only 10-day ensemble forecasts are produced, whereas the operational system produces forecasts up to 16 days.

The SVs were computed using nonlinear, tangent-linear, and adjoint versions of the GEM model configured with a 120 \( \times \) 60 horizontal grid (3\( ^\circ \) latitude and longitude grid spacing) and 28 vertical levels. The tangent-linear and adjoint models, as partially described by Tanguay and Polavarapu (1999), are employed with a simplified planetary boundary layer parameterization (Laroche et al. 2002). The initial conditions used to obtain the linearization trajectory for the SV calculation were the same EnKF-derived mean analyses used to specify the EPS ensemble mean.

The ensemble perturbations computed using the SVs differ qualitatively from those obtained with the EnKF. Buehner and Zadra (2006) performed a comparison of SV perturbations with perturbations from a preoperational version of the EnKF which is relevant for the present study. They demonstrated that SV perturbations have a higher relative contribution from potential energy than from kinetic energy at initial time with the kinetic energy growing more quickly during the first
24–48 h of the forecasts. In contrast, the EnKF perturbations have a relative distribution between kinetic and potential energy that is more similar and remains stable during the forecasts. In addition, the average vertical structure of SV and EnKF perturbations were shown by Buehner and Zadra (2006) to be quite different with SVs concentrated around 600–700 hPa and the EnKF perturbations around 350 hPa. As shown in Fig. 11a of Buehner and Zadra (2006), the SVs have a vertical tilt in the zonal direction that is characteristic of perturbations growing by baroclinic processes. Such a tilt is much less evident in the EnKF perturbations, consistent with Fig. 4 of Buizza et al. (2008) that compares perturbations computed with either SVs or their ensemble 4D-Var approach.

b. Verification tools

Various verification scores are computed to evaluate the relative performance of the approaches described above. To evaluate the average growth rate and spatial distribution of the forecast ensemble spread, it is compared against the RMS error of the ensemble mean forecast (relative to the 4D-Var operational analysis) at various lead times.

The usefulness of ensemble forecasts can be measured both in terms of the overall statistical consistency between predicted probabilities and observed frequencies and the ability to correctly capture variations in the probability distribution (relative to the climatological distribution) over subsets of cases. These two aspects are referred to as the reliability and resolution, respectively. A measure of the quality of probabilistic forecasts for scalar variables that includes both reliability and resolution is the continuous ranked probability score (CRPS; Stanski et al. 1989). The CRPS is defined as the mean squared error in the predicted cumulative distribution. Following the procedure summarized by Candille et al. (2007), the CRPS is computed relative to radiosonde observations and is obtained separately over the northern extratropics, tropics, and southern extratropics. The statistical significance of differences in the CRPS between pairs of experiments is also computed using a bootstrap resampling technique. The CRPS can be decomposed into separate terms representing the reliability and resolution (Hersbach 2000). The smallest possible value for the CRPS (and its reliability and resolution components) is zero, occurring only for a perfect deterministic system.

4. Results with SV and EnKF initial ensemble perturbations

The first set of results demonstrates the impact of using different ensemble perturbations for initializing the ensemble forecasts. All experiments use the same initial ensemble mean, obtained from the EnKF ensemble mean analysis. The perturbations are generated using either 20 members of the EnKF analysis ensemble (experiment EnKF-EnKF), random linear combinations of SVs (experiment EnKF-SV), or no perturbations (experiment EnKF-NP).

a. Impact on the mean ensemble spread

The approaches are first evaluated in terms of the mean ensemble spread. Figure 1 shows the ensemble spread (standard deviation) averaged over the 42 cases during the boreal winter of 2007–08 and over three geographical regions: the northern extratropics (NH-X), tropics (TR), and southern extratropics (SH-X). In the two extratropical regions, the spread is shown for geopotential height at 500 hPa and, in the tropics, it is shown for temperature at 850 hPa. Also shown is the RMS error in the ensemble mean forecast from each experiment. A perfect ensemble prediction system would produce an ensemble spread that is statistically consistent with the error in the ensemble mean. The 90% confidence interval is also shown for each quantity as computed with a bootstrap resampling (with replacement) of the data from the 42 cases using 1000 random samples.

The results in Figs. 1a,c,d,f show that the mean ensemble spread generally agrees with the error in the ensemble mean over the two extratropical regions when using initial perturbations derived from either the EnKF analyses or the SVs. Note that the perturbations obtained using SVs were tuned to give a similar mean ensemble spread as the experiment with EnKF perturbations at 48-h lead time. For earlier lead times, the SV perturbations produce less ensemble spread and, for later times, they produce slightly more ensemble spread than the EnKF perturbations in the extratropics. However, these differences in ensemble spread for later times are mostly

<table>
<thead>
<tr>
<th>Expt</th>
<th>Ensemble mean</th>
<th>Ensemble perturbations</th>
</tr>
</thead>
<tbody>
<tr>
<td>EnKF-EnKF</td>
<td>EnKF ensemble mean analysis</td>
<td>EnKF ensemble perturbations</td>
</tr>
<tr>
<td>EnKF-SV</td>
<td>EnKF ensemble mean analysis</td>
<td>SV perturbations computed with the dry total energy norm</td>
</tr>
<tr>
<td>EnKF-NP</td>
<td>EnKF ensemble mean analysis</td>
<td>No perturbations</td>
</tr>
<tr>
<td>4DVar-EnKF</td>
<td>High resolution 4D-Var analysis</td>
<td>EnKF ensemble perturbations</td>
</tr>
</tbody>
</table>
FIG. 1. The average ensemble spread (i.e., the ensemble standard deviation; black lines) and RMS error of the ensemble mean forecast (gray lines) of the 500-hPa geopotential height for the (a),(d),(g) northern extratropics; (c),(f),(i) southern extratropics; and for the (b),(e),(h) temperature at 850 hPa in the tropics as a function of lead time. Results are shown for the (top) EnKF-EnKF, (middle) EnKF-SV, and (bottom) EnKF-NP experiments. The vertical line segments represent the 90% confidence interval computed from a bootstrap resampling procedure. These results are computed from 42 cases over boreal winter 2007–08.
not statistically significant according to the 90% confidence interval. This indicates that, as expected, the SV perturbations have a higher growth rate than the EnKF perturbations. This difference is most apparent during the first day, during which the EnKF perturbations appear to grow much slower than the ensemble mean error and have a larger spread than the error in the ensemble mean. For lead times larger than 24 h, the EnKF-EnKF and EnKF-SV experiments both produce slightly less spread than the error in the ensemble mean. Such a result would also occur for a perfect ensemble forecast because of the contribution of error in the verifying analysis. Comparing results from the extratropical regions of the two hemispheres after the first day, the growth rate appears to be consistent with the growth in the ensemble mean error in the winter hemisphere (Figs. 1a,d), whereas in the summer hemisphere the ensemble spread grows at a slightly lower rate than the ensemble mean error (Figs. 1c,f). As expected, the ensemble spread obtained when using no initial perturbations (Figs. 1g–i) is smaller than when using EnKF or SV perturbations for all lead times. Perhaps surprisingly, though, the results from this experiment show that the initial perturbations are responsible for less than half of the RMS ensemble spread after 48 h and at a lead time of 10 days the ensemble spread obtained with and without initial perturbations is nearly equal.

Results for the tropics (Figs. 1b,e) also show a higher growth rate for the SV perturbation than for the EnKF perturbations, especially over the first 48 h. During this period the SV perturbations appear to greatly overestimate the growth rate of the ensemble mean error. After the first 24 h, it appears that the initial perturbations are responsible for a much smaller fraction of the ensemble spread than in the extratropics. Already at day 2, the ensemble spread of the EnKF-NP experiment is only about 20% smaller than the spread in the EnKF-EnKF experiment. This suggests that the techniques, described earlier, to simulate model errors play a more important role in the tropics than in the extratropics for determining the growth in the ensemble spread. Also, the very narrow confidence intervals for the ensemble spread indicate that very little day-to-day variations occur in the spatially averaged ensemble spread over the tropics. Overall, the growth rate of ensemble spread and ensemble mean error of temperature at 850 hPa in the tropics is much lower than for geopotential height at 500 hPa in the extratropics.

Similar results are shown in Fig. 2, but for a period during the boreal summer of 2007. Again the SV perturbations produce higher growth rates for all three regions. Like for the boreal winter period, the growth rate of ensemble spread in the extratropics of the winter hemisphere (Fig. 2, right panels) is more consistent with the growth of the ensemble mean error, whereas for the summer hemisphere (Fig. 2, left panels) the ensemble spread somewhat underestimates the growth of the ensemble mean error at longer lead times. The results for the tropics (Fig. 2, middle panels) are very similar to those for the boreal winter period.

In Fig. 3, the geographical distribution of the RMS error of the 48-h ensemble mean forecast (i.e., the average of the 20 ensemble forecast members) from the EnKF-EnKF experiment (Fig. 3a) is shown together with the geographical distributions of the ensemble spread (i.e., the ensemble standard deviation) for the three ensemble prediction experiments. The ensemble mean error of 500-hPa geopotential height is much larger in the extratropics than in the tropics. Also, locally large errors are seen over the North Pacific and North Atlantic in the Northern Hemisphere and over much of the Southern Hemisphere between 45° and 75°S. The ensemble spread obtained with no perturbations (Fig. 3b) is much smaller than the ensemble mean forecast error with the largest values seen over both polar regions and a slightly larger spread over the North Pacific than over Asia and Europe. The results when using either SV (Fig. 3c) or EnKF (Fig. 3d) perturbations are remarkably similar. Consistent with the large-scale structure of the ensemble mean error, both show locally large spread over the North Pacific, North Atlantic, and in the region north of Antarctica close to 180° longitude.

Figure 4 shows the same results as the previous figure, but for the ensemble mean forecast error and ensemble spread obtained from the 120-h forecasts. Again, the ensemble spread obtained using the SV (Fig. 4c) and EnKF (Fig. 4d) perturbations are very similar and both correspond quite well with the ensemble mean error (Fig. 4a), though with a smaller amplitude and less small-scale detail. Even when using no initial perturbations, the mean ensemble spread (Fig. 4b) of the 120-h forecasts shows some of the same patterns as when using initial perturbations, such as the locally large spread over the North Pacific, North Atlantic, and across the Southern Hemisphere between 45° and 75°S.

b. Probabilistic forecast scores

A more comprehensive measure of the quality and utility of the probabilistic forecasts is given by the CRPS, as described in section 3b. The CRPS for the boreal winter period computed using radiosonde observations over the northern extratropics, tropics, and southern extratropics is shown in Fig. 5. Similar to Fig. 5, Fig. 6 shows the difference between the CRPS from the EnKF-EnKF and EnKF-SV experiments together with the 90% confidence interval as determined by the bootstrap
resampling procedure described by Candille et al. (2007). Therefore, only when the zero line is outside of the confidence interval do we consider the difference between the two experiments to be significant. Negative values of the difference correspond to better ensemble forecasts from the EnKF-EnKF experiment. In both extratropical regions the CRPS for the 500-hPa geopotential height is similar for the EnKF-EnKF and EnKF-SV experiments. In the northern extratropics (Figs. 5a and 6a), the differences are only statistically significant for the first 2 and the last 2 days of the 10-day forecasts during which the EnKF perturbations produce better probabilistic...

FIG. 2. As in Fig. 1, but for boreal summer 2007.
forecasts (lower values of CRPS). For lead times between 3 and 7 days the SVs produce slightly better forecasts, but the differences are too small to be statistically significant. In the southern extratropics (Figs. 5c and 6c), the EnKF perturbations produce improved forecasts that are statistically significant, but only for lead times up to 3 days. For later lead times, differences in the CRPS from the two approaches are not statistically significant. When using no initial perturbations (EnKF-NP), the CRPS is significantly higher than either of the other two experiments for almost all lead times in both extratropical regions. In the tropics (Fig. 5b), the use of SV perturbations results in CRPS values similar to using no perturbations for the lead times up to 2 days and values significantly higher than with EnKF perturbations up to 6 days. For later lead times, the difference in CRPS from using SV or EnKF perturbations is not statistically significant.

Figure 7 shows similar results as the Fig. 5, but for the boreal summer period. Also, like Fig. 6, Fig. 8 shows the difference between the CRPS from the EnKF-EnKF and EnKF-SV experiments and the 90% confidence interval, but for the boreal summer period. These results are generally similar to those obtained for the boreal winter period. Again, the EnKF perturbations produce lower values of CRPS than the SV perturbations, especially at early lead times. For the northern extratropics, however, this difference is statistically significant for all lead times (Figs. 7a and 8a) whereas it is only significant up to day 2 in the southern extratropics (Figs. 7c and 8c). For the northern extratropics, the use of no initial perturbations produces surprisingly similar results as the other two experiments. This seems to indicate a higher importance of model error relative to initial condition error during the summer as compared with the winter period.

The CRPS was also computed for other variables and levels, including temperature, geopotential height, and both wind components at 250, 500, 850, and 925 hPa in both extratropical regions and the tropics. Overall, the results from the full set of variables and levels are generally consistent with the results already discussed.

To help understand the differences seen in the CRPS, it is decomposed into its reliability and resolution components as shown in Fig. 9 only for the northern extratropics during the boreal winter period. The differences in CRPS between the EnKF-EnKF and EnKF-SV experiments (Fig. 6a) during the first 2 days of the forecasts appear to be due to the reliability component (Fig. 9a), whereas the differences for the last 2 days are due to the resolution component (Fig. 9b). The poor reliability of the ensemble forecasts with SV perturbations at early lead times is likely related to the small ensemble
spread that is, on average, much smaller than the ensemble mean error (Fig. 1d). By including evolved SVs in the calculation of the initial perturbations, as done at ECMWF (Leutbecher and Palmer 2008), the reliability for early lead times would likely be improved. The reliability is, however, improved when using the SV perturbations as compared to using the EnKF perturbations for days 4–7. The resolution component (Fig. 9b) is very similar when using SV or EnKF perturbations for all lead times except for days 9 and 10 during which significantly lower values are obtained with the EnKF perturbations. The use of no initial perturbations clearly degrades the reliability component over all lead times, whereas the resolution component is very similar to the

---

**FIG. 4.** As in Fig. 3, but for the 120-h ensemble forecasts.

**FIG. 5.** The CRPS of 500-hPa geopotential height for (a) northern extratropics, (c) southern extratropics, and (b) temperature at 850 hPa in the tropics. The results are shown for the EnKF-EnKF (solid), EnKF-SV (dashed), and EnKF-NP (dashed–dotted) experiments during boreal winter 2007–08.
other experiments for the early portion of the forecasts. The significant improvement in CRPS from using EnKF perturbations relative to SV perturbations during the boreal summer in the northern extratropics is mostly due to the resolution component beyond a lead time of 2 days (not shown). Such differences in the medium range for the resolution component are more difficult to explain than the large differences in ensemble spread during the first 48 h. It seems to indicate that the spatial distribution and structure of the initial perturbations can affect the quality of probabilistic scores well into the medium range. This is also supported by the results of Buizza et al. (2008, see their Fig. 7), which show that the ensemble 4D-Var approach (which has some resemblance to the EnKF approach) results in perturbations that explain a larger fraction of forecast error than SV perturbations up to day 4 in the extratropics even though the ensemble spread is much smaller than the error in the ensemble mean.

Figure 10 shows the sample mean RMS error of the ensemble mean conditioned on the predicted spread (standard deviation). This is computed following the approach outlined by Leutbecher and Palmer (2008) in which the ensemble mean error and spread are first ordered according to the value of ensemble spread and binned into 20 equally populated groups. The average RMS ensemble mean error and ensemble spread are then both calculated for each of the 20 bins. The resulting
average RMS error is shown in Fig. 10 as a function of the average ensemble spread for lead times of 2 (Fig. 10a), 5 (Fig. 10b), and 10 days (Fig. 10c). Similar to the results of Leutbecher and Palmer (2008), this shows that the ensemble forecasts spread provides useful information on the uncertainty in the ensemble mean forecast (i.e., all curves in Fig. 10 have a positive slope). However, when using both the EnKF and SV perturbations the ensemble spread generally overestimates the RMS ensemble mean error when the error is large. This is seen for all three lead times shown. In general, the results from using EnKF and SV perturbations are quite similar, but the SV perturbations result in a slightly larger range of values for both the ensemble mean error and the ensemble spread at the lead time of 2 days. Even with no initial perturbations, the ensemble spread already provides useful information at a lead time of 2 days, though the ensemble spread systematically underestimates the ensemble mean error, as already noted. At a lead time of 10 days, the result from using no initial perturbations is very similar to the other two experiments.

5. Results with the 4D-Var analysis as initial ensemble mean

The second set of results focus on the impact of using different approaches to specify the ensemble mean when initializing the ensemble forecasts. All experiments use the same initial ensemble perturbations, obtained from the EnKF analysis ensemble. The ensemble mean is obtained either from the EnKF analysis ensemble mean (experiment EnKF-EnKF) or from the spatially interpolated 4D-Var deterministic analysis (experiment 4DVar-EnKF). Because the EnKF analysis ensemble mean is an average of the 96 EnKF analyses, it may be expected to be spatially smoother than the deterministic analysis.

a. Verification of deterministic forecasts

The quality of the analyses used for specifying the initial ensemble mean of the EPS is first evaluated in the context of deterministic forecasts. The EnKF analysis ensemble mean and the interpolated 4D-Var analysis were each used to initialize a set of 42 deterministic forecasts for both the boreal winter and summer periods. The same forecast model configuration was used for all forecasts. The quality of the resulting forecasts is measured by using radiosonde observations to compute both standard deviations and bias. Figure 11 shows the standard deviation and bias of forecast error computed relative to radiosonde observations for 500-hPa geopotential height in the extratropics and temperature at 850 hPa in the tropics. The level of statistical significance of the differences between the two experiments is also shown when the value is at least 95%. The results for the two experiments are generally quite similar for the boreal winter period with the exception of small but statistically significant differences favoring the forecasts initialized with 4D-Var analyses in the northern extratropical stratosphere (not shown). Figure 12 shows the same verification scores as the previous figure, but now for the boreal summer period. In the northern extratropics, statistically significant differences generally favoring the forecasts initialized with 4D-Var analyses are seen at most lead times. In the southern extratropics a consistent positive impact from using 4D-Var analyses is seen (for the standard deviation) only up to a lead time of 3 days. Same as for the boreal winter period, the results from the 2 experiments are very similar in the
tropics, but with a smaller temperature bias for the first 2 days when using the 4D-Var analyses.

The results for the boreal winter period of December 2007–January 2008 can be compared to those obtained by Buehner et al. (2010b) for February 2007. In that study, implementations of the EnKF and 4D-Var systems were used in which both used the same vertical levels and assimilated the same observations. This is in contrast to the present study for which the 4D-Var analysis system has significantly higher vertical resolution than the EnKF. As mentioned earlier, some additional types of observations are also assimilated in the 4D-Var relative to those used in the EnKF. This may partly explain why the positive results for the EnKF relative to 4D-Var obtained by Buehner et al. (2010b) are not reproduced in the present set of experiments. Specifically, that study found significant differences in favor of the EnKF in the extratropical troposphere at medium range and in the tropics for all lead times. Nonetheless, some qualitative differences between the analyses produced by the two systems as shown in the experiments of Buehner et al. (2010a,b) are also applicable to the present study. Specifically, the EnKF ensemble mean analyses fit the radiosonde observations significantly less closely than the 4D-Var analyses and the EnKF is able to capture flow-dependent features of the background-error covariances better than the 4D-Var system with its relatively simple background-error covariance representation.

b. Probabilistic forecast scores

The quality of the probabilistic forecasts produced by the EnKF-EnKF and 4DVar-EnKF experiments are evaluated by again computing the CRPS using radiosonde observations. Figure 13 shows the CRPS as a function of forecast lead time for the boreal winter period of the 500-hPa geopotential height for the extratropics and temperature at 850 hPa for the tropics. Since the two experiments only differ with respect to the specification of the initial ensemble mean and nearly neutral results were obtained from the deterministic forecasts, it is not surprising that the CRPS is also very similar for the two experiments. Statistically significant differences are only obtained in the northern extratropics (Fig. 13a) for day 1 and in the southern extratropics (Fig. 13c) for days 1 and 9, favoring the use of 4D-Var analysis in each case. For the tropics (Fig. 13b), a significant difference is obtained for day 2, in favor of the EnKF analysis.

Figure 14 shows the same probabilistic forecast scores as the previous figure, but for the boreal summer period. For the northern extratropics (Fig. 14a), a significant difference in favor of using the 4D-Var analysis is obtained for lead times of 1 and 2 days, while a significant difference in favor of the EnKF analysis is obtained at day 5. For the southern extratropics (Fig. 14c) the use of the 4D-Var analysis results in a significantly improved CRPS for forecast days 1, 3, and 5–7. A similar improvement is also seen in the tropics where a significant difference in favor of the 4D-Var analysis is obtained for forecast days 1–4.
6. Conclusions

The initial ensemble mean and perturbations in the Canadian operational EPS are obtained from the EnKF. In this study, several alternative strategies were evaluated for specifying both the ensemble mean and perturbations of the 20 initial EPS members. All experiments were performed over 2-month periods during both the boreal summer and winter using a system very similar to the global ensemble prediction system that became operational on 10 July 2007.

Relative to the operational configuration that relies on the EnKF, the use of SVs to compute initial perturbations results in a nearly neutral impact on probabilistic
forecast scores in the winter hemisphere extratropics during both seasons, but statistically significant differences in favor of the EnKF both in the tropics and, for a limited set of lead times, in the summer hemisphere extratropics. Both approaches lead to significantly better ensemble forecasts than with no initial perturbations, though results are quite similar in the tropics when using either SVs or no perturbations. The poor results from using SVs in the tropics is partially due to simply using the same approach to compute the SVs as in the extratropics, which has previously been shown to produce unrealistic perturbation growth (Barkmeijer et al. 2001). In addition, the use of an initial-time norm that does not include information on analysis uncertainty and the lack of linearized moist processes in the calculation of the SVs are two factors that limit the overall quality of the resulting SV-based ensemble forecasts. On the other hand, the ability of the EnKF-based perturbations to accurately describe analysis uncertainty is likely limited by the relatively large random perturbations.

FIG. 13. The CRPS of 500-hPa geopotential height for the (a) northern extratropics, (c) southern extratropics, and for (b) temperature at 850 hPa in the tropics computed from the 42 cases during boreal winter 2007–08 for the EnKF-EnKF (solid) and 4DVar-EnKF (dashed) experiments.
added to the EnKF analysis ensemble members (to provide sufficient ensemble spread in the early medium range). Such significant approximations in the implementations of both approaches should be considered when interpreting the results. Future research on both approaches may lead to improvements with respect to these and other approximations and could therefore affect the conclusions of such a comparison.

Relative to the operational configuration, use of the 4D-Var analysis to specify the initial ensemble mean results in a positive impact on the probabilistic forecast scores during the boreal summer period in the southern extratropics and tropics, but a near-neutral impact otherwise. However, recent improvements in the vertical resolution of the operational EnKF would likely reduce the potential improvement from using the 4D-Var analysis to specify the initial ensemble mean state as demonstrated in this study.

As already mentioned, no effort was made to better capture the errors in the tropics with SVs constrained to the regions around tropical cyclones or to capture the errors that grew in the past by using evolved SVs. Instead, it was decided to direct any possible future research effort on SVs toward combining the EnKF and SV approaches. This is supported by the better performance of the EnKF in the tropics and more realistic error growth for early lead times from using EnKF perturbations. Also, the results of Buizza et al. (2008) suggest that the use of an ensemble data assimilation approach better captures the error from structures that grew in the past than the use of evolved SVs. The perturbations from the EnKF can be incorporated in two different ways with SVs for computing initial-time perturbations, as described next.

The spatial and multivariate structure and growth rate of SVs are strongly influenced by the initial- and final-time norms used in their calculation. Consequently, this can potentially have a significant impact on the quality of ensemble forecasts produced from SV-based initial perturbations. As demonstrated theoretically by Ehrendorfer and Tribbia (1997), the optimal initial-time norm for predicting the evolution of forecast uncertainty in a linear context is based on the inverse of the analysis-error covariance matrix. Buehner and Zadra (2006) demonstrated the impact on SVs from using a flow-dependent estimate of the analysis-error covariances from the EnKF analysis ensemble to specify such an initial-time norm. A natural extension to the present study would be to evaluate such a hybrid SV-EnKF approach in the context of ensemble forecasting and to compare this with the EnKF and SV approaches examined here. Such an approach may lead to improved probabilistic forecasts relative to the results with SV perturbations presented in this study, but possibly only for early lead times. This approach is somewhat similar to the Hessian singular vector approach examined by Barkmeijer et al. (1999), which was found to produce worse ensemble forecasts than when using SVs computed with a total energy norm. However, their approach used an estimate of analysis error covariances from 3D-Var, whereas the EnKF analysis ensemble provides a more flow-dependent estimate of these covariances. Alternatively, a simpler approach would be to combine perturbations from the SVs, as computed in the present study, with perturbations from the EnKF. Such an approach was evaluated by Buizza et al. (2008) by summing SV perturbations and perturbations from an ensemble of perturbed 4D-Var data assimilation experiments and was shown to provide better results than

![Fig. 14. As in Fig. 13, but for boreal summer 2007.](unauthenticated)
when using either of the two types of perturbations individually.

Acknowledgments. The authors thank Martin Charron and Ayrton Zadra for helpful comments on an earlier version of the manuscript. The official reviews of Roberto Buizza and an anonymous reviewer also led to significant improvements. A.M.’s work was partly supported by the government of Canada Program for International Polar Year, through the project ThorpeX Arctic Weather and Environmental Prediction Initiative (TAWEPI).

REFERENCES


