Four-Dimensional Variational Data Assimilation for the Canadian Regional Deterministic Prediction System

MONIQUE TANGUAY, LUC FILLION, AND ERVIG LAPALME
Meteorological Research Division, Environment Canada, Dorval, Québec, Canada

MANON LAJOIE
Canadian Meteorological Centre, Environment Canada, Dorval, Québec, Canada

(Manuscript received 29 June 2011, in final form 10 November 2011)

ABSTRACT

As a second step in the development of the Canadian Regional Data Assimilation System following Fillion et al., this study extends the approach to the four-dimensional variational data assimilation (4D-Var) context. Emphasis is first put on illustrating the importance of controlling lateral boundary conditions (LBCs). The use in the minimization of a horizontal grid over a domain exceeding the horizontal grid of the high-resolution nonlinear model is then proposed. The authors examine the performance of this 4D-Var formulation as an upcoming upgrade to the currently operational regional three-dimensional variational data assimilation (3D-Var) system. Forecast verifications against radiosonde data for 118 winter cases and 118 summer cases were performed. Results indicate a slight positive impact up to 48 h against North American radiosondes, but with a significant positive impact (especially for winds) at mid- and high latitudes during the summer. Accumulated precipitation scores over 24 h, whether during the first or second day of the forecasts, are slightly improved. The regional 4D-Var analysis system described in this study can run within current real-time “regional run” allocation for operations at the Canadian Meteorological Center (CMC). Future improvements of this system are briefly mentioned especially regarding the upcoming computer upgrade at CMC.

1. Introduction

The Canadian Meteorological Center’s (CMC) operational regional data assimilation and forecasting system was upgraded on 20 October 2010. The previous global variable-resolution forecasting approach was replaced by a limited-area nested forecasting approach. Both systems are based on the Global Environmental Multi-scale (GEM) model (Côté et al. 1998), which can be run either in a global uniform, a limited-area, or a global variable-grid configuration. Under the constraints of the same data assimilated, similar background-error statistics modeling assumptions and equal computer resources, Fillion et al. (2010, hereafter referred to as F10) demonstrated that the new limited-area regional data assimilation and forecasting system (hereafter REG-3D) performs as well as the former global variable-resolution system.

The four-dimensional variational data assimilation (4D-Var) approach is a temporal extension of the three-dimensional variational data assimilation (3D-Var) approach by including the model integration as part of the observation operator. We measure the distance between the analyzed state and distributed observations at their appropriate time, instead of using the static analysis as in 3D-Var. This requires the tangent-linear (TL) and adjoint (AD) of the forecast model. For earlier reviews of 4D-Var, see Le Dimet and Talagrand (1986) or Courtier et al. (1994). During the last few years, 4D-Var has been implemented at many centers in a global context (e.g., Rabier et al. 2000; Gauthier et al. 2007; Rawlins et al. 2007) and in limited-area systems (e.g., Huang et al. 2002; Wlasak et al. 2004; Honda et al. 2005; Huang et al. 2009). In this paper, we consider the extension of REG-3D to 4D-Var (hereafter REG-4D).

Limited-area models are more economical than global models for a given horizontal resolution. They can produce more accurate regional forecasts through the use of a finer resolution exclusively for the region of primary interest without tremendously increasing computer
time. The expense we pay for these advantages is the need for a careful formulation of the lateral boundary conditions (LBCs). The specification of model variables at open boundaries is required for numerical integrations. Seminal theoretical reviews of how to prescribe appropriate LBCs are available in Sundström and Elvius (1979) and Arakawa (1984). Those studies discuss the influence of errors generated at or transmitted through the boundaries, the choice of mathematical and numerical boundary conditions, the stability problems, and the effects of overspecification. It is shown how rapidly the slope of the characteristics spreads the influence of the initial and boundary data through the domain. At inflow, boundary conditions should be prescribed whereas at outflow, the solution is determined entirely by the interior solution being advected out through the boundary. If we prescribe a boundary condition at these points, we do not get a well-posed problem (Kreiss and Oliger 1973; Oliger and Sundström 1978). Limited-area models that are mathematically ill posed must adopt various types of diffusion and/or use a sponge layer near the boundaries to artificially reduce wave reflections (see e.g., Davies 1976). Furthermore, LBCs are usually provided by solutions from a model with a larger domain, with much coarser resolution and simpler physical parameterizations. Those inconsistencies represent sources of forecasting errors. Those errors may propagate into the interior domain and hence contaminate model solutions. Warner et al. (1997) proposes some guidelines for helping to minimize the negative impact of imperfect LBCs.

Limited-area prediction constitutes a mixed initial boundary value problem and imposes additional complexity on 4D-Var data assimilation. A careful study of the impact of initial and lateral boundary conditions using an adjoint operator approach was performed by Errico et al. (1993) and underlined the strong dependence on winds. Specifically, the treatment of boundaries in the forward model may induce difficulties when we consider the associated adjoint model. Regions of physical forward inflow become regions of gradient information outflow during the adjoint model integration. The adjoint solution without lateral boundary control is therefore entirely prescribed at its outflow boundary and this is known to be detrimental. This may lead for instance to gravity wave noise (Gustafsson et al. 1998). In a 4D-Var context, Zou and Kuo (1996) made a first attempt to determine a boundary forcing along with the interior solution. The trend at the boundary was optimized and this had a major impact on the quality of retrieved fields. Similar control of LBCs was adopted in many 4D-Var systems over a limited-area domain as will be discussed further in the following sections.

Lu and Browning (2000) corrected the negative impact linked to observational errors at the outflow boundary by incorporating it as part of the control variables. Further results on the problem of limited-area data assimilation are collected in Park and Županski (2003).

In the context of REG-4D, we propose to use a TL/AD grid beyond the initial forecasting area. In section 2, the impact of various options to control LBCs is reviewed using identical twin assimilation experiments. In section 3, we compare 4D-Var assimilations of a single observation: one with a global grid and one with a limited-area domain embedded in the global grid, in order to carefully determine the size of the extended domain. In section 4, we compare this REG-4D strategy against REG-3D. A summary is given in section 5.

2. Identical twin assimilation experiments

The GEM model used in all experiments employs a two time step implicit, semi-Lagrangian scheme on a latitude–longitude Arakawa C-grid staggering and an unstaggered sigma-pressure hybrid vertical grid. A complete description is available in Côté et al. (1998). The specification of LBCs for the limited-area configuration of the GEM model (hereafter GEM-LAM) is based on Thomas et al. (1998). The nature of Arakawa C-grid discretization of the basic equations combined with the implicit formulation requires that only normal wind components are to be specified in order to properly close the mathematical problem. Using a 1D schematic model, the definition of LBCs, driving zone and core zone as used in GEM are given in the appendix. It is also stressed that the semi-Lagrangian aspect brings further considerations on the driving zone since the fields within that zone (not only at the border of the core zone) influence the circulation inside the core zone. In addition, a blending zone (based on Davies 1976) is incorporated in order to relax the forced boundary specifications to the actual model inner solution.

In this section, we present identical twin assimilation experiments illustrating the importance of controlling LBCs in a 4D-Var environment. GEM-LAM is run here in a configuration where the horizontal grid mesh is an exact subgrid of a global uniform latitude–longitude grid of the driving model (denoted GEM-GLOBAL). Thus, both models have the same horizontal resolution; that is, 200 km over North America (grid $200 \times 100$ GLOBAL; $54 \times 54$ LAM) and both have the same 58 vertical levels with a pressure top at 10 hPa. Both models use the same model time step: $\Delta t = 45$ min. The frequency input from the driving model to the limited-area model is done at every time step and the driving zone covers seven grid
points. No blending is used here. We are therefore close to an “acid test” setting (as defined in Staniforth 1997) where the solution obtained over a limited area should well match that of an equivalent-resolution model integrated over a much larger domain. The experiment setup is chosen specifically to minimize the contamination from inappropriate LBCs.

As is typically the case for such twin assimilation experiments (e.g., Thépaut and Courtier 1991), the cost function is based on the energy norm. It is therefore simpler than the one in the REG-4D analysis system since no background term appears in the functional. Another difference with REG-4D is that, without loss of generality, we do not consider the incremental formulation. So the cost function measures the difference between the full fields of the nonlinear model and the observations, not the difference between the TL increments and the innovations (i.e., the discrepancy between the observations and the trajectory). The cost function is defined over the full limited-area domain (i.e., it covers the core zone plus the driving zone; this defines the “analysis area”). The “observations” are extracted from a GEM-LAM reference integration (denoted REF) and are available at each time step over the 6-h assimilation period. The initial conditions of REF are the same as the ones used in REF (i.e., in this experiment case C experiments. Finally, the performance of the extended TL/AD grid chosen in the REG-4D analysis system is covered in case D.

a. Case A: No control of LBCs + right driver

In the first experiment henceforth referred to as case A, we minimize the cost function without control of LBCs. The latter are kept fixed over the driving zone and are identical to the ones used in REF (i.e., in this experiment, no errors are introduced in the driving zone). To simulate a realistic level of error structure and amplitudes, the minimization is initialized with fields over the core zone coming from the truth (i.e., REF fields), but valid at \( t = t_1 \) (where \( t_1 = t_0 + 6 \) h is the end of the assimilation period). The descent of the cost function is illustrated in Fig. 1 (solid line) where we observe a reduction of at least five orders of magnitude over the course of the minimization. Figure 2 (top panels) presents errors in temperature with respect to REF at model level close to 250 hPa over the full domain at \( t = t_0 \) (left) and \( t = t_1 \) (right) coming from the first iteration of the minimization. The wind vectors of REF are superimposed and may be used to locate the error structure’s inflow and outflow regions. Figure 2 (bottom panels) is equivalent to Fig. 2 (top panels), but from the last iteration obtained through the minimization process.

It is seen that the maximum errors in temperature are reduced by three orders of magnitude and the minimization has no problem recovering the correct initial state over the core zone.

b. Case B: No control of LBCs + wrong driver

In experiment case B, we start the minimization with the same initial conditions over the core zone as in case A. To simulate errors in the driving conditions, the LBCs at each driving time step are set equal to the LBCs used by REF at \( t = t_1 \). Results of the minimization are given in Fig. 3 (same as Fig. 2, but for case B). At \( t = t_0 \), errors in temperature at LBCs are the same for the first and last iterations of the minimization since the LBCs do not belong to the control vector of the minimization. These errors are clearly visible over the driving zone of the last iteration (the contour interval is 10 times smaller than the one used in the first iteration plot). At \( t = t_0 \) and \( t = t_1 \), errors in temperature in the core zone are decreased by one order of magnitude, except for some areas. Errors at \( t = t_0 \) over outflow areas move out of the domain during the integration whereas larger errors in two inflow areas penetrate farther inside the core zone. The descent of the cost function is shown in Fig. 1 (dashed line) and we note a plateau after 10 simulations. We conclude that the wrong driving conditions in the inflow region are unable to supply the correct information to the core zone during the integration, and thus hinder the recovery of the REF solution.

c. Case C: Control of LBCs with various options + wrong driver

In case C experiments, we evaluate four different approaches to control LBCs. The basic control variables are, as mentioned previously, wind components, temperature, and logarithm of surface pressure at initial time \( t = t_0 \) over the core zone. All approaches examined here include these variables over the driving zone in the control vector of the minimization. However, we may vary the degrees of freedom in time, depending on the chosen option. As boundary values are correlated in
space and time, that is, they are not all independent, their amount can be reduced. Seiler (1993) suggested reducing the number of control variables by filtering high-frequency variability from the boundary values. Here, we use a Taylor series expansion in time and increase the number of driving time steps to be controlled, thus leading to more accurate time derivative estimates. Table 1 gives a description of the studied options in case C. In option 0, the control of LBCs is done at each time step. In option 1, only LBCs at \( t = t_0 \) are part of the control variables and LBCs at other time steps are prescribed using persistence. In option 2, LBCs at \( t = t_0 \) and \( t = t_1 \) are included in the control variables. The LBCs at other time steps are given using linear interpolation. This option is used in Ishikawa and Koizumi (2002) and Gustafsson (2006). It is equivalent to the one in Zou and Kuo (1996) and Zhang et al. (2010) where LBCs at \( t = t_0 \) and time tendency \( \partial(LBCs)/\partial(t) \) at \( t = t_0 \) are included in the control variables. We thus estimate the time tendency over the complete assimilation period. In option 3, LBCs at \( t = t_0, t = t_m = (t_0 + t_1)/2, \) and \( t = t_1 \) are part of the control variables. The LBCs at other time steps are given using quadratic interpolation.

In case C experiments, we start the minimization with the same initial conditions over the core zone and the same erroneous LBCs at all time steps as in case B. The errors in temperature with respect to REF over the full domain at \( t = t_0 \) and \( t = t_1 \) are thus the same as Fig. 3 (top panels). In option 0, the control of LBCs is done at each time step. At the last iteration, the errors in temperature (see Fig. 4) at \( t = t_0 \) and \( t = t_1 \) are reduced by three orders of magnitude. Large errors in the core zone at \( t = t_1 \) downstream from inflow areas are no longer visible when we allow the control of LBCs in the minimization. The descent of the cost function is also illustrated in Fig. 1 (dashed line; diamonds) and we observe a reduction in the functional of at least five orders of magnitude. We note that case C (option 0) exhibits at least initially, a slower rate of convergence compared to case A, which has no control of LBCs. This is to be expected when we increase the size of the control vector in the minimization. The convergence with the other options is presented as well [dashed lines; square (option 1), cross (option 2), triangle (option 3)]. All cases are characterized by monotonic descent and the degree of convergence increases with the implicit order of accuracy of the temporal discretization of their respective Taylor series expansion, as described in Table 1.

**d. Case D: Extended TL/AD grid + wrong driver**

Results presented so far show the importance of controlling LBCs during 4D-Var. However, none of the proposed options described previously was adopted stand alone within the REG-4D data assimilation system.
Although feasible, they necessitate a careful design of background-error covariances when implementing these LBCs as extra control variables for the minimization. Only option 1 (persistence) could benefit from these background-error covariances over the full limited-area domain. However, option 1 does not allow a propagation of the analysis increments in the driving zone since it assumes a stationary state. We present in the following an alternative approach. We recall that the analysis area is defined as the area covered by the core zone plus the driving zone. In addition, note that we refer here to “TL/AD grid” even if we do not use an incremental formulation of the cost function. In the current context, this refers to the minimization grid, whereas in the next section, it will truly correspond to the actual TL/AD grid used in the incremental formulation of REG-4D.

In experiment case D, we minimize the cost function with an extension of the TL/AD grid beyond the analysis area and control of LBCs based on option 1. To understand the rationale behind this approach, we refer the reader to the paper by F10 where the overall operational data assimilation and modeling context is described. Note that an extensive summary will be given in the following section. We only need at this point to recall a few elements. Due to the fact that we have available global fields from the driving model outside the analysis area, together with the fact that our 3D-Var analysis REG-3D system uses a global Gaussian grid and associated
control variables, we adopt a specific 4D-Var strategy to exploit these facts. In essence, we allow a larger GEM-LAM TL/AD domain than the analysis area in order to propagate analysis increments during the 4D-Var minimization so as to match accurately the actual propagation that would occur in, for example, a global TL/AD model (same spatial resolution) over the time assimilation window (e.g., here 6 h). The design of such a grid extension cannot be done in isolation, but rather involves many other considerations like the presence of observational

### Table 1. Description of options for control of LBCs in case C.

<table>
<thead>
<tr>
<th>Case C options</th>
<th>Control LBCs</th>
<th>Taylor series expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>All $t$</td>
<td>None</td>
</tr>
<tr>
<td>1</td>
<td>$t_0$</td>
<td>$LBCs(t) = LBCs(t_0)$</td>
</tr>
<tr>
<td>2</td>
<td>$t_0, t_1$</td>
<td>$LBCs(t) = a_0(t)LBCs(t_0) + a_1(t)LBCs(t_1)$ with $a_0(t) = \frac{(t-t_1)}{(t_0-t_1)}$, $a_1(t) = \frac{(t-t_0)}{(t_1-t_0)}$</td>
</tr>
<tr>
<td>3</td>
<td>$t_0, t_m, t_2$</td>
<td>$LBCs(t) = b_0(t)LBCs(t_0) + b_1(t)LBCs(t_1) + b_2(t)LBCs(t_m)$ with $b_0(t) = \frac{(t-t_m)(t-t_1)}{(t_0-t_m)(t_0-t_1)}$, $b_1(t) = \frac{(t-t_0)(t-t_1)}{(t_m-t_0)(t_m-t_1)}$, and $b_2(t) = \frac{(t-t_0)(t-t_m)}{(t_1-t_0)(t_1-t_m)}$</td>
</tr>
</tbody>
</table>
errors, dynamical balance, minimization aspects related to the actual choice of control variables, just to mention the most obvious. Using the operational REG-3D setup, single observation and full data assimilation tests were performed to adequately define the required extension of the TL/AD grid and results are described in details in section 3b.

Here, we demonstrate the performance of our approach against cases A, B, and C examined previously. It suffices to say for the moment that the TL/AD grid (74 × 74 for identical twin experiments) overlaps the analysis grid and its total size is increased by approximately 30% in each X and Y directions. The observations are the same as in case A to C [i.e., on the full limited-area of GEM-LAM (54 × 54)]. Option 1 is adopted so the control variables of the minimization cover the core zone of GEM-LAM (74 × 74) plus LBCs at t = t0. The other time steps are obtained from a persistence approximation. The descent of the cost function is illustrated in Fig. 1 (dotted line). We observe that the approach proposed here performs better (with a reduction close to three orders of magnitude) than all previous experiments except for two cases: case A where the true driving conditions are given and case C (option 0) where control of LBCs is performed at every time step. However, the added performance of case A and case C (option 0) is for very small analysis increments well below the level of approximation where the solution is sought in practical operational implementations of 4D-Var. Typically, a factor of 2 in the reduction of the cost function and two order reductions in the norm of the gradient are a general standard for operational contexts.

3. Design of REG-4D

In this section, we first describe the particular differences between the currently operational REG-3D and the new 4D-Var Regional analysis, REG-4D. Second, we carefully examine the optimal extension of the TL/AD horizontal grid so as to accurately propagate analysis increments during the 6-h time integration of the data assimilation window. Here, a global 4D-Var analysis at the same space and time resolution and using the same analysis control variables is used as the reference. Once this validation step of REG-4D is finalized, extensive CMC-type objective evaluations are performed and is the subject of section 4.

a. REG-3D versus REG-4D

The primary objective of the REG-3D system is the production of tropospheric 48-h forecasts over the North American continent. The details of our limited-area and global analysis components were given in F10 (section 3b). We summarize here the essential aspects. As demonstrated in F10, due to the large horizontal extent of the regional domain considered (Fig. 5, domain in dark blue), a spherical-harmonics spectral representation of background-error statistics was preferred over a biFourier representation. The latter is more justified for smaller domains and is kept as an option for upcoming use in the kilometric-scale local analysis domains in support of CMC operations in various targeted regions within Canada. The use of a global Gaussian grid and a spherical-harmonics representation of the homogeneous and isotropic background-error correlations are described in F10 together with the definition of the
analysis control variables. Analysis increments are produced on a global Gaussian grid at 100-km horizontal resolution (with $400 \times 200$ grid points). One subtlety that is worth noting is the use of the same $400 \times 200$ analysis Gaussian grid as used in REG-3D [see section 3d(1) in F10], but the poles are rotated in REG-4D. The rotation angles of the analysis grid are the same as the rotation angles used in the configuration of the GEM-LAM TL/AD (and high-resolution GEM-LAM model). Since the GEM-LAM TL/AD grid is an exact subdomain of this global Gaussian (rotated) grid, the communication of information (i.e., analysis increments and adjoint sensitivities) is direct; that is, no spatial interpolations in the horizontal (and vertical) are involved. Since by construction, the horizontal background-error correlations are homogeneous and isotropic, it is still valid in the rotated frame of reference defining the rotated Gaussian grid. This means the computer code for that aspect remains intact. It is also important to stress that the use of triangular spectral truncation maintains isotropic spatial resolution. However, background-error standard deviations are spatially rotated (i.e., all of them are treated as true scalars since Helmholtz’s functions for winds are used as control variables together with temperature, logarithm of specific humidity, and surface pressure).

As shown in Fig. 6, a global 4D-Var analysis (e.g., G206; G606 being a separate surface analysis) is routinely performed every 6 h, from which the initial conditions are produced (R206) for the GEM-LAM at 15 km (with $649 \times 672$ grid points in the horizontal) integrated for 9 h to provide a background trajectory necessary for the REG-4D analysis (R112). In parallel to this, initial conditions for the global driving model GEM-GLOBAL (55 km) named here D206 are prepared from the G206 analysis and serve in a 9-h global forecast to provide the background fields for the parallel global driving analysis and for the LBCs of the 9-h GEM-LAM at 15 km. The former is a global 3D-Var analysis with a triangular truncation at T-108 (with the smallest horizontal wavelength of about 180 km). It is performed for the driving model at the same analysis time as the REG-3D analysis, but with all data available.

Fig. 5. The operational regional model grid at 15-km horizontal resolution is shown in dark blue. The tangent-linear/adjoint model grid at 100-km horizontal resolution is shown in red and exceeds the 15-km grid in dark blue.
over the globe and the same data cutoff time. This synchronous driving analysis was found to be beneficial since it allows observations outside the GEM-LAM analysis area to influence the GEM-LAM forecast through the LBCs (see F10, section 3b). Once this global analysis is ready, a 48-h global run with the 55-km global model is performed and serves as the driving conditions (every hour) for the regional 15-km model run. For the regional analysis, only observations over the GEM-LAM analysis area are assimilated. At the end of the minimization, the regional analysis increments (100-km resolution) are added to the first-guess GEM-LAM fields (at 15-km resolution) to complete the REG-3D analysis. We stress that there is a difference between the time of validity of REG-3D and REG-4D analyses. After the minimization, the REG-4D analysis is obtained 3 h before the synoptic time $T$. The REG-3D analysis is already available at this time. So a 3-h GEM-LAM forecast is needed to carry this REG-4D analysis up to time $T$, using the same driving conditions as the 9-h GEM-LAM background trajectory. This represents the REG-4D analysis at time $T$. There is a potential mismatch however between the driving conditions used for GEM-LAM between $[T - 3 h, T]$ and between $[T, T + 48 h]$. This was not evaluated in the current configuration of REG-4D. All model configurations involve a model lid at 0.1 hPa with 80 vertical levels (see F10’s Fig. 3).

The REG-4D is a temporal extension of REG-3D and operates with an incremental formulation (Courtier et al. 1994). It shares with REG-3D the general formulation of the cost function with its background term $J_b$ (weighting the analysis increment with prescribed error covariances) and its observation term $J_o$. In REG-3D, $J_o$ measures the weighted difference between the static analysis increment and the innovations. In REG-4D, $J_o$ measures the weighted difference between the time-evolved analysis increments (obtained from the tangent-linear GEM-LAM forecast at 100-km horizontal resolution) and the innovations at the appropriate time of the observations. The innovations are defined as the difference between observations and forecasts at 15-km horizontal resolution issued from the first-guess GEM-LAM at initial time. The REG-4D is computationally more demanding as compared to REG-3D since it requires the integration of the tangent-linear model for $J_o$ and the adjoint model for $J_o$.

b. Design of the extended TL/AD computational grid

We describe here the procedure used to define an adequate extension strategy of the TL/AD grid as introduced in section 2, to be used in a full fledged 4D-Var data assimilation context afterward. We recall that the goal here is to allow a larger grid than the analysis grid in order to propagate analysis increments during the 4D-Var minimization so as to match accurately the actual propagation occurring in a control global model (same spatial resolution) over the time assimilation window (e.g., here 6 h). Such an extension of the grid will be mostly governed by horizontal advection considerations (e.g., strong horizontal winds around the jet level in the model). To get a good indication on the amount of extension needed, we perform single observation (1-Obs) data-assimilation type experiments where a single 1.0 m s$^{-1}$ $u$ component of wind field innovation at 250 hPa is assimilated at the middle of the assimilation window with standard deviation observation error of 1 m s$^{-1}$ and with the same background-error statistics as the REG-3D system.
We compare the increments computed by three “1-Obs” 4D-Var assimilation experiments that differ in the formulation of the tangent-linear and adjoint models. The first model is GEM-GLOBAL (Gaussian grid 400 × 200). The other two GEM-LAM model grids are exact subdomains where grid points are collocated with the global Gaussian grid. The second model (grid 104 × 104) covers almost identically the same extension as the analysis area and the third model (grid 134 × 134) is the extended TL/AD grid we are testing. All TL/AD models are at 100-km resolution. The horizontal diffusion and vertical sponge layer are different between the GEM-GLOBAL and GEM-LAM configurations since faster algorithms in the GEM-LAM model can be adopted due to the less-severe aspect ratio of the latitude–longitude grid over the limited area being used. For both GEM-LAM configurations, the driving zone is fixed at seven grid points and Davies blending is activated over two grid points. The driving model is a GEM-GLOBAL but at 55-km horizontal resolution. It provides LBCs at each 90-min interval to the nonlinear trajectory on which GEM-LAM TL/AD is based within the incremental 4D-Var procedure. Since GEM-LAM TL/AD has a model time step of 45 min, linear temporal interpolation is used to estimate missing driving time steps for the nonlinear trajectory at analysis resolution (100 km). Differences in horizontal resolution and time steps between the driving model and GEM-LAM have an impact on the nonlinear trajectory. It should be borne in mind here that our procedure relies on the well posedness of our forecast problem with regard to the specification of LBCs (i.e., errors introduced through the LBCs are gradually propagated inside the extended GEM-LAM domain at a speed governed predominantly by advective processes). The GEM-LAM TL/AD specification of LBCs is different. Option 1 described in Table 1 was adopted. That means that the structure of the analysis increments outside the core zone is governed by the background-error covariances and is held fixed in time.

In Fig. 7, we compare the increments of the three 4D-Var assimilation experiments at a model vertical level close to 250 hPa. The zonal wind increments at the beginning of the assimilation period (at t = −3 h) are mainly advected by the basic-state wind of the trajectory over 6 h. The GEM-GLOBAL (400 × 200 grid points) background winds are superimposed on all panels as a reference. The respective GEM-LAM TL/AD core grid boundaries are shown as a solid line (middle and bottom panels) and the boundary of the analysis area is delimited by a dashed line.

It is important to mention here that, as discussed in F10, the operational global 4D-Var system uses a choice of control variables which is different from the one used in the REG-3D and REG-4D systems. For the purpose of validating our approach here, we perform the global 4D-Var test in “regional mode” (i.e., using the same control variables and background-error statistics as the regional system). Although the latter, by design (see F10), emphasizes the North American region, it is necessary here to perform such a control experiment in a clean way in order to validate our strategy adequately.

Based on the results of Fig. 7 (middle panels) with GEM-LAM (104 × 104 grid), it is clear that the analysis increments are inappropriately evolved near the core grid boundary of the TL/AD models. This is due to the fact that the control of LBCs follows option 1 (i.e., the analysis increments that appear between the core grid boundary and the boundary of the analysis area are kept fixed in time while integrating the TL model). We can conclude that for innovations located within a certain distance from the core grid boundaries of the TL/AD, significant errors will appear in the initial and subsequent spatial structure of the analysis increments resulting from 4D-Var under such a treatment of LBCs. Some limited-area 4D-Var systems at other centers impose zero increments at the boundary (clearly detrimental). This corresponds to option 1 with a persistence of 0.

Results from our approach of extending the TL/AD (134 × 134) grid to permit better evolution of the analysis increment for innovations in the vicinity of the lateral boundaries are presented in Fig. 7 (bottom panels). By displacing the lateral boundaries of the TL/AD core grid sufficiently, it is seen that the initial and final structure of the resulting 4D-Var analysis increments precisely match the control experiment with the global configuration (top panels) over the analysis area. Practically, after some testing, we decided to fix the extension criteria as 30% in each direction of the analysis grid area.

We now extend the validation of our strategy against the global reference by using all the observational data normally assimilated within the operational REG-3D system. Figure 8 compares the 250-hPa temperature analysis increments of two 4D-Var assimilations, one with GEM-GLOBAL (Gaussian 400 × 200; top panels) and the other with GEM-LAM (134 × 134; bottom panels). The solid lines represent the boundaries of the core TL/AD grid designed previously and the dashed lines represent the boundaries of the analysis area within which all data are assimilated. We conclude that the proposed extension of the GEM-LAM TL/AD model grid allows us to accurately simulate global results. This is valid in inflow as well as in outflow areas. It is therefore adopted in REG-4D.

4. Objective evaluations

We now present the results of extensive objective evaluations of this REG-4D data assimilation system. We describe in sequence the dataset assimilated and the
Fig. 7. 4D-Var zonal wind increment (at analysis resolution) at model level close to 250 hPa resulting from a single 250-hPa zonal wind observation at 40°N, 60°W with a 1 m s$^{-1}$ innovation and observation error for (top) GEM-GLOBAL Gaussian 400 x 200, (middle) GEM-LAM 104 x 104, (bottom) GEM-LAM 134 x 134. (left) At $t = -3$ h and (right) at $t = +3$ h. GEM-GLOBAL background winds are superimposed. Solid and dashed lines delimit the TL/AD core zone and analysis area, respectively. Contours are plotted from −0.25 to 0.25; the contour interval is 0.025. Solid and dashed contours indicate positive and negative values, respectively.
objective evaluations of two-day forecasts against radiosonde data for winter and summer cases, verifications against global analyses, and impact on precipitation scores. The REG-4D only has one outer loop with 25 iterations. To allow a second outer loop, we would need a mixing of the updated GEM-LAM analysis and initial conditions of the driving model since the full domain of GEM-LAM TL/AD (134 x 134 grid) overlaps the GEM-LAM analysis area. However, this is not done yet since the current system fits tightly into the analysis 20-min real-time constraint normally used in operations at CMC. The analysis increments at low resolution (100 km) are added to the high-resolution GEM-LAM first guess (15 km). The tangent-linear and adjoint models include the same dynamical processes as the high-resolution forecast model, but with only a few physical processes: simplified vertical diffusion, simplified grid-scale condensation, and moist convective adjustment. Only simplified vertical diffusion is used, however, in our single inner loop due to accuracy limitations of our linearized physics. In 4D-Var, observations are assimilated at the appropriate time over the whole 6-h assimilation window. The preprocessing of observations includes a background check quality control and data thinning. The selection of observations is based on the 4D screening scheme of Rabier et al. (2000),
which represents a significant increase in the volume of data ingested by 4D-Var assimilation as compared to REG-3D. The background-error statistics are the same as those used by REG-3D (see section 3a for details). Table 2 lists all the observations assimilated by the currently operational REG-3D system and the REG-4D version examined here. Note also that both systems use the same data types as the 3D-Var global data assimilation system of the driving model, but only data inside the GEM-LAM (15 km; dark blue domain in Fig. 5) are used for the regional analyses.

Figure 9 shows the evaluation of 48-h forecasts for 118 summer 2008 cases against North American radiosondes. The standard deviation and bias errors of winds, temperature, geopotential, and dewpoint depressions are shown as a function of the vertical pressure coordinate. The control operational REG-3D is shown compared with the REG-4D system. We notice an improvement in REG-4D winds between 200 and 500 hPa, but only slight improvement for the mass fields (temperature and geopotential). The moisture field is left essentially unchanged. Similar results apply for winter 2009 cases (118 cases also, results not shown). By carefully examining the regions of verification, we note that the largest improvement obtained by implementing a 4D-Var approach came from summer cases at mid- and high latitudes. Figure 10 shows verifications of 36-h forecasts against Arctic radiosondes. There is a very significant improvement at 36 h of REG-4D over REG-3D for winds and a slight improvement for temperature over the Arctic region (summer 2008 cases). The improvements are noticeable even for the short-range forecasts (e.g., 12-h forecasts, results not shown) and are maintained even up to 48 h (Fig. 11).

To identify the regions of greatest improvement, we performed an evaluation of forecasts against CMC global 4D-Var analyses every 6 h. Figure 12 shows clearly the widespread regions of improvement (regions in dark blue) for summer and winter cases. It is also apparent in these figures that during winter, the regions of improvements tend to appear farther south in the domain whereas during the summer, the improvements are mostly confined to latitudes above 40°N. This behavior may be due to the increased importance of moist physical processes during the summer as compared to dynamical contributions, combined with the fact that our TL/AD schemes only involve dry physical processes. It is shown in Fig. 13 (top panel) that 48-h geopotential forecasts at 500 hPa are improved with REG-4D and can be seen as a function of time for all 118 summer cases examined. A similar conclusion applies for winter cases (results not shown). Finally, to stress the importance of the short-term improvement in wind forecasts with REG-4D, we show in Fig. 13 (bottom panel) the 12-h wind forecast evaluations as a function of time for summer cases. The improvement is significant and strongly suggests that in the future development of the REG-4D system, it would be beneficial to consider a few analysis cycles before launching 48-h regional forecasts. This would allow the use of an improved trial field within the regional data assimilation.

We now describe the impacts of the REG-4D system on the accumulated precipitation scores. We present results against the Cooperative Observer Program (COOP) precipitation dataset (hereafter referred to as SHEF data, see F10 for details). Verifications against the surface synoptic observations lead to similar conclusions (results not shown). Overall, the new system has a neutral impact

<table>
<thead>
<tr>
<th>Family</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper air</td>
<td>Radiosondes Air Report, Aircraft Meteorological Data Relay</td>
</tr>
<tr>
<td>Aircraft (temperature,</td>
<td></td>
</tr>
<tr>
<td>pressure, and humidity)</td>
<td></td>
</tr>
<tr>
<td>Surface Profiler</td>
<td>Geostationary Operational Environmental Satellite (GOES)</td>
</tr>
<tr>
<td></td>
<td>Multifunctional transport satellite</td>
</tr>
<tr>
<td>Satellite winds</td>
<td><em>Aqua/Terra</em></td>
</tr>
<tr>
<td></td>
<td>Advanced Scatterometer (ASCAT)</td>
</tr>
<tr>
<td>Satellite radiances</td>
<td>NOAA Advanced Microwave Sounding Unit (AMSU)-A</td>
</tr>
<tr>
<td></td>
<td>NOAA AMSU-B</td>
</tr>
<tr>
<td></td>
<td><em>Aqua</em> AMSU-A</td>
</tr>
<tr>
<td></td>
<td>Atmospheric Infrared Sounder (AIRS)</td>
</tr>
<tr>
<td></td>
<td>Infrared Atmospheric Sounding Interferometer (IASI)</td>
</tr>
<tr>
<td></td>
<td>Special Sensor Microwave Image/Sounder (SSMI/S)</td>
</tr>
<tr>
<td></td>
<td>Clear-sky radiances <em>GOES-11, Meteosat-7(-9)</em></td>
</tr>
</tbody>
</table>
according to this measure except for two specific aspects that we describe in the following. In Fig. 14, we notice that the bias of the accumulated precipitation in summer test cases over North America (0–24 or 12–36 h, left and right panels, respectively) has been modified in response to 1) the use of REG-4D rather than REG-3D and 2) the increased number of observations in REG-4D. It is one persistent feature of the combined use of REG-4D with
increased assimilated data in the temporal assimilation window that such bias modification up to two days can be accomplished. Normally, changes in the model physics or increased spatial resolution are at the source of such precipitation bias changes and are rarely caused by more limited modifications in other components of the data assimilation scheme. Overall, the threat scores appear neutral between REG-3D and REG-4D experiments. We note, however, that the reduced bias of accumulated precipitation for small amount categories is welcome in the context of the regional system since it is already well known that operational Canadian regional systems of the past 10 years have suffered from a noticeable bias here and REG-4D brings a beneficial correction.

FIG. 10. As in Fig. 9, but for forecasts verified at 36 h against Arctic radiosondes (i.e., north of 58°).
(left panel) shows a slight improvement in threat scores for the winter test cases over North America against the SHEF network for almost all categories. Finally, still against SHEF data but focusing on the east coast of North America during the summer, it can be seen in Fig. 15 (right panel) that the precipitation bias has been significantly reduced for all categories (again a desirable aspect) and we observe a slight improvement of the threat score.

5. Conclusions

The new Canadian regional 3D-Var data assimilation and forecasting system was described in Fillion et al. (2010). This system became operational on 20 October 2010 at the Canadian Meteorological Center (CMC). Following this, many possible upgrades could be envisaged. Among them, the extension from 3D-Var to 4D-Var data...
FIG. 12. 48-h forecast verifications of geopotential at 500 hPa. (top) 118 winter cases from 0000 UTC 1 Jan 2009 to 1200 UTC 28 Feb 2009, every 12 h. (bottom) 118 summer cases described in Fig. 9. The verification measure is the difference between rms differences of the REG-3D forecasts against CMC operational 4D-Var global analyses and the corresponding quantity for the REG-4D forecasts. Blue: REG-4D is better; Green: REG-3D is better.
assimilation has been finalized. This paper reports on the design and evaluation of the regional 4D-Var system, referred to as REG-4D. In this context, we have shown first of all the influence of boundary conditions on the accuracy of tangent-linear and adjoint models. Based on results of previous studies on the subject, we found it appropriate to examine identical twin assimilation experiments to study the lack of control of LBCs since the latter may induce large errors at an inflow boundary, which are advected farther inside the core zone (i.e., the horizontal domain defining the predictive variables in GEM-LAM). These errors are corrected when LBCs are included at each time step in the control vector of the minimization. To reduce the size of the control vector, a possible way to limit the number of time steps was shown. We observed that the degree of convergence is linked to the implicit order of accuracy of temporal discretization of Taylor series expansion employed.

Although feasible, these techniques necessitate the careful design of background-error correlations to accommodate the additional control variables. For the Canadian regional data assimilation system, however, we designed a new approach where the TL/AD horizontal grid is extended past the nonlinear high-resolution limited-area forecast model domain (see Fig. 5). We demonstrate that this extension allows an exact propagation of analysis increments as compared to the control global tangent-linear model. This strategy has been validated with single-observation tests near the boundaries of the model and within a full-fledged data assimilation context (i.e., the same observational data as used in CMC operations). An appropriate domain size extension of 30% in each direction was then chosen within this complex data assimilation system, which typically uses a 6-h data assimilation window and extends vertically to include most of the stratosphere.

Fig. 13. (top) Time series verification of REG-3D (blue line) and REG-4D (red line) 48-h geopotential forecasts at 500 hPa against CMC operational global 4D-Var analyses for the summer cases described in Fig. 9. (bottom) As in (top), but for the 12-h wind forecasts at 500 hPa.
Having passed this validation phase, the REG-4D system was extensively tested using objective evaluation scores normally used at CMC. Systematic evaluations were done using 118 winter 2009 test cases and 118 summer 2008 test cases. The REG-4D system was shown to produce slightly better verification scores against radiosonde data over the North American continent up to 48 h (both seasons). Particularly noticeable is the significant improvements in wind forecasts over the Arctic region. This improvement was shown to occur very early in the forecast and to persist over the two-day range. To complement these results, verifications against operational CMC Global 4D-Var analyses revealed almost uniquely improvements over the forecast region. For the summer test cases, improvements from REG-4D are seen to occur mostly at mid- and high latitudes. This result could possibly be related to the current absence of tangent-linear moist processes (e.g., deep convection) in our REG-4D system, thus significantly impacting the verification scores for the summer test cases over the southern part of the forecast domain. Overall, it was shown that the accuracy of precipitation forecasts (bias and threat scores) were slightly improved both for summer and winter cases.

It is relevant to mention here that this REG-4D analysis system does meet the 20-min real-time allocation available in the operational context to finalize the production of the regional analysis step. The treatment of observations within the variational analysis code has been made more efficient during the last two years with clear benefits. REG-4D can be run efficiently with 512 CPUs using message passing interface (MPI), which is affordable for operations, given the current computer power available at CMC. Further improvements are underway to improve resource sharing under MPI. For future developments of this deterministic REG-4D system, we are currently considering the use of appropriate regional ensemble forecasts to represent the flow-dependent part of the background-error covariances in addition to the currently used homogeneous and isotropic correlations. In the near future, we expect to finalize a REG-4D version of the system where the high-resolution nonlinear model will be used at 10 km rather than 15 km (as in this study) and some improvements in the physics package. A vertically staggered grid is currently under testing within the REG and global systems and should be the version used in future operational implementations. Ground-based GPS data are already known to be beneficial within the current REG-3D
system and work is under way to extend these improvements to the REG-4D context. Finally, as we will soon go through a significant computer upgrade at CMC, impact from higher spatial and temporal resolution of the 4D-Var analysis increments will be examined.

Acknowledgments. The authors thank Michel Desgagné and Vivian Lee for their constant support of new versions of the GEM model during the development phase of this project. The authors thank Drs. Stéphane Bélair and Paul Vaillancourt for their help on the evaluations of the precipitation results obtained in this study and Dr. Mateusz Reszka for providing an internal review of the paper.

APPENDIX

LBCs in GEM-LAM: 1D Schematic Model

The specification of LBCs in GEM-LAM is the same as in Thomas et al. (1998). Because GEM is a two time-step implicit, semi-Lagrangian, gridpoint model with a latitude–longitude Arakawa C-grid staggering in the horizontal, normal wind components are specified to properly close the mathematical problem. This is illustrated using a 1D linearized shallow-water model, following McDonald (2000):

\[
\begin{align}
\frac{\partial \phi}{\partial t} + U \frac{\partial \phi}{\partial x} + \Phi \frac{\partial u}{\partial x} &= 0, \\
\frac{\partial u}{\partial t} + U \frac{\partial u}{\partial x} + \frac{\partial \phi}{\partial x} &= 0,
\end{align}
\]

where \( u = u(x, t) \) is the wind velocity, \( \phi = \phi(x, t) \) is the geopotential and \( U, \Phi \) are the corresponding mean constant values. We assume \( U > 0 \) and \( U < \sqrt{\Phi} \). The distribution of variables on the \( x \) axis of an Arakawa C-grid is as follows:

\[
\begin{align}
\phi_d^{n+1} + \Phi \frac{\Delta t}{2\Delta x} (u_d - u_{d+1})^{n+1} &= R^\phi_i = (X^\phi i)^n, \\
\phi_d^{n+1} + \Phi \frac{\Delta t}{2\Delta x} (u_d + u_{d+1})^{n+1} &= 0,
\end{align}
\]

where \( 2\Delta \) is the size of the driving zone and \( M \) is the size of the core zone. A two-time-step implicit, semi-Lagrangian discretization of Eqs. (A1)–(A2) using this distribution gives the following:

\[
\begin{align}
\phi_d^{n+1} + \Phi \frac{\Delta t}{2\Delta x} (u_d - u_{d+1})^{n+1} &= R^\phi_i = (X^\phi i)^n, \\
\phi_d^{n+1} + \Phi \frac{\Delta t}{2\Delta x} (u_d + u_{d+1})^{n+1} &= 0,
\end{align}
\]
where \( n \) is a time level, \( d + 2 \leq i \leq d + M - 1 \) in Eq. (A4) and \( d + 1 \leq i \leq d + M - 1 \) in Eq. (A5). Here \( X_i \) refers to semi-Lagrangian interpolation of \( X \) and

\[
(X^\phi_i) = \phi_i - \frac{\Delta t}{2\Delta x} (u_i - u_{i-1})^n, \quad (A6)
\]

\[
(X^{u_i}) = u_i - \frac{\Delta t}{2\Delta x} (\phi_{i+1} - \phi_i)^n. \quad (A7)
\]

Note that \((X^\phi)^n\) and \((X^{u})^n\) are also evaluated inside the driving zone to allow semi-Lagrangian interpolation. Having computed \( R^\phi \) and \( R^u \), Eq. (A5) is substituted in Eq. (A4) resulting in

\[
\phi_i^{n+1} - \Phi \left( \frac{\Delta t}{2\Delta x} \right)^2 (\phi_{i+1} - 2\phi_i + \phi_{i-1})^{n+1} = R^\phi_i - \frac{\Delta t}{2\Delta x} (R^\mu_i - R^\mu_{i-1}), \quad (A8)
\]

with \( d + 2 \leq i \leq d + M - 1 \). To close the system, we have adopted well-posed LBCs (Oliger and Sundström 1978) by imposing normal winds \( u_{i+1}^n \) and \( u_{i+M}^n \) on the grid distribution shown in Eq. (A3). The resulting equations for \( \phi_{d+1} \) and \( \phi_{d+M} \) are as follows:

\[
\phi_{d+1}^{n+1} = \Phi \left( \frac{\Delta t}{2\Delta x} \right)^2 (\phi_{d+2} - 2\phi_{d+1} + \phi_{d})^{n+1} = R^\phi_{d+1} - \frac{\Delta t}{2\Delta x} (R^\mu_{d+1} - u_{d+1}^n), \quad (A9)
\]

\[
\phi_{d+M}^{n+1} = \Phi \left( \frac{\Delta t}{2\Delta x} \right)^2 (-\phi_{d+M+1} + \phi_{d+M})^{n+1} = R^\phi_{d+M} - \frac{\Delta t}{2\Delta x} (u_{d+M}^n - u_{d+M+1}). \quad (A10)
\]

Having solved the system in Eqs. (A8)–(A10) for \( \phi_{i}^{n+1} \), \( u_{i}^{n+1} \) can be computed from Eq. (A5).

REFERENCES


Seiler, U., 1993: Estimation of open boundary conditions with the
Staniforth, A., 1997: Regional modeling: A theoretical discussion.
Sundström, A., and T. Elvius, 1979: Computational problems re-
lated to limited-area modeling. GARP Publications Series,
Thépaut, J.-N., and P. Courtier, 1991: Four-dimensional vari-
tional data assimilation using the adjoint of a multilevel
1225–1254.
Thomas, S. J., C. Girard, R. Benoit, M. Desgagne, and P. Pellerin,
1998: A new adiabatic kernel for the MC2 model. Atmos.–
Ocean, 36, 241–270.
on lateral boundary conditions as a basic and potentially serious
limitation to regional numerical weather prediction. Bull. Amer.
4D-Var over a North Atlantic and European domain. Joint
SRNWP/Met Office/HIRLAM Workshop on Variational As-
similation: Towards 1-4km Resolution, Exeter, Devon, United
Kingdom, Met Office, 1 p. [Available online at http://research.
metoffice.gov.uk/research/nwp/external/srnwp/workshop_
 nov2004/Presentations/Posters/wlasak.pdf.]
Zhang, X., X.-Y. Huang, Y.-R. Guo, N. Gustafsson, and M. Zhang,
tutorial, 38 pp. [Available online at http://www.mmm.ucar.edu/wrf/
Zou, X., and Y.-H. Kuo, 1996: Rainfall assimilation through an
optimal control of initial and boundary conditions in a limited-