Implementation of Deterministic Weather Forecasting Systems Based on Ensemble–Variational Data Assimilation at Environment Canada. Part I: The Global System

MARK BUEHNER,* RON MCTAGGART-COWAN,+ ALAIN BEAULNE,# CÉCILIEN CHARETTE,* LOUIS GARAND,* SYLVAIN HEILLIETTE,* ERVIG LAPALME,* STÉPHANE LAROCHE,* STEPHEN R. MACPHERSON,* JOSÉE MORNEAU,# AND AYRTON ZADRA+

* *Data Assimilation and Satellite Meteorology Research Section, Environment Canada, Dorval, Quebec, Canada*
+ *Numerical Weather Prediction Research Section, Environment Canada, Dorval, Quebec, Canada*
# *Data Assimilation and Quality Control Development Section, Environment Canada, Dorval, Quebec, Canada*

(Manuscript received 6 October 2014, in final form 12 February 2015)

ABSTRACT

A major set of changes was made to the Environment Canada global deterministic prediction system during the fall of 2014, including the replacement of four-dimensional variational data assimilation (4DVar) by four-dimensional ensemble–variational data assimilation (4DEnVar). The new system provides improved forecast accuracy relative to the previous system, based on results from two sets of two-month data assimilation and forecast experiments. The improvements are largest at shorter lead times, but significant improvements are maintained in the 120-h forecasts for most regions and vertical levels. The improvements result from the combined impact of numerous changes, in addition to the use of 4DEnVar. These include an improved treatment of radiosonde and aircraft observations, an improved radiance bias correction procedure, the assimilation of ground-based GPS data, a doubling of the number of assimilated channels from hyperspectral infrared sounders, and an improved approach for initializing model forecasts. Because of the replacement of 4DVar with 4DEnVar, the new system is also more computationally efficient and easier to parallelize, facilitating a doubling of the analysis increment horizontal resolution. Replacement of a full-field digital filter with the 4D incremental analysis update approach, and the recycling of several key variables that are not directly analyzed significantly reduced the model spinup during both the data assimilation cycle and in medium-range forecasts.

1. Introduction

For more than a decade, numerous operational numerical weather prediction (NWP) centers have had significant improvements in analysis and forecast accuracy by adopting the four-dimensional variational data assimilation (4DVar) approach (e.g., Rabier et al. 2000; Rabier 2005; Rawlins et al. 2007; Gauthier et al. 2007). In parallel with these developments, a significant research and development effort has led to the successful application of the ensemble Kalman filter (EnKF) to NWP for ensemble prediction (e.g., Houtekamer et al. 2014). Combinations of variational data assimilation with the EnKF have recently been used to obtain additional improvements in deterministic analysis and forecast accuracy. For example, the Met Office in the United Kingdom recently modified their operational global 4DVar system to include flow-dependent background-error covariances estimated from EnKF ensembles of background states (Clayton et al. 2013). Similarly, the National Centers for Environmental Prediction in the United States also modified their operational global three-dimensional variational data assimilation (3DVar) system to include ensemble background-error covariances (Wang et al. 2013; Kleist and Ide 2015). These approaches result in significant improvements when compared with the same 3DVar or 4DVar system that only employs the climatological and highly parameterized models of background-error covariance that are typically used (e.g., Derber and Bouttier 1999; Gauthier et al. 1999).

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/MWR-D-14-00354.1

Unlike the case of only adding ensemble covariances to an existing variational system, the data assimilation approach used at Environment Canada (EC) for global NWP was recently changed in a more substantial way. The previous system, based on 4DVar with climatological background-error covariances, was replaced with the four-dimensional ensemble–variational data assimilation [4DEnVar; see Lorenc (2013) for the recommended nomenclature for such methods]. Unlike 4DVar, the new 4D assimilation approach does not require the tangent linear and adjoint version of the forecast model, but instead relies on 4D ensemble covariances to explicitly estimate the full temporal–spatial background-error covariances over the assimilation time window. These covariances are estimated from the ensemble of background states from the existing operational global EnKF at EC [a modified version of the system described by Houtekamer et al. (2014)]. By removing the use of the tangent linear and adjoint of the forecast model from the data assimilation procedure, the 4DEnVar approach is more computationally efficient and easier to parallelize than 4DVar. Moreover, results from a previous study by Buehner et al. (2013) demonstrate sufficiently improved analysis and forecast accuracy to justify replacing 4DVar with 4DEnVar. This change was combined with numerous other modifications, including an improved satellite radiance bias correction procedure, an improved treatment of radiosonde and aircraft observations, a new procedure for initializing the forecast model, and the assimilation of a large number of additional observations. The goal of this study is to describe and evaluate the impact of these recent changes to the operational Global Deterministic Prediction System (GDPS) at EC. A nearly identical data assimilation system is also now used for the operational Regional Deterministic Prediction System (RDPS) as detailed in the companion paper by Caron et al. (2015).

The next section provides a brief description of all changes, including a summary of the individual impacts of each change on analysis and forecast accuracy in cases where they were already evaluated in previous studies. In section 3, results are given from numerical experiments designed to evaluate the impact of only the new satellite radiance bias correction procedure. Similarly, section 4 provides results from numerical experiments designed to evaluate the impact of only the changes to the forecast model and its initialization. Results showing the combined impact of all modifications in the new global 4DEnVar-based system relative to the previous 4DVar-based system are the focus of section 5. Finally, some conclusions are given in section 6.

2. Description of the system modifications

A 4DVar-based system has been used operationally at EC to provide the initial conditions for global forecasts since 15 March 2005 (Gauthier et al. 2007). During the years that followed, significant improvements were made to both the forecast model and the data assimilation system. Some significant changes include the following: extending the model and assimilation domain in the vertical to fully include the stratosphere starting on 22 June 2009 (Charron et al. 2012); improving model forecasts of tropical cyclones starting on 12 July 2011 (Zadra et al. 2014a); adding new observations [Infrared Atmospheric Sounding Interferometer (IASI), SSMIS, and a reduced thinning of all radiances] starting on 16 November 2011; reducing the model grid spacing from 0.45° × 0.3° to 0.35° × 0.23°, reducing the grid spacing for the analysis increment from 1.5° × 1.5° to 0.9° × 0.9°, and introducing vertical staggering starting on 13 February 2013 (Zadra et al. 2014b; Girard et al. 2014). The set of changes relative to this 4DVar-based system that were recently implemented at EC are summarized in Table 1 and described in more detail in the remainder of this section.

a. *Replacement of 4DVar with 4DEnVar*

The current 4DVar approach was replaced by 4DEnVar using a configuration very similar to what was described and evaluated by Buehner et al. (2013). In that study, the impact of only changing the data assimilation approach from 4DVar to 4DEnVar was evaluated. Compared with the experimental configuration of 4DEnVar tested in that study, the current 4DEnVar-based system uses a higher horizontal resolution for the analysis increment (grid spacing in latitude and longitude of 0.45° × 0.45° instead of 0.6° × 0.6°) and EnKF covariances estimated from a slightly larger ensemble size (256 instead of 192 members). These differences are due to changes made in the operational EnKF of the Global Ensemble Prediction System (GEPS) tested in combination with the new GDPS [the previous EnKF configuration is described by Houtekamer et al. (2014)]. In addition, whereas the ensemble members used by Buehner et al. (2013) were obtained from a previously run EnKF experiment, the experiments with the complete new system are run in coupled mode. Specifically, the EnKF assimilates observations with quality control and bias correction performed by the GDPS and the GDPS uses the ensemble of background states produced by the EnKF experiment to specify the background-error covariances for 4DEnVar. This two-way dependence requires that the GDPS and GEPS be run together with the appropriate flow of information

TABLE 1. Summary of system components that were modified in the new 4DEnVar-based system relative to the 4DVar-based system that became operational on 13 Feb 2013.

Component modified	4DVar-based system	4DEnVar-based system
Assimilation approach	4DVar, two outer-loop iterations (with 35 and 30 inner loop iterations, respectively), increments computed every 18 min	4DEnVar, no outer-loop (70 inner loop iterations), increments computed every hour
Analysis increment horizontal grid	0.9° × 0.9° Gaussian grid	0.45° × 0.45° Gaussian grid
Background-error covariances	Lagged-forecast difference method covariances are only specified at the beginning of the 6-h assimilation window.	From the surface to ~40 hPa, the average of lagged-forecast difference method covariances and 4D ensemble covariances from 256 ensemble members every hour over the 6-h assimilation window. Above ~10 hPa, only the 3D lagged-forecast difference method covariances.
Instruments with assimilated radiance data (No. of channels)	AMSU-A (11), AMSU-B (4), MHS (4), SSMIS (7), geostationary imagers (1), AIRS (87), and IASI (62)	AMSU-A (11), AMSU-B (4), MHS (4), SSMIS (7), geostationary imagers (1), AIRS (142), and IASI (142)
Other satellite data	GPS-RO refractivity, atmospheric motion vectors (AMVs), and scatterometer winds	GPS-RO refractivity, AMVs, scatterometer winds, and ZTD from GB-GPS over North America
Treatment of radiosonde and aircraft data	Radiosonde uses launch time and station location for entire profile and aircraft has no temperature bias correction.	Radiosonde uses appropriate measured or computed time and horizontal position for each pressure level, aircraft has static temperature bias correction, and both radiosonde and aircraft have increased vertical resolution.
Version of RTTOV	8.7	10.2
Satellite radiance bias correction	Coef are computed from obs-minus-background, based on the last 7 days, 4 times per day, except static for AMSU-A channels 11–14.	Coef are computed from obs-minus-analysis using a separate 3DVar analysis that does not include radiances, based on the last 7 days, 2 times per day, except static for AMSU-A channels 13–14.
Model initialization scheme	DF is applied to full model fields during the first 6 h of model integration (i.e., from $T + 0$ to $T + 6$ h for $T = 0000$, 0600, 1200, or 1800 UTC).	4DIAU is applied over 6-h assimilation window (i.e., from $T - 3$ to $T + 3$ h for $T = 0000$, 0600, 1200 or 1800 UTC).
Recycled physics variables	None (all physics variables are initialized using values diagnosed from the analyzed state)	Total condensate, turbulent kinetic energy, turbulence regime, mixing length, friction velocity, and PBL height
Sea ice concentration analysis	Based on averaging gridded observations with 35-km grid spacing, one analysis per day, ice concentration from one satellite and manual analyses from Canadian Ice Service	Based on a 3DVar system with 10-km grid spacing, 4 analyses per day, ice concentration from four satellites and manual analyses from Canadian Ice Service

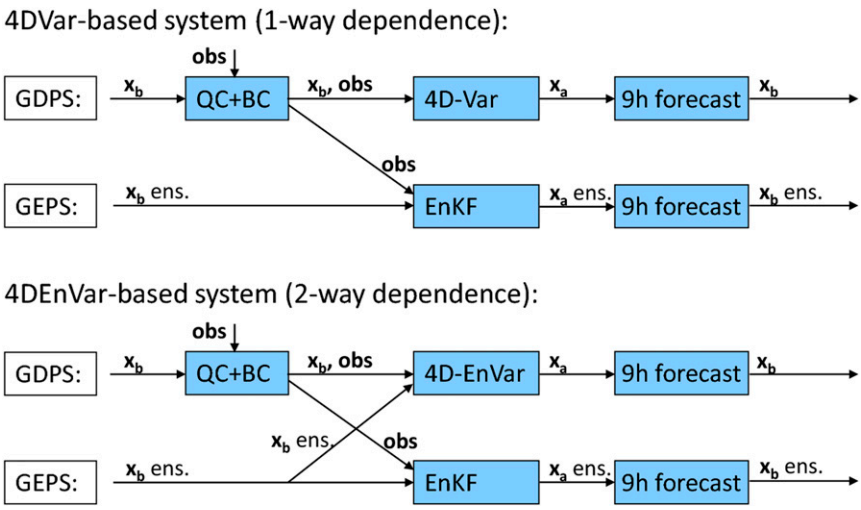

FIG. 1. Dependencies between the deterministic and ensemble prediction systems (GDPS and GEPS) in the (top) previous 4DVar-based system and (bottom) new 4DEnVar-based system. Note that QC refers to quality control and BC to bias correction, both applied to the observations before assimilation in both systems. The vectors $\mathbf{x}_b$ and $\mathbf{x}_a$ refer to the background and analysis states, respectively, and ens. refers to the EnKF ensemble of either background or analysis states.

between the systems at each analysis time, as shown in Fig. 1. These relatively small differences in the 4DEnVar configuration likely have only a minor effect on the measured impact of replacing 4DVar with 4DEnVar and, therefore, the results from Buehner et al. (2013) can be helpful in interpreting how the impact of the complete set of changes is related to just changing the data assimilation approach.

Based on six-week data assimilation experiments for both winter and summer periods, Buehner et al. (2013) demonstrated that forecasts from 4DEnVar analyses have either similar or better scores in the tropical troposphere and the winter extratropics than the forecasts from 4DVar analyses. In contrast, the medium-range forecast scores for the summer extratropics with 4DEnVar are either similar to or worse than those of 4DVar in the troposphere. The seasonal differences in extratropical medium-range forecast quality are largest in the southern extratropics, where the forecasts from 4DEnVar are significantly worse during February–March and better during July–August than those from 4DVar. Relative to radiosonde observations, the 6-h forecasts from the 4DEnVar experiments are significantly better than those from 4DVar for both periods and in all regions.

b. *Improved treatment of radiosonde and aircraft observations*

Many changes were made to the processing and assimilation of radiosonde and aircraft data. In the 4DVar-based system, radiosonde profiles are interpolated to 27 prescribed pressure levels from 1050 to 10 hPa. Aircraft data are thinned by selecting one set of observations (i.e., wind, temperature, and humidity if available) every 15 min within latitude–longitude boxes of $1° \times 1°$ with 50-hPa depth in the troposphere and 25-hPa depth above 300 hPa. In the 4DEnVar-based system, the same approach is used for the vertical selection of radiosonde and aircraft data. One set of observations per model layer is selected, which increases the volume of assimilated data by roughly 50% for radiosonde and 35% for aircraft. A static bias correction based on the reported pressure level is applied to all temperature observations from aircraft reports, without consideration of aircraft type or phase of flight. This adjustment (based on a one-month data assimilation experiment in which aircraft data were not assimilated) reduces the temperature by 0.1 K between 700 and 500 hPa, 0.2 K between 500 and 400 hPa, 0.3 K between 400 and 300 hPa, and 0.5 K between 300 and 100 hPa. The observation-error statistics have also been revised for both radiosonde and aircraft data, following the approach developed by Desroziers et al. (2005) in which the covariance between observation minus background and observation minus analysis provides, under specific assumptions, an estimate of observation-error variance.

Another important improvement made to radiosonde data processing in the 4DEnVar-based system is the recovery of the horizontal position and time of observations following the method proposed by Laroche and

Sarrazin (2013). In all previous data assimilation systems used operationally at EC, including the 4DVar-based system, the upper-air station location and launch time are assigned to all radiosonde data. This is because the actual horizontal position and time of observations are not provided in the radiosonde reports in alphanumeric codes available through the Global Telecommunication System (GTS). Laroche and Sarrazin (2013) estimate the balloon trajectory and the acquisition time from the observed wind and a representative elapsed ascending time profile. In the future, direct metadata will be reported in Binary Universal Form for Representation of Meteorological Data (BUFR) code that will progressively replace the current alphanumeric codes for transmitting radiosonde data. Consequently, the modifications made here to the radiosonde data processing will facilitate the ongoing transition from alphanumeric codes to BUFR code.

The impact of these changes to the assimilation of radiosonde data on forecast accuracy is positive and statistically significant at short range in the upper troposphere and lower stratosphere where the balloon drift can be large (as much as 200 km in the Northern Hemisphere winter season). The impact of the new static temperature bias correction from aircraft reports, not shown here, is significant over North America and Europe where a warm bias between 300 and 100 hPa can be clearly seen when assimilating the uncorrected observations. This bias correction also yields a small improvement in short-range temperature forecast scores within that layer.

c. *Assimilation of ground-based GPS data*

Ground-based GPS (GB-GPS) zenith tropospheric delay (ZTD) observations are introduced as a new observation type in the 4DEnVar-based system. Macpherson et al. (2008, hereafter MAC2008) describe how GB-GPS ZTD observations are obtained from ground-based GPS receivers and how they provide accurate all-weather measurements of integrated water vapor when combined with measurements of surface pressure and temperature. Assimilation of ZTD data in an analysis system supplies valuable information on water vapor in the lower to midtroposphere at locally high spatial resolution and frequency compared to radiosonde observations.

The assimilated GB-GPS data come from the NOAA Global Systems Division GPS network, covering North America and Hawaii. Observations are available every 30 min and include ZTD plus collocated (or nearby) surface observations of pressure, temperature, and relative humidity. The impact of assimilating these observations was tested in a previous version of the RDPS (based on 3DVar) as described in MAC2008. Several important differences exist between that initial implementation of GB-GPS assimilation and the strategy adopted for the new 4DEnVar-based system.

The ZTD observation operator has been updated and optimized. Profiles of ZTD are first computed on model levels starting at the model top using horizontally interpolated values of pressure, temperature, and specific humidity. The value of ZTD at the observation height is then obtained through vertical interpolation of the model ZTD (or extrapolation if the GPS receiver antenna is located below the model's surface). The new operator gives similar results to those obtained with the operator described in MAC2008 but is simpler, more efficient, and easier to maintain.

Only ZTD and collocated surface pressure data are assimilated in the 4DEnVar-based GDPS. The additional surface pressure data are not assimilated at sites close to existing synoptic surface stations to avoid potential data duplication. In MAC2008, 9 of the 12 total 30-min GB-GPS observations per site in the 6-h assimilation window are assimilated, while in the new system the GB-GPS observations are assimilated every 2 h, giving only 3 of the 12 observations per site per 6-h window. As the observations are more separated in time, it was felt that there was no longer a need to account for the temporal correlations in observation error as done in MAC2008. Also, the ZTD data are not bias corrected as in MAC2008, but ZTD data are rejected for the few sites with significant biases as determined from monitoring of differences between the observations and the background state. Growth in the NOAA network since 2004 has increased the number of sites for which data are assimilated (after 50-km spatial thinning is applied) from 200 in the original study to approximately 300 here.

The impact of GB-GPS data assimilation in the new 4DEnVar-based GDPS was evaluated over the two-month periods of February–March and July–August 2011, with results qualitatively similar to those presented in MAC2008. As expected, impacts are most obvious over North America and downstream over the Atlantic Ocean. The positive impact of the GB-GPS data is most obvious when verifying the precipitable water (PW) forecasts against PW retrieved from GB-GPS observations (Fig. 6 in MAC2008). Forecast error measured against retrieved PW is reduced in terms of both standard deviation and bias. In particular, the GB-GPS data help reduce a systematic moist bias in 1200 UTC analyses and the resulting short-range forecasts. The impact of assimilating GB-GPS data is neutral when evaluating forecasts against radiosonde observations. Precipitation verification results reveal overall a

TABLE 2. Bias model predictors for satellite radiance observations.

Type of radiance data	4DVar-based system	4DEnVar-based system
Microwave radiances (AMSU-A or AMSU-B, MHS, and SSMIS)	The 1000–300-, 200–50-, and 50–5-hPa geopotential thicknesses. In addition, for only AMSU-A channels 13 and 14, the 10–1-hPa thickness.	As in 4DVar-based system
Infrared radiances (AIRS, IASI, and geostationary imagers)	Observed brightness temperature	The 1000–300-, 200–50-, and 50–5-hPa geopotential thicknesses.

small and mixed impact of the GB-GPS data, with improvement in precipitation forecast skill noted for some regions, lead times, and amount thresholds. The overall impact of GB-GPS data is reduced in winter compared to summer due to lower average water vapor over North America.

d. *Improved satellite radiance bias correction procedure*

It is essential to employ an effective procedure for estimating and removing the error bias for all satellite radiance observations before assimilation (Derber and Wu 1998; Dee and Uppala 2009). Most procedures rely on other unbiased (or less biased) observations to determine the bias correction of radiance observations. The 4DVar-based system uses an adaptive approach in which the coefficients of a bias model are estimated using the last seven days of differences between the observed brightness temperature and the brightness temperature computed from the background state. This approach essentially assumes that the background state provides an unbiased estimate of the atmospheric state. However, if even a small bias is introduced during the short-term forecast that produces the background state, the satellite radiance observations will be adjusted to be consistent with this biased state. Given the increasingly dominant role played by satellite radiance observations, it is unlikely that other types of observations can act to reduce this bias when assimilated in combination with the large volume of radiances and, therefore, the resulting analysis will retain the bias. It is this known limitation of the approach that necessitates the use of static bias model coefficients for AMSU-A channels 11–14 to avoid a long-term drift of the mean upper-stratospheric temperatures.

The bias correction procedure was modified for use in the new 4DEnVar-based system to allow the other types of observations to have a more direct influence on determining how the radiance observations are adjusted. This is accomplished by performing an extra 3DVar analysis in which no radiance observations are assimilated. This analysis is then used instead of the background state to serve as the reference state for estimating the bias model coefficients. Because the radiance observations are not assimilated in the extra 3DVar analysis, the other types of observations can more easily change the mean temperature and humidity in the background state. This is especially the case for observing systems with extensive coverage over the globe, namely GPS radio occultation (GPS-RO) observations and to a lesser extent radiosonde and aircraft observations. Because of the increased influence of other observation types, the bias model coefficients for only AMSU-A channels 13 and 14 are now held static (instead of channels 11–14 in the previous system). The static coefficients for these two channels were recomputed for the complete 4DEnVar-based system presented in section 5. Results from experiments that demonstrate the impact of changes to the satellite radiance bias correction procedure in isolation are presented in section 3.

The choice of predictors used in the bias model was slightly modified in the new system to reduce the inconsistency between microwave and infrared observation bias estimates (Table 2). Also, to increase the influence of radiosonde observations on the estimated bias, the bias model coefficients are now estimated over the last seven days only from the assimilation windows centered at 0000 and 1200 UTC instead of using all four analysis times per day. However, it was recently realized that using data from only these two analysis times results in an incomplete spatial sampling of the bias for each sensor since the orbits at 0000 and 1200 UTC generally cover similar areas. Consequently, this change is currently being reevaluated.

e. *Additional AIRS/IASI channels assimilated*

In the existing 4DVar-based system, only 87 Atmospheric Infrared Sounder (AIRS; from *Aqua*) and 62 IASI (from *MetOp-A*) channels were assimilated. No IASI moisture channels in the 5.5–7.5-μm spectral region were assimilated. In the 4DEnVar-based system, 142 channels are assimilated for each instrument, covering the spectral domain 4.4–15.4 μm and, in particular, about 30 high-peaking temperature channels (70–150 hPa) were added (Garand et al. 2013). The impact of this change to the number of assimilated AIRS and IASI channels was evaluated in combination with changes to the specified observation-error standard deviation for

AIRS and IASI in experiments with a preliminary version of the 4DEnVar-based system over the two-month period February–March 2011. The additional channels increased the combined number of assimilated AIRS and IASI observations from about 1.2 million to about 3.8 million day^{-1}. Results show a modest improvement in the accuracy of temperature and midtropospheric humidity forecasts up to day 5 in the extratropics, but a neutral impact in the tropics.

f. Other changes to radiance assimilation

Satellite radiance observation-error statistics for assimilation purposes are typically estimated using the standard deviation over a fixed period of time of differences between observed brightness temperature and the brightness temperature computed from the background state. The resulting channel-dependent values of observation-error standard deviation are often adjusted to change the weight given to the observations and improve the resulting analysis by compensating for unknown spatiotemporal error correlations and representativeness errors. Previously, there was no well-defined method or procedure for computing and updating the satellite radiance observation-error statistics used at EC. Such a procedure was developed and implemented for the 4DEnVar-based systems as described by Garand et al. (2013). In the new system, errors are defined by a predetermined instrument- and channel-dependent factor F multiplied by the standard deviation of the observation–background differences. For microwave temperature sounding channels (AMSU-A), the factor ranges from 1.4 to 3.7 with the higher factors applied to the higher peaking channels. The factor is set to 2.0 for microwave water vapor channels [AMSU-B, microwave humidity sounder (MHS), SSMIS, and geostationary satellite imagers]. For hyperspectral infrared channels (AIRS and IASI), F is set to 1.6 for temperature channels and 2.0 for all other channels. The chosen values for F result in assigned standard deviation values for most channels that are similar to those used in the previous system. Consequently, data assimilation tests show a mostly neutral impact from this change and, therefore, are not shown. The main advantage of the new approach is that observation-error standard deviation values for new sensors, especially those with a large number of channels, can be computed more efficiently and with less chance of human error while providing consistency with existing sensors.

The fast radiative transfer model used for radiance assimilation, Radiative Transfer for the Television and Infrared Observation Satellite Operational Vertical Sounder (RTTOV; Matricardi et al. 2004), was updated from version 8.7 to 10.2. This allowed the use of a more up-to-date spectroscopy and the improved sea surface microwave Fast Emissivity Model, version 4 (FASTEM-4; Liu et al. 2011). This change affected the bias of some channels before bias correction; however, after bias correction, the overall fit of the background state to the observations was almost unchanged. The impact of this change on the forecast scores was also neutral.

g. Changes to the forecast model and its initialization

The configuration of the atmospheric model component of the 4DEnVar-based system remains essentially unchanged from that described by Zadra et al. (2014b) and used in the 4DVar-based system, although extensive modifications have been made to the initialization procedure. The Global Environmental Multiscale (GEM) model (Côté et al. 1998a,b) is run on a global cylindrical equidistant grid with a horizontal grid spacing of 0.35° × 0.225° in the meridional and zonal directions, respectively. For the 9-h forecasts used to generate the background state, the step length of the two time-level implicit, semi-Lagrangian core has been reduced from 540 to 450 s so that steps coincide with times for which the ensemble covariances are available from the EnKF (Houtekamer et al. 2014). The time step remains 720 s in the medium-range forecast integration.

The only appreciable modifications to the model's physical parameterizations reside in the calculations of snow density and the thickness of the first prognostic layer. The former involves a correction to a unit conversion of the analyzed value, leading to an approximate doubling of the snow density seen by the model. The latter results in an increased first-layer thickness that enhances the activity of the shallow convective scheme. The impacts of these corrections on the model climate and near-surface forecasts are discussed in more detail in section 4a.

The focus of model development for the 4DEnVar-based system was on the reduction of spinup[1] in the system. A key step in the development of a system that smoothly transitions between assimilation windows was the replacement of the existing full-field digital filter (Lynch and Huang 1992; Fillion et al. 1995), with a four-dimensional incremental analysis update (4DIAU) technique. Unlike the traditional incremental analysis update (IAU) approach, in which a single analysis increment is distributed across the assimilation window (Bloom et al. 1996), time-varying increments for 4DIAU

[1] The term spinup is used generically here to imply rapid evolution of the model state early in the integration and thus includes effects sometimes referred to separately as spindown.

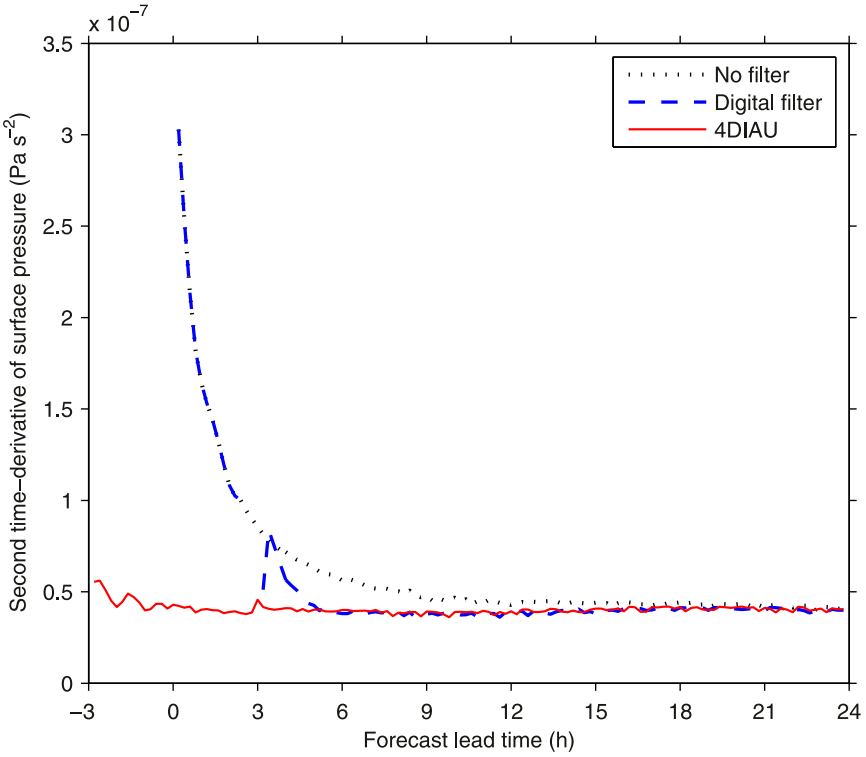

FIG. 2. The global-mean second time derivative of surface pressure is shown for each forecast lead time, for the initialization techniques shown in the legend. The ability of the different initialization techniques to limit spurious wave activity can be assessed using the second time derivative of surface pressure, a quantity that reflects the departure of the barotropic modes from the Charney (1955) balance because of its direct relationship to the time tendency of column-integrated divergence.

are computed at hourly intervals by 4DEnVar. Lorenc et al. (2015) demonstrate that the filtering properties of IAU described by Polavarapu et al. (2004) apply in an empirical sense to 4DIAU in a hybrid-4DEnVar system because the waves generated by imbalances introduced by the analysis increments are uncorrelated with the slowly varying increments themselves. A confirmation of this result for the 4DEnVar-based system is shown in Fig. 2. Throughout the integration using 4DIAU, gravity wave activity remains near baseline levels, a degree of balance achieved only after 4 h of integration in a comparable simulation initialized with a full-field digital filter.

The smooth transitions between assimilation windows created by the IAU technique (Rienecker et al. 2011) allow for the potential of spinup reduction by recycling of the complete state vector, including physical components. However, the availability of updated surface analyses (not introduced incrementally in the 4DEnVar-based system) and storage limitations make the recycling of the full state impractical (a full checkpoint represents over 750 GB of data and is dependent on the internal memory structure of the model). Instead, the 4DEnVar-based system only recycles a few key variables between integrations: horizontally staggered winds, condensate mixing ratio, and fields related to the boundary layer parameterization. A description of the tests that led to the selection of these variables is provided in section 4c, where it is shown that the adoption of this strategy succeeds in reducing spinup in the 4DEnVar-based system.

h. Other changes

A new global sea ice concentration analysis is used to specify the fixed ice concentration for the atmospheric forecasts. The new system uses a 3DVar approach to assimilate ice concentration retrieved from passive microwave satellite observations (SSM/I and SSMIS) and from the manual ice charts produced by the Canadian Ice Service [more details are given by Smith et al. (2015)]. A set of two-month data assimilation experiments with a previous version of 4DEnVar over February–March 2011 revealed localized changes to the mean surface air temperatures of up to 1 K, most notably

in Baffin Bay; however, there was no significant impact on the short- and medium-range forecasts.

As part of an ongoing effort to unify the work flow software used in research and operations, job management in the 4DEnVar-based GDPS is performed by the EC-developed Maestro sequencer. The unification of the sequencing tool and all of the scripts significantly simplified the technological transfer of this system from research to operations despite the complexity and intersystem dependencies that accompanied the use of 4DEnVar.

3. Evaluation of the modified satellite radiance bias correction procedure

A pair of data assimilation and forecast experiments are used to evaluate the impact of a modified procedure, described in section 2d, for satellite radiance bias correction in isolation. To accelerate the testing, these experiments both use 3DVar with first guess at the appropriate time. Since modification of the satellite radiance bias correction is expected to affect mostly the very large spatial and temporal scales of the analysis increments, the relative impact of this modification with 3DVar should be very similar to the impact with 4DEnVar. The experiments were performed over the two-month period of February–March 2011, with the only difference between them being the procedure used for satellite radiance bias correction. The experiment with the new approach is slightly different from that used in the 4DEnVar-based system, since it is not applied to radiances from SSMIS and geostationary satellite imagers (the previous approach is used for observations from these sensors). An additional difference is that the bias model coefficients are estimated from the four analysis times per day instead of only at 0000 and 1200 UTC. Based on the results from shorter experiments, the first difference was found to affect only the bias of humidity and the second difference has only a very small overall impact.

Figure 3 shows results from the 3DVar experiments using the previous bias correction approach (blue) and the new approach (red), both compared against ERA-Interim (Dee et al. 2011). The radiance observations assimilated in ERA-Interim are bias corrected using a variational approach that also allows all other observations (including those from GPS-RO and radiosondes) to determine the error bias for radiances (Dee and Uppala 2009). The mean temperature and geopotential height (dashed curves) are both significantly[2] closer to ERA-Interim when the new radiance bias correction procedure is used, most notably above 20 hPa. As stated earlier, the two highest peaking AMSU-A channels (13 and 14) are corrected using static coefficients in the bias model and the same coefficients are used in the two experiments presented in this section. This likely explains why a larger negative temperature bias is seen in the analyses for both experiments above 5 hPa (Fig. 3b). Below this, the improvements result from cooling in the lower troposphere and warming in the upper troposphere and throughout most of the stratosphere introduced in the analysis (Figs. 3a,b) and maintained in the 48-h forecast (Figs. 3c,d). Small but significant improvements in the standard deviation of temperature and geopotential height are seen at layers throughout the column in both the analysis and short-range forecast.

4. Evaluation of changes to the forecast model and its initialization

Modifications to the forecast model component of the 4DEnVar-based system can be broken into three categories as outlined in section 2g: physics, initialization, and recycling. The impact of these changes on the forecasts is assessed using different sets of experiments designed to isolate the contribution from each element. The period of investigation is February–March 2011, with experiments consisting of either forecast-only or data assimilation cycles, as necessary. Whenever possible, the results of experiments whose configurations are close to that of the 4DEnVar-based system are employed; however, some impacts are evaluated in experiments with somewhat different configurations. Most notably, some experiments employ a Yin–Yang GEM model grid with 0.15° grid spacing (Qaddouri and Lee 2011) instead of the global grid chosen for the 4DEnVar-based system. The change in grid definition is not expected to affect the results presented here in a meaningful way, since each set of experiments is self-consistent in this respect.

a. Impact of physics changes

A correction to the calculation of the thickness of the first prognostic layer increases the activity of the shallow convection scheme (Bélair et al. 2005) by eliminating a dry-adiabatic base in the plume model component of the parameterization. A series of forecasts shows no impact from this correction when evaluated against surface or radiosonde observations, but yields a 2% increase in global cloud cover and a 1% increase in precipitation. An analysis of the planetary albedo suggests that this adjustment is small and potentially beneficial given that comparisons against satellite and surface observations

[2] Significance is computed using permutation tests for both bias and std dev [see chapter 15 of Efron and Tibshirani (1993)].

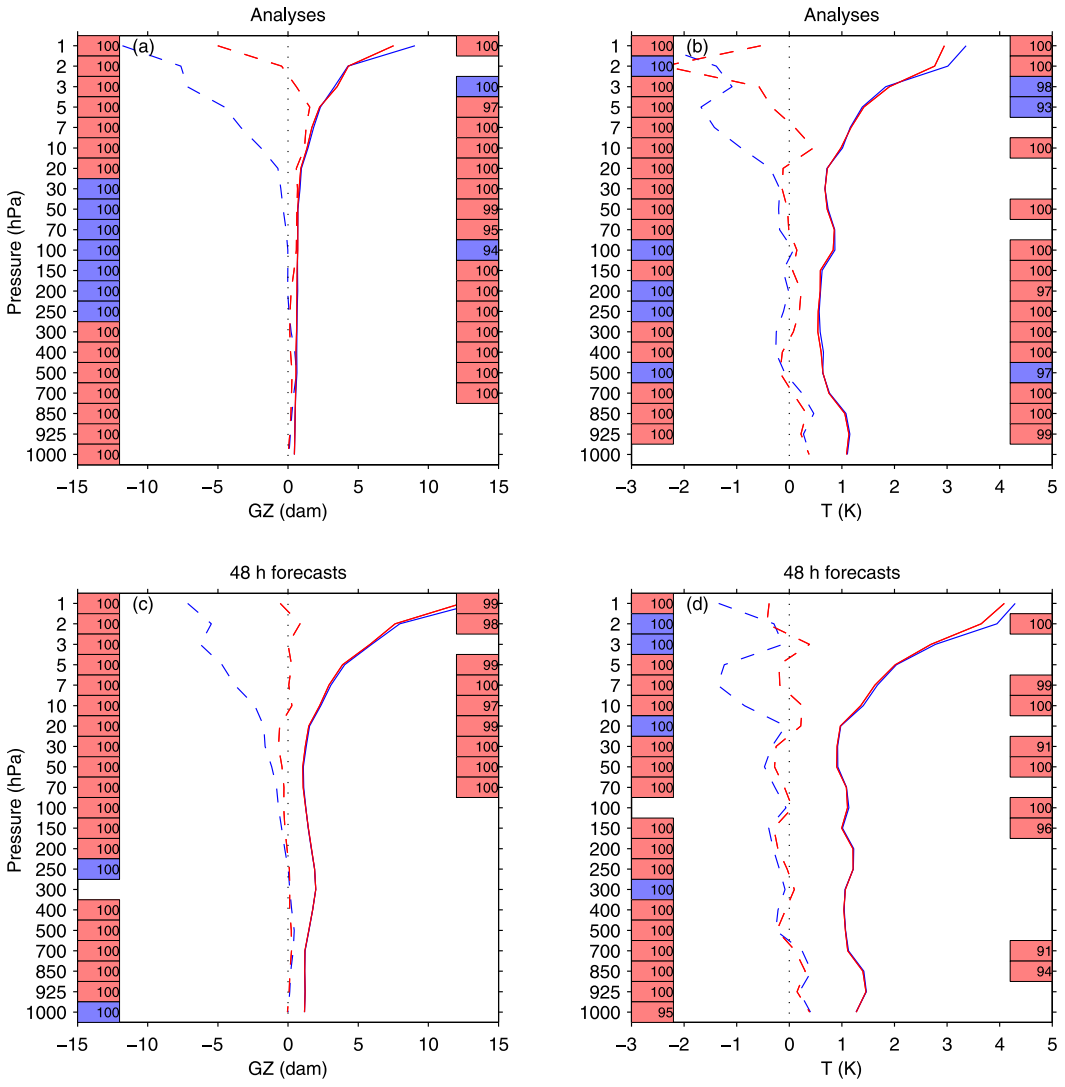

FIG. 3. The bias (dashed curves) and std dev (solid curves) from two 3DVar experiments evaluating the impact of the new satellite radiance bias correction procedure (red curves) relative to the previous approach (blue curves). The (a),(b) analyses and (c),(d) 48-h forecasts for (left) geopotential height and (right) temperature are verified against ERA-Interim over the global domain for February–March 2011. The numbers contained within the shaded boxes on the left (right) are the level of confidence in percent that the biases (std dev) from the two experiments are different, with the color indicating which experiment has the better score.

show that the GEM model produces insufficient cloud coverage (Paquin-Ricard et al. 2010; Garand et al. 2011).

The other potentially relevant physics modification affects the calculation of the maximum snow density. The effect of this correction to an error in a unit conversion is a 2–2.5-fold increase in the density of the snowpack (Bélair et al. 2003). This is expected to have an impact on forecasts through the snow albedo (Carmagnola et al. 2013) and the thermal conductivity of the snowpack. To evaluate the impact of this modification in isolation, the erroneous snow density calculation was reintroduced in the 4DEnVar-based system and tested over an eight-week period. The effect of the corrected snow density calculation on near-surface temperature and dewpoint predictions is shown in Fig. 4. Surface temperature standard deviation is improved by about 0.2 K in the overnight hours, while biases are decreased significantly only in the first 12 h of the forecasts (Figs. 4a,c). This first overnight period is also characterized by improved dewpoint forecasts, while results at longer lead times are mixed (Figs. 4b,d).

b. *Impact of modified initialization*

The introduction of analysis increments using the 4DIAU technique described in section 2g replaces the digital filter that is applied to the full model fields in

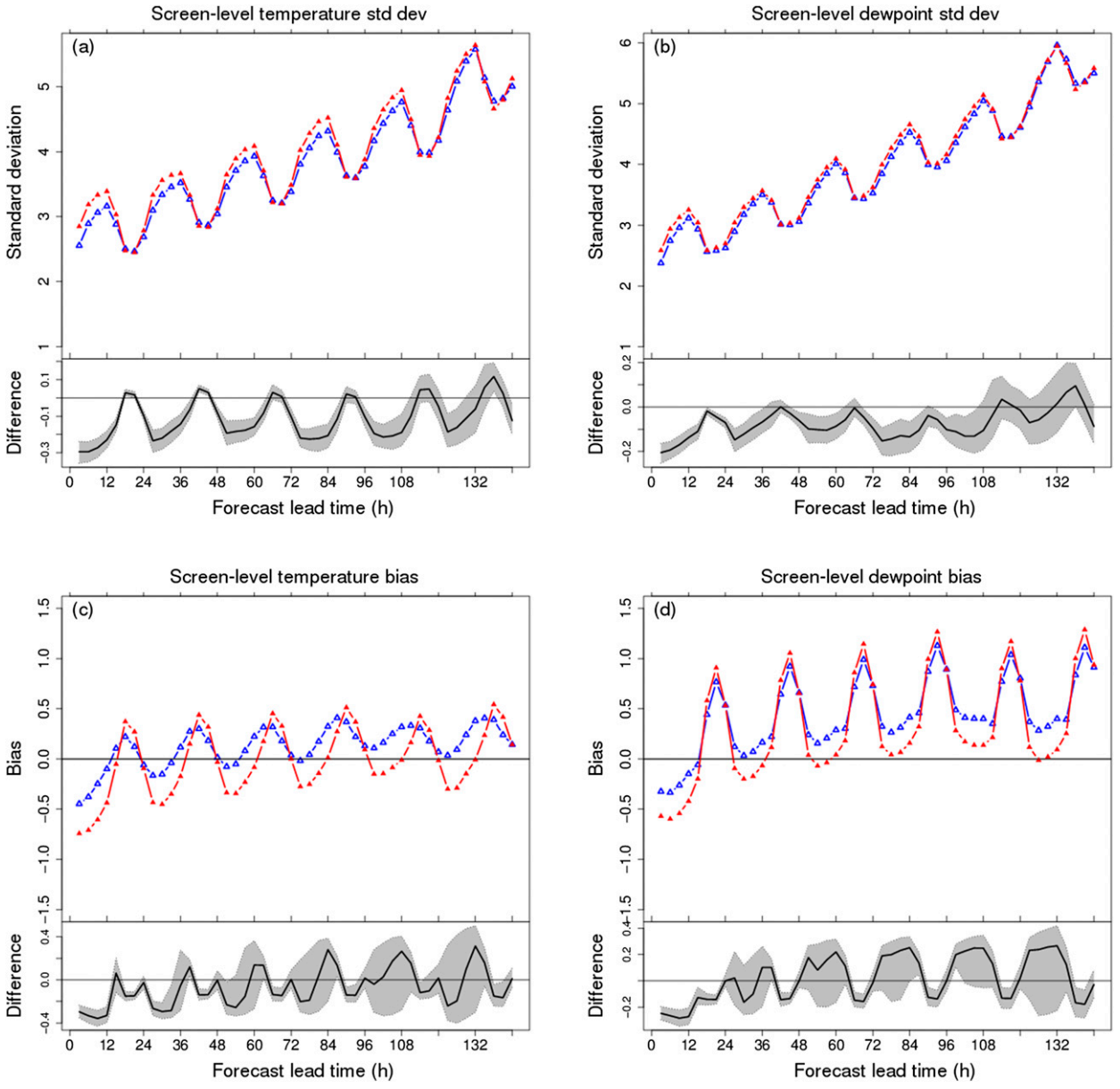

FIG. 4. The (a),(b) std dev and (c),(d) bias relative to near-surface (also referred to as screen level) (left) temperature and (right) dewpoint observations over North America of forecasts from the 4DEnVar-based system with the original erroneous unit conversion for snow density (red curves) and with the correct conversion (blue curves). Results are presented for February–March 2011 forecasts initialized at 0000 UTC. The difference in either std dev or bias between the two experiments is shown below and the shading represents the 90% confidence interval based on a bootstrap analysis.

the 4DVar-based system. To assess the impact of this change in isolation, the results from a pair of two-month data assimilation experiments using 4DEnVar and the Yin–Yang model configuration are presented here. The experiments differ with respect to both their initialization procedure (digital filter vs 4DIAU) and their recycling of the model state (to be discussed in more detail in section 4c). Although the majority of the impact described here is consistent with documented issues related to the application of a digital filter to the full fields, the potential for additional contribution from state recycling cannot be ruled out.

After 24 h of integration, the impact on the global standard deviation and bias relative to ERA-Interim is neutral for winds and moisture (not shown) and significantly improved for geopotential height (Fig. 5a) despite

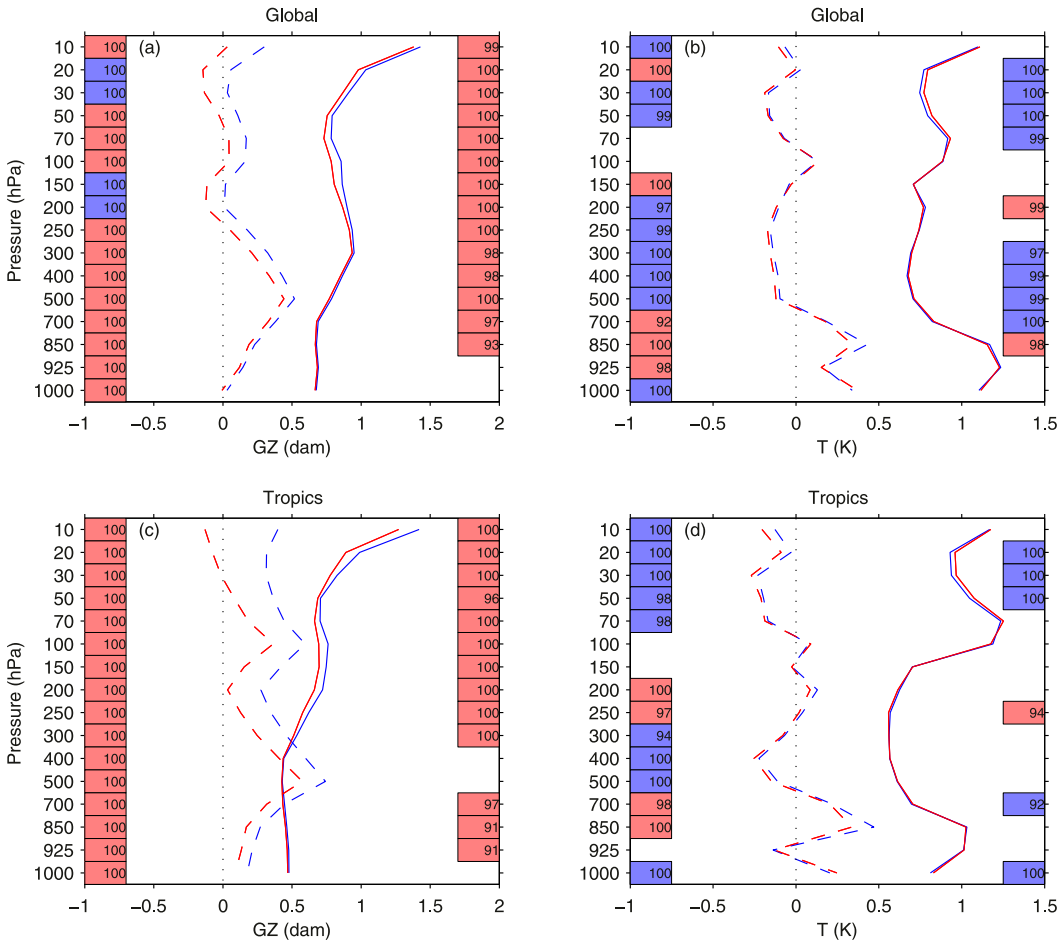

FIG. 5. As in Fig. 3, but for 24-h forecasts from a pair of 4DEnVar experiments evaluating the impact of 4DIAU (red) compared with a nonincremental digital filter (DF; blue). Results are shown for the (a),(b) global domain or (c),(d) only the tropics (20°S–20°N).

being slightly negative for temperature (Fig. 5b). The bulk of the improvement in the height field is focused in the tropics as shown in Fig. 5c, where biases are reduced by almost 3 m in the upper troposphere. The improvement in standard deviation decays with forecast length, but the bias improvement remains significant throughout the forecast.

The source of the increased skill in tropical height forecasts is an improved representation of the semidiurnal tide (Dai and Wang 1999). A clear wave-2 structure appears in the mean difference between 24-h forecasts of 1000-hPa geopotential height from the two experiments (Fig. 6a), with centers of action in the Pacific and Indian Oceans that are in phase with the semidiurnal tide at the valid times of the forecasts (0000 and 1200 UTC). Geopotential heights from the experiment using 4DIAU are in better agreement with ERA-Interim than those from the digital filter–based experiment (cf. Figs. 6b and 6c). Although the response function of the digital filter [Fig. 2a of Polavarapu et al. (2004)] suggests a 15% power reduction of this mode on each application, the cumulative effect of filtering is far greater in an assimilation cycle (Sankey et al. 2007). Seaman et al. (1995) demonstrate that it is this repeated filtering of the full fields, rather than the analysis increments alone, that leads to the degradation of the tidal signal. The vertical profiles of height prediction improvements shown in Fig. 5 are consistent with those related to the introduction of an incremental filtering technique [e.g., Fig. 5 of Seaman et al. (1995)], increasing with height in the troposphere in a manner commensurate with the amplitude of a well-represented semidiurnal tide.

The impact of 4DIAU on model spinup is best evaluated from an energetics perspective. Sankey et al. (2007) demonstrate that the replacement of full-field digital filtering with an incremental approach has a significant impact on mesoscale divergent kinetic energy

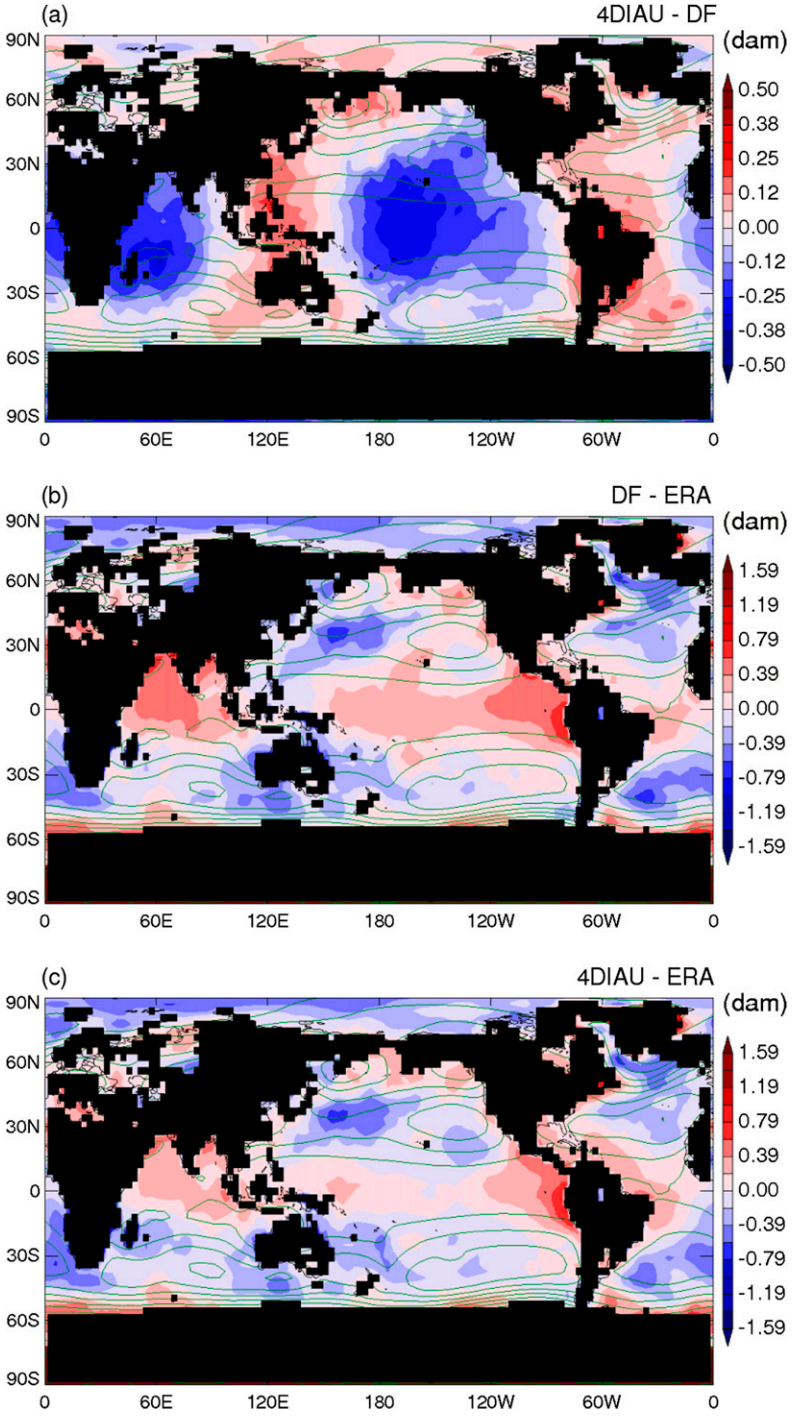

FIG. 6. (a) Difference in 1000-hPa geopotential height in the 24-h forecast (0000 and 1200 UTC) between 4DEnVar experiments using 4DIAU or a nonincremental DF. (b) As in (a), but for experiments using DF or ERA-Interim. (c) As in (a), but for experiments using 4DIAU or ERA-Interim. Note that areas where the surface pressure is less than 1000 hPa are shown in black. The mean 1000-hPa geopotential heights are shown in green contours at 3.2-dam intervals.

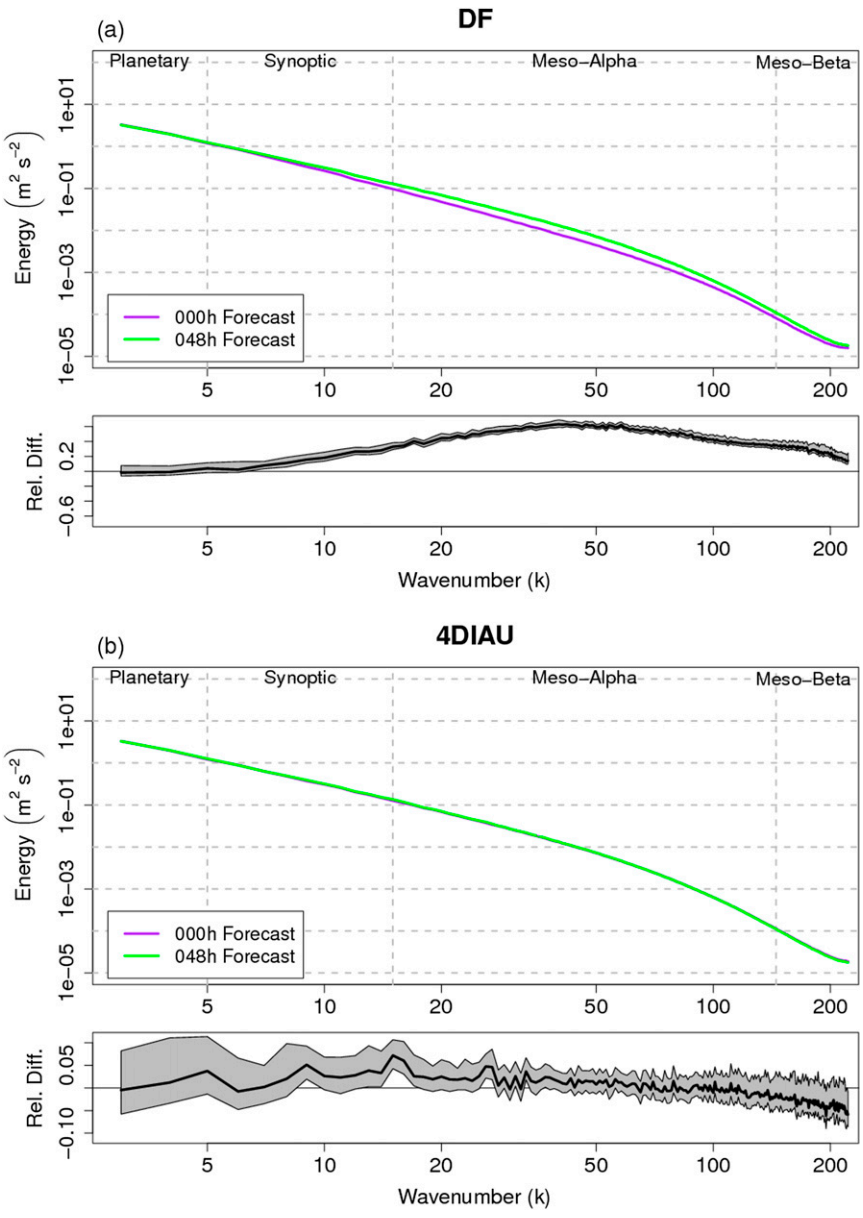

FIG. 7. Spinup of kinetic energy at 500 hPa in short-range forecasts from (a) DF- and (b) 4DIAU-based cycles in a Yin–Yang model configuration. The mean power spectrum in the 0- (purple) and 48-h (green) forecasts is displayed above, with scales corresponding to wavenumber-based divisions identified with gray dashed vertical lines. The relative difference between the power spectra, defined as the difference divided by the 0-h value, is shown below. The range of the ordinate scales to the max relative difference and is, therefore, different for (a) and (b). A 1000-member bootstrap of the difference is used to estimate the 95% confidence interval, which is plotted with gray shading below.

following repeated filter applications in an assimilation cycle. This effect can be seen in Fig. 7 for a pair of data assimilation experiments using the different initialization techniques. The mesoscale portion of the energy spectrum is significantly depleted when the full-field digital filter is employed, with kinetic energy associated with motions on these scales reduced by over 50% in the analysis as compared with a short-range forecast. This leads to a dramatic spinup over the first two days of integration. The replacement of the digital filter with 4DIAU ensures that these scales of motion are retained in the analysis. Phenomena with characteristic time

scales of 6–10 h, largely eliminated by the nonselective digital filter, include frontal waves, mesoscale convective systems, diabatic Rossby vortices, and polar lows. Such features are at least partially resolved by the 4DEnVar-based system and can help to establish predictability in the 1–2-day range (Lorenz 1969; Gray 2001; Zhang et al. 2007). Although the standard deviation of short-range forecasts relative to observations or analyses is slightly increased with 4DIAU due to a more fully developed mesoscale spectrum, the retention of these scales should improve the representation of error growth and the reverse energy cascade (Charron et al. 2010; Waite and Snyder 2013).

c. *Impact of recycling*

An important advantage of IAU-based techniques compared to digital filter or normal mode initialization strategies (Daley 1981) is the creation of a continuous integration path along which the short-term forecasts blend seamlessly between assimilation windows without any abrupt changes. This allows for the potential recycling of the full model state to reduce spinup as described in section 2g. A system without appreciable spinup will remain closer to the model climate [M-Climate; Graham et al. (2013)], with persistent increments indicative of true biases in the model or observations. This attribute of the 4DEnVar-based system is expected to be beneficial for future model development.

A selection of the key components of the model state whose recycling significantly reduces model spinup is required, since full recycling is impractical (section 2g). A perfect-model framework is adopted, based on a continuous reference integration initialized at 1200 UTC 5 February 2011. The same period is covered with a set of integrations that restart every 6 h to mimic an increment-free data assimilation cycle. A comparison of the restarted results with those of the reference integration provides an estimate of the impact of imperfect recycling on the data assimilation cycle. The global-mean absolute difference in surface pressure is used to assess the magnitude of these differences. Similar measures using other fields and levels yield qualitatively similar results. A set of otherwise-perfect restarts perturbed with a bit flip in the sixth significant digit of the state variables is used as a baseline for comparison of difference growth: any combination of recycled variables that leads to an integration whose results approach this level of difference from the continuous reference integration is considered acceptable.

A set of restarted integrations was first performed in which the model's physical state is perfectly recycled (Fig. 8a). In previous systems, including the 4DVar-based system, the analyzed wind components were computed on the same horizontal grid as the mass variables, instead of on the model's horizontally staggered Arakawa C grid (Arakawa 1988). The interpolations required for this grid transformation result in differences that exceed those from the bit-flip case by 50%, whereas recycling winds directly on the model grid yields bit-flip-level differences (Fig. 8a). Further testing also shows that, at least in the hydrostatic configuration considered here, recycling of the model-coordinate and real vertical motion fields (Girard et al. 2014) does not appear to be necessary.

Differences when none of the model's physical state is recycled (as in the 4DVar-based system) exceed the bit-flip case by an order of magnitude (Fig. 8b). The total condensate field [liquid water and ice combined; Sundqvist et al. (1989)] is readily identified as relevant, but corrects only a fraction of the observed differences. The remaining key variables for recycling are related to the boundary layer parameterization (Mailhot and Benoit 1982), in which the higher-order closure depends on predicted turbulent kinetic energy, turbulence regime, and a time-averaged estimate of the mixing length (McTaggart-Cowan and Zadra 2015). The friction velocity and boundary layer height fields are also recycled in the 4DEnVar-based system because these single-level fields represent a small volume of additional data that yield a small marginal benefit in these tests. The prevalence of quantities relevant to the boundary layer is somewhat surprising given that the lower atmosphere typically adjusts on hourly time scales (Stull 1988). However, these variables contain information about the history of the boundary layer state that cannot be reconstructed diagnostically during successive restarts.

The leading-order impact of selective recycling on clouds and precipitation in a data assimilation experiment is shown in Fig. 9. The amplitude of the spinup evident in the cloud field over the first day of the digital filter–based forecast is reduced in the 4DIAU integration with selective recycling (Fig. 9a). The slow adjustment of the latter beyond day 2 appears to be related to upper-tropospheric drying in the analysis that results from a change in how the humidity increments are computed, a complicating factor that is independent of initialization technique. A nearly steady $\sim$0.38 mm (3 h)$^{-1}$ global-mean precipitation rate in the 4DIAU-based system (Fig. 9b) is indicative of an overall reduction in spinup and the relative proximity of the system to the M-Climate. The initial precipitation peak in the 4DIAU-based system is the result of enhanced convective activity across the intertropical convergence zone. The difference between precipitation accumulations over the 0–3- and 24–27-h periods (Fig. 9c) shows a dramatic spindown of tropical convection in both systems. This

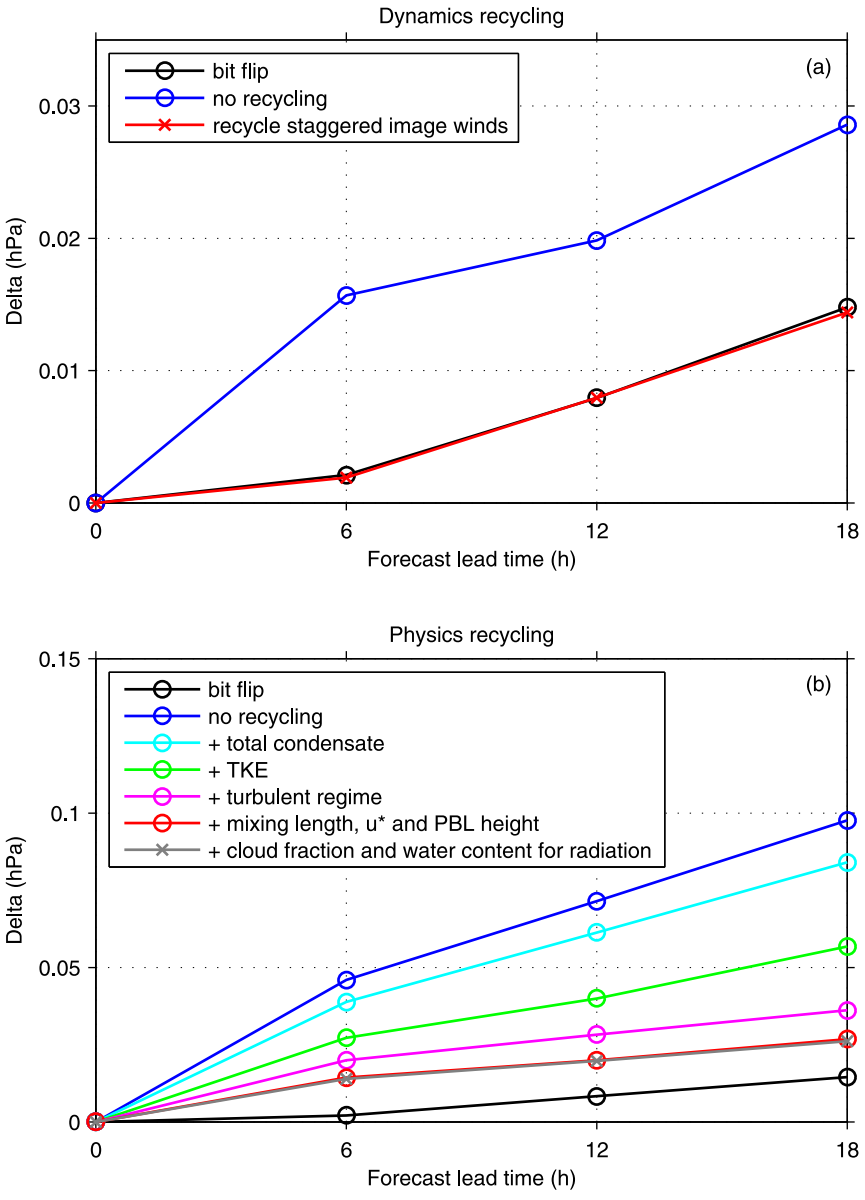

FIG. 8. Globally averaged mean absolute differences in surface pressure between the continuous reference integration and selective recycling experiments as identified in the legends. The tests of dynamics variable recycling shown in (a) are undertaken with perfect physical state recycling. The selective physics variable recycling tests shown in (b) are cumulative from the original no recycling integration as indicated by plus signs in the legend. The turbulent kinetic energy is abbreviated as TKE, and u^* represents the friction velocity in (b). Results of the bit-flip experiment are shown with a black line.

behavior appears to be the result of an unexpected increase in moisture at the top of the boundary layer in the 0.15° Yin–Yang grid configuration that is not observed on the implemented 0.35° × 0.225° global grid. The digital filter–based system suffers from a larger convective spindown, but the associated initial peak is masked by the dramatic spinup of stratiform precipitation over the rest of the globe. The relative lack of spinup in the 4DIAU-based system means that the increment-induced convective peak becomes apparent in the global-mean precipitation time series.

5. Evaluation of combined impact from all changes

Experiments were performed with the new 4DEnVar-based system, including all modifications described in

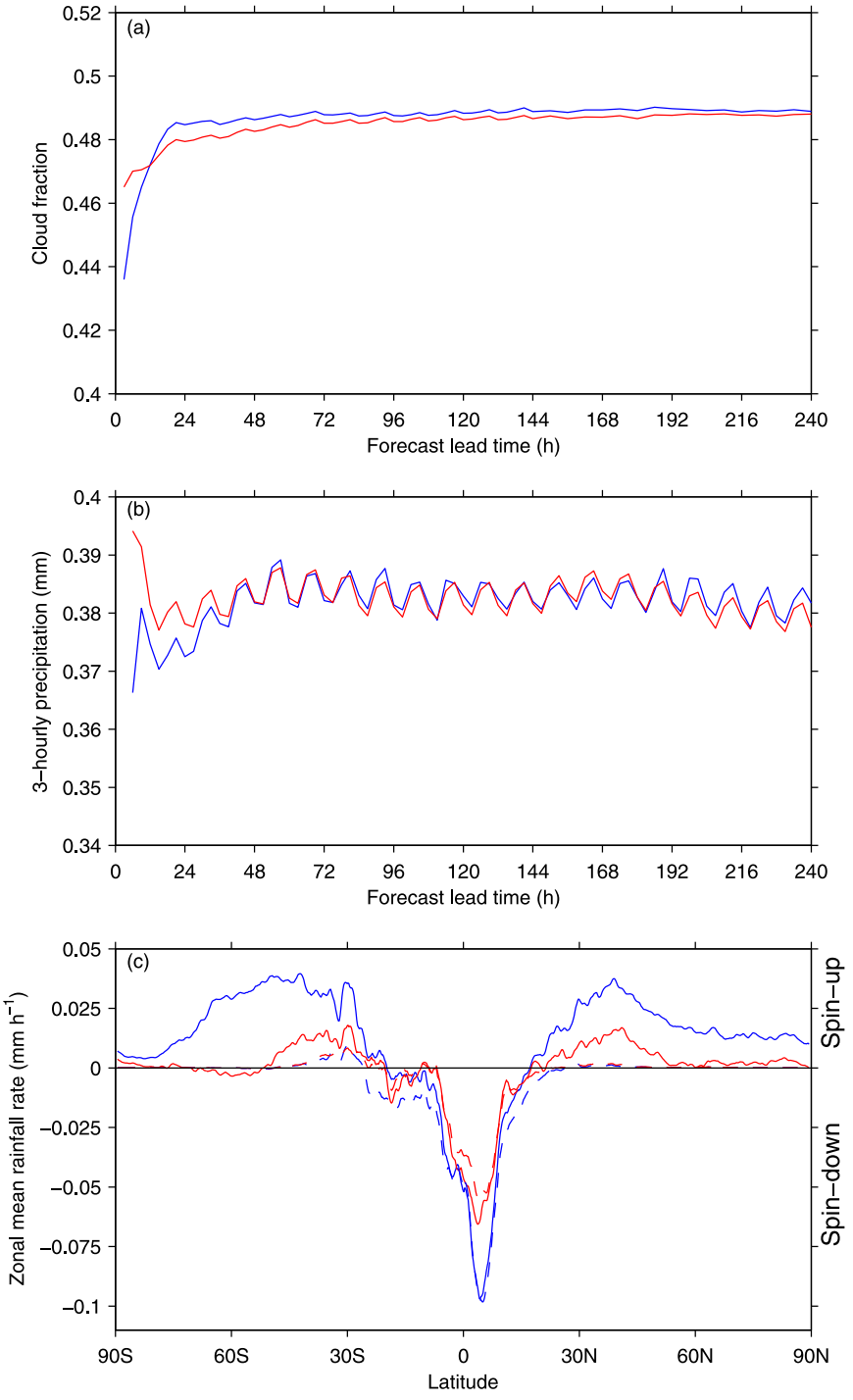

FIG. 9. Time evolution of (a) global-mean cloud cover and (b) 3-hourly precipitation accumulations during the boreal winter period when using the DF (blue) and 4DIAU with selective recycling (red) in 10-day forecasts with the 0.15° Yin–Yang model configuration. (c) The difference between zonal-mean 3-h-averaged precipitation rates over the 24–27- and 0–3-h periods. Negative values are indicative of spindown and positive values are indicative of spinup. Total precipitation is shown with solid lines while the convective component is shown with dashed lines when using the DF (blue) and 4DIAU with selective recycling (red).

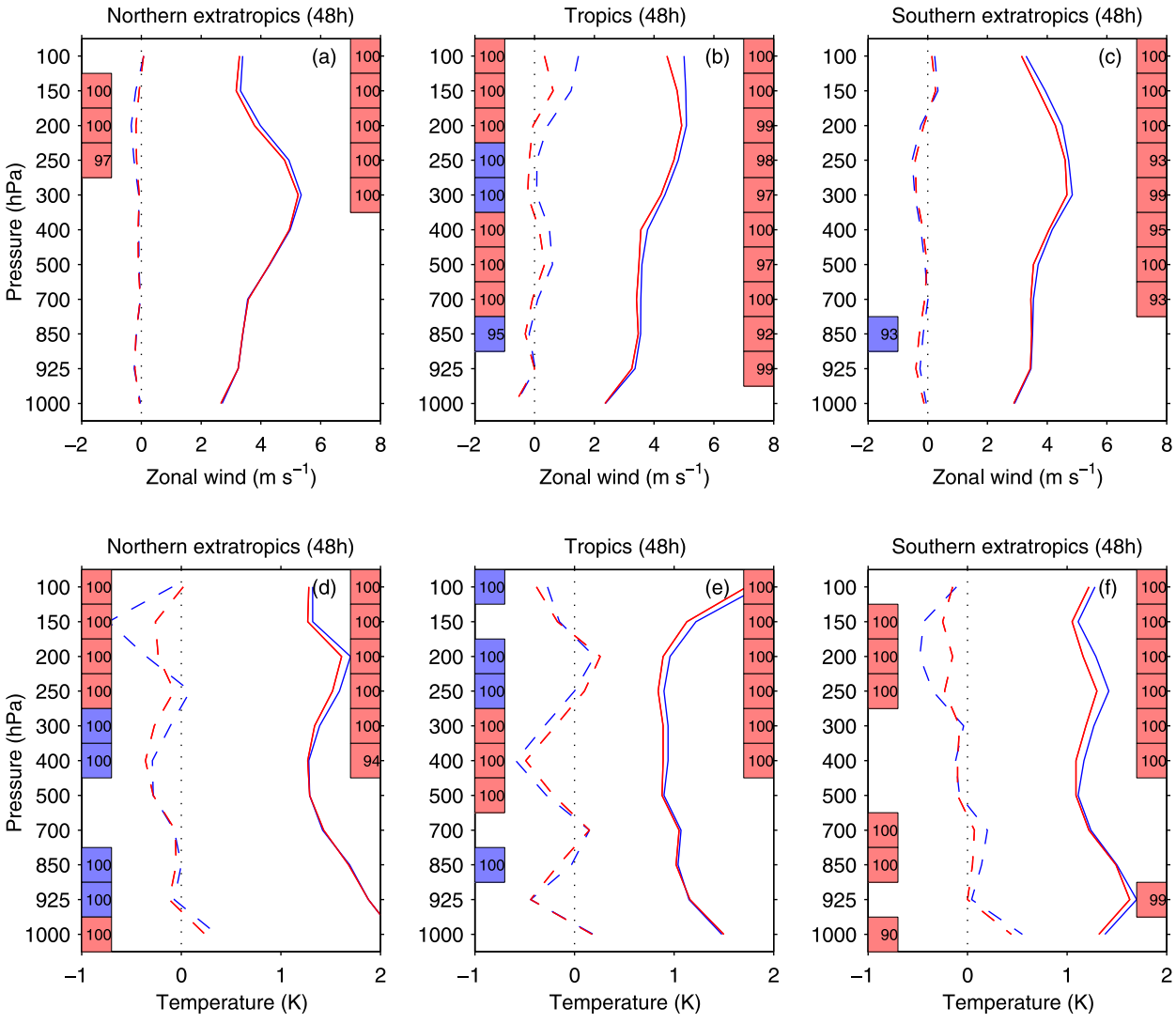

FIG. 10. The bias (dashed curves) and std dev (solid curves) relative to radiosonde observations of 48-h forecasts from the 4DEnVar-based system including all improvements (red curves) and the 4DVar-based system (blue curves). Results are shown for both (a)–(c) zonal wind and (d)–(f) temperature over the (a),(d) northern extratropics (north of 20°N); (b),(e) tropics (between 20°N and 20°S); and (c),(f) southern extratropics (south of 20°S) during February–March 2011. The numbers contained within the shaded boxes on the left (right) are the level of confidence in percent that the biases (std dev) from the two experiments are different, with the color indicating which experiment has the better score.

section 2, over the two-month period from 1 February to 31 March 2011 (referred to hereafter as FebMar) and from 1 July to 31 August 2011 (referred to hereafter as JulAug). The results were compared against those from a 4DVar-based system equivalent to the system that became operational on 13 February 2013 (Zadra et al. 2014b). Given the practical advantages of 4DEnVar over 4DVar discussed by Buehner et al. (2013), the results from these experiments demonstrated a sufficient improvement in analysis and forecast accuracy from the new 4DEnVar-based system to justify its operational implementation during the fall of 2014.

a. *Verifications against radiosonde observations*

First, forecast quality is assessed by comparison with radiosonde temperature and zonal wind observations over three regions (northern extratropics, tropics, and southern extratropics). The statistical significance[3] of the difference between the experiments in bias and standard deviation is shown on left and right sides of Figs. 10a–f,

[3] Significance is computed using the Student's *t* test for bias and the *F* test for std dev.

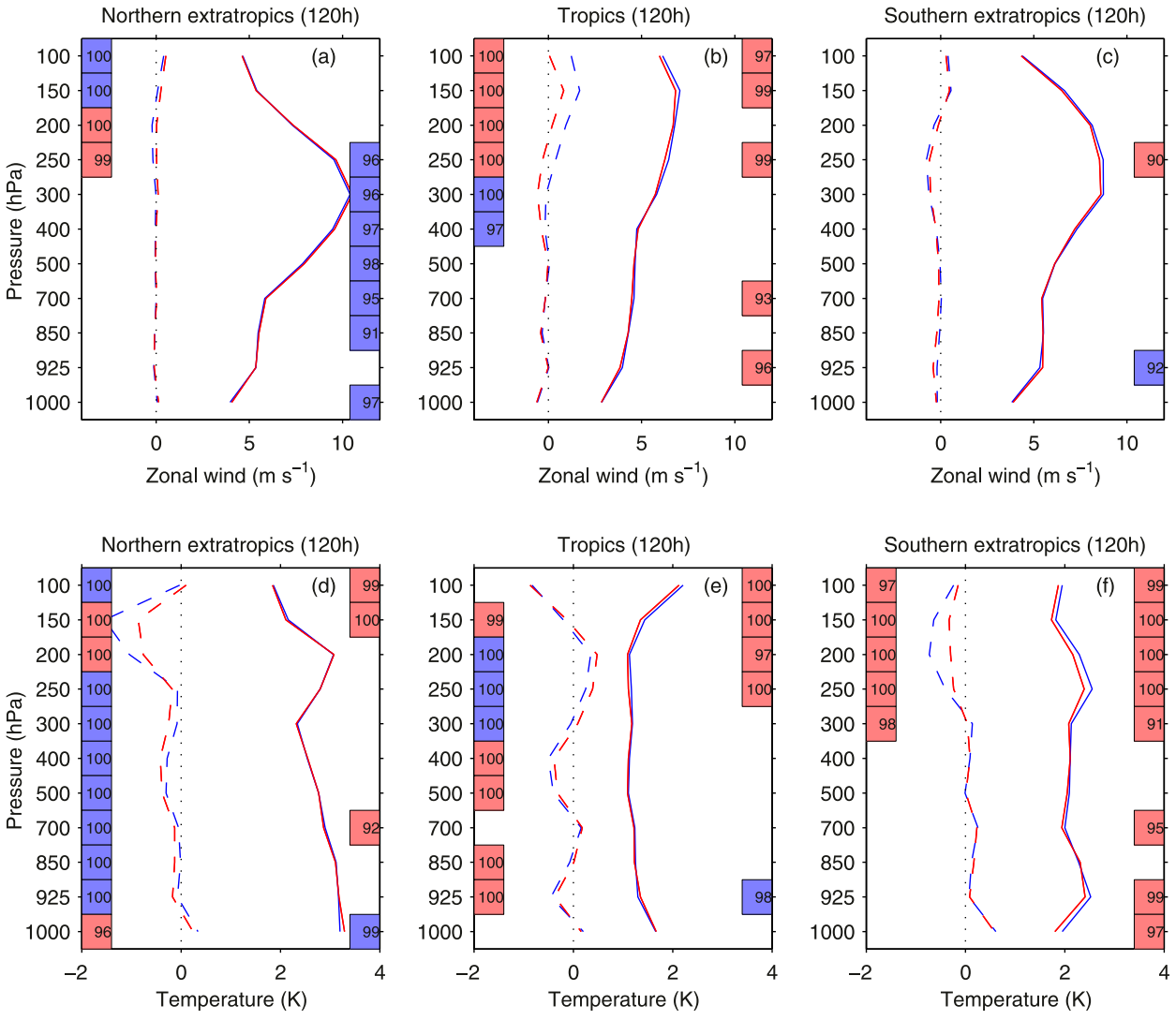

FIG. 11. As in Fig. 10, but for 120-h forecasts.

respectively. When greater than or equal to 90%, confidence levels are contained within colored boxes that are either red or blue when the score favors the new or old system, respectively.

The standard deviation is reduced in the 4DEnVar-based system (red) compared with the 4DVar-based system (blue) for the 48-h FebMar forecasts of both zonal wind and temperature in the upper troposphere and lower stratosphere in all three regions and down to the surface for winds in the tropics (Fig. 10). The bias for wind is similar for the two systems in the extratropics and generally improved in the tropics with the new system. For temperature, the bias is improved and degraded in different vertical layers, except in the southern extratropics where it is mostly improved using the new system. Compared with the 48- and 120-h forecasts shown here, the temperature bias with respect to the 6-h forecasts shows a more systematic improvement for all levels. Consistent with the expected impact of the new radiance bias correction procedure (section 3), the temperature bias in the 4DEnVar-based system varies less in the vertical, most noticeably in the two extratropical regions.

The differences in standard deviation between the experiments tend to diminish with forecast lead time (Fig. 11). By 120 h, the new system has a similar or reduced standard deviation except for zonal wind in the northern extratropics. The small, but just statistically significant increase in standard deviation for zonal wind in this region (Fig. 11a) is likely misleading (i.e., a type-I error) since the same scores for the 72-, 96-, and 144-h forecasts (and also the standard

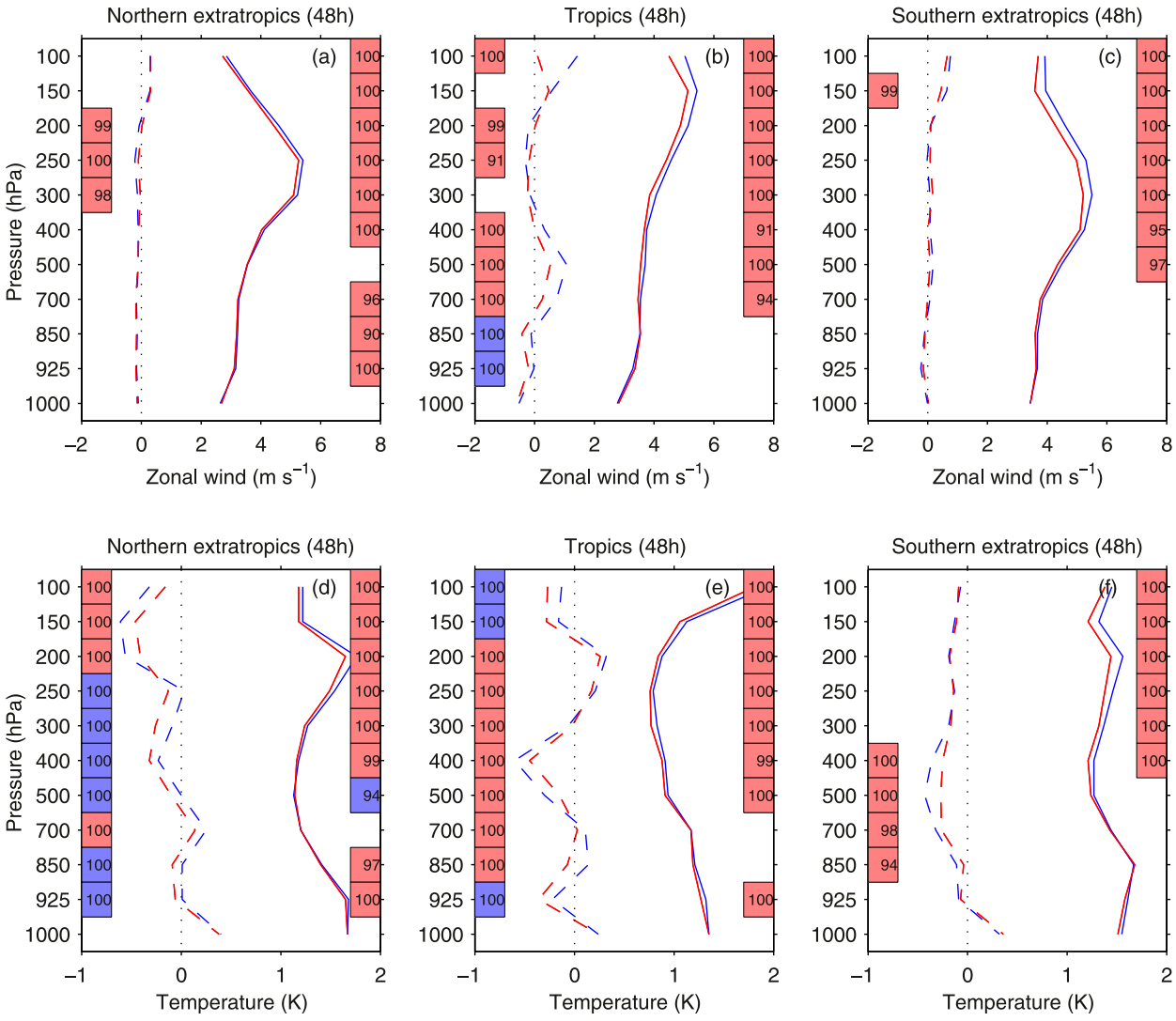

FIG. 12. As in Fig. 10, but for July–August 2011.

deviation of 120-h northern extratropical forecasts relative to ERA-Interim) all show either a similar or reduced standard deviation for the new system. In the southern extratropics, the standard deviation maintains much of the reduction seen in the 48-h forecasts. The bias for zonal wind and temperature and the relative differences between the two experiments are quite similar at short and medium forecast ranges (cf. Figs. 10 and 11).

Figures 12 and 13 show the same results as Figs. 10 and 11, respectively, except for the JulAug period. The short-range JulAug forecast scores generally resemble those of the FebMar period (cf. Figs. 10 and 12). However, the results suggest a slightly more positive impact from the new system in the lower troposphere in the northern extratropics when compared to the FebMar period. Results for medium-range forecasts are also similar to those seen for FebMar, with a positive or neutral impact from the new system in the southern extratropics and tropics, but a near-neutral impact in the northern extratropics (cf. Figs. 11 and 13).

b. *Verifications against independent analyses*

The two systems were also evaluated through comparisons with an independent set of analyses from the ERA-Interim (Fig. 14). A general improvement in agreement with ERA-Interim is evident for the new system during both periods, between the surface and 10 hPa and for all forecast lead times, with the exception of a small degradation around 300 hPa at 96 and 120 h in the FebMar period. The largest improvements generally occur during the first two days of the forecasts, as may be

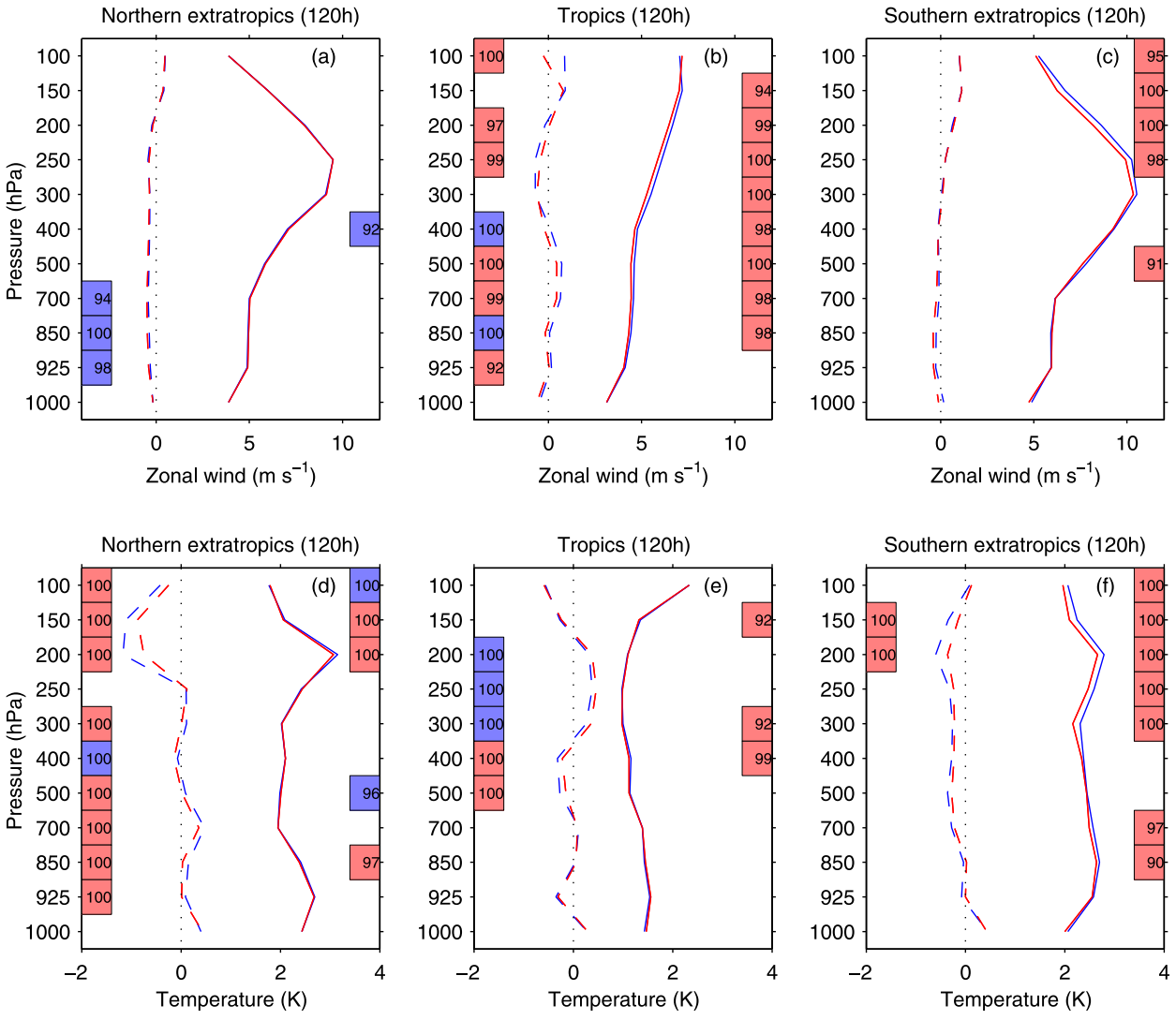

FIG. 13. As in Fig. 11, but for July–August 2011.

expected from changes mainly to the data assimilation system and the model initialization procedure. In addition to a reduced temperature standard deviation, zonal wind standard deviation and biases of both fields are significantly reduced between the surface and 10 hPa in the short- to medium-range FebMar forecasts (Fig. 15). Similar to the verification against radiosonde observations, the impact of the changes in the new system are somewhat smaller in forecasts at longer lead times, but still generally positive.

c. *Verifications against GPS-RO observations*

The analyses from the FebMar period were compared against GPS-RO refractivity observations to provide an alternative measure of the accuracy of the mean temperature, which is largely determined by the satellite radiance bias correction procedure (Fig. 16). Even though these refractivity observations were assimilated in both systems, because of their relatively small number compared with satellite radiance observations, it is primarily the radiances that determine the mean temperature (and, therefore, also refractivity) in the analyses. Up to 30 km (around 10 hPa), the analysis bias is in much better agreement with the GPS-RO observations for the 4DEnVar-based system than with the 4DVar-based system. This is consistent with the new bias correction procedure giving GPS-RO observations a more significant role in determining the correction of satellite radiances, and consequently the mean analysis temperature. Above this level, the biases in the two systems increase in amplitude, but with opposite sign. It should be recalled that AMSU-A

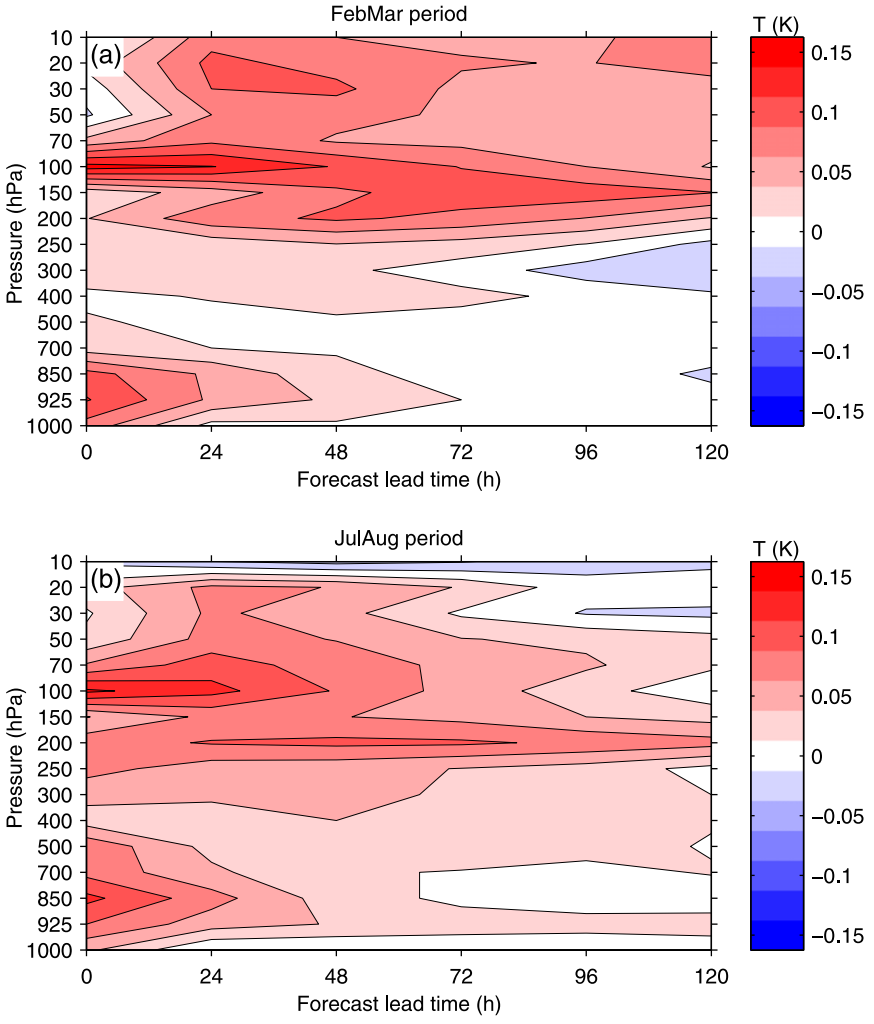

FIG. 14. Differences in temperature std dev of the forecasts from both systems relative to ERA-Interim as a function of pressure and forecast lead time, such that a positive difference (red shading) corresponds with a smaller std dev for the 4DEnVar-based system than for the 4DVar-based system. The results are shown averaged over the entire global domain during (a) February–March and (b) July–August 2011.

channels 13 and 14 (with maximum sensitivity to temperature at around 6 and 2 hPa, respectively) in the 4DEnVar-based system and channels 11–14 in the 4DVar-based system have static bias model coefficients. Therefore, the increased bias in GPS-RO refractivity in the new system above 10 hPa is likely related to how the static bias model coefficients used in the experiment were computed.

d. Verifications against surface observations

Near-surface temperature forecasts from the 4DEnVar-based system are significantly improved at most lead times compared to those of the 4DVar-based system when evaluated against in situ observations taken approximately 1.5 m above the surface from North America during the FebMar period (Fig. 17). The forecast impact varies with lead time with the smallest improvements near solar noon (near 1800 UTC). The reduction in standard deviation is more statistically significant and slightly larger during the first day (Fig. 17a). The temperature bias (Fig. 17c) has a similar diurnal variation to the standard deviation, with a significant reduction in the bias amplitude except for valid times around solar noon. The 50% reduction in the cold bias over the first night of the forecasts initialized at 0000 UTC, accompanied by a 0.5-K reduction in standard deviation, implies that the reduction of spinup in the system has a strong positive impact on short-range

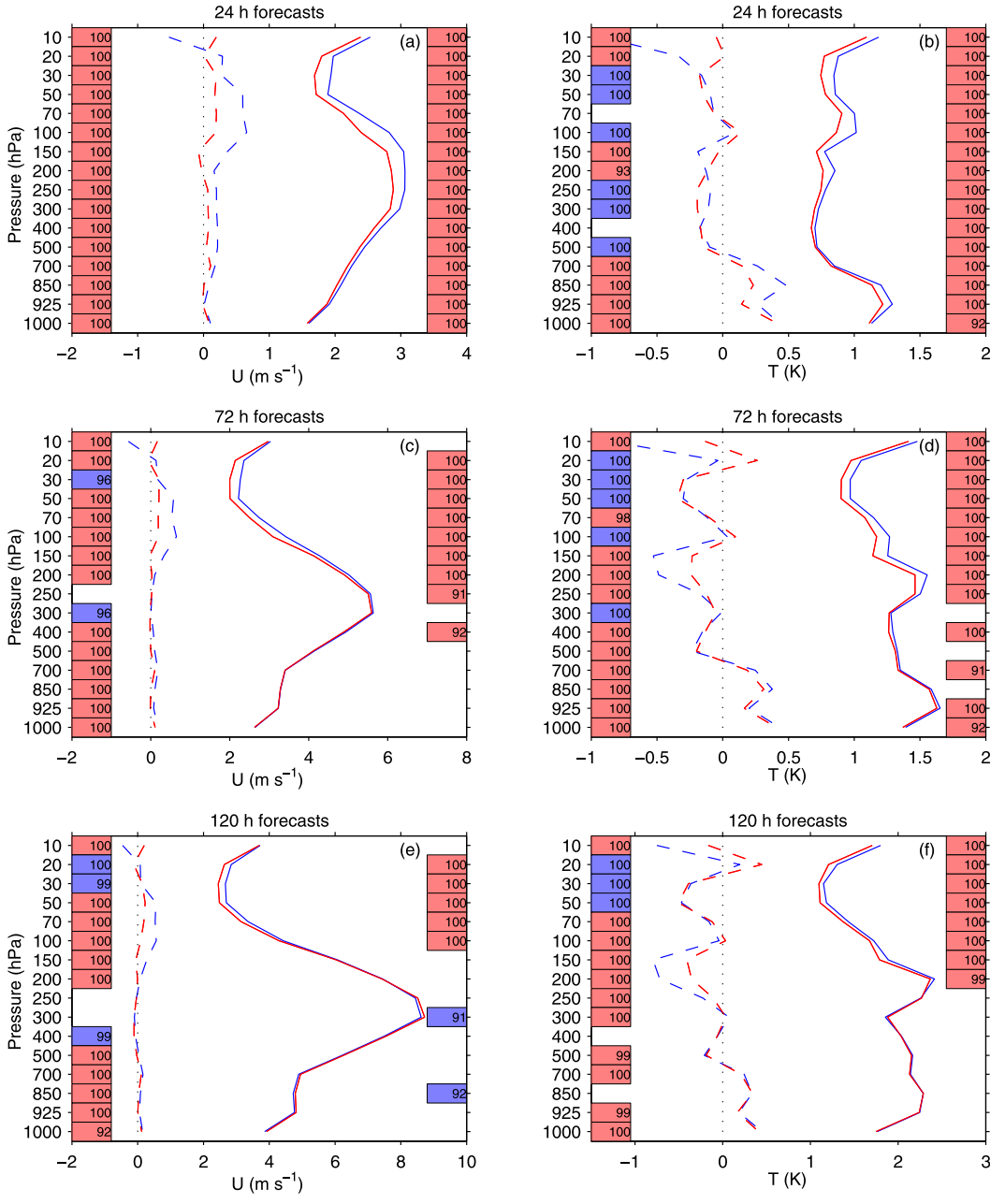

FIG. 15. The bias (dashed curves) and std dev (solid curves) relative to ERA-Interim, similar to Fig. 3, for forecasts from the 4DEnVar- (red curves) and 4DVar-based (blue curves) systems. Results from the (a),(b) 24-; (c),(d) 72-; and (e),(f) 120-h forecasts for (left) zonal wind and (right) temperature are shown for the global domain during February–March 2011.

near-surface predictions. At longer lead times, the improved standard deviation of temperature is primarily due to the correction related to snow density (cf. Figs. 4a and 17a). Dewpoint forecasts, however, are affected by the reduction of spinup in both the first overnight period and at longer range (Figs. 17b,d). The latter is consistent with improvements in the estimate of soil properties (particularly soil moisture), which exert a strong influence on near-surface quantities and have some of the longest memory of any field in the assimilation system (Betts et al. 1996). Similar improvements over the JulAug period, most notably a significant reduction in dewpoint standard deviation over the first three days of integration, further demonstrate the positive impact of

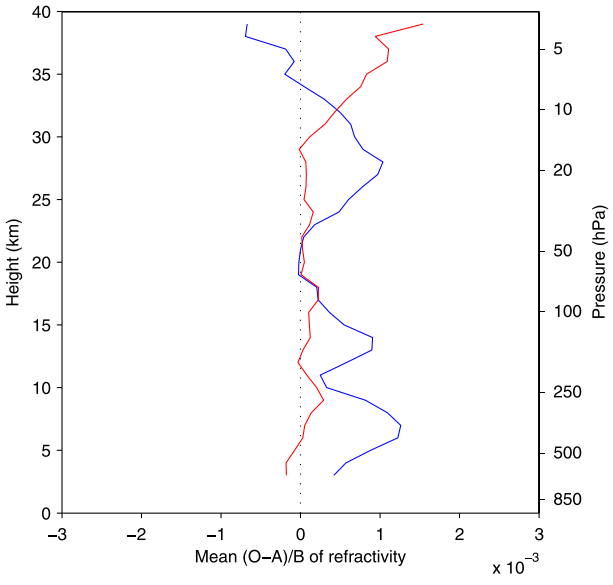

FIG. 16. The globally averaged difference between the observed GPS-RO refractivity and the same quantity computed from the analyses of the 4DEnVar- (red curves) and the 4DVar-based (blue curves) systems. The differences are normalized by the refractivity computed from the background state. Results are shown for February–March 2011.

spinup reduction on near-surface quantities within the system.

e. *Assessment of changes in model spinup*

The introduction of analysis increments using 4DIAU and the selective recycling of key components of the physical state were shown to reduce model spinup in section 4. A similar comparison shows that the 4DEnVar-based system predicts a larger cloud fraction than the 4DVar-based system throughout the 10-day forecast integration, with spinup over the first day reduced (Fig. 18a). A gradual increase in cloud cover in the 4DIAU-based system results from a change in the treatment of humidity in the assimilation cycle as noted in section 4c. The nearly complete elimination of the total precipitation spinup in the 4DEnVar-based system is remarkable, as the relatively stable M-Climate rain rate is approached only after 7–8 days of integration in the 4DVar-based system (Fig. 18b). Differences in the spinup of precipitation shown in Figs. 9b and 18b are a by-product of an unexpected sensitivity in the tropical moisture profile to the model grid. The surplus of moisture at the top of the tropical boundary layer in the 0.15° Yin–Yang configuration that led to excessive convective activity during the increment period is absent from the lower-resolution global grid configuration that was finally adopted in the 4DEnVar-based system.

6. Conclusions and discussion

The updated version of the GDPS provides improved forecast accuracy relative to the previous system, based on results from two sets of two-month data assimilation and forecast experiments. The improvements are largest at shorter lead times, but significant improvements are maintained in the 120-h forecasts for most regions and vertical levels. The improvements were shown to result from the combined impact of numerous changes, notably, the replacement of 4DVar by 4DEnVar, an improved treatment of radiosonde and aircraft observations, an improved radiance bias correction procedure, assimilation of new radiance and GB-GPS observations, and an improved approach for initializing model forecasts. Because of the replacement of 4DVar with 4DEnVar, the new system is more computationally efficient and easier to parallelize, facilitating a doubling of the analysis increment horizontal resolution that would have been prohibitively expensive with the 4DVar-based system. The replacement of the full-field digital filter with 4DIAU, and the recycling of several key variables that are not directly analyzed have significantly reduced the model spinup both during the data assimilation cycle and in medium-range forecasts.

The new 4DEnVar-based system was tested in a real-time configuration in parallel with the operational 4DVar-based system during the summer and early fall of 2014. Compared with the configuration used for the 2011 tests presented in this study, several changes were made to the system, including the assimilation of SSMIS data from *DMSP-17* and *-18* satellites (only data from *DMSP-16* were used previously), and the assimilation of several types of data from the *MetOp-B* satellite: AMSU-A, MHS, IASI, Advanced Scatterometer (ASCAT), and GPS-RO. Based on the results obtained during this period, generally consistent with those from the 2011 tests, the decision was made to implement the complete 4DEnVar-based system on 18 November 2014 for operational global deterministic prediction at the same time as the new 4DEnVar-based system for the RDPS, as described by Caron et al. (2015).

Future improvements to the 4DEnVar approach are expected to be achieved through improvements to the 4D ensemble background-error covariances from the EnKF. In the current system, an equal weighting of the ensemble and climatological covariances is used from the surface up to about 40. As shown by Lorenc et al. (2015), reliance on the climatological covariances represents a significant limitation of 4DEnVar when compared with 4DVar. Ideally, the covariances in 4DEnVar could be completely determined by 4D

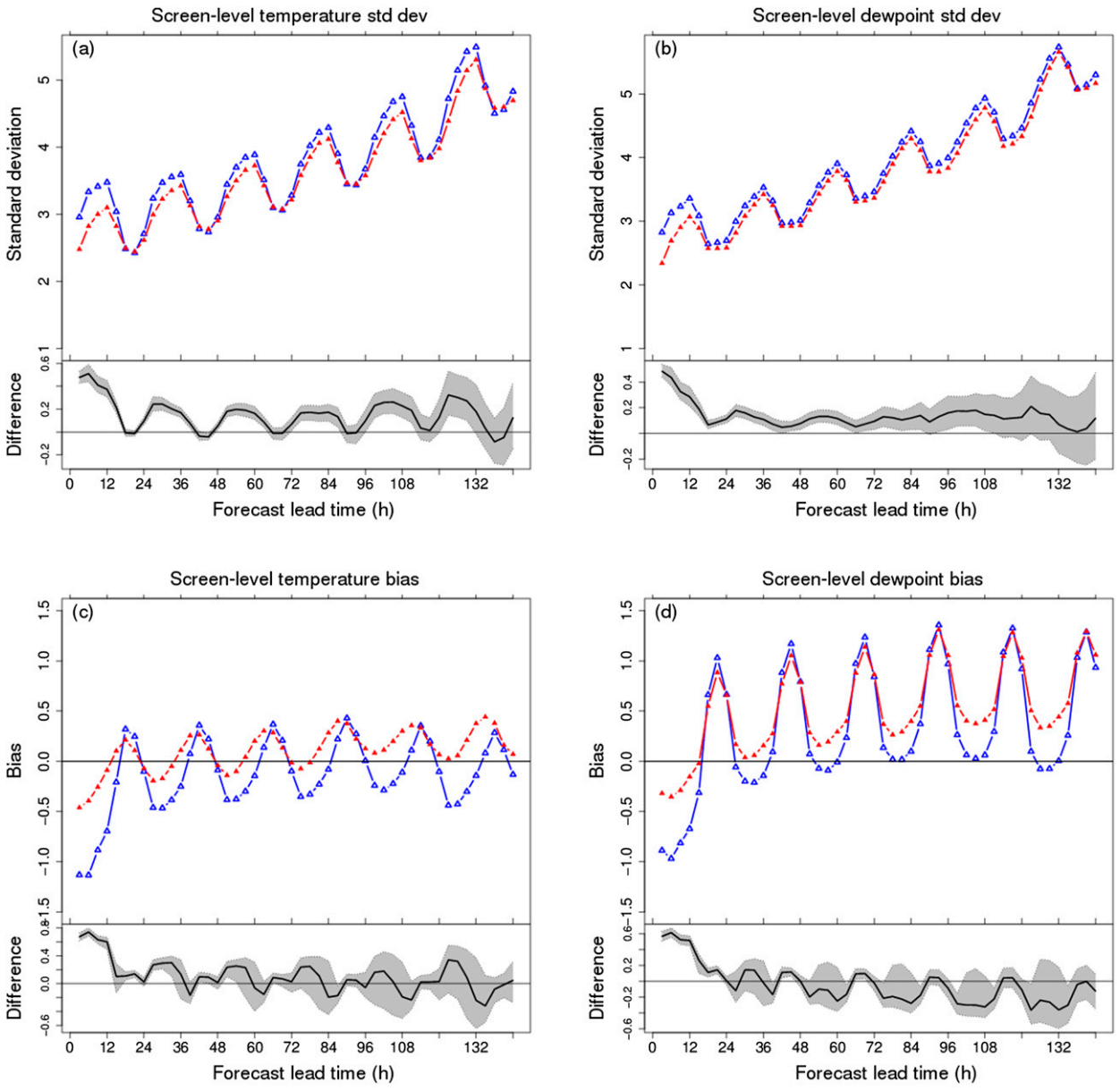

FIG. 17. As in Fig. 4, but for the 4DEnVar- (red) and 4DVar-based (blue) systems.

flow-dependent ensembles from the EnKF. However, experiments at EC with increased weight for the ensemble covariances (and a corresponding reduced weight for the climatological covariances) have yet to result in forecast improvements (Buehner et al. 2013). This suggests that some aspect of the ensemble covariances produced by the current EnKF, or the way in which they are used in 4DEnVar, requires improvement before a benefit can be obtained from reducing the reliance on the climatological covariances.

Additional future improvements to the GDPS will be realized by improving the forecast model. In all previous versions of the GDPS, the short-term forecasts performed in the data assimilation cycle were in a constant state of severe spinup. This was due to the repeated application of the full-field digital filter and the lack of important information transferred from one forecast to the next, especially related to the clouds and turbulence parameterizations. The large reduction in model spinup in the new system should make the model's short-range

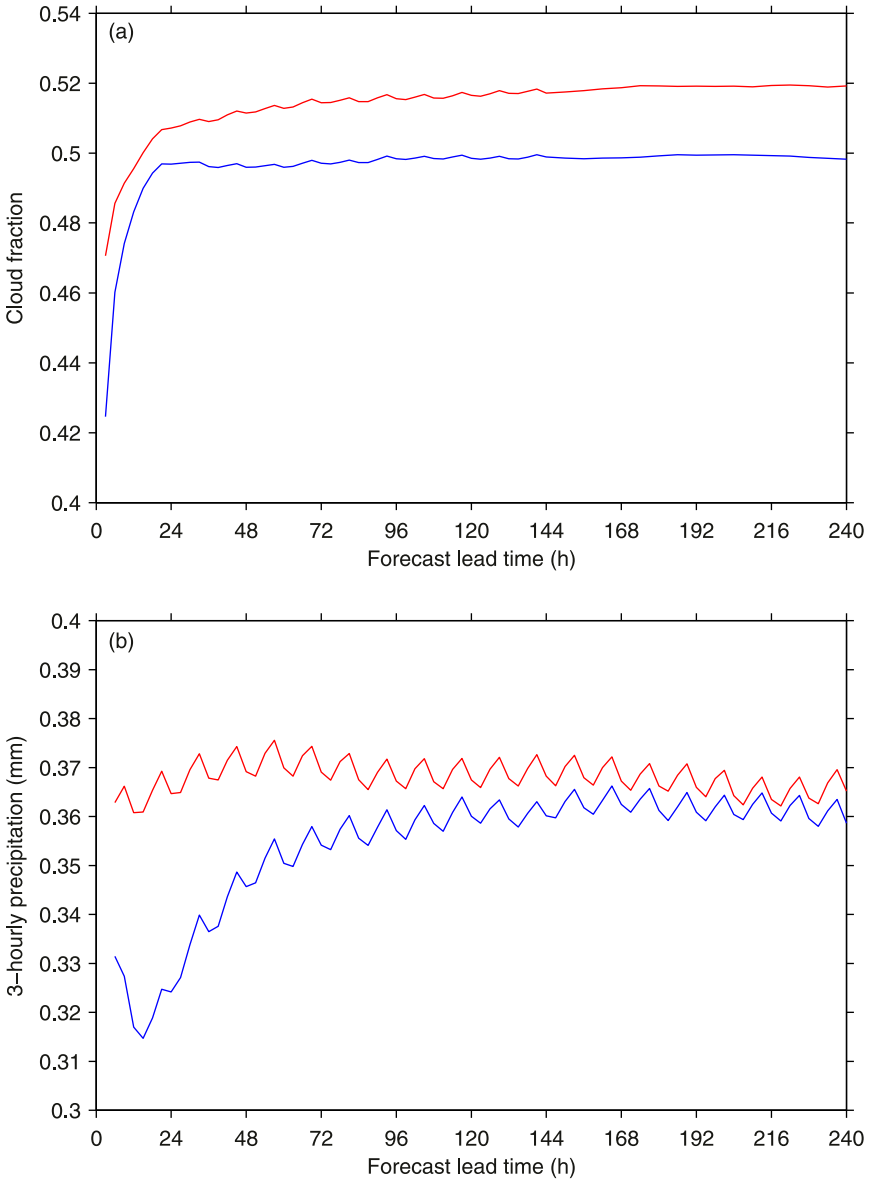

FIG. 18. As in Fig. 9, but for the 4DVar- (blue) and 4DEnVar-based (red) systems.

response to modifications more predictable and consistent with the model climatology.

Acknowledgments. The authors thank the large number of people in the research, development, and operations divisions at the Canadian Meteorological Centre that made this project possible, starting with Martin Charron for initiating the project. Thanks to Yves Chartier, Michel Van Eeckhout, Reine Parent, Vincent Vu, and Martin Charron for performing the coupled GPDS–GEPS experiments. Thanks also to Dominic Racette and Sua Lim for help with the Maestro sequencer. Several others are thanked for providing significant contributions: Jean-François Caron, Stéphane Chamberland, Michel Desgagné, Peter Houtekamer, Pierre Koclas, Manon Lajoie, Vivian Lee, Alain Patoine, Michel Roch, Judy St. James, Paul Vaillancourt, and Marcel Vallée. Finally, the authors thank three reviewers whose comments helped to improve an earlier version of the paper.

REFERENCES

Arakawa, A., 1988: Finite-difference methods in climate modeling. *Physically-Based Modeling and Simulation of Climate and Climate Change,* Vol. I, M. Schlesinger, Ed., D. Reidel, 79–168.

Bélair, S., R. Brown, J. Mailhot, B. Bilodeau, and L.-P. Crevier, 2003: Operational implementation of the ISBA land surface scheme in the Canadian regional weather forecast model. Part II: Cold season results. *J. Hydrometeor.,* **4,** 371–386, doi:10.1175/1525-7541(2003)4<371:OIOTIL>2.0.CO;2.

——, J. Mailhot, C. Girard, and P. Vaillancourt, 2005: Boundary layer and shallow cumulus clouds in a medium-range forecast of a large-scale weather system. *Mon. Wea. Rev.,* **133,** 1938–1960, doi:10.1175/MWR2958.1.

Betts, A. K., J. H. Ball, A. C. M. Beljaars, M. J. Miller, and P. A. Viterbo, 1996: The land surface–atmosphere interaction: A review based on observational and global modeling perspectives. *J. Geophys. Res.,* **101,** 7209–7225, doi:10.1029/95JD02135.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.,* **124,** 1256–1271, doi:10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2.

Buehner, M., J. Morneau, and C. Charette, 2013: Four-dimensional ensemble–variational data assimilation for global deterministic weather prediction. *Nonlinear Processes Geophys.,* **20,** 669–682, doi:10.5194/npg-20-669-2013.

Carmagnola, C. M., and Coauthors, 2013: Snow spectral albedo at Summit, Greenland: Measurements and numerical simulations based on physical and chemical properties of the snowpack. *The Cryosphere,* **7,** 1139–1160, doi:10.5194/tc-7-1139-2013.

Caron, J.-F., T. Milewski, M. Buehner, L. Fillion, M. Reszka, S. Macpherson, and J. St-James, 2015: Implementation of deterministic weather forecasting systems based on ensemble–variational data assimilation at Environment Canada. Part II: The regional system. *Mon. Wea. Rev.,* **143,** 2560–2580, doi:10.1175/MWR-D-14-00353.1.

Charney, J., 1955: The use of the primitive equations of motion in numerical prediction. *Tellus,* **7A,** 22–26, doi:10.1111/j.2153-3490.1955.tb01138.x.

Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Toward random sampling of model error in the Canadian Ensemble Prediction System. *Mon. Wea. Rev.,* **138,** 1877–1901, doi:10.1175/2009MWR3187.1.

——, and Coauthors, 2012: The stratospheric extension of the Canadian Global Deterministic Medium-Range Weather Forecasting System and its impact on tropospheric forecasts. *Mon. Wea. Rev.,* **140,** 1924–1944, doi:10.1175/MWR-D-11-00097.1.

Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.,* **139,** 1445–1461, doi:10.1002/qj.2054.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998a: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.,* **126,** 1373–1395, doi:10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2.

——, J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998b: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part II: Results. *Mon. Wea. Rev.,* **126,** 1397–1418, doi:10.1175/1520-0493(1998)126<1397:TOCMGE>2.0.CO;2.

Dai, A., and J. Wang, 1999: Diurnal and semidiurnal tides in global surface pressure fields. *J. Atmos. Sci.,* **56,** 3874–3891, doi:10.1175/1520-0469(1999)056<3874:DASTIG>2.0.CO;2.

Daley, R., 1981: Normal mode initialisation. *Rev. Geophys. Space Phys.,* **19,** 450–468, doi:10.1029/RG019i003p00450.

Dee, D. P., and S. Uppala, 2009: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Quart. J. Roy. Meteor. Soc.,* **135,** 1830–1841, doi:10.1002/qj.493.

——, and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.,* **137,** 553–597, doi:10.1002/qj.828.

Derber, J. C., and W.-S. Wu, 1998: The Use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.,* **126,** 2287–2299, doi:10.1175/1520-0493(1998)126<2287:TUOTCC>2.0.CO;2.

——, and F. Bouttier, 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus,* **51A,** 195–221, doi:10.1034/j.1600-0870.1999.t01-2-00003.x.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.,* **131,** 3385–3396, doi:10.1256/qj.05.108.

Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap.* Chapman & Hall/CRC, 436 pp.

Fillion, L., H. L. Mitchell, H. Ritchie, and A. Staniforth, 1995: The impact of a digital filter finalization technique in a Global Data Assimilation System. *Tellus,* **47A,** 304–323, doi:10.1034/j.1600-0870.1995.t01-2-00002.x.

Garand, L., O. Pancrati, and S. Heillette, 2011: Validation of forecast cloud parameters from multispectral AIRS radiances. *Atmos.–Ocean,* **49,** 121–137, doi:10.1080/07055900.2011.567379.

——, M. Buehner, S. Heilliette, S. R. Macpherson, and A. Beaulne, 2013: Satellite radiance assimilation impact in new Canadian ensemble–variational system. *Proc. 2013 EUMETSAT Meteorological Satellite Conf.,* Vienna, Austria, EUMETSAT, Session 12, Paper 4.

Gauthier, P., M. Buehner, and L. Fillion, 1999: Background-error statistics modelling in a 3D variational data assimilation scheme: Estimation and impact on the analyses. *Proc. ECMWF Workshop on Diagnosis of Data Assimilation Systems,* Reading, United Kingdom, ECMWF, 131–145.

——, M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau, 2007: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Wea. Rev.,* **135,** 2339–2354, doi:10.1175/MWR3394.1.

Girard, C., and Coauthors, 2014: Staggered vertical discretization of the Canadian Environmental Multiscale (GEM) model using a coordinate of the log-hydrostatic-pressure type. *Mon. Wea. Rev.,* **142,** 1183–1196, doi:10.1175/MWR-D-13-00255.1.

Graham, R., T. Alcott, N. Hosenfeld, and R. Grumm, 2013: Anticipating a rare event utilizing forecast anomalies and a situational awareness display: The western U.S. storms of 18–23 January 2010. *Bull. Amer. Meteor. Soc.,* **94,** 1827–1836, doi:10.1175/BAMS-D-11-00181.1.

Gray, M. E. B., 2001: The impact of mesoscale convective-system potential-vorticity anomalies on numerical-weather-prediction forecasts. *Quart. J. Roy. Meteor. Soc.,* **127,** 73–88, doi:10.1002/qj.49712757105.

Houtekamer, P. L., X. Deng, H. L. Mitchell, S.-J. Baek, and N. Gagnon, 2014: Higher resolution in an operational ensemble Kalman filter. *Mon. Wea. Rev.,* **142,** 1143–1162, doi:10.1175/MWR-D-13-00138.1.

Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP

GFS. Part I: System description and 3D-hybrid results. *Mon. Wea. Rev.,* **143,** 433–451, doi:10.1175/MWR-D-13-00351.1.

Laroche, S., and R. Sarrazin, 2013: Impact of radiosonde balloon drift on numerical weather prediction and verification. *Wea. Forecasting,* **28,** 772–782, doi:10.1175/WAF-D-12-00114.1.

Liu, Q., F. Weng, and S. English, 2011: An improved fast microwave water emissivity model. *IEEE Trans. Geosci. Remote Sens.,* **49,** 1238–1250, doi:10.1109/TGRS.2010.2064779.

Lorenc, A. C., 2013: Recommended nomenclature for EnVar data assimilation methods. WGNE 2013 Blue Book 1-07, 2 pp.

——, N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, 2015: Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Wea. Rev.,* **143,** 212–229, doi:10.1175/MWR-D-14-00195.1.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus,* **21,** 289–307, doi:10.1111/j.2153-3490.1969.tb00444.x.

Lynch, P., and X.-Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.,* **120,** 1019–1034, doi:10.1175/1520-0493(1992)120<1019:IOTHMU>2.0.CO;2.

Macpherson, S. R., G. Deblonde, J. M. Aparicio, and B. Casati, 2008: Impact of NOAA ground-based GPS observations on the Canadian Regional Analysis and Forecast System. *Mon. Wea. Rev.,* **136,** 2727–2746, doi:10.1175/2007MWR2263.1.

Mailhot, J., and R. Benoit, 1982: A finite-element model of the atmospheric boundary layer suited for use with numerical weather prediction models. *J. Atmos. Sci.,* **39,** 2249–2266, doi:10.1175/1520-0469(1982)039<2249:AFEMOT>2.0.CO;2.

Matricardi, M., F. Chevallier, G. Kelly, and J.-N. Thépaut, 2004: An improved general fast radiative transfer model for the assimilation of radiance observations. *Quart. J. Roy. Meteor. Soc.,* **130,** 153–173, doi:10.1256/qj.02.181.

McTaggart-Cowan, R., and A. Zadra, 2015: Representing Richardson number hysteresis in the NWP boundary layer. *Mon. Wea. Rev.,* **143,** 1232–1258, doi:10.1175/MWR-D-14-00179.1.

Paquin-Ricard, D., C. Jones, and P. A. Vaillancourt, 2010: Using ARM observations to evaluate cloud and clear-sky radiation processes as simulated by the Canadian regional climate model GEM. *Mon. Wea. Rev.,* **138,** 818–838, doi:10.1175/2009MWR2745.1.

Polavarapu, S., S. Ren, A. Clayton, D. Sankey, and Y. Rochon, 2004: On the relationship between incremental analysis updating and incremental digital filtering. *Mon. Wea. Rev.,* **132,** 2495–2502, doi:10.1175/1520-0493(2004)132<2495:OTRBIA>2.0.CO;2.

Qaddouri, A., and V. Lee, 2011: The Canadian Global Environmental Multiscale model on the Yin–Yang grid system. *Quart. J. Roy. Meteor. Soc.,* **137,** 1913–1926, doi:10.1002/qj.873.

Rabier, F., 2005: Overview of global data assimilation developments in numerical weather-prediction centres. *Quart. J. Roy. Meteor. Soc.,* **131,** 3215–3233, doi:10.1256/qj.05.129.

——, H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.,* **126,** 1143–1170, doi:10.1002/qj.49712656415.

Rawlins, F., S. Ballard, K. Bovis, A. Clayton, D. Li, G. Inverarity, A. Lorenc, and T. Payne, 2007: The Met Office four-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.,* **133,** 347–362, doi:10.1002/qj.32.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate,* **24,** 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Sankey, D., S. Ren, S. Polavarapu, Y. J. Rochon, Y. Nezlin, and S. Beagley, 2007: Impact of data assimilation filtering methods on the mesosphere. *J. Geophys. Res.,* **112,** D24104, doi:10.1029/2007JD008885.

Seaman, R., W. Bourke, P. Steinle, T. Hart, G. Embery, M. Naughton, and L. Rikus, 1995: Evolution of the Bureau of Meteorology's Global Assimilation and Prediction system. Part I: Analysis and initialisation. *Aust. Meteor. Mag.,* **44,** 1–18.

Smith, G. C., and Coauthors, 2015: Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System. *Quart. J. Roy. Meteor. Soc.,* doi:10.1002/qj.2555, in press.

Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology.* Kluwer Academic Publishers, 670 pp.

Sundqvist, H., E. Berge, and J. E. Kristjánsson, 1989: Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Mon. Wea. Rev.,* **117,** 1641–1657, doi:10.1175/1520-0493(1989)117<1641:CACPSW>2.0.CO;2.

Waite, M. L., and C. Snyder, 2013: Mesoscale energy spectra of moist baroclinic waves. *J. Atmos. Sci.,* **70,** 1242–1256, doi:10.1175/JAS-D-11-0347.1.

Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.,* **141,** 4098–4117, doi:10.1175/MWR-D-12-00141.1.

Zadra, A., R. McTaggart-Cowan, P. A. Vaillancourt, M. Roch, S. Bélair, and A.-M. Leduc, 2014a: Evaluation of tropical cyclones in the Canadian Global Modeling System: Sensitivity to moist process parameterization. *Mon. Wea. Rev.,* **142,** 1197–1220, doi:10.1175/MWR-D-13-00124.1.

——, and Coauthors, 2014b: Improvements to the Global Deterministic Prediction system (GDPS) (from version 2.2.2 to 3.0.0), and related changes to the Regional Deterministic Prediction System (RDPS) (from version 3.0.0 to 3.1.0). Canadian Meteorological Centre Tech. Note, 88 pp. [Available online at http://collaboration.cmc.ec.gc.ca/cmc/CMOI/product_guide/docs/lib/op_systems/doc_opchanges/technote_gdps300_20130213_e.pdf.]

Zhang, F., N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.,* **64,** 3579–3594, doi:10.1175/JAS4028.1.