



## Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations

MARK BUEHNER, P. L. HOUTEKAMER, CECILIE CHARETTE, HERSCHEL L. MITCHELL, AND BIN HE

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

(Manuscript received 17 July 2009, in final form 3 November 2009)

### ABSTRACT

An intercomparison of the Environment Canada variational and ensemble Kalman filter (EnKF) data assimilation systems is presented in the context of producing global deterministic numerical weather forecasts. Five different variational data assimilation approaches are considered including four-dimensional variational data assimilation (4D-Var) and three-dimensional variational data assimilation (3D-Var) with first guess at the appropriate time (3D-FGAT). Also included among these is a new approach, called Ensemble-4D-Var (En-4D-Var), that uses 4D ensemble background-error covariances from the EnKF. A description of the experimental configurations and results from single-observation experiments are presented in the first part of this two-part paper. The present paper focuses on results from medium-range deterministic forecasts initialized with analyses from the EnKF and the five variational data assimilation approaches for the period of February 2007. All experiments assimilate exactly the same full set of meteorological observations and use the same configuration of the forecast model to produce global deterministic medium-range forecasts.

The quality of forecasts in the short (medium) range obtained by using the EnKF ensemble mean analysis is slightly degraded (improved) in the extratropics relative to using the 4D-Var analysis with background-error covariances similar to those used operationally. The use of the EnKF flow-dependent error covariances in the variational system (4D-Var or 3D-FGAT) leads to large (modest) forecast improvements in the southern extratropics (tropics) as compared with using covariances similar to the operational system (a gain of up to 9 h at day 5). The En-4D-Var approach leads to (i) either improved or similar forecast quality when compared with the 4D-Var experiment similar to the currently operational system, (ii) slightly worse forecast quality when compared with the 4D-Var experiment with EnKF error covariances, and (iii) generally similar forecast quality when compared with the EnKF experiment.

### 1. Introduction

This is the second part of a study to intercompare the variational and ensemble Kalman filter (EnKF) data assimilation approaches in the context of global deterministic numerical weather prediction (NWP). In the first part (Buehner et al. 2010, hereafter Part I), a detailed description of the experimental configurations is presented together with results from single-observation experiments using the different data assimilation approaches considered. The standard four-dimensional variational data

assimilation (4D-Var) and EnKF approaches differ in several important respects: the use of a deterministic versus an ensemble data assimilation approach, the use of an iterative variational algorithm versus a sequential (with respect to batches of observations) solution algorithm, differences in the approach for applying spatial covariance localization, the use of static versus flow-dependent background-error covariances, and different techniques for temporally evolving the error covariances within the assimilation window. Other configurations of the variational data assimilation approach considered in this study are more similar to the EnKF than the standard 4D-Var approach, thus allowing the impact of some of these differences to be evaluated. In Part I, results from a set of single-observation experiments demonstrate that the differences in the application of spatial localization and in

---

*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

the evolution of background-error covariances can result in relatively large differences in analysis increments.

This second part of the study focuses on results from realistic data assimilation experiments that span a 1-month period and use a full set of real meteorological observations, including satellite radiances. Experiments using five different configurations of the variational data assimilation approach are performed in addition to the EnKF. One of the variational data assimilation configurations is a relatively new approach, called Ensemble-4D-Var (En-4D-Var), that uses 4D ensemble background-error covariances. A similar approach was proposed by Liu et al. (2008). Several different types of verification scores are presented from a series of 6-day deterministic forecasts initialized with analyses obtained from each of the six data assimilation experiments. For the EnKF experiment, the ensemble-mean analysis was used to initialize the deterministic forecasts. All forecasts were produced using the same configuration of the Global Environmental Multiscale (GEM) model (Côté et al. 1998), which is very similar to the model used operationally for global deterministic forecasts at the Canadian Meteorological Centre (CMC) between 28 May 2008 and 22 June 2009 (Bélair et al. 2009). Consequently, this intercomparison uses configurations of the forecast model and a network of assimilated observations that make the results entirely relevant for operational NWP.

In the next section, a brief description of the configurations of the variational and EnKF data assimilation systems considered in this study is given (a more detailed description is provided in Part I). In section 3, results are presented from a series of 1-month data assimilation experiments and the resulting 6-day deterministic forecasts. Several different combinations of experiments are compared to highlight the impact of specific aspects that differentiate the experiments. Finally, some conclusions are given in section 4.

## 2. Brief description of data assimilation configurations

The data assimilation systems considered here are based on the operational configurations of the 4D-Var and EnKF systems that were implemented on 28 May 2008. At that time both operational systems were modified to expand the set of assimilated observations. It should be noted that both operational systems have undergone further significant modification during 2009, though these are not considered in this study. For the purpose of this study, numerous changes were made to the 2008 operational configurations of the 4D-Var and EnKF systems. The goal was to eliminate as many of the

differences as possible so that only the differences that are fundamental to each of the approaches remain. A detailed description of the configurations used in this study is given in Part I.

Five configurations of the variational data assimilation system are considered in addition to the EnKF. Experiments based on both 4D-Var and 3D-Var with first guess at the appropriate time (3D-FGAT) are included, each employing prescribed background-error covariances similar to those used operationally (referred to as 4D-Var-Bnmc and 3D-FGAT-Bnmc). These covariances are obtained with the “NMC method” (Parrish and Derber 1992), are static in time, and employ horizontally homogeneous and isotropic correlations. In addition, both 4D-Var and 3D-FGAT experiments are performed that use flow-dependent background-error covariances computed from the EnKF background ensembles with spatial covariance localization applied (referred to as 4D-Var-Benkf and 3D-FGAT-Benkf). The approach for applying spatial localization to ensemble background-error covariances in a variational data assimilation system was described and evaluated in the context of 3D-Var using a preoperational version of the EnKF by Buehner (2005). The final variational data assimilation experiment uses the En-4D-Var approach. In this approach, 4D flow-dependent background-error covariances estimated from EnKF ensembles are employed to produce a 4D analysis without the need for tangent-linear or adjoint versions of the forecast model. This is equivalent to how background-error covariances are evolved in the EnKF analysis procedure. A summary of the five variational data assimilation experiments is given in Table 1 of Part I.

All configurations of the variational data assimilation system use the incremental approach to produce their deterministic analysis on the same grid as that used for the deterministic medium-range forecasts ( $800 \times 600 \times 58$  L) while the analysis increment is computed on a lower-resolution grid ( $400 \times 200 \times 58$  L). To produce deterministic medium-range forecasts for the EnKF experiment, the ensemble-mean analysis is used to initialize the same configuration of the forecast model as used for the variational data assimilation experiments. Consequently, the mean of the EnKF analysis ensemble must be interpolated from its lower-resolution grid ( $400 \times 200 \times 58$  L) to the model grid ( $800 \times 600 \times 58$  L). This may cause the early portion of the resulting medium-range forecasts to be negatively affected because of the adjustment of the model fields to the higher-resolution grid and surface topography field.

Following the implementation of 28 May 2008, the types of observations assimilated operationally in the 4D-Var system were wind, temperature, and humidity

from radiosondes; wind and temperature from aircraft; wind, temperature, pressure, and humidity from in situ surface observations; wind from profilers over the United States; atmospheric motion wind from geostationary and polar-orbiting satellites; surface wind over water from the Quick Scatterometer (QuikSCAT); and radiances from the Advanced Microwave Sounding Unit A and B (AMSU-A/B), the Atmospheric Infrared Sounder (AIRS), the Special Sensor Microwave Imager (SSM/I), and geostationary satellites. In the operational EnKF, the same observations were assimilated as those used in the 4D-Var system except for the radiances from AIRS, SSM/I, and geostationary satellites. Another difference between the systems is that in the operational 4D-Var system, radiosonde humidity observations up to 70 hPa were assimilated, whereas in the EnKF they were only assimilated up to 200 hPa.

In the 2008 operational 4D-Var system, radiance observations were assimilated using version 8 of the Radiative Transfer for (A)TOVS (RTTOV) model (Saunders et al. 2006) and a vertical interpolation algorithm that ensures all relevant vertical levels of the analysis grid participate in the calculation of the brightness temperatures [same as the algorithm included in version 9 of RTTOV and described by Rochon et al. (2007)]. The vertical interpolation is needed to map temperature and humidity from the analysis grid vertical levels to the predefined pressure levels used by RTTOV. In the operational EnKF system, the same version of RTTOV was used as in the 4D-Var, but in conjunction with a simple linear interpolation with respect to the logarithm of pressure. Another difference in the assimilation of radiance observations is that the 4D-Var system performed the vertical interpolation on the natural logarithm of specific humidity before converting to specific humidity, whereas the EnKF interpolated specific humidity directly.

The two highest peaking AMSU-A channels assimilated (9 and 10) had different assigned values for the observation-error variance in the two operational systems. The values used in the 4D-Var system were artificially large when compared with the variances of the observation-minus-background differences. These large values were originally chosen to reduce problems caused by large analysis increments of temperature near the model top from assimilating these observations and an associated degradation in forecast quality. Because of the way vertical covariance localization is applied in the EnKF, large increments do not occur near the model top from assimilating AMSU-A observations (as demonstrated in Part I). Consequently when these data were first introduced into the EnKF, it was determined that using smaller, more realistic observation-error variances for these two channels improved the quality of the analyses.

Several changes related to the assimilated observations were made to the 2008 operational configurations of the two systems to remove most of the differences just described. To ensure that both systems assimilate exactly the same observations, the radiances from AIRS, SSM/I, and geostationary satellites were eliminated from the variational data assimilation configurations. Radiosonde observations of humidity were also eliminated above 200 hPa in the variational experiments to be consistent with the EnKF. The vertical interpolation algorithm used with RTTOV in the EnKF was changed to be the same as in 4D-Var. Conversely, the variational data assimilation system was modified so that this vertical interpolation is applied to specific humidity instead of its natural logarithm. The observation-error variances for the two highest peaking AMSU-A channels were left unchanged in the EnKF and variational data assimilation configurations. However, for the 4D-Var, 3D-FGAT, and En-4D-Var configurations that use background-error covariances derived from the EnKF background ensembles, the same observation-error variances were used as in the EnKF. Because of an important difference in how spatial covariance localization is applied, as discussed in section 5 of Part I, the analysis increments from the EnKF and variational data assimilation systems can be significantly different, even when using the same background and observation-error statistics.

The observation bias correction and all quality control decisions were extracted from an independent 4D-Var data assimilation experiment with a configuration very similar to the 2008 operational system. Consequently, all bias correction and quality control procedures were deactivated in the experimental configurations of the variational and EnKF data assimilation systems.

### 3. Results from 1-month analysis-forecast experiments

A series of 1-month analysis-forecast experiments were performed using the EnKF and the five configurations of the variational data assimilation system described above. All experiments assimilate the same observations over the month of February 2007. Analyses from each experiment valid at 0000 and 1200 UTC each day are used to produce 56 six-day forecasts. The forecast quality is measured by comparing mass and wind fields with the radiosonde observations that were accepted for assimilation by the operational 4D-Var data assimilation system and that also appear on a recent list of 637 high-quality radiosonde stations adopted by the World Meteorological Organization (WMO) for standard NWP verification. Verification scores computed against analyses are also presented using analyses from the 4D-Var-Bnmc

experiment to verify the results from all experiments. This was considered to be more appropriate than the usual approach of using the analyses from each experiment to verify its own forecasts because the analyses from the different experiments may differ significantly and the lower-resolution EnKF ensemble mean analyses likely have different statistical properties than the deterministic analyses. To confirm that this choice does not affect the conclusions, the EnKF ensemble mean analyses were also used to compute the verification scores for all experiments (not shown). In the two extratropical regions for which the scores are presented, use of the EnKF ensemble mean analyses resulted in the same relative ranking of the experiments as when using the 4D-Var-Bnmc analyses. Most verification scores are computed separately for the two extratropical regions and the tropics and for which  $20^\circ$  latitude is used to define the transition between tropics and extratropics. The maximum number of radiosonde stations available for verification in the three regions is 446, 104, and 87 for the northern extratropics, tropics, and southern extratropics, respectively.

*a. EnKF and 4D-Var-Benkf versus 4D-Var-Bnmc*

The first set of comparisons considers the EnKF, 4D-Var-Bnmc, and 4D-Var-Benkf experiments. The EnKF and 4D-Var-Bnmc experiments are most similar to the operational systems at CMC. These two systems differ with respect to the many aspects discussed in Part I. The 4D-Var-Benkf experiment is included in these comparisons to demonstrate the importance of one of these differences: the use of spatially localized and flow-dependent background-error covariances (EnKF and 4D-Var-Benkf) versus the use of stationary and highly parameterized covariances obtained using the NMC method (4D-Var-Bnmc).

The standard deviation and bias of the difference between all assimilated radiosonde observations and the corresponding analyses (observation minus analysis) over the entire month for both the EnKF and 4D-Var-Bnmc experiments are shown in Fig. 1. Note that even though the model extends up to 10 hPa, these and most results that follow are only shown up to 100 hPa. Results are not shown above 100 hPa primarily because large errors near the model top, mostly due to model error, make it difficult to discern impacts on the tropospheric scores (an example of such a result is shown in section 3d). From the larger values of standard deviation it is clear that the EnKF ensemble mean analyses fit the assimilated radiosonde observations less closely than the 4D-Var-Bnmc analyses for zonal wind, temperature, and dewpoint depression. For geopotential height, which is not an assimilated quantity in either the variational or EnKF systems, and for the biases of all observed

variables, no significant differences are seen. Overall, a similar set of results is obtained when comparing the fit of the 4D-Var-Benkf analyses to the 4D-Var-Bnmc analyses (not shown) and therefore the decreased fit to observations is likely due to differences in the background-error covariances.

The difference in verification scores between the forecasts launched from the EnKF and 4D-Var-Bnmc analyses are shown in Fig. 2. The standard deviations were first computed of the observed-minus-forecasted values for zonal wind  $U$ , temperature  $T$ , and geopotential height  $GZ$  radiosonde observations over the entire month for the northern extratropics (NH-X), tropics (TR), and southern extratropics (SH-X). Then this verification score for the EnKF experiment was subtracted from the same score for the 4D-Var-Bnmc experiment, which serves as the control experiment for this comparison. The result is shown as contour plots with respect to pressure and forecast lead time from days 1 to 6. Therefore, a positive value of the plotted quantity corresponds with a higher standard deviation for the 4D-Var-Bnmc experiment as compared with the EnKF experiment. We will refer to this as a positive impact for the EnKF experiment relative to the 4D-Var-Bnmc experiment.

The level of statistical significance of these differences in standard deviation is shown in Fig. 3. This was obtained using a bootstrap resampling procedure (with replacement) applied to the verification scores computed separately for 14 sets of nonoverlapping 2-day periods (obtained from the 28 days in February 2007, twice per day). Motivated by the results of Candille et al. (2007, section 4b), the resampling procedure was applied using 2-day periods, instead of 12-h periods, to reduce the negative impact that temporal correlations of the observation-minus-forecast differences have on the accuracy of the estimated significance levels. The verification scores, specifically the variances, were averaged over a random sample of 14 sets of 2-day periods chosen from the original 14 sets. This procedure was repeated for 10 000 random samples and the relative frequency that the EnKF experiment verified closer to the radiosonde observations than the 4D-Var-Bnmc experiment was computed. A significance level was then obtained from this frequency that represents the probability that the two standard deviations are distinct. Therefore, a significance level with a magnitude of 100% means the standard deviation for one experiment is always lower than the other in the 10 000 random samples. Similarly, a significance level of 0% corresponds with an equal frequency for one experiment having a lower standard deviation than the other, that is, a 0% probability that the verification scores are different. In Fig. 3, only the situations with a significance level above 90% or 95%

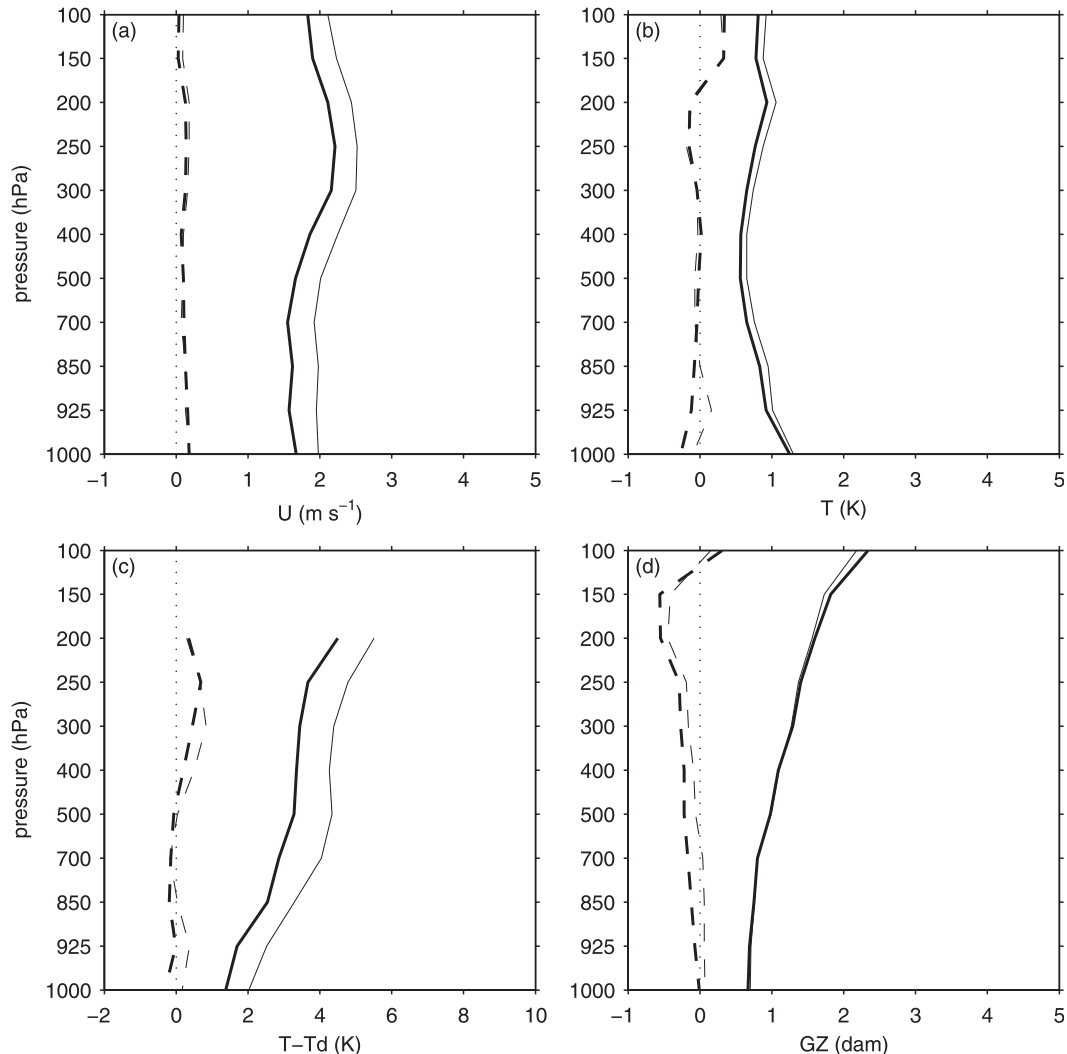


FIG. 1. Standard deviation (solid lines) and bias (dashed lines) of the analyses with respect to all radiosonde observations (observation minus analysis) for the EnKF ensemble mean analysis (thin lines) and the 4D-Var-Bnmc analysis (thick lines) with respect to (a)  $U$ , (b)  $T$ , (c) dewpoint depression  $T - T_d$ , and (d)  $GZ$ .

are indicated. A plus sign (filled circle) is used to denote situations in which the EnKF experiment has a significantly lower (higher) standard deviation than the 4D-Var-Bnmc experiment.

The results in Fig. 2 for the northern extratropics (left panels) generally show slightly negative impacts for the EnKF experiment relative to the 4D-Var-Bnmc experiment in the troposphere for forecasts up to a lead time of 3 days. For forecast days 5 and 6 the impacts become generally positive (with the maximum difference exceeding  $0.2 \text{ m s}^{-1}$ ,  $0.15 \text{ K}$ , and  $0.3 \text{ dam}$ , for zonal wind, temperature, and geopotential height, respectively). The corresponding panels in Fig. 3 confirm that the negative impacts for temperature and geopotential height for lead times up to day 3 are significant (to the 95% level);

however, the positive impacts at longer lead times are only significant for a very limited set of vertical levels and only for day 5. In the tropical region (middle panels of Fig. 2), a more systematic positive impact is obtained for the zonal wind (exceeding  $0.5 \text{ m s}^{-1}$ ). The corresponding panels in Fig. 3 show that the impact on zonal wind is significant and that significant (small) positive impacts also occur for temperature and geopotential height for the early portion of the forecasts. In the southern extratropics (right panels of Fig. 2) positive impacts are seen for all three variables with the largest impact on forecast day 5 (exceeding  $0.4 \text{ m s}^{-1}$ ,  $0.2 \text{ K}$ , and  $0.4 \text{ dam}$ ). However the results in Fig. 3 indicate that these impacts for zonal wind and temperature are statistically significant only over a restricted set of vertical levels. It is



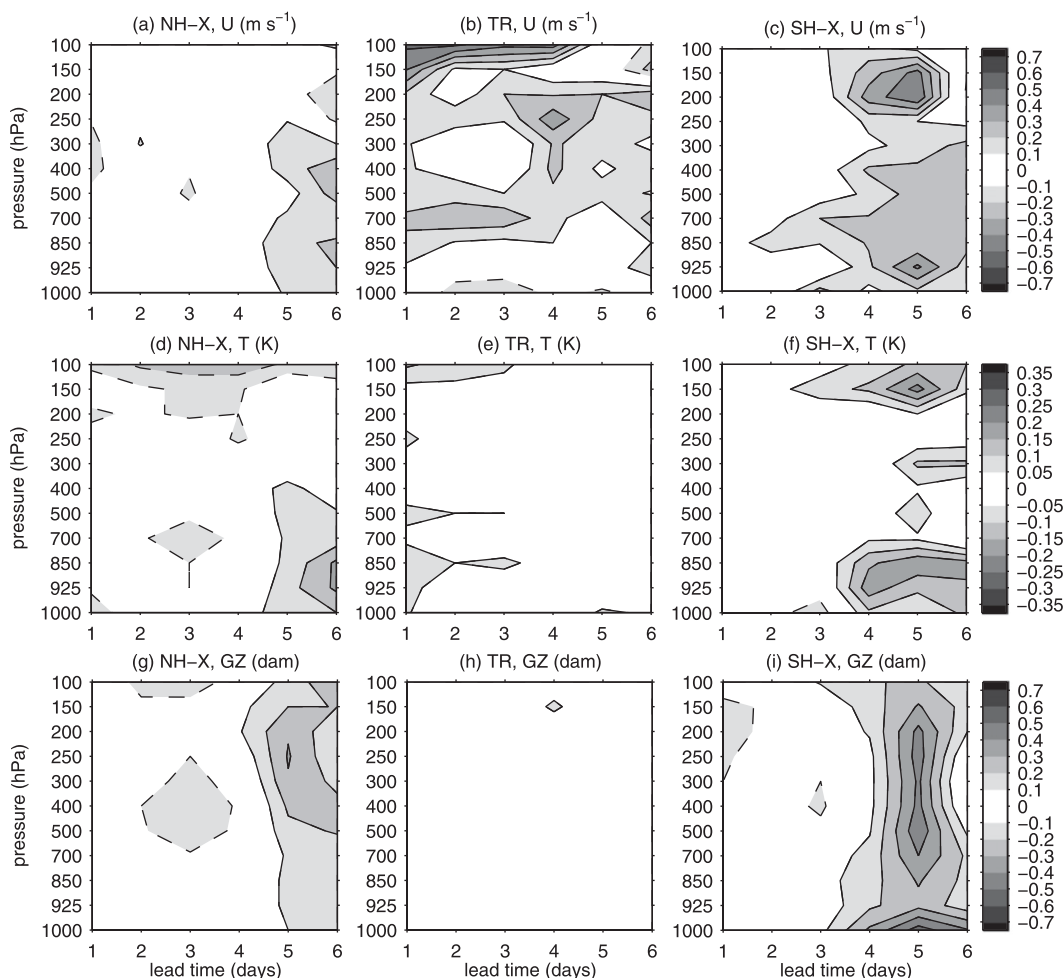


FIG. 2. Impact on the verification scores of the 6-day forecasts from the EnKF experiment with respect to the 4D-Var-Bnmc experiment as a function of lead time and pressure. The plotted quantity is the difference in the standard deviations of the forecasts verified against radiosonde observations such that positive values correspond with a positive impact (i.e., a reduction in the standard deviation) for the EnKF experiment relative to the 4D-Var-Bnmc experiment. Positive (negative) values are denoted by solid (dashed) contours. The impact is shown for (top)  $U$ , (middle)  $T$ , and (bottom)  $GZ$  for (left) NH-X, (middle) TR, and (right) SH-X.

possible that the negative impacts seen in the northern extratropics (and to a lesser extent in the southern extratropics) during the early stages of the forecasts are associated with an adjustment process caused by the difference in horizontal resolution and surface topography used for the lower-resolution EnKF analyses and the higher-resolution medium-range deterministic forecasts. In addition, any bias in radiosonde observations may give an apparent advantage to the 4D-Var-Bnmc approach since its analyses fit these observations more closely (as seen in Fig. 1).

The next set of verification scores presented are for the medium-range forecasts from the 4D-Var-Benkf experiment again relative to the 4D-Var-Bnmc experiment. This comparison shows the effect on the 4D-Var analysis of

replacing the temporally static and highly parameterized prescribed background-error covariances (similar to those used operationally) with the flow-dependent ensemble covariances estimated from the EnKF background ensembles. The same type of contour plot of differences in the observation-minus-forecast standard deviations, as shown for the previous comparison, is presented in Fig. 4. The statistical significance of these differences computed using the bootstrap resampling procedure is shown in Fig. 5. Generally neutral impacts are seen in the northern extratropics for the three variables and for all lead times and pressure levels, except for a small positive impact (positive contoured values) from the 4D-Var-Benkf experiment on zonal wind around 250 hPa during the first 3 forecast days. As seen in Fig. 5, statistically significant

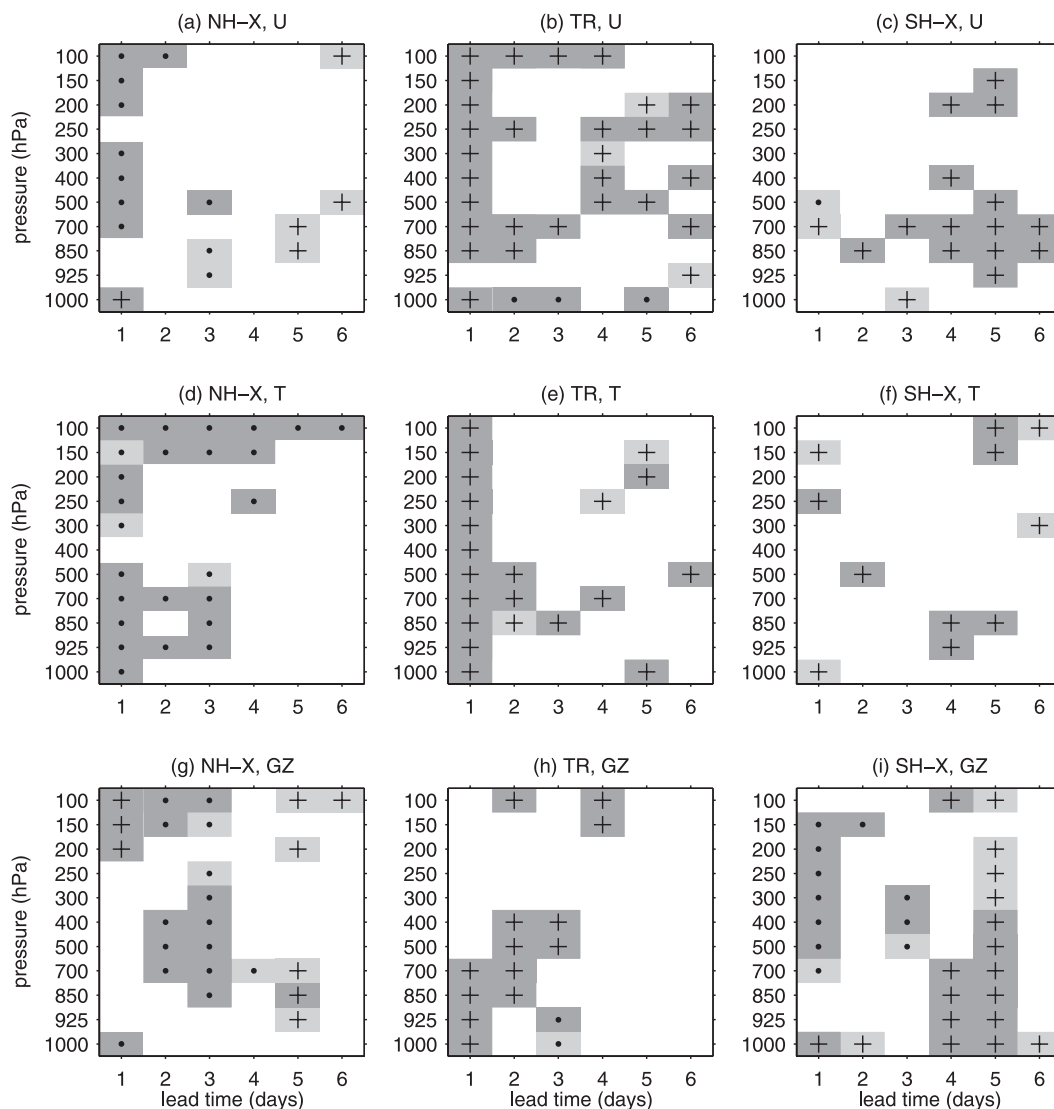


FIG. 3. Significance level for the difference in the standard deviations shown in Fig. 2 between the EnKF and 4D-Var-Bnmc experiments. Situations for which the significance level is  $>90\%$  are denoted with either a plus sign, when the EnKF experiment has the lower standard deviation, or a filled circle, when the 4D-Var-Bnmc experiment has the lower standard deviation. The light shading corresponds to a significance level  $>90\%$  and the dark shading corresponds to a significance level  $>95\%$ .

positive impacts are also obtained for temperature and geopotential height during the early portion of the forecasts. In the tropical region, like for the EnKF experiment, a positive impact is seen for the zonal wind (exceeding  $0.4 \text{ m s}^{-1}$ ) and also small positive impacts for temperature ( $0.05 \text{ K}$ ) and geopotential height ( $0.1 \text{ dam}$ ). Again, the results in Fig. 5 confirm the significance of this positive impact on zonal wind and also show that small, yet significant positive impacts (denoted by a plus sign) occur for temperature and geopotential height over a larger range of vertical levels and lead times than seen in Fig. 4. The largest positive impacts occur in the southern

extratropics for all three variables with maximum impact between forecast days 4 and 6 (exceeding  $0.7 \text{ m s}^{-1}$ ,  $0.25 \text{ K}$ ,  $0.6 \text{ dam}$ , for zonal wind, temperature, and geopotential height, respectively). The largest positive impacts for both zonal wind and geopotential height occur around 200 hPa. Positive impacts for the three variables in the southern extratropics are shown in Fig. 5 to be statistically significant over almost all lead times and vertical levels.

The three experiments just considered are now compared with respect to the anomaly correlation computed using analyses from the 4D-Var-Bnmc experiment. The anomaly correlation for the 500-hPa geopotential

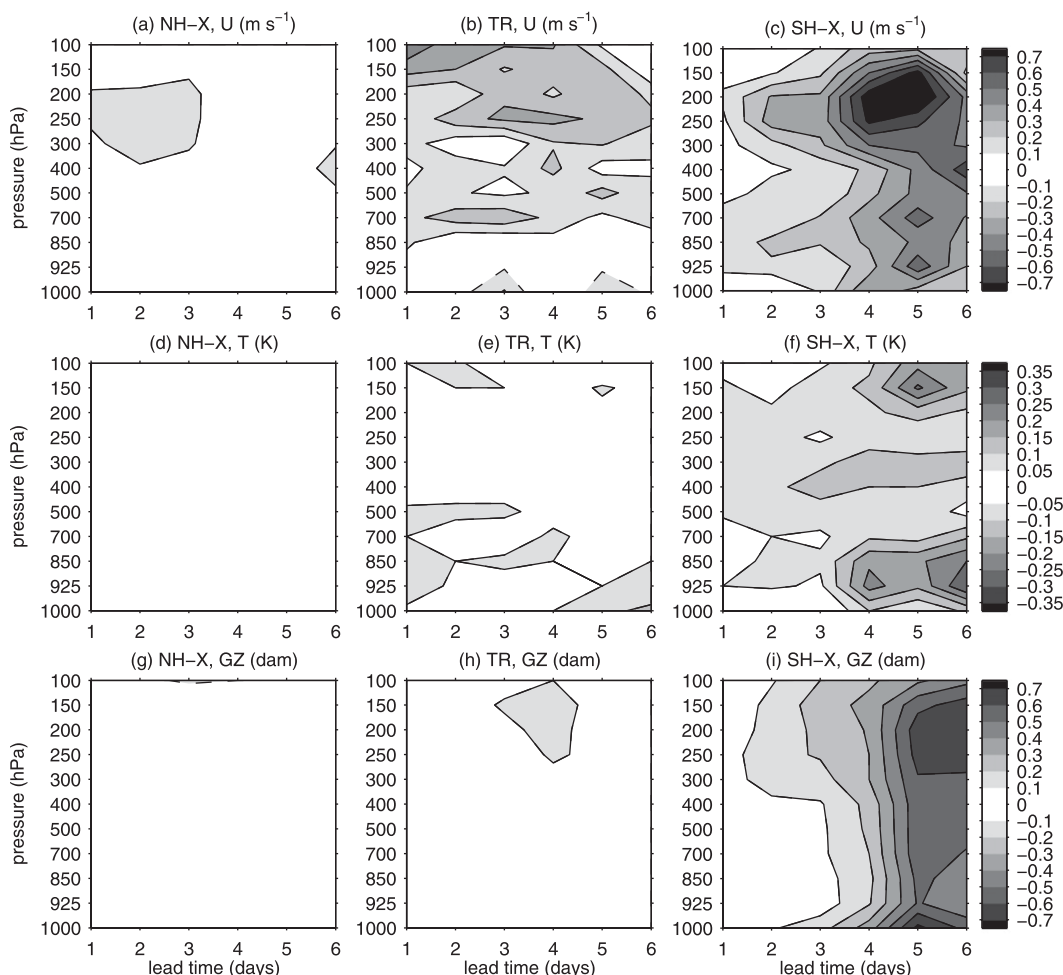


FIG. 4. As in Fig. 2, but showing the impact from the 4D-Var-Benkf experiment with respect to the 4D-Var-Bnmc experiment. Positive values (solid contours) correspond with a positive impact (i.e., a reduction in the standard deviation) for the 4D-Var-Benkf experiment relative to the 4D-Var-Bnmc experiment.

height is shown for the northern (Fig. 6a) and southern (Fig. 6b) extratropical regions. These results are generally consistent with the verification scores obtained using radiosonde observations. In the northern extratropics, using the EnKF ensemble mean analyses to initialize medium-range forecasts leads to a slightly negative impact up to forecast day 4 and a slightly positive impact for days 5 and 6, relative to the 4D-Var-Bnmc experiment. At day 5 this improvement is approximately equivalent to a gain of 1 h. Forecasts produced from the 4D-Var-Benkf experiment have slightly larger anomaly correlations of 500-hPa geopotential height up to day 4. For forecast days 5 and 6 the scores for the 4D-Var-Benkf and EnKF experiments are similar. For the southern extratropics, the EnKF experiment again leads to a slight negative impact up to day 4 and a slight positive impact thereafter, relative to the 4D-Var-Bnmc experiment. The analyses from the 4D-Var-Benkf experiment result in

forecasts of significantly better quality than the 4D-Var-Bnmc and EnKF experiments for all forecast lead times. The improvement steadily increases with the length of the forecast and at day 5 is approximately equivalent to a gain of 9 h with respect to the 4D-Var-Bnmc experiment.

An evaluation of the resulting precipitation forecasts from the three experiments just considered is now presented using the global  $1^\circ$  daily precipitation analyses produced by the Global Precipitation Climatology Project (Huffman et al. 2001). The evaluation was performed by computing both the bias and the equitable threat score (ETS) for the same three latitude bands used for the results already presented. These scores were computed for 24-h accumulated precipitation for the first 3 forecast days for several precipitation threshold values (where scores for a particular threshold include all precipitation events equal to or larger than that threshold value). For precipitation, the bias, for a given



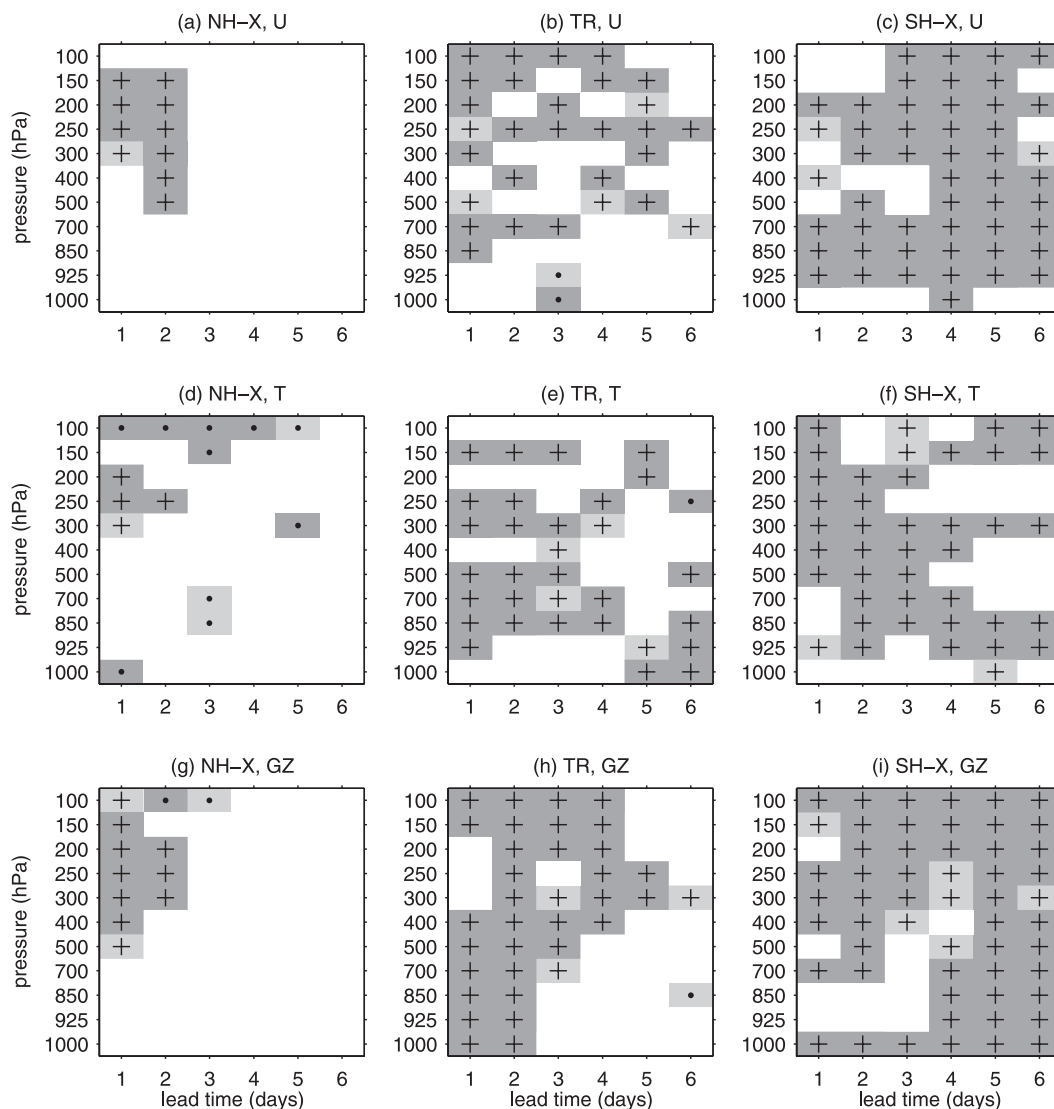


FIG. 5. As in Fig. 3, but showing the 90% and 95% significance levels for the difference in the standard deviations shown in Fig. 4 between the 4D-Var-Benkf and 4D-Var-Bnmc experiments. Situations for which the 4D-Var-Benkf experiment has a significantly lower (higher) standard deviation are denoted with a plus sign (filled circle).

latitude band and forecast lead time, is defined as the ratio of the total number of predicted events to the total number of observed events. A value of 1 represents no bias. The ETS, proposed by Schaefer (1990) in which it was called the Gilbert skill score, is widely used within the NWP community. It is essentially the ratio of the number of correctly predicted events to the total number of events predicted and/or observed. In addition, the number of correctly predicted events associated with random chance is removed from consideration when computing the ETS. Larger values of the ETS correspond with improved precipitation forecasts.

Verification scores for precipitation in both the northern and southern extratropical regions showed

little differences in the bias and ETS among the three experiments. However results from the tropics, shown in Fig. 7, have significant differences between the EnKF, 4D-Var-Bnmc, and 4D-Var-Benkf experiments during the early portion of the forecasts. During the first 24 h of the forecasts, the precipitation forecasts from the EnKF experiment show a much lower frequency of precipitation events (Fig. 7a) than for the two 4D-Var experiments (and the observations), especially for the smallest threshold values. Whereas the 4D-Var experiments lead to an overestimation of precipitation for most threshold values, the EnKF experiment leads to an underestimation. For forecast days 2 and 3, the bias (Figs. 7c,e, respectively) from the EnKF becomes slightly larger than for the

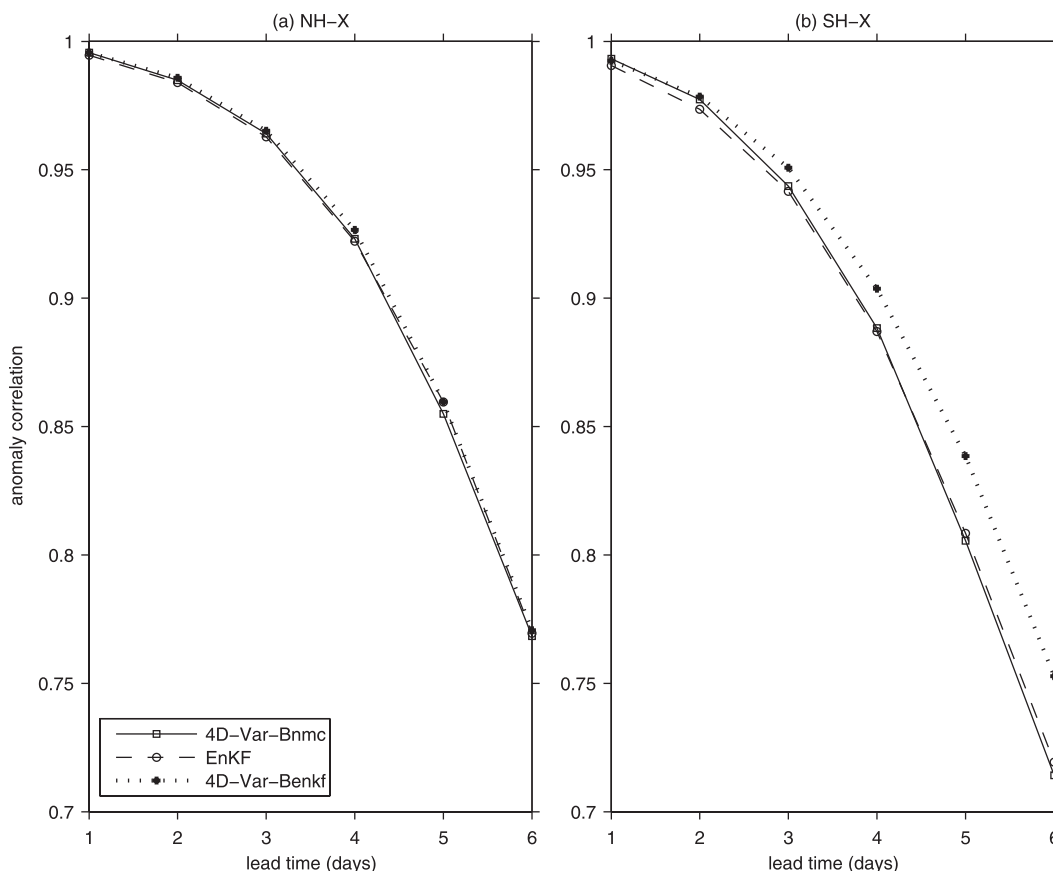


FIG. 6. Anomaly correlation of 500-hPa geopotential height for the (a) northern extratropics and (b) southern extratropics. Results from the 4D-Var-Bnmc (solid), EnKF (dashed), and 4D-Var-Benkf (dotted) experiments are shown.

4D-Var experiments for thresholds larger than 2 mm. Similarly, the bias from the 4D-Var-Benkf experiment is also lower than that from the 4D-Var-Bnmc experiment during the first forecast day and then becomes larger for the second and third forecast days, but the magnitude of the difference is much smaller than for the EnKF experiment. Considering now the ETS for the first 24 h of the forecasts, the EnKF experiment has lower values (representing lower quality forecasts) than those from the two 4D-Var experiments, except for the lowest threshold values. This is likely the period when the deterministic forecasts from the EnKF mean analyses are most severely impacted by the change in horizontal resolution. However, for forecast days 2 and 3, results from the EnKF experiment produce the highest ETS values except for threshold values above 10 mm (for which the number of cases is much smaller and therefore the results less robust). The 4D-Var-Benkf experiment leads to slightly improved ETS scores relative to 4D-Var-Bnmc for almost all threshold values and for all lead times.

The large differences in the precipitation bias during the first 24 h of the forecasts from the EnKF and 4D-Var experiments (as seen in Fig. 7) are also seen in the evolution of the 3-h accumulated precipitation averaged over all cases and spatially averaged over the two extratropical regions and the tropics (Fig. 8). In all three regions the EnKF experiment has the lowest value of 3-h accumulated precipitation at the beginning of the forecasts, but after 12–18 h attains larger values than the 4D-Var experiments. This behavior is most evident in the tropics (Fig. 8b) where the initial mean precipitation is 0.1 mm for the EnKF experiment and 0.6 mm for the 4D-Var-Bnmc experiment. However, after about 18 h, the forecasts from the EnKF experiment produce slightly higher levels of precipitation than the 4D-Var experiments. In the southern extratropics (Fig. 8c), the initial difference in the mean 3-h accumulated precipitation between the experiments is smaller than in the tropics; in the northern extratropics (Fig. 8a), the difference is smallest. Since the chosen period is February, it is possible that these differences are related to

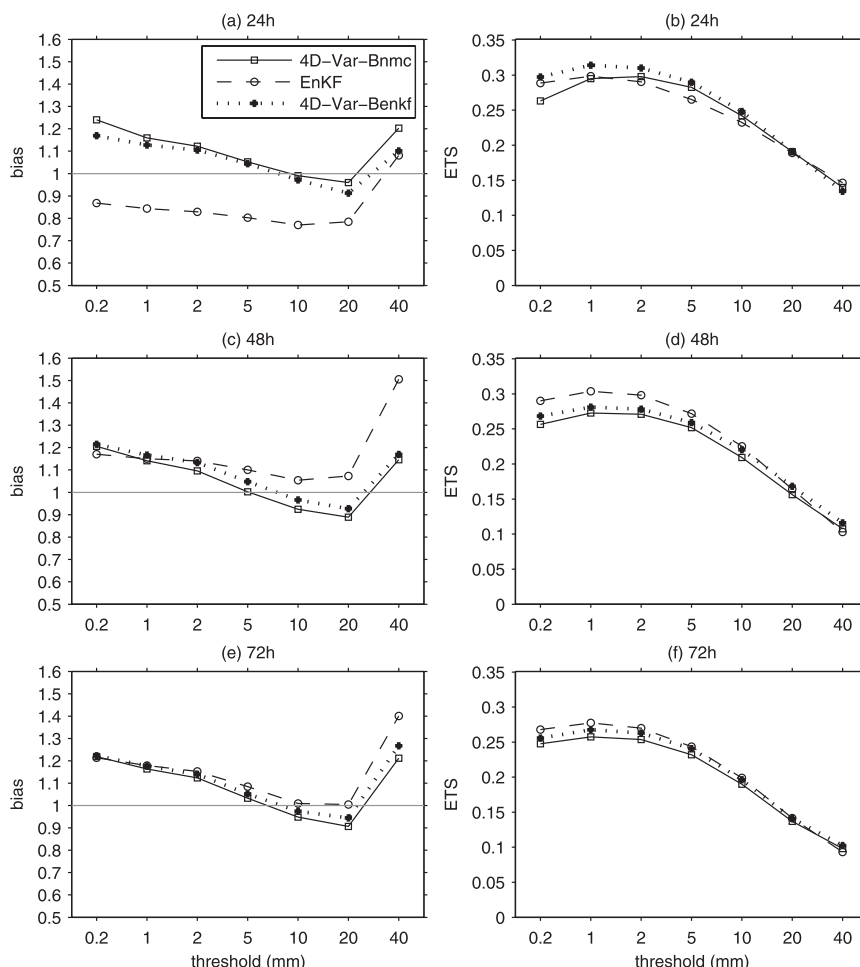


FIG. 7. (top–bottom) Verification scores for forecasts of 24-h accumulated precipitation in the tropics for the first three forecast days. The (a),(c),(e) bias and (b),(d),(f) ETS from the 4D-Var-Bnmc (solid), EnKF (dashed), and 4D-Var-Benkf (dotted) experiments.

differences in the initial amount of convection, which is most active in the tropics and southern extratropics. The use of the ensemble mean from the EnKF at a lower horizontal resolution to initialize the deterministic forecasts may produce vertical temperature and humidity profiles that are more stable than those produced by the deterministic analysis systems at the same resolution as the forecast model. This may be expected, since convection parameterizations are often tuned differently as a function of horizontal resolution such that convection is more easily activated in lower-resolution model configurations. For the tropical region, it is also interesting to note that while the 4D-Var-Bnmc experiment leads to an initially high precipitation intensity that decreases during the first 12 h of the forecasts, the 4D-Var-Benkf experiment has a much more consistent level of precipitation throughout the first 48 h. This suggests that using EnKF background-error covariances within the 4D-Var leads to analyses that are in

better balance with the model physics than using covariances similar to those in the operational system.

### b. 3D-FGAT-Benkf versus 3D-FGAT-Bnmc

Verification scores from the two 3D-FGAT experiments are now examined to determine if the impact obtained by replacing the background-error covariances similar to those used operationally (3D-FGAT-Bnmc) with those computed from EnKF background ensembles (3D-FGAT-Benkf) is similar to that obtained in the 4D-Var experiments. The changes in the standard deviations computed with respect to radiosonde observations for the 3D-FGAT-Benkf and 3D-FGAT-Bnmc experiments are shown in Fig. 9 and are comparable with the results shown in Fig. 4 for the 4D-Var experiments. In the northern extratropics, modest positive impacts are obtained in the troposphere for zonal wind (exceeding

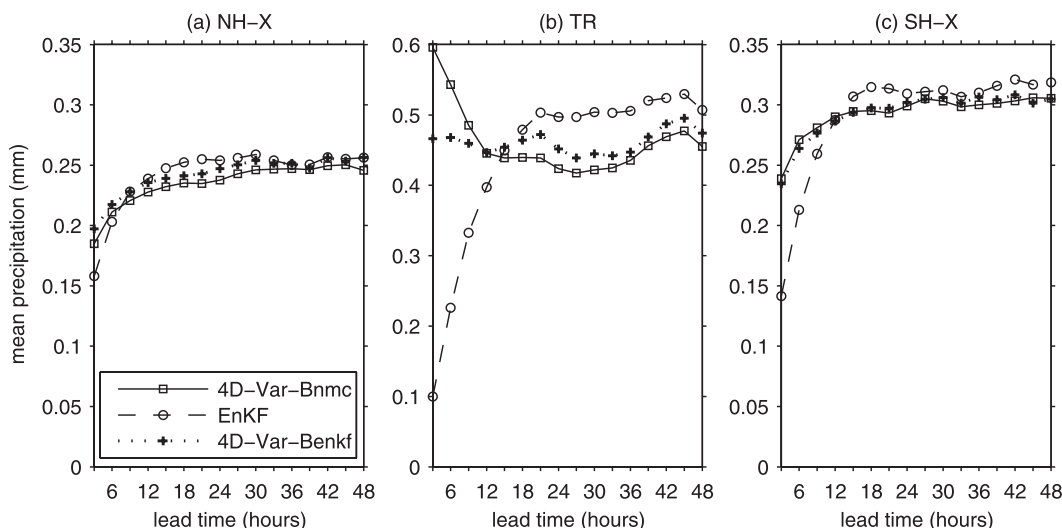


FIG. 8. Evolution of the forecasted 3-h accumulated precipitation averaged over all cases and over the (a) northern extratropics, (b) tropics, and (c) southern extratropics. Results shown are from the 4D-Var-Bnmc (solid), EnKF (dashed), and 4D-Var-Benkf (dotted) experiments.

$0.3 \text{ m s}^{-1}$ ), temperature (exceeding  $0.05 \text{ K}$ ), and geopotential height (exceeding  $0.1 \text{ dam}$ ). This is larger than the impact obtained in the context of 4D-Var. In the tropics, the impacts are very similar to those seen for the 4D-Var experiments. Finally in the southern extratropics, the large positive impacts obtained with the 4D-Var experiments are again seen with the 3D-FGAT experiments, though slightly smaller in amplitude for zonal wind and geopotential height around  $200 \text{ hPa}$  (exceeding  $0.5 \text{ m s}^{-1}$ ,  $0.3 \text{ K}$ , and  $0.6 \text{ dam}$ , for zonal wind, temperature, and geopotential height, respectively).

### c. En-4D-Var versus 3D-FGAT-Benkf and 4D-Var-Benkf

The next set of results compares the quality of forecasts from the En-4D-Var experiment relative to both the 3D-FGAT-Benkf and 4D-Var-Benkf experiments. All three variational data assimilation experiments employ flow-dependent background-error covariances obtained from the EnKF background ensembles. The three experiments primarily differ in how the error covariances are evolved in time through the 6-h assimilation time window, as discussed in section 6 of Part I. The ensemble covariances valid at the middle of the assimilation window are used for the 3D-FGAT-Benkf experiment with no time evolution. For the 4D-Var-Benkf experiment, the ensemble covariances valid at the beginning of the assimilation window are used and evolved implicitly throughout the assimilation window to obtain 4D error covariances using the tangent linear and adjoint versions of the forecast model. For the En-4D-Var experiment, the EnKF background ensemble members, each evolved with the nonlinear

forecast model and sampled at five time levels throughout the assimilation window, are used to obtain an estimate of the 4D error covariances. Thus, only 3D covariances are used for the 3D-FGAT-Benkf experiment and 4D covariances are used for both the En-4D-Var and 4D-Var-Benkf experiments, but with the temporal evolution being performed differently.

In Fig. 10 the impacts on the standard deviations with respect to radiosonde observations are shown for the En-4D-Var experiment relative to the 3D-FGAT-Benkf experiment in the same format as for several of the previous figures. Positive impacts are obtained for the northern extratropics, which is an improvement in the En-4D-Var forecasts relative to the 3D-FGAT-Benkf forecasts, for zonal wind ( $>0.3 \text{ m s}^{-1}$ ), temperature ( $>0.1 \text{ K}$ ), and geopotential height ( $>0.2 \text{ dam}$ ) beyond forecast days 3 or 4. In the tropics, the impacts are neutral for all three variables. A similar level of positive impact is obtained in the southern extratropics as in the northern extratropics. The statistical significance of these differences in the verification scores was also computed using the same bootstrap resampling procedure described earlier. The results (not shown) indicate that additional small, yet statistically significant positive impacts are obtained for forecast days 1 and 2 in both extratropical regions.

Next, the quality of the forecasts from the En-4D-Var experiment is again evaluated, but this time relative to the 4D-Var-Benkf forecasts, as shown in Fig. 11. In both extratropical regions negative impacts are generally obtained and especially so in the southern extratropics. A neutral impact is seen in the tropics. This means that the way the error covariances are temporally evolved in the

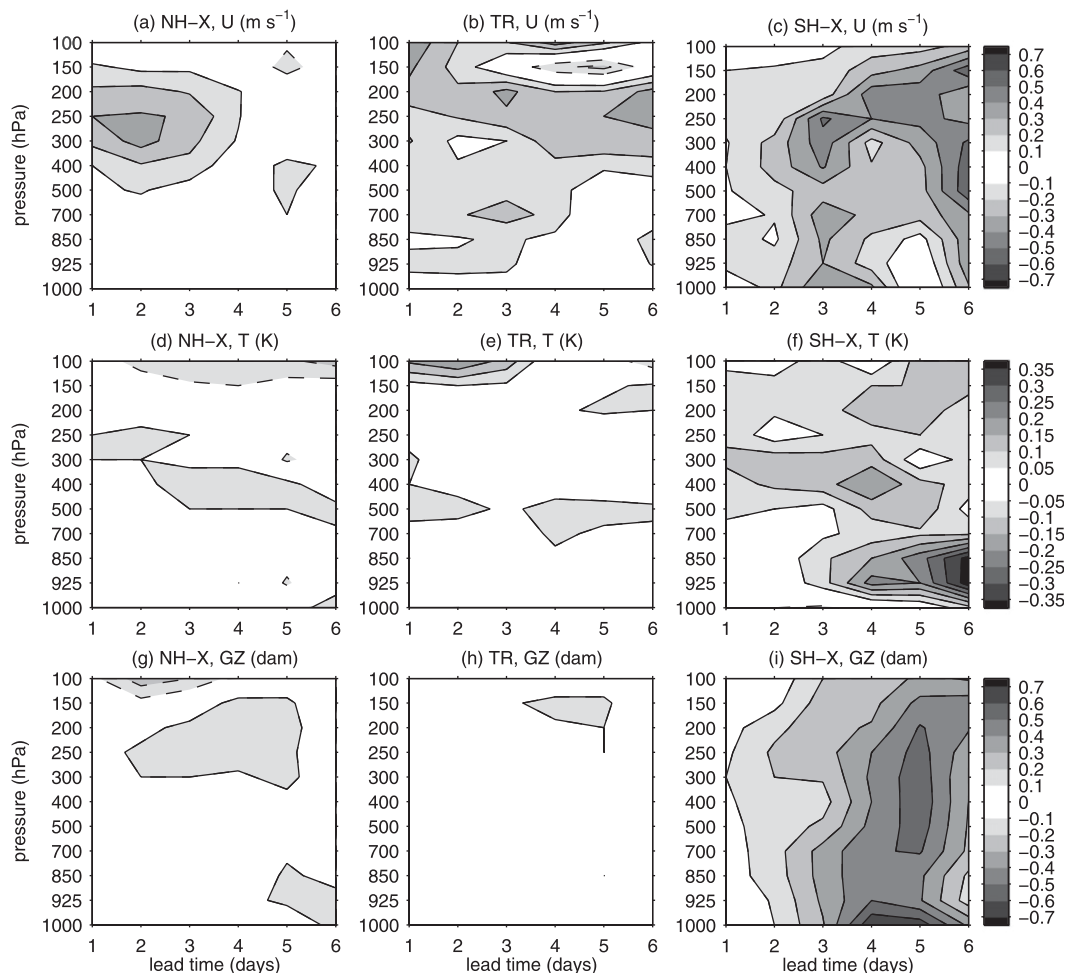


FIG. 9. As in Fig. 2, but showing the impact from the 3D-FGAT-Benkf experiment with respect to the 3D-FGAT-Bnmc experiment. Positive values (solid contours) correspond with a positive impact (i.e., a reduction in the standard deviation) for the 3D-FGAT-Benkf experiment relative to the 3D-FGAT-Bnmc experiment.

En-4D-Var experiment leads to worse forecasts than with the 4D-Var-Benkf experiment. These negative impacts are very small in the troposphere of the northern extratropics. In the southern extratropics, the negative impacts are larger for zonal wind ( $>0.3 \text{ m s}^{-1}$ ), temperature ( $>0.1 \text{ K}$ ), and geopotential height ( $>0.3 \text{ dam}$ ). Results from applying the bootstrap resampling procedure (not shown) indicate that significant negative impacts are obtained for temperature and geopotential height starting from forecast day 1 in both extratropical regions. In the southern extratropics, the negative impacts relative to the 4D-Var-Benkf experiment are similar in amplitude to the positive impacts relative to the 3D-FGAT-Benkf experiment. In other words, the quality of the forecasts from the En-4D-Var experiment is approximately halfway between the quality of the forecasts from the 3D-FGAT-Benkf and 4D-Var-Benkf experiments. However, for the northern extratropics,

the improvement in the En-4D-Var experiment relative to the 3D-FGAT-Benkf experiment is larger than the degradation relative to 4D-Var-Benkf. The temporal evolution of the background-error covariances seems to have much less impact in the tropics where all three experiments produce similar quality forecasts.

Verification results obtained by comparing forecasts from the three experiments just considered against 500-hPa geopotential height analyses are shown in Fig. 12 in terms of anomaly correlation. Up to forecast day 5 for the northern extratropics (Fig. 12a), the 4D-Var-Benkf experiment produces the best forecasts and the En-4D-Var experiment produces forecasts of slightly lower quality. For day 6 the forecasts from the En-4D-Var experiment are slightly better than those from the 4D-Var-Benkf experiment. Both experiments produce forecasts of higher quality than those from the 3D-FGAT-Benkf experiment for all forecast lead times in the northern



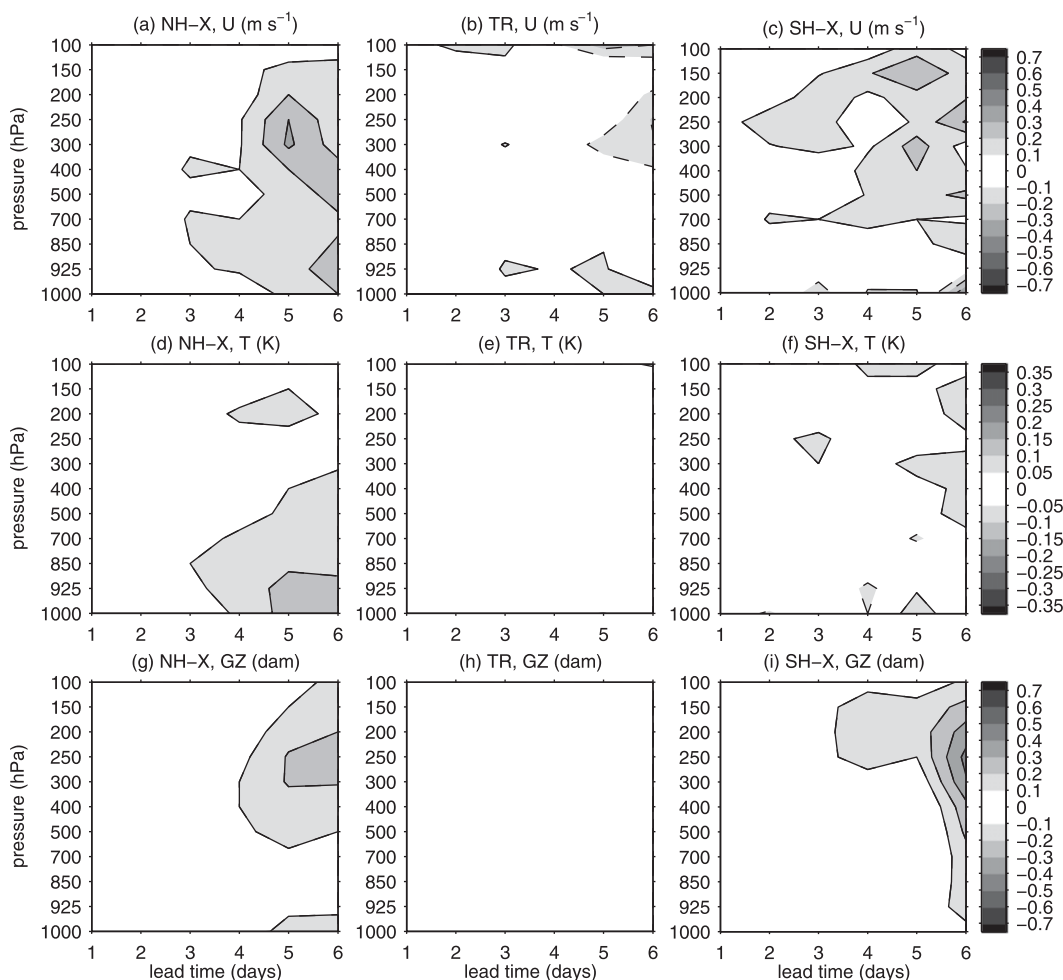


FIG. 10. As in Fig. 2, but showing the impact from the En-4D-Var experiment with respect to the 3D-FGAT-Benkf experiment. Positive values (solid contours) correspond with a positive impact (i.e., a reduction in the standard deviation) for the En-4D-Var experiment relative to the 3D-FGAT-Benkf experiment.

extratropics. These results are generally consistent with those shown in Figs. 10 and 11. In the southern extratropics (Fig. 12b), the 4D-Var-Benkf experiment again leads to the best forecasts, in this case for all forecast lead times. The forecasts from the En-4D-Var experiment are consistently better than those from the 3D-FGAT-Benkf experiment and worse than those from the 4D-Var-Benkf experiment. Again, this is generally consistent with the results shown in Figs. 10 and 11. The improvement of the 4D-Var-Benkf experiment relative to the 3D-FGAT-Benkf can be compared with a similar improvement observed by Laroche et al. (2007) when comparing 4D-Var and 3D-FGAT experiments in the context of the first operational implementation of 4D-Var at CMC.

#### d. En-4D-Var versus 4D-Var-Bnmk and EnKF

The final set of comparisons is chosen to demonstrate the quality of the forecasts from the En-4D-Var approach

relative to the experiments that most closely correspond with the currently operational versions of the 4D-Var and EnKF systems. As discussed earlier, the En-4D-Var approach incorporates the EnKF ensemble covariances within the variational data assimilation system in a very similar way as in the EnKF system itself.

Impacts on the verification scores with respect to radiosonde observations are shown in Fig. 13 for the En-4D-Var experiment relative to the 4D-Var-Bnmk experiment. In the northern extratropics the impacts are very small, possibly slightly positive for zonal wind, but slightly negative for temperature and geopotential height. The impacts in the tropics are positive for zonal wind,  $>0.3 \text{ m s}^{-1}$  around 200 hPa, and generally neutral for temperature and geopotential height. Large positive impacts are obtained in the southern extratropics. The pattern of the impacts are similar to the positive impacts seen for the 4D-Var-Benkf experiment relative to the 4D-Var-Bnmk

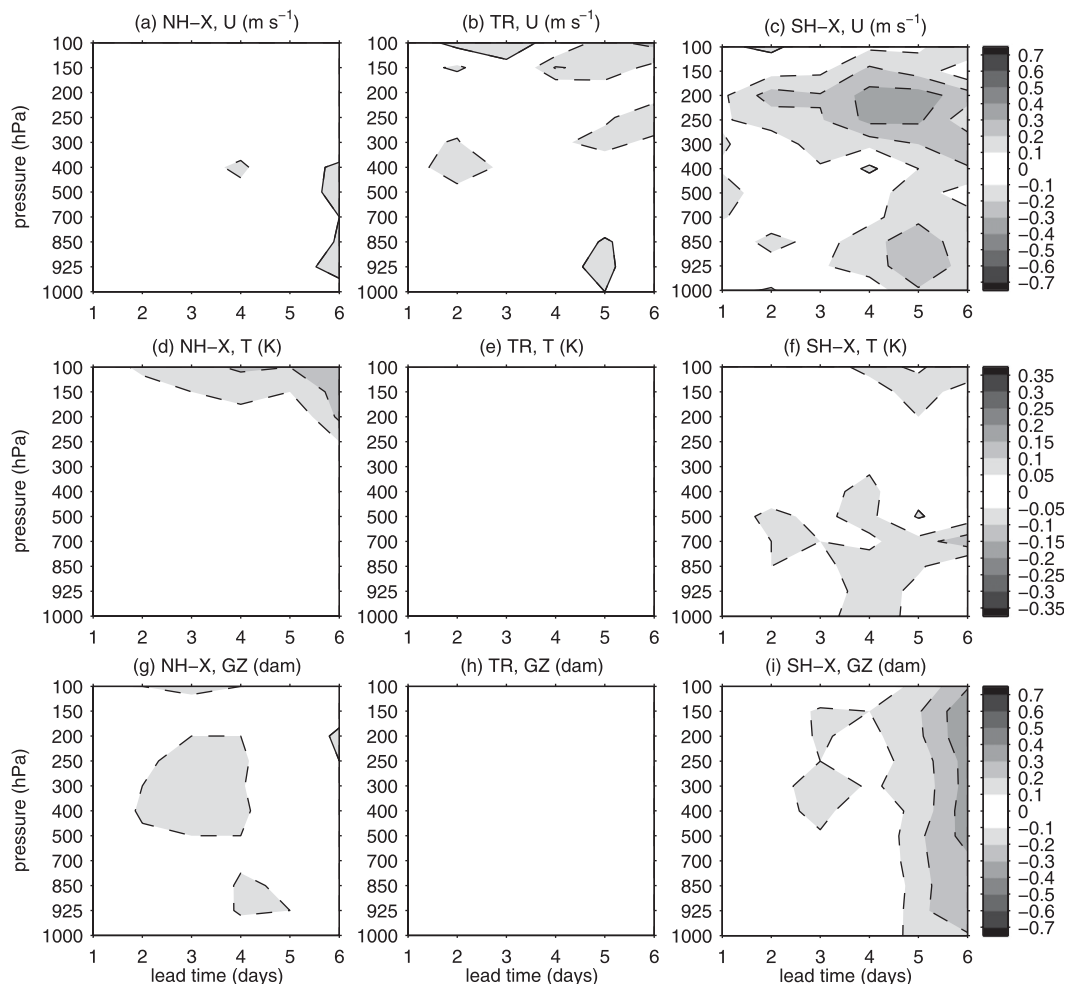


FIG. 11. As in Fig. 2, but showing the impact from the En-4D-Var experiment with respect to the 4D-Var-Benkf experiment. Positive values (solid contours) correspond with a positive impact (i.e., a reduction in the standard deviation) for the En-4D-Var experiment relative to the 4D-Var-Benkf experiment.

experiment in Fig. 4, but the amplitudes are smaller (greater than  $0.5 \text{ m s}^{-1}$ ,  $0.25 \text{ K}$ , and  $0.4 \text{ dam}$  for zonal wind, temperature, and geopotential height, respectively).

Impacts on the same verification scores are shown in Fig. 14 for the En-4D-Var experiment relative to the EnKF experiment. Even though this configuration of the variational data assimilation system uses the EnKF background-error covariances in a similar way to how they are used in the EnKF itself, several differences still remain between the two systems (as discussed in Part I), which lead to differences in the quality of the resulting medium-range forecasts. Small negative impacts occur in the northern extratropics starting at forecast days 4 or 5 for zonal wind ( $>0.1 \text{ m s}^{-1}$ ), temperature ( $>0.1 \text{ K}$ ), and geopotential height ( $>0.2 \text{ dam}$ ). However, results from applying the bootstrap resampling procedure described earlier suggest that a significant positive impact is obtained for the En-4D-Var experiment for the very

early portion of the forecasts (not shown). The impacts in the tropics are essentially neutral for the three variables and all forecast lead times. In the southern extratropics, small positive impacts are generally seen for zonal wind ( $>0.2 \text{ m s}^{-1}$ ), temperature ( $>0.1 \text{ K}$ ), and geopotential height ( $>0.3 \text{ dam}$ ) for all lead times with the largest impacts occurring for day 6.

Figure 15 shows the standard deviations of the 24-h forecasts from the 4D-Var-Bnmc, EnKF, and En-4D-Var experiments relative to radiosonde observations of temperature over the entire vertical domain of the model. These results are shown to illustrate the impact of the difference in covariance localization between the EnKF and variational data assimilation systems when using the EnKF covariances. As demonstrated earlier in Part I, this impact is most pronounced for the assimilation of radiance observations that have the highest sensitivity to temperature near the top of the model domain.

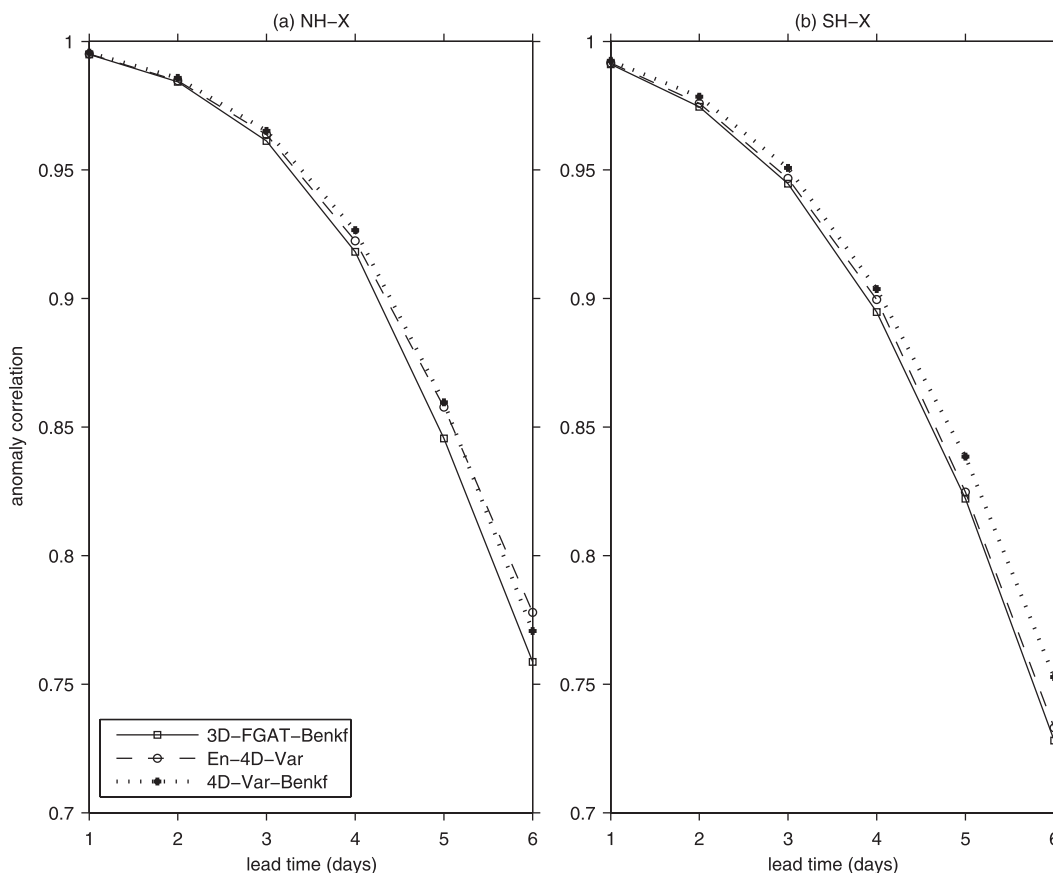


FIG. 12. As in Fig. 6, but for the 3D-FGAT-Benkf (solid), En-4D-Var (dashed), and 4D-Var-Benkf (dotted) experiments.

The verification scores are quite similar for the EnKF and 4D-Var-Bnmc experiments in all three regions: northern extratropics (Fig. 15a), tropics (Fig. 15b), and southern extratropics (Fig. 15c). However, the scores for the En-4D-Var experiment are significantly degraded above 50 hPa relative to the EnKF and 4D-Var-Bnmc experiments for both the tropics and southern extratropics. In the same regions and at the same vertical levels, a similar amount of degradation is also obtained for the 4D-Var-Benkf and 3D-FGAT-Benkf experiments and for other forecast lead times (not shown). It is therefore clear that the artificially inflated values of observation-error variance for AMSU-A channels 9 and 10 as used in the operational 4D-Var are necessary in all configurations of the variational data assimilation system because of the difference in the way spatial covariance localization is applied in the two systems. Despite the large degradation obtained for the upper levels when using the ensemble background-error covariances in the variational data assimilation system, this does not appear to negatively impact the forecasts below 50 hPa.

#### 4. Conclusions

Five configurations of a variational data assimilation system and the EnKF were compared in a near-operational context. The data assimilation systems considered differ in several important ways, including the use of a deterministic versus an ensemble data assimilation approach, the use of a variational versus a sequential solution algorithm, the approach for temporally evolving the error covariances within the assimilation window, the use of static versus flow-dependent background-error covariances, and the approach for applying spatial covariance localization. By comparing the results from various combinations of the experiments, it was demonstrated how these differences can have a significant impact on the resulting forecast quality.

The 4D-Var configuration similar to the system currently operational at CMC (4D-Var-Bnmc) and the EnKF produce deterministic forecasts that are generally of similar quality. For the early portion of the forecasts, the 4D-Var-Bnmc experiment leads to slightly better forecasts than

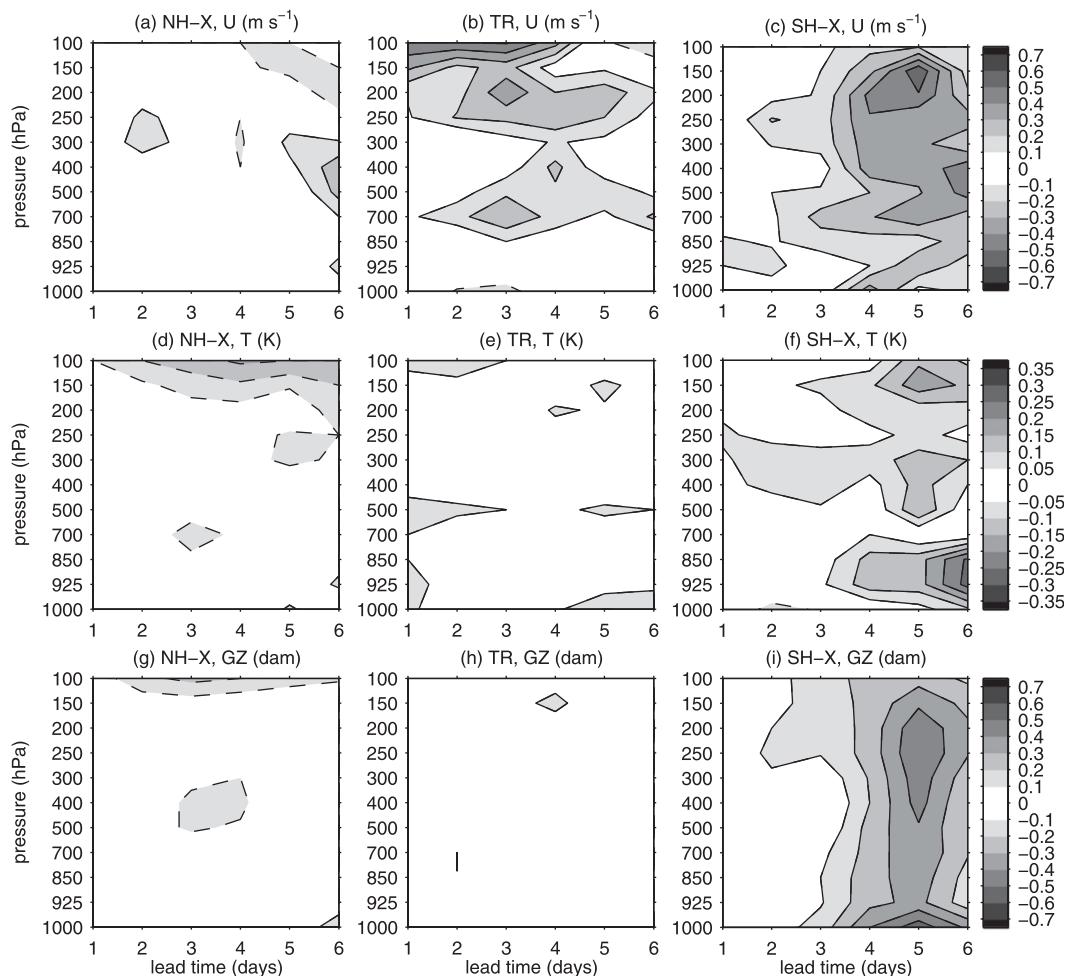


FIG. 13. As in Fig. 2, but showing the impact from the En-4D-Var experiment with respect to the 4D-Var-Bnmc experiment. Positive values (solid contours) correspond with a positive impact (i.e., a reduction in the standard deviation) for the En-4D-Var experiment relative to the 4D-Var-Bnmc experiment.

the EnKF ensemble mean analysis in the extratropical regions and the opposite is true in the medium range up to day 6. The degradation in the short range may result from the coarser resolution of the EnKF analyses that are used to initialize the high-resolution forecast model. In addition, any bias in the radiosonde observations used for verification may give an apparent disadvantage to the EnKF since its analyses fit these observations less closely. In the tropics, use of the EnKF leads to consistently better forecasts, especially for zonal wind.

In the context of both 4D-Var and 3D-FGAT experiments, the use of flow-dependent background-error covariances estimated from the EnKF ensembles instead of covariances similar to those used operationally leads to significant improvements in the resulting forecast quality. This positive impact is largest in the southern extratropics and the tropics. In the context of 4D-Var, a gain of approximately 9 h is obtained for the 5-day forecasts in

the southern extratropics (as measured by the anomaly correlation with respect to 500-hPa geopotential height analyses). It should be noted that no effort was made to tune any parameters related to the use of the EnKF covariances in the variational data assimilation system (e.g., covariance localization parameters or the overall variance) since this would have made the intercomparison with the EnKF more difficult.

When using the EnKF error covariances, the 4D-Var experiment (4D-Var-Benkf) also produces significantly better forecasts in the southern extratropics than when initializing the deterministic forecasts with the EnKF ensemble mean analysis. Though these two experiments both use the same ensemble error covariances at the beginning of the assimilation window, the two approaches differ in many ways, including how the covariances are evolved through time and how covariance localization is applied. Comparisons with the En-4D-Var approach help to

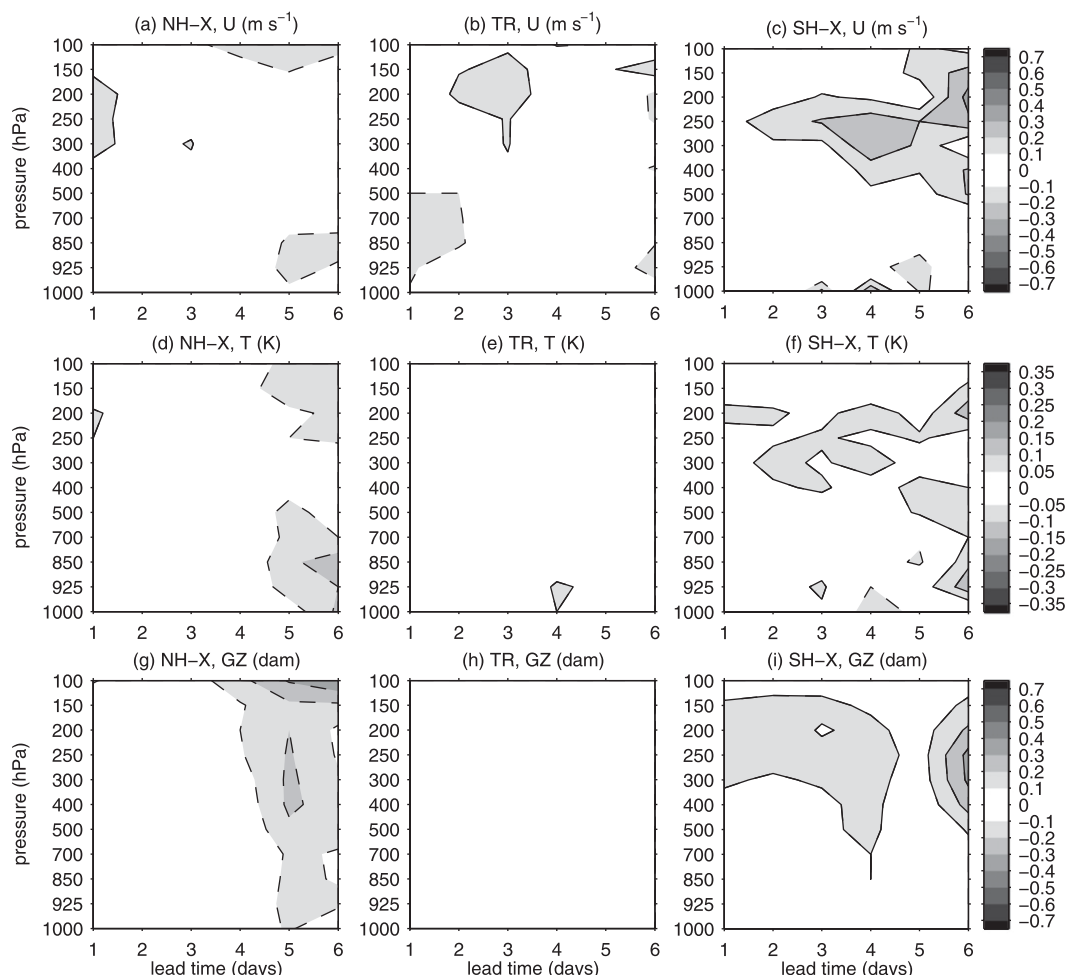


FIG. 14. As in Fig. 13, but showing the impact from the En-4D-Var experiment with respect to the EnKF experiment.

evaluate the importance of the difference in covariance evolution, since the En-4D-Var approach evolves the covariances within the variational data assimilation system in the same way as the EnKF, but is otherwise similar to the 4D-Var-Benkf experiment. In the southern extratropics, the En-4D-Var approach produces forecasts that are slightly worse than the 4D-Var-Benkf experiment, but slightly better than the EnKF. These results suggest that the approach for evolving the error covariances in 4D-Var only partially explains the improvement in the southern extratropics for the 4D-Var-Benkf experiment relative to the EnKF experiment. In both the northern extratropics and the tropics, the 4D-Var-Benkf, EnKF, and En-4D-Var experiments all produce forecasts that are more similar in quality. The similarity of the results from the 3D-FGAT-Benkf, En-4D-Var, and 4D-Var-Benkf experiments in the tropics suggests that the forecast quality in this region is insensitive to the use of any covariance time evolution.

The forecast degradation seen in the En-4D-Var experiment relative to the 4D-Var-Benkf experiment

(especially in the southern extratropics) may result from several factors related to the treatment of the background-error covariances. First, as described in Part I, the covariances are temporally interpolated to the time of the observations differently in the two experiments. In the En-4D-Var (and EnKF) experiment, they are linearly interpolated to the time of the observations from covariances separated by 90 min. In the 4D-Var-Benkf experiment, the covariances are implicitly obtained every 45 min, but then the covariances closest to the observation time are used without interpolation. Also, the 4D-Var-Benkf experiment incorporates nonlinearity in the covariance evolution by linearizing the forecast model a second time, during the second iteration of the outer loop, about a model trajectory that provides a better fit to the observations than the background trajectory. In contrast, in the En-4D-Var approach (and the EnKF) each background ensemble member is evolved with the nonlinear model with no equivalent influence from the observations. Another difference is that,



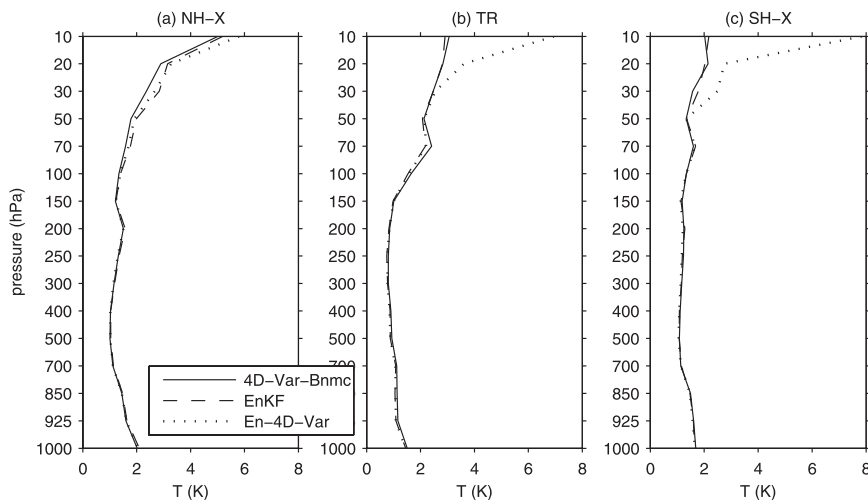


FIG. 15. Standard deviation of the 24-h forecasts with respect to radiosonde observations of temperature over the entire vertical domain of the model in the (a) northern extratropics, (b) tropics, and (c) southern extratropics for the 4D-Var-Bnmc (solid), EnKF (dashed), and En-4D-Var (dotted) experiments.

because spatial localization is performed before temporal evolution, the covariances are evolved (implicitly) in the 4D-Var-Bnmc experiment within a much higher dimensional space than for the En-4D-Var experiment for which spatial localization is applied after temporal evolution. Finally, in the 4D-Var-Bnmc experiment, the spatial localization is only applied to the 3D covariances at the beginning of the assimilation window. Consequently, it is not necessary to apply spatial localization to the between-time cross covariances, for which the most appropriate approach is not obvious (Bishop and Hodyss 2009). This is in contrast to the En-4D-Var (and EnKF) approach that requires spatial localization to be applied to these cross covariances.

Differences in the relative impacts in the northern and southern extratropical regions for several of the comparisons considered in this study may be due to differences in the observational network. Because of the lower volume of radiosonde, wind profiler, and aircraft observations in the southern extratropics, there is a relatively higher reliance on satellite radiance observations in this region as compared to the northern extratropics. Since these radiances are only indirectly related to the analysis variables (temperature and humidity), the background-error covariances play a more significant role in determining the resulting analysis increment than when assimilating wind, temperature, and humidity observations. Similarly, the reduced number of wind observations in the southern extratropics results in the wind increments being largely determined by the background-error cross covariances between the mass and wind fields. Consequently, this may explain why the use of more accurate *flow-dependent*

background-error covariances from the EnKF results in a larger impact in the southern extratropics. However, the meteorology is also different in winter (Northern Hemisphere) and summer (Southern Hemisphere) seasons, which may also affect the impact of various differences between the approaches. An extension of this work, already begun, includes an evaluation of the same data assimilation approaches, but for a boreal summer period to help explain these differences. Preliminary results show that the use of EnKF covariances within 4D-Var again leads to improvements mostly in the southern extratropics and tropics.

For the variational data assimilation experiments that use background-error covariances estimated from the EnKF ensembles, no relationship is imposed between the background state and the EnKF background ensemble. Consequently, the possibility exists that the EnKF ensemble mean may differ significantly from the deterministic background state, thus causing the background-error covariances to possibly be less useful when assimilating observations to correct this background state. An approach evaluated in an idealized context by Zhang et al. (2009) addresses this by replacing the EnKF ensemble mean analysis with the deterministic analysis from the variational data assimilation system.

An approach to produce a high-resolution deterministic analysis with the EnKF system was also developed in the context of our study. The incremental approach, similar to that used in the variational data assimilation system, was used to produce a single high-resolution analysis from a high-resolution background state and a lower-resolution

analysis increment. The analysis increment was computed at the same resolution as the ensemble members using unperturbed observations and the background-error covariances estimated from the entire 96-member background ensemble. This approach was expected to reduce the adjustment of the low-resolution analysis fields to the higher-resolution model grid and surface topography during the early stages of the forecasts that likely occurs in the EnKF experiment considered in this study. However, for unidentified reasons, it was found that the quality of the resulting high-resolution analysis was not better than that of the EnKF ensemble mean analysis.

Additional experiments may also be conducted with the variational data assimilation system to evaluate various techniques for modeling the ensemble-based error covariances. For instance, to reduce the imbalance from applying covariance localization to wind components (Mitchell et al. 2002), it has been shown by Kepert (2009) that application of covariance localization instead to the streamfunction and velocity potential, results in less degradation of the geostrophic balance present in the ensemble members. Applying this variable transformation within the variational data assimilation system is likely easier than within the EnKF system. The possibility of applying additional filtering to the ensemble-based covariances may also be explored, such as the technique of localizing the covariances in spectral space, as evaluated in an idealized context by Buehner and Charron (2007).

**Acknowledgments.** The authors thank Luc Fillion and other members of the “WWRP/THORPEX Workshop on 4D-Var and Ensemble Kalman Filter Intercomparisons” Program Committee for providing the initial impetus for this work and Gilbert Brunet and Godelieve Deblonde for their continued support of this project. The authors also thank Craig Bishop, Eugenia Kalnay, and Jeff Whitaker for their helpful official reviews.

## REFERENCES

- Bélair, S., M. Roch, A.-M. Leduc, P. A. Vaillancourt, S. Laroche, and J. Mailhot, 2009: Medium-range quantitative precipitation forecasts from Canada's new 33-km deterministic global operational system. *Wea. Forecasting*, **24**, 690–708.
- Bishop, C. H., and D. Hodyss, 2009: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus*, **61A**, 97–111.
- Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043.
- , and M. Charron, 2007: Spectral and spatial localization of background-error correlations for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133**, 615–630.
- , P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.*, **138**, 1550–1566.
- Candille, G., C. Côté, P. L. Houtekamer, and G. Pellerin, 2007: Verification of an ensemble prediction system against observations. *Mon. Wea. Rev.*, **135**, 2688–2699.
- Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998: The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.*, **126**, 1373–1395.
- Huffman, G. J., R. F. Adler, M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeorol.*, **2**, 36–50.
- Kepert, J. D., 2009: Covariance localisation and balance in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 1157–1176.
- Laroche, S., P. Gauthier, M. Tanguay, S. Pellerin, and J. Morneau, 2007: Impact of the different components of 4DVAR on the global forecast system of the Meteorological Service of Canada. *Mon. Wea. Rev.*, **135**, 2355–2364.
- Liu, C., Q. Xiao, and B. Wang, 2008: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.*, **136**, 3363–3373.
- Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791–2808.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Rochon, Y. J., L. Garand, D. S. Turner, and S. Polavarapu, 2007: Jacobian mapping between vertical coordinate systems in data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133**, 1547–1558.
- Saunders, R., P. Brunel, S. English, P. Bauer, U. O'Keeffe, P. Francis, and P. Rayner, 2006: RTTOV-8—Science and validation report. Met. Office Publ. NWPSAF-MO-TV-007, 46 pp.
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575.
- Zhang, F., M. Zhang, and J. A. Hansen, 2009: Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, **26**, 1–8.