14B.3 Evaluation of the operational 4D-Var at the Meteorological Service of Canada

S. Laroche*, P. Gauthier, M. Tanguay, S. Pellerin, J. Morneau
Meteorological Service of Canada

1. Introduction

The three-dimensional variational data assimilation (3D-Var) system was introduced in the operational suite of the Canadian Meteorological Centre (CMC) in 1997 in preparation for the direct assimilation of satellite radiances (Gauthier et al., 1999a; Chouinard et al., 2001). Recently, this system has been extended to 4D-Var in which the observations are now compared to the background field at the exact time and the background error statistics are implicitly flow-dependent. This relaxes the assumption of stationarity implicit in 3D-Var. Moreover, 4D-Var is now producing an analysis that fits globally a time series of observations which extracts tendency information that is difficult to achieve in any sequential data assimilation scheme.

This paper describes the characteristics of the upgraded variational data assimilation system developed at the Meteorological Service of Canada (MSC) and summarizes the results of an extensive inter-comparison between the operational 3D-Var and the new 4D-Var. These experiments aim at better understanding and quantifying the contribution of each component of 4D-Var. The 4D-Var system was implemented in the global (medium-range) forecasting system of CMC on 15 March 2005.

2. The variational data assimilation system

The incremental formulation as proposed by Courtier et al., (1994) is adopted. The analysis increment is calculated at a lower resolution (1.5^oL28 in our system) than the forecast model, which is 0.9^oL28 for the current global forecasting system (Côté et al., 1998). The analysis increment is obtained by minimizing the cost function

$$J = \frac{1}{2}\left[\delta\hat{x}_0^{(k)} - (\hat{x}_0^b - \hat{x}_0^{(k)})\right]^T \mathbf{B}^{-1}\left[\delta\hat{x}_0^{(k)} - (\hat{x}_0^b - \hat{x}_0^{(k)})\right] + \frac{1}{2}\sum_{i=0}^{n}\left[\mathbf{H}_i\delta\hat{x}_i^{(k)} - d_i^{(k)}\right]^T \mathbf{R}^{-1}\left[\mathbf{H}_i\delta\hat{x}_i^{(k)} - d_i^{(k)}\right],$$

where $d_i^{(k)} = y_i - H_i(x_i^{(k)})$ is the innovation vector , k is the outer loop index, y_i is the observation vector in the time interval i, **B** and **R** are the background and observation error covariance matrices respectively, H_i is the observation operator that map the full-resolution trajectory $x_i^{(k)} = M(t_o,t_i)x_0^{(k)}$ into the observation space, M represents the full-resolution model integration. The high resolution analysis update is obtained from

$$x_0^{(k)} = x_0^{(k-1)} + \mathbf{S}^{-I}(\delta\hat{x}_0^{(k-1)}),$$

where $\mathbf{S}^{-I}$ is the interpolation operator from low to high resolution, while $\delta\hat{x}_i^{(k)} = \mathbf{M}(t_o,t_i)\delta\hat{x}_0^{(k)}$ represents the low-resolution analysis increment propagated in time with the tangent linear model (TLM) **M** linearized around M, during the iterative minimization. This process is referred to as the inner loop.

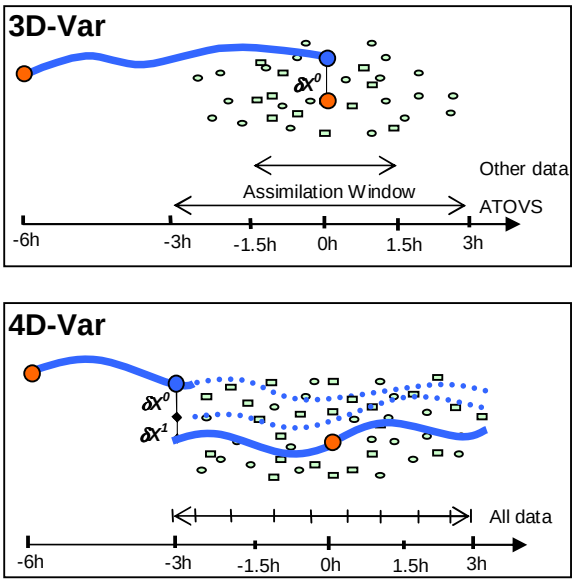

Fig. 1: Schematics of 6-h data assimilation cycles for the global forecasting system: the former 3D-Var (top panel) and the new implemented 4D-Var (bottom panel).

* *Corresponding author address:* Stéphane Laroche, Meteorological service of Canada, 2121 Route Trans-Canadienne, Dorval, QC, Canada, *H9A 1J3*; e-mail: stephane.laroche@ec.gc.ca

The minimization is done with a quasi-Newton algorithm developed by Gilbert and Lemaréchal (1989). The initial state of the low-resolution trajectory is also updated after each outer loop as follows:

$$\hat{x}_0^{(k)} = \mathbf{S}x_0^{(k)} \ ,$$

where $\hat{x}_0^{(0)} = \hat{x}_0^b$ is the background field at low resolution. This formulation is valid for both 3D-Var and 4D-Var.

In the formerly operational 3D-Var, the background field is a 6-h forecast from the previous analysis, as shown in Fig. 1. A single outer loop is performed (k = 0) with only one time interval (n = 0) covering the whole assimilation window. The analysis increment is not evolved in time and is estimated at the center of the assimilation window, corresponding to the synoptic time (t_0 = 0h in Fig. 1). The minimization is stopped when the gradient of the cost function is reduced by two orders of magnitude, which is usually achieved within 90 inner loops. The assimilation time window is 3h, centered at the synoptic time, for most observations, except for the ATOVS radiances which are assimilated over the full 6-h time window.

In the new 4D-Var scheme, the background field now corresponds to a trajectory covering the whole assimilation window (-3h < t_0 < +3h in Fig.1) and is obtained from a 9-h forecast from the previous analysis (valid at t_0 = -6h in Fig. 1). The TLM of the global environmental multi-scales (GEM) model and its adjoint (Tanguay and Polavarapu, 1999) are employed in the inner loop to propagate the analysis increment and the gradient of the cost function over the assimilation window. The analysis is obtained after two outer loops (k = 0,1). In the first outer loop, 40 inner loops are performed with only the vertical diffusion as simplified linearized physics in **M** (Laroche et al., 2002). After updating the full-resolution and low-resolution trajectories, 30 more inner loops are done with a comprehensive set of simplified physical parameterizations including vertical diffusion, subgrid-scale orographic effects, large-scale precipitation and deep moist convection (Zadra et al., 2004; Mahfouf, 2005). The data selection process has been modified for all observation types except surface reports. The assimilation time window is now 6h for most observations. It is divided into 9 time intervals (n = 8) of 45 min, except 22.5 min at both ends. For each interval, the data are temporally thinned to retain the observation closest to the middle of the time interval. This observation processing is referred to as *4D thinning*. This has considerably increased the number of frequently reported data such as aircraft, Atmospheric Motion Vectors (AMV) and profiler data as shown in Fig. 2. The number of TOVS radiances has also increased, especially over the high-latitudes where several satellite orbits overlap. Overall, the number of observations has increased by 60% with 4D-Var. Note that the spatial thinning as well as the treatment for satellite data is the same as in 3D-Var (Wagneur et al., 2004, Hallé and Chouinard, 2003). The upper-air analysis with 4D-Var is the result of a final 3-h integration of the full-resolution model. The surface analysis valid at synoptic time is then added to get the complete 4D-Var analysis. Finally, the resolution of the analysis increment (1.5^oL28), background error statistics (Gauthier et al., 1999b) and the data quality control remain the same as in 3D-Var (Gauthier et al., 2003).

A variant of 3D-Var has also been tested in which the first guess at the appropriate time (FGAT) from the full-resolution model trajectory is used to calculate the misfit to the observations.

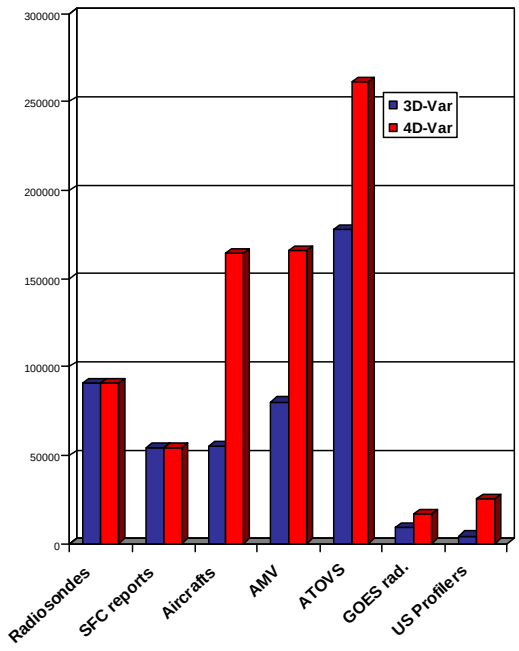

Fig 2: Average amount of data assimilated per day in 3D-Var and 4D-Var at the CMC.

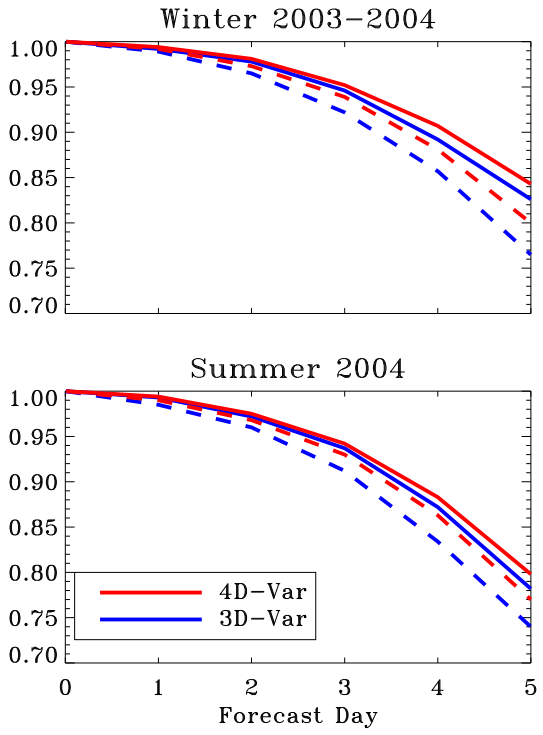

Fig. 3: Anomaly correlation scores for 500 hPa geopotential height for North America (solid) and the southern hemisphere (dashed).

3. Results from the trials

An extensive pre-implementation evaluation of 4D-Var against the operational 3D-Var was conducted. Results from two-month assimilation periods in winter 2003-2004 and summer 2004 show a consistent improvement in the extratropics with 4D-Var, for both periods. Fig. 3 shows the anomaly correlation scores of the geopotential height at 500 hPa obtained during these periods for 1 to 5-day forecasts from 3D-Var (in blue) and 4D-Var (in red) over North America and the southern hemisphere. The improvement in the southern hemisphere represents nearly 9-h gain in predictability. The scores have also improved over North America for both seasons with 4D-Var. The improvement is greatest over the west coast of North America at shorter ranges while it is best over the eastern part of the continent beyond 2 day forecast. This suggests that the improvement over the Pacific Ocean propagates eastward inland. The impact in the tropics is rather neutral (not shown). There is little improvement in the forecasts of tropical storms as well as in their extra tropical transition in the first three forecast days. Beyond that, 4D-Var seems too amplify the storm speed, already systematically to fast with 3D-Var. The origin of this problem stems from the relatively coarse resolution (0.9^{o}) and long time step (45 min) of the global forecast model, which is unable to properly simulate tropical storms. This indicates that weather elements in the full-resolution model should be well represented to get the full benefit from 4D-Var.

The spectra of the transient component of the 500 hPa geopotential height forecast errors from 3D-Var and 4D-Var (against their own analyses) at day-1, day-3 and day-5 for January 2004 are displayed in Fig. 4. A consistent improvement in the largest scales up to wavenumber 30 and a slight degradation in scales beyond wavenumber 100 are clearly seen for all forecast ranges. The difference at these small scales is explained by the use of a digital filter (Fillion et al., 1995) during the 6-h forecast used as background field in 3D-Var. Thus the 3D-Var analysis is relatively free of high-frequency gravity-wave noise since the addition of the low-resolution analysis increments to the background field does not modify the atmospheric state beyond wavenumber 100. On the other hand, the digital filter is not activated in the final 3-h forecast to generate the 4D-Var analysis. This increase of forecast error in the small scales from 4D-Var (when verified against its own analysis) is overall negligible since the forecast error is dominant in the synoptic scales (wavenumber between 10 and 40). However, the digital filter is applied to the full model state at the beginning of the medium-range forecast (Côté et al. 1998). Although we found that the use of this digital filter has a negligible impact on forecast scores in the context of 3D-Var, it is known that it alters the semidiurnal cycle. This problem will be addressed in the near future by implementing an incremental digital filter, as proposed by Gauthier and Thépaut (2001).

4. Impact of the new 4D-Var components

The contribution of each 4D-Var component to the improvement has been assessed. This is done by performing data assimilation cycles over a period of one-month (August 2004) with various configurations ranging in complexity from our former 3D-Var to the new 4D-Var. Table 1 summarizes the 6 different configurations used in this study. The individual features evaluated are: the implementation of the FGAT, the 4D thinning process (providing 60% more observations than with the 3D thinning), the use of the TLM and its adjoint of the GEM model in the assimilation window, the update of the trajectories between the two outer loops, and the use of a comprehensive set of simplified physics (better) in the second outer loop.

Fig. 5 shows the 500 hPa geopotential height scores over the southern hemisphere for the 6

configurations. The root-mean-square (RMS) forecast errors for this region are used to assess the impact of the new components because the score differences between the former 3D-Var to the new 4D-Var are the largest. Thus consistent reduction of errors when including the new features can clearly be seen. There are several reasons why the difference in scores between 3D-Var and 4D-Var are the greatest. Among them, the southern hemisphere is in the winter season in August, thus more dynamically active. Also, satellite data, distributed over the whole assimilation window, are largely dominant. To summarize the results, Fig. 6 displays the rankings based on the average RMS forecast errors over day-1 to day-5, for each configuration. The improvements are expressed in percentages, 100% being the improvement of the new 4D-Var with respect to the former 3D-Var.

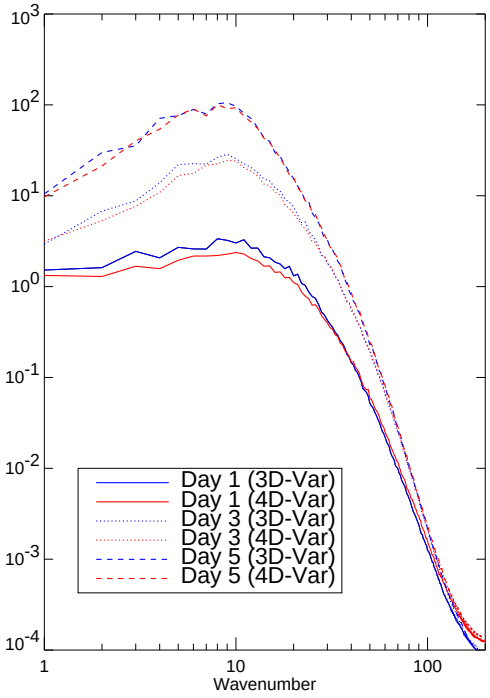

Fig. 4: Transient error spectra for forecasts of the 500 hPa geopotential height for days 1, 3 and 5 (blue from 3D-Var and red from 4D-Var).

Type	Outer loops	Simplified Physics	Temporal thinning
3D-Var	1	-	3D
3D-Var (FGAT)	1	-	3D
4D-Var (1 loop)	1	(simpler)	4D
4D-Var (simpler)	2	(simpler, simpler)	4D
4D-Var (3D-thin)	2	(simpler, better)	3D
4D-Var	2	(simpler, better)	4D

Table 1: Configurations used to assess the impact of the new components of 4D-Var. The number of outer loops, the complexity of the simplified physics in the first and second outer loops (if applicable) and the temporal thinning are indicated. The use of only the vertical diffusion as simplified physics is referred to as 'simpler' physics, as opposed to the 'better' physics which include all the simplified physical parameterizations.

The impact of each new element of 4D-Var can be assessed from the difference in performance between two configurations, as shown in Fig. 6. The most significant source of improvement (50%) is from the use of the TLM and its adjoint to propagate the information in the assimilation window. This mechanism can also be interpreted as the implicit flow dependent propagation of the covariance error statistics (Thépaut *et al.*, 1996). 36% of the improvement is explained by the introduction of the 4D thinning process while the implementation of the FGAT contributes for 14%. The use of a comprehensive set of simplified physics improves the results by 6%. The update of the trajectories between the first and second outer loop leads to very little improvement (3%). This may be explained by the small difference in horizontal resolution between the simplified model used in the inner loops (1.5^o) and the current high-resolution model (0.9^o) in the global forecasting system. Note that the improvement from using the TLM and its adjoint deduced here also includes this slight contribution from the trajectory updates.

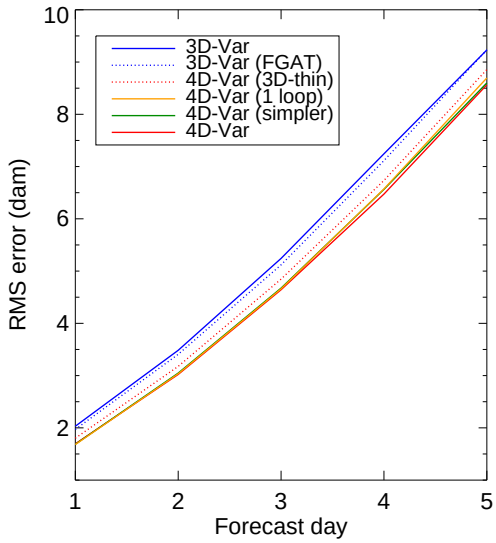

Fig. 5: Root-mean-square forecast errors against own analyses for the 500 hPa geopotential height average over August 2004 for southern hemisphere. See text for a detailed description of the experiments shown.

Finally, it is important to note that the contributions of each feature are not necessarily independent in the sense that the improvement obtained by combining two components may be greater than the sum of their individual impact.

5. Impact of the delayed cutoff time

Based on the good performance of 4D-Var obtained from the trials, a parallel suite with 4D-Var was initiated at the CMC in December 2004 for the final evaluation before implementation. The first 4D-Var results from this parallel suite were not quite as good as those from the pre-implementation experiments. After investigation, it was found that the performance of 4D-Var is more sensitive to the delayed cutoff time than 3D-Var.

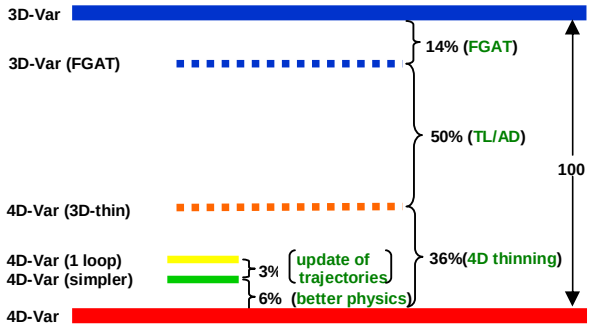

Fig. 6: Contribution of the various components of 4D-Var to the improvements over 3D-Var.

Fig. 7 shows the average percentage of observations received after the synoptic time for the aircraft, the AMV and ATOVS radiance data. These data span the whole 6-h assimilation time window. Operationally, the analysis for the medium-range forecast at 00 and 12 UTC is performed after a 3-h cutoff time for availability of observations. For the pre-implementation experiments, the forecasts used in the evaluation were initiated from analyses with a 9-h cut-off time, corresponding to that of the final data assimilation cycling. The mean percentage of AMV and ATOVS data available 3h after the synoptic time is between 50 and 60%, while most data in assimilation time window are received 9h after the synoptic time. Most of the missing data with 3-h cutoff time are in the second half period of the assimilation window.

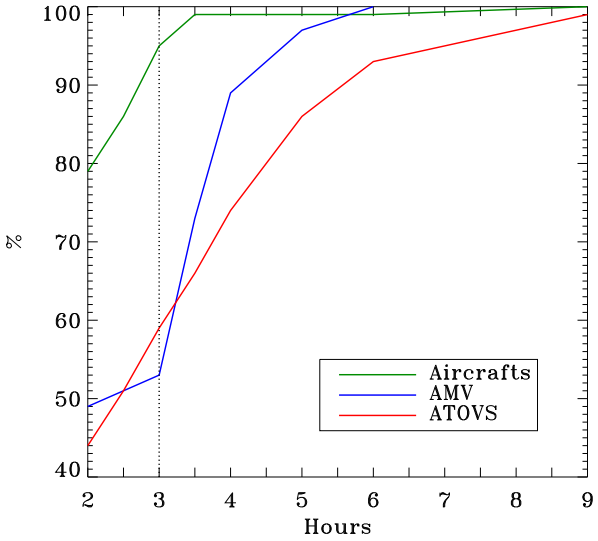

Fig. 7: Average percentage of observations in the global data assimilation system received after the synoptic time in January 2004.

Fig. 8 shows the 500 hPa geopotential height scores for the southern hemisphere from 3D-Var and 4D-Var with 3-h and 9-h cutoff times. The use of a shorter cutoff time is less detrimental for 3D-Var than 4D-Var. The reduction of the forecast performance with a reduced cutoff time is nearly 23% with 4D-Var whereas it is only about 5% with 3D-Var. This indicates that the observations at the end of the assimilation period play an important role in 4D-Var. On the other hand, the stationary assumption may lessen the benefit of assimilating extra observations in 3D-Var. Nevertheless, the performance of 4D-Var remains much superior to 3D-Var with a 3-h cutoff time.

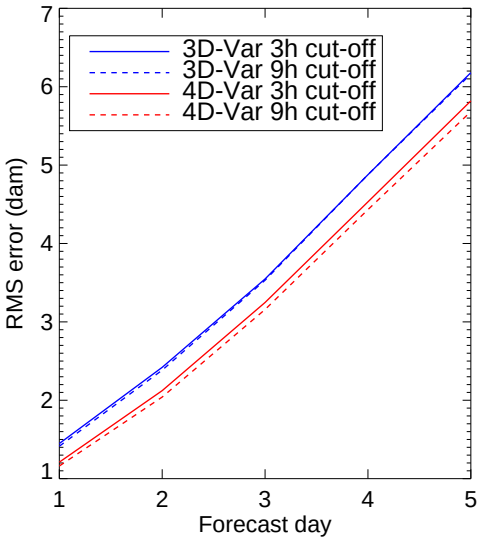

Fig. 8: Root-mean-square forecast error against own analyses for the 500 hPa geopotential height average over January 2005 for southern hemisphere.

6. Concluding remarks

A 4D-Var data assimilation system was implemented in the CMC operational suite on 15 March 2005. This system gives consistent forecast improvement in the extratropics with respect to those from 3D-Var, particularly in the southern hemisphere. The improvements are in the large atmospheric scales (wavenumbers less than 30). The use of the TLM and adjoint of GEM to propagate the information in the assimilation window and the 4D thinning are the two features of the new 4D-Var that contributes the most to the forecast improvement. A shorter delayed cutoff time, as the one used in the operational context for the 00 and 12 UTC forecasts, is more detrimental to 4D-Var. This is because 4D-Var is better than 3D-Var at extracting information from observations at the end of the assimilation window.

The implementation of 4D-Var was a major step towards a more accurate global forecasting system at the CMC. This will permit the optimal use of the current and future remote sensing data, the volume of which is foreseen to increase significantly in the near future. The next important step will be the implementation of a higher resolution forecast model with improved physical parameterizations. The resolution (currently 0.9^{o}L28) will increase to 0.35^{o}L58, but with the same model lid at 10 hPa. Preliminary results with this model show a much better handling of tropical storms, which will make the use of 4D-Var in the tropics more valuable. All the efforts on 4D-Var are now focused to make this new forecasting system affordable. This implies a revision of the incremental strategy and improving the computational efficiency of the simplified model, in which the number of vertical levels will increase from 28 to 58. Also the minimization algorithm and quality control process should be revised. Finally, these significant changes will warrant a re-estimation of background and observation error statistics following the method presented in Buehner (2005) and Buehner *et al.* (2005).

7. References

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.*, **131**, 1013-1043.

Buehner, M., P. Gauthier and Z. Liu, 2005: Evaluation of new estimates of background and observation error covariances for variational assimilation. (submitted to *Q. J. R. Meteorol. Soc.*).

Chouinard C., C. Charette, J. Hallé, P. Gauthier, J. Morneau and R. Sarrazin, 2001: The Canadian 3D-Var analysis scheme on model vertical coordinate. 14th AMS conference on numerical weather prediction, Fort Lauderdale, 30 July-2 August 2001, 14-18.

Côté, J., S. Gravel, A. Staniforth, A. Patoine, M. Roch and A. N. Staniforth, 1998: The operational CMC-MRB global environmental multiscale (GEM) model. *Mon. Wea. Rev.*, **126**, 1373-1395.

Courtier, P., J.-N. Thépaut and A. Hollingsworth, 1994 : A strategy for operational implementation of 4D-Var using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**, 1367-1387.

Fillion, L., H. L. Mitchell, H. R. Ritchie, and A. N. Staniforth, 1995: The impact of a digital filter finalization technique in a global data assimilation system. *Tellus*, **47A**, 304-323.

Gauthier, P., C. Charette, L. Fillion, P. Koclas, and S. Laroche, 1999a. Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: the global analysis. *Atmosphere -Ocean*, **37**, 103-156.

Gauthier, P., M. Buehner and L. Fillion, 1999b: Background-error statistics modeling in a 3D variational data assimilation scheme: Estimation and impact on the analysis. In proceedings, ECMWF Workshop on diagnosis of data assimilation systems. ECMWF, 131-145.

Gauthier, P. and J-N. Thépaut, 2001: Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Wea. Rev.*, **129**, 2089-2102.

Gauthier, P., C. Chouinard and B. Brasnett, 2003:

Quality Control: Methodology and Application. Data Assimilation for the Earth System. R. Swinbank et al., Kluwer Academic Publishers. 177-187.

Gilbert, J. C. and C. Lemaréchal, 1989: Some numerical experiments with variable-storage quasi-Newton algorithms. *Mathematical Programming*, **45**, 407-435.

Hallé, J. and C. Chouinard, 2003: Use of AMSU-B radiances in the Canadian Meteorological Centre 3D-Var system. 12th conf. on sat. meteor. and ocean., 8-13 Feb. 2003, Long Beach, CA.

Laroche, S., M. Tanguay and Y. Delage, 2002: On the linearization of a simplified planetary boundary layer parameterization. *Mon. Wea. Rev.*, **127**, 551-564.

Mahfouf, J-F., 2005: Linearization of a simple moist convection scheme for large-scale NWP models. To appear in *Mon. Wea. Rev*.

Tanguay, M. and S. Polavarapu, 1999: The adjoint of the semi-Lagrangian treatment of the passive tracer equation. *Mon. Wea. Rev.*, **127**, 551-564.

Thépaut, J-N., P. Courtier, G. Belaud and G. Lemaître, 1996: Dynamical structure functions in a four-dimensional variational assimilation: a case-study. *Q. J. R. Meteorol. Soc.*, **122**, 535-561.

Wagneur, N., C. Chouinard, L. Garand, J. Hallé, R. Sarrazin, J. St-James and G. Verner, 2004: Impact of newly available types of remote sensing data at MSC. 13th conf. on sat. meteor. and ocean., 19-23 Sept. 2004, Norfolk, VA.

Zadra, A, M. Buehner, S. Laroche and J-F. Mahfouf, 2004: Impact of the GEM model simplified physics on the extratropical singular vectors. *Q. J. R. Meteorol. Soc.*, **130**, 2541-2569.