# Impact of a Revised Analysis Algorithm on an Operational Data Assimilation System

HERSCHEL L. MITCHELL, CLÉMENT CHOUINARD, AND CÉCILIEN CHARETTE

*Direction de la Recherche en Météorologie, Atmospheric Environment Service, Dorval, Québec, Canada*

RICHARD HOGUE

*Canadian Meteorological Centre, Atmospheric Environment Service, Dorval, Québec, Canada*

STEVEN J. LAMBERT

*Canadian Centre for Climate Modelling and Analysis, Atmospheric Environment Service, Victoria, British Columbia, Canada*

(Manuscript received 30 May 1995, in final form 6 October 1995)

ABSTRACT

The improvement of analysis and data assimilation techniques can have a large impact, as shown here in the context of the Canadian global and regional data assimilation systems. Both of these systems utilize the same analysis component that was recently changed as follows: (a) a completely 3D algorithm replaced the previous split 3D scheme, which involved separate vertical and horizontal steps; (b) the assimilation of SATEM data was revised and is now done in terms of thicknesses over relatively thick layers; (c) an additional analysis level (at 925 hPa) was added and a derived temperature analysis replaced the former temperature analysis; (d) observation and forecast error statistics were revised; and (e) a correction procedure was introduced for certain types of radiosondes to offset the negative impact of solar and longwave radiation.

While many of these changes are interrelated, preventing a systematic evaluation of each in isolation, it is shown that the revised 3D algorithm eliminates a problem that sometimes occurred in areas of dense surface data, SATEM data have a large positive impact in the Southern Hemisphere, and the radiosonde bias-correction scheme very significantly reduces the geopotential height bias observed previously in the upper atmosphere over certain regions, such as western North America.

The overall evaluation of the analysis changes shows that in general the new analysis results in more accurate 6-h forecasts, with the largest improvements in the Tropics and especially in the Southern Hemisphere. In conjunction with these forecast gains, the evaluation of the general circulation statistics for August also show significant changes: the new analyses are more energetic, exhibiting a substantially stronger Hadley circulation and stronger zonal winds about Antarctica. The global forecasts from the revised analysis system consistently exhibit a significantly more rapid spinup of global precipitation as compared to the previous system.

## 1. Introduction

The importance of data assimilation has increasingly been recognized in the last two decades, not only in numerical weather prediction (NWP) but also in other areas of atmospheric and oceanic science. Using a numerical forecast model in the analysis process provides a direct link between the observations and the model's dynamics and physical parameterizations. This focuses attention on the differences between the observations and the model forecast, resulting in a more efficient use of the observations and hence in better analyses and forecasts. In addition to their primary function, the data-assimilation systems developed for NWP are capable of detecting problems in the observations (Hollingsworth et al. 1986), useful for calibrating satellite observing systems (Anderson et al. 1987), employed to reanalyze decades of data to monitor climate variability (Schubert et al. 1993), and used to study the distribution of stratospheric ozone (Rood et al. 1992). A subject of current research, 4D data assimilation techniques based on the variational approach and/or Kalman filtering, involves further strengthening the link between the observations and the numerical models (e.g., Cohn and Parrish 1991; Daley 1991; Ghil and Malanotte-Rizzoli 1991; Thépaut et al. 1993).

At the Canadian Meteorological Centre (CMC), a global data assimilation system was implemented in 1991 (Mitchell et al. 1993, hereafter MCLHC) and a regional spinup cycle has been developed for the Canadian regional model (Chouinard et al. 1994). Both the global and regional systems utilize the same multivariate, 3D, statistical interpolation procedure. Each is configured to produce analyses on the appropriate domain: globally, as of June 1993 on a 240 × 120

*Corresponding author address:* Dr. H. L. Mitchell, Atmospheric Environment Service, DRM 2121 voie de Service nord, Porte 500, Route Trans-canadienne, Dorval, PQ H9P 1J3, Canada.

(Gaussian) latitude–longitude grid (resolution about 1.5°), and regionally, as of November 1993, on a 50-km mesh covering North America (Mailhot et al. 1995).

In addition to increasing the horizontal resolution, it was suggested in the conclusion of MCLHC that several other aspects of the data assimilation system could be improved. With regard to the large impact of the initialization component of the system, a reduction of its frequency cutoff, implemented in June 1993, resulted in improvements to the tropical circulation (Fillion et al. 1995). More recently, the analysis component of the system has been changed. A fully 3D analysis procedure has been developed to replace the previous scheme, and a new ensemble of prediction and observation error statistics has been derived. Satellite soundings obtained from the *TIROS-N* Operational Vertical Sounding (TOVS) instruments (known as SATEMs) are now assimilated in terms of thickness, as reported, instead of temperature, and the observation errors for these data have been substantially revised. These changes, and several others as discussed below, were implemented in the CMC global and regional data assimilation systems on 21 September 1994. The purpose of this paper is to briefly describe these modifications and their impact.

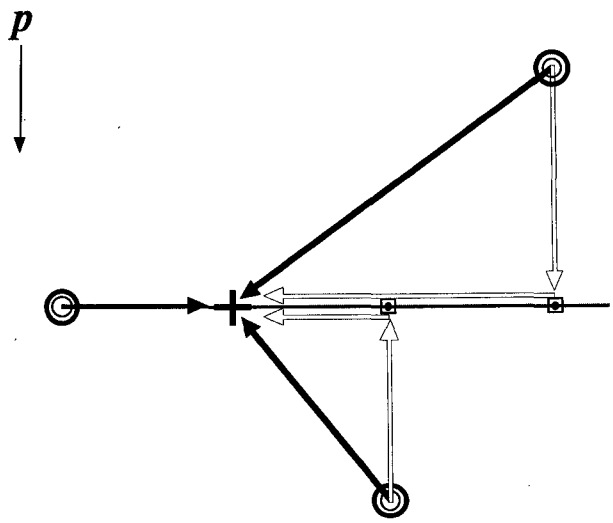


FIG. 1. Schematic diagram illustrating the direct use of off-level data in the new 3D analysis (solid arrows) as compared to the two-step approach of the split analysis (white arrows). The cross symbol indicates a grid point where the analysis is to be calculated and the circles indicate different types of observations: an observation on the analysis level (e.g., from a radiosonde) and two off-level reports (e.g., from an aircraft and a surface station). The squares are intermediate vertically interpolated values of the split scheme.

## 2. The revised analysis algorithm

### *a. Analysis procedure*

For many years, the analysis was based on the split vertical–horizontal scheme of Rutherford (1976). Unlike a fully 3D algorithm, in which observations in a 3D volume surrounding the analysis point are used directly in a single analysis step, the split scheme consisted of a vertical, followed by a horizontal, step. The vertical step served to propagate information in the vertical (i.e., produce a correction) for subsequent use in the horizontal, multivariate statistical interpolation step. As the volume and quality of single-level data and satellite data has improved over the last decade, some of the weaknesses of this approach became evident. One way to address this problem is, as in this study, to adopt a one-step 3D approach such as at the European Centre for Medium-Range Weather Forecasts (ECMWF) (Shaw et al. 1987; ECMWF 1992) and the National Meteorological Center (NMC; DiMego 1988). Unlike the ECMWF volume approach, our 3D analysis continues to proceed point by point. This allows correlation widths to vary in the vertical as observed (Hollingsworth and Lönnberg 1986; Lönnberg and Hollingsworth 1986; Mitchell et al. 1990, hereafter MCCB; Rabier and McNally 1993) and as discussed further in section 2d.

The previous two-step analysis procedure is represented by the white arrows in Fig. 1: the vertical statistical interpolation step generated interpolated values for subsequent use at a given analysis level together with the observations available on that level. The split scheme correctly handled one isolated single-level observation, yielding results identical to those of a fully 3D scheme. However, the split scheme did not account properly for the correlation between two or more vertically interpolated values. As a result, nearly coincident vertically interpolated values had very small cross correlations and effectively behaved as isolated observations. It follows that projection in the vertical could produce an amplified effect, as discussed and illustrated below. This effect was enhanced when observed increments and/or single-level data densities were large. As increasing numbers of observations from surface stations and commercial aircraft were added in support of higher-resolution analyses, this deficiency became more obvious. By contrast, the fully 3D analysis (indicated by the solid arrows) properly accounts for interdata correlations by *simultaneously* utilizing observations located above or below the level currently being analyzed without generating intermediate vertically interpolated values.

Figure 2 illustrates the effect of this change at a single analysis point. For this test, both the split and new analyses were given the same background field and all observations, except surface reports, were withheld. It can be seen that the effect of the moderately dense surface data is projected quite differently by the two analyses. In the split analysis, the effect of the surface data initially increases upward and then is set abruptly to zero where the estimated analysis error from the vertical step reaches 99% of the background error. In the fully 3D analysis, there is no need to impose such a

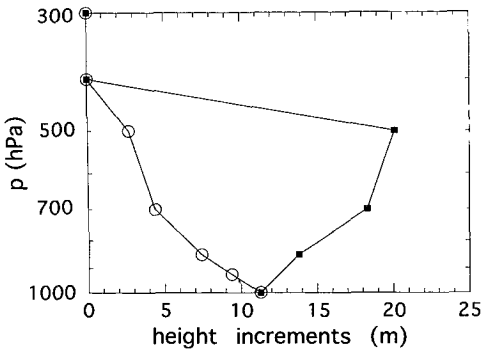


FIG. 2. Analyzed height increments as a function of pressure at a single analysis point in an area of moderately dense surface reports (ten reports within a circle of radius 350 km) for the split (solid squares) and the new 3D (open circles) analysis schemes.

condition, and the effect of the surface data decreases in the vertical following the vertical correlations.

*b. Assimilation of SATEM data*

SATEM satellite data are transmitted in the form of layer thickness profiles (1000–850, 1000–700, 1000–500 hPa, etc.) derived from *NOAA-11* and *NOAA-12* radiances. In the previous analysis procedure, these were converted to temperatures at all analysis levels for assimilation. In the revised analysis procedure, each SATEM profile is reduced to a set of seven thicknesses (1000–700, 700–500, 500–300, 300–100, 100–50, 50–30, 30–10 hPa) for direct use in the multivariate analysis, following Kelly and Pailleux (1988). As in the previous analysis procedure, SATEM data are not used below 100 hPa over continental areas. SATEM observation error variance estimates, which had not been updated for many years, were expressed in terms of thickness and reduced to reflect recent improvements introduced at NESDIS (National Environmental Satellite, Data, and Information Service) in the preparation of SATEMs (Reale et al. 1993). Depending on the indicator flag in each SATEM report (clear or cloudy), a different set of observation errors are utilized, as specified in Table 1. The seven-layer vertical correlations of SATEM observation errors are specified following ECMWF (1992). Unlike the previous scheme, the revised analysis procedure allows for the horizontal correlation of SATEM observation errors. It is assumed that 80% of the observation error is correlated and, as in Andersson et al. (1994), the horizontal correlation is modeled by a Gaussian function with a length scale of 350 km. While this feature may not be important for assimilating 500-km resolution SATEMs (as is done currently), it has proven of value in recent tests with 250-km resolution SATEM data (not shown).

*c. Vertical resolution and derived temperature analysis*

In the previous scheme, both height and temperature were analyzed. The height increments were used to produce the surface pressure $p_s$ increments, while both height and temperature increments were used in the vertical-interpolation-of-increments step to produce the model mass variable (virtual temperature $T_v$) in $\sigma$ ($=p/p_s$) coordinates. Historically, the main motivation for the temperature analysis was for assimilating the SATEM data. Since in the new analysis SATEMs are assimilated in terms of thicknesses instead of temperature and given that the temperature data in a radiosonde report are integrated to produce the geopotential height data, there is little need to reuse temperature data from soundings in the analysis. Moreover, while the temperature analysis was of value in the vertical-interpolation-of-increments procedure (by providing information about the slope of the height increments profile), it was felt that increasing the vertical resolution of the height analysis would be equally, if not more, beneficial. As a first step in this direction, 925 hPa was added as an analysis level. Consequently, the analysis of temperature and the use of temperature observations were dropped. While at present this implies a loss of some information (e.g., from aircraft reports), single-level temperatures could be assimilated as thicknesses (analogous to our treatment of SATEM data) over the layer formed by the analysis level above and below each observation. However, this possibility has not yet been investigated.

For operational purposes, a diagnostic temperature "analysis" is produced by interpolating the $\sigma$-level virtual temperatures to $p$ coordinates. In mountainous regions, the virtual temperature below the surface is derived hydrostatically from the isobaric height analyses. In each case, a humidity correction is applied yielding the final isobaric temperature analysis.

*d. Error statistics*

The three-dimensional correlations required in the 3D analysis are evaluated assuming horizontal–vertical separability. The fact that the 3D analysis proceeds point by point in the vertical allows for vertical variation of the horizontal correlation function width, that is, at a given level the analysis uses the function width appropriate for that level, as in Rutherford (1976). This is in contrast to the volume approach that gives better

TABLE 1. Standard deviation (m) of SATEM thickness observation errors for each of the seven layers assimilated.

| Thickness layer (hPa) | Clear | Cloudy |
|---|---|---|
| 10–30 | 32 | 32 |
| 30–50 | 16 | 16 |
| 50–100 | 16 | 16 |
| 100–300 | 24 | 30 |
| 300–500 | 24 | 40 |
| 500–700 | 14 | 25 |
| 700–1000 | 14 | 25 |

horizontal and vertical consistency, but, when coupled with the separability hypothesis, requires that the horizontal correlation width be constant in the vertical. Ideally, one would like to use a nonseparable correlation model (see, e.g., Bartello and Mitchell 1992), in conjunction with the volume approach.

Horizontal correlations continue to be modeled using autoregressive functions, as described in MCCB, but these were reparameterized using current innovations (i.e., differences between observations and background) from the global assimilation system. Similarly, forecast and radiosonde observation errors and vertical correlations were updated. These had originally been determined from 1986–87 innovations of a lower-resolution hemispheric system, as described in MCCB, and were later applied in the global system as described in MCLHC. An attempt was made to determine revised error statistics not only for the Northern Hemisphere but also for the Tropics and the Southern Hemisphere of the global system. Due to a problem with radiosonde observations in the upper atmosphere over North America (see next subsection), the Northern Hemisphere statistics were not determined solely on the basis of North American radiosonde innovation sequences, as in MCCB. Also since our error statistics are determined using the available radiosonde network, our error estimates in the radiosonde-sparse Tropics and Southern Hemisphere are much less reliable than in the Northern Hemisphere.

The changes to the error statistics are illustrated in Tables 2 and 3, which show both previous and revised, height and wind, forecast and radiosonde observation errors, respectively. The tables show midlatitude values in both hemispheres for January and July at a selected number of analysis levels. Looking first at the January statistics for the Northern Hemisphere, it can be seen that overall there has been a significant reduction in the background errors, in conjunction with a much smaller increase in the observation errors. In July, the Northern Hemisphere errors are also reduced: substantially in the upper levels and moderately below the tropopause. These reductions are indicative of, and consistent with, the improved accuracy of 6-h forecasts since 1986–87. The error statistics used previously in the Southern Hemisphere had been set to those of the Northern Hemisphere, as described in MCLHC. As can be seen from the tables, the revised errors for January are, in general, significantly smaller in the upper levels and

TABLE 2. Forecast-error standard deviations for height (m) and wind (m $s^{-1}$) at a selected number of analysis levels for midlatitude (30°–60°) Northern and Southern Hemispheres for January and July. Both the previous (denoted old) and revised (denoted new) standard deviations are shown. Wind errors are for wind components.

| | Northern Hemisphere | | | | Southern Hemisphere | | | |
|---|---|---|---|---|---|---|---|---|
| | Height (m) | | Wind (m $s^{-1}$) | | Height (m) | | Wind (m $s^{-1}$) | |
| Level (hPa) | Old | New | Old | New | Old | New | Old | New |
| | | | | January | | | | |
| 50 | 39.5 | 32.1 | 5.3 | 3.5 | 33.6 | 18.4 | 4.0 | 3.2 |
| 100 | 26.3 | 24.3 | 4.8 | 3.7 | 26.6 | 17.6 | 3.9 | 4.5 |
| 300 | 22.9 | 17.5 | 6.9 | 6.0 | 17.3 | 19.4 | 6.0 | 6.4 |
| 500 | 14.6 | 10.9 | 4.9 | 4.4 | 10.3 | 12.4 | 3.5 | 4.8 |
| 700 | 9.7 | 7.6 | 3.8 | 3.3 | 7.9 | 9.6 | 3.5 | 4.0 |
| 1000 | 9.8 | 7.2 | 2.9 | 3.2 | 7.8 | 8.5 | 3.0 | 3.7 |
| | | | | July | | | | |
| 50 | 33.6 | 21.6 | 4.0 | 2.3 | 39.5 | 18.4 | 5.3 | 3.2 |
| 100 | 26.6 | 17.2 | 3.9 | 2.7 | 26.3 | 17.6 | 4.8 | 4.5 |
| 300 | 17.3 | 14.8 | 6.0 | 5.7 | 22.9 | 19.4 | 6.9 | 6.4 |
| 500 | 10.3 | 9.2 | 3.5 | 3.5 | 14.6 | 12.4 | 4.9 | 4.8 |
| 700 | 7.9 | 6.7 | 3.5 | 3.1 | 9.7 | 9.6 | 3.8 | 4.0 |
| 1000 | 7.8 | 6.1 | 3.0 | 2.6 | 9.8 | 8.5 | 2.9 | 3.7 |

TABLE 3. As in Table 2 but for radiosonde-observation-error standard deviations.

| | Northern Hemisphere | | | | Southern Hemisphere | | | |
|---|---|---|---|---|---|---|---|---|
| | Height (m) | | Wind (m $s^{-1}$) | | Height (m) | | Wind (m $s^{-1}$) | |
| Level (hPa) | Old | New | Old | New | Old | New | Old | New |
| | | | | January | | | | |
| 50 | 23.2 | 26.8 | 3.5 | 3.5 | 24.1 | 16.7 | 2.3 | 2.3 |
| 100 | 18.7 | 21.0 | 3.5 | 3.6 | 19.1 | 14.0 | 2.1 | 3.2 |
| 300 | 10.8 | 13.5 | 3.6 | 3.8 | 12.4 | 11.1 | 3.2 | 4.3 |
| 500 | 7.6 | 8.8 | 3.1 | 3.3 | 7.4 | 7.4 | 3.2 | 3.9 |
| 700 | 5.4 | 6.7 | 2.8 | 3.1 | 4.9 | 4.8 | 2.6 | 3.2 |
| 1000 | 5.9 | 5.9 | 2.6 | 3.1 | 5.6 | 6.6 | 2.6 | 3.0 |
| | | | | July | | | | |
| 50 | 24.1 | 17.7 | 2.3 | 1.9 | 23.2 | 16.7 | 3.5 | 2.3 |
| 100 | 19.1 | 14.1 | 2.1 | 2.1 | 18.7 | 14.0 | 3.5 | 3.2 |
| 300 | 12.4 | 11.1 | 3.2 | 2.9 | 10.8 | 11.1 | 3.6 | 4.3 |
| 500 | 7.4 | 6.9 | 3.2 | 2.4 | 7.6 | 7.4 | 3.1 | 3.9 |
| 700 | 4.9 | 5.0 | 2.6 | 2.4 | 5.4 | 4.8 | 2.8 | 3.2 |
| 1000 | 5.6 | 5.0 | 2.6 | 2.3 | 5.9 | 6.6 | 2.6 | 3.0 |

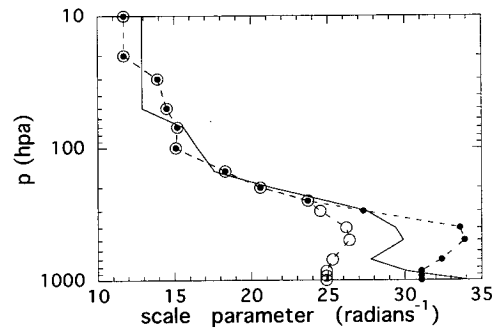


FIG. 3. Vertical profiles of the horizontal scale parameter. The solid curve represents the midlatitude values used previously in both hemispheres, whereas the dashed curves joining the dots and open circles show the revised scale parameters of the Northern and Southern Hemispheres, respectively. Units are inverse radians.

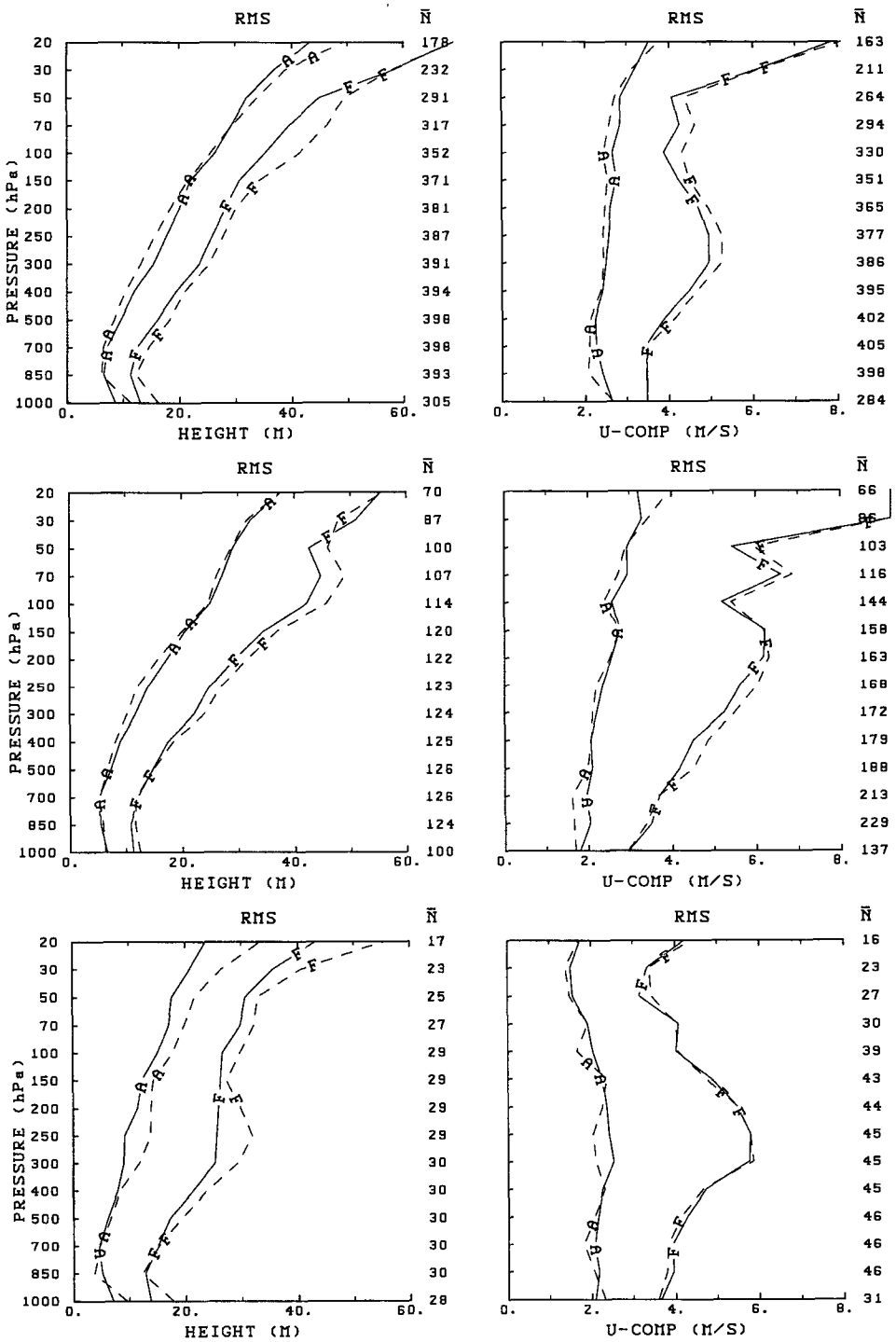


FIG. 4. Vertical profiles of total rms differences of O-A (denoted A) and O-F (denoted F) for radiosonde observations for the period 17–27 December 1993. Left-hand panels are for height (m) and right-hand panels are for zonal wind component (m $s^{-1}$). Upper panels are for the Northern Hemisphere (30°–90°N), middle panels are for the Tropics (30°S–30°N), and bottom panels are for the Southern Hemisphere (30°–90°S). Dashed and solid curves are from the assimilation cycles with the previous and revised assimilation systems, respectively. For each level, $\bar{N}$ indicates the average number of radiosonde observations over the corresponding region per main synoptic hour.

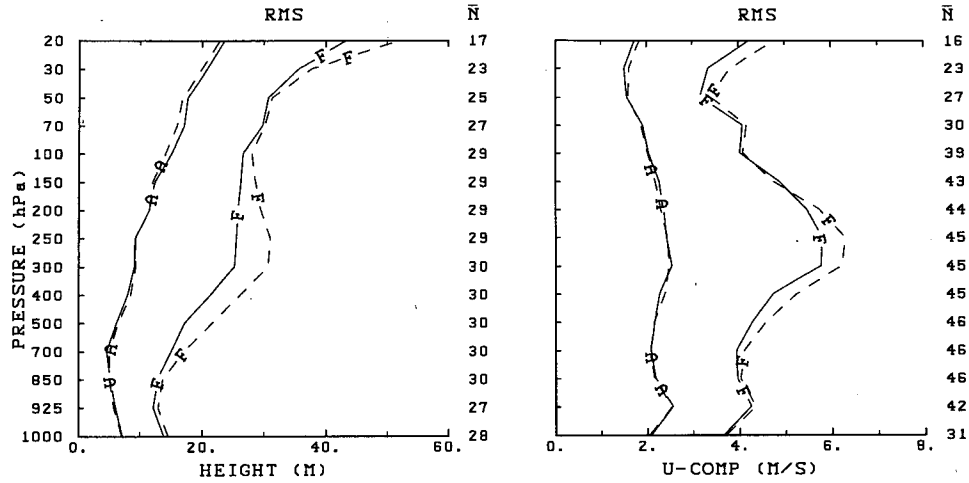


FIG. 5. As in Fig. 4 but solid curves are from the assimilation cycle with SATEM reports, while dashed curves are from the assimilation cycle with no SATEM reports. The region is 30°–90°S and all results were obtained using the revised assimilation system.

somewhat larger below. Revised July error statistics for the Southern Hemisphere were also generated, but tests with these yielded results that were inferior to the results obtained with January statistics, and consequently it was decided to use the same Southern Hemisphere statistics all year round.

Observation errors for other types of observations are as specified in Table 1 of MCCB except, to make the 1000-hPa analysis more responsive to surface reports, the observation error assigned to such reports was reduced to 6 m.

Figure 3 illustrates the impact that reparameterizing the statistics had on the horizontal scale parameter. This parameter, which determines the horizontal scale of an observation's influence, appears in the expression for the autoregressive function used to represent the horizontal correlations, as discussed in MCCB. Since, in fact, this parameter is inversely proportional to the horizontal correlation width, it can be seen from the figure that below 250 hPa the horizontal correlations have become narrower in the Northern Hemisphere and broader in the Southern Hemisphere. Note, in general, how much broader the horizontal correlations are in the stratosphere as compared to the troposphere.

### *e. Corrections to radiosonde reports for radiation effects*

Early tests of the revised analysis highlighted an unrealistic diurnal oscillation in the analyses of the upper atmosphere over certain regions. This was found to be due to the effects of solar and infrared radiation on the actual radiosonde instruments (e.g., McMillin et al. 1992, and references therein). Only reports from stations using certain types of radiosonde equipment showed this effect, which varied in magnitude with solar angle. This is well known to radiosonde instrument manufacturers and some (e.g., Vaisala, see Turtiainen 1993) apply a correction to the raw measurements. The effect was found to be particularly large over western North America. Consequently, while the 3D analysis was being tested in parallel at the CMC, a radiosonde bias-correction scheme was introduced into the parallel run. The scheme, based on the one developed at ECMWF, corrects only radiosonde heights originating from stations using certain types of radiosonde equipment (B. Strauss, ECMWF, and G. Verner, CMC 1994, personal communication). It subsequently became part of the revised analysis system.

## 3. Results

The effect of the revised analysis in the data assimilation system is evaluated in a manner similar to that of MCLHC. That is, we consider the impact (a) on verifications against radiosonde data of the 6-h forecasts used as the background (or prior estimate) and the analyses, (b) on various aspects of the global general circulation as defined by the analyses, and (c) on the forecasts beyond 6 h. As in MCLHC, O denotes the observations and A and F denote the analyses and 6-h forecasts interpolated to the observation locations. A T119 global spectral model with 21 levels (Ritchie and Beaudoin 1994) is used to drive the global assimilation system. A 50-km version of the regional finite-element model with 25 levels (Mailhot et al. 1995) provides the 6-h forecasts for the regional spinup cycle.

### *a. Verifications of the analyses and 6-h forecasts against radiosonde data*

The first set of results are from a test period during Northern Hemisphere winter (13–27 December 1993). All results for this period use radiosonde ob-

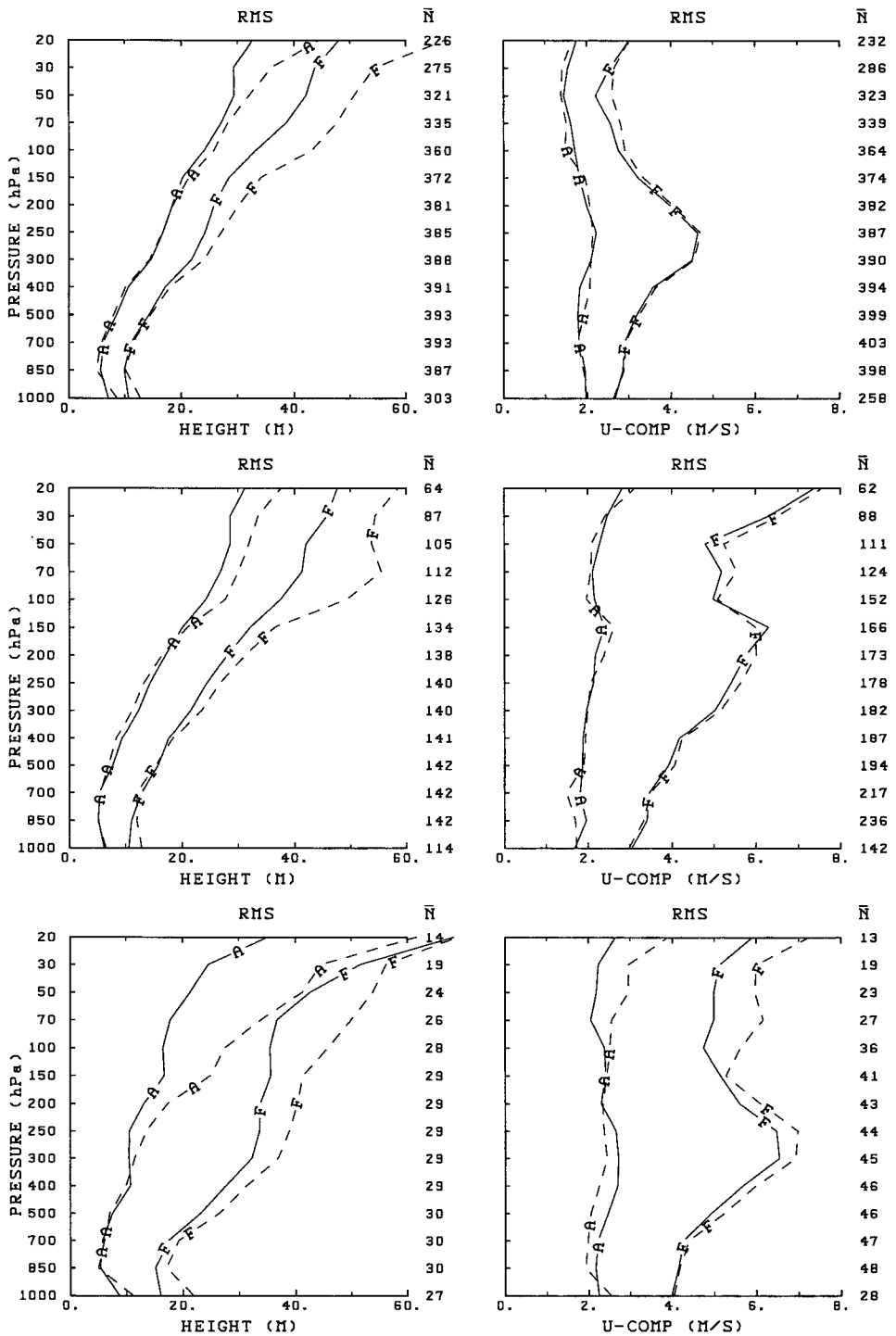


FIG. 6. As in Fig. 4 but for the month of August 1994.

servations *uncorrected* for radiation effects, which allows an evaluation of the impact of the revised analysis algorithm itself. Results are shown for the last 10 days of the period to allow the system a few days to adjust to the revised analysis. Figure 4 shows vertical profiles of root-mean-square (rms) differences of O-A and O-F for radiosonde observations for this period. Results from the assimilation cycles with the previous and revised systems are shown for both height and zonal wind for the regions: 30°–90°N (top panels), 30°S–30°N

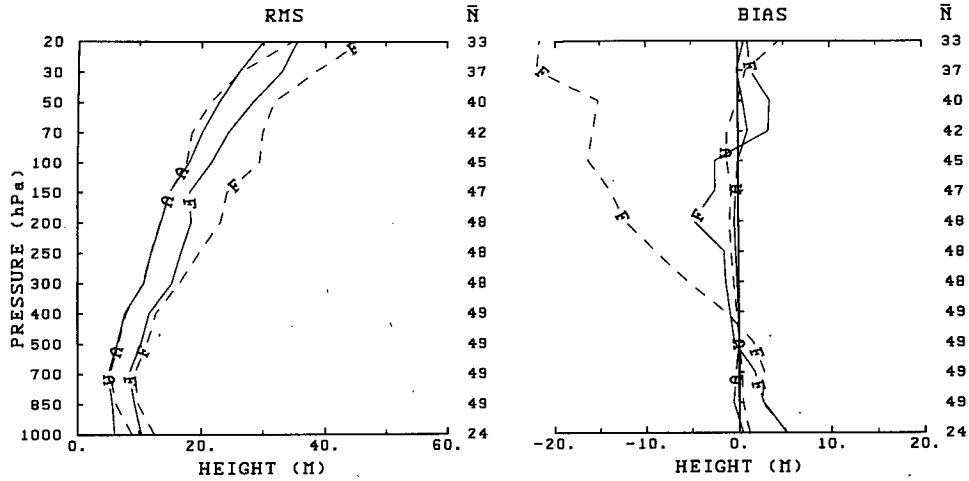


FIG. 7. Vertical profiles of O-A (denoted A) and O-F (denoted F) height (m) statistics for western North American (30°–60°N) radiosonde observations at 1200 UTC for August 1994. Left-hand panel is for total rms differences and right-hand panel is for the bias component. Dashed and solid curves are from the assimilation cycles with the previous and revised assimilation systems, respectively. For each level, $\bar{N}$ indicates the average number of radiosonde observations over the region per main synoptic hour.

(middle panels), and 30°–90°S (bottom panels). The differences between the two assimilation cycles for the meridional wind component (not shown) are similar to those for the zonal component.

It can be seen from Fig. 4 that the O-A statistics (i.e., the fit to the radiosonde data) have not been improved in all regions (e.g., Northern Hemisphere), consistent with the previously noted reduction in the forecast error and relatively smaller increase in the observation error. Overall there has been a significant improvement in the 6-h forecasts, which is especially evident in the Northern and Southern Hemisphere heights. As argued in MCCB, the O-F statistic is a better indicator of assimilation system performance, as reductions in this statistic generally lead to improved forecasts beyond 6 h. Note also the substantial improvement in the height analyses and 6-h forecasts at 1000 hPa. We attribute this to the improved assimilation of single-level (surface) reports and the reduced surface observation errors and not to the additional analysis level at 925 hPa, on the basis of the results of a separate assimilation cycle (not shown).

The revisions to the treatment of SATEM data motivated a particular interest in the impact of these data. However, there are difficulties in evaluating the impact of these data using radiosonde verifications, since over land SATEM data are only used above 100 hPa. These difficulties are especially pronounced in the Northern Hemisphere with its large continents and relatively dense radiosonde network. By contrast, the Southern Hemisphere is dominated by oceans, and therefore the volume and potential impact of SATEM data is much larger and more easily detected by the relatively sparse radiosonde network.

In fact, SATEM data were found to be of benefit in both hemispheres above 100 hPa, and their overall impact was particularly striking in the Southern Hemisphere. This is illustrated in Fig. 5, which shows results for the 30°–90°S region. It can be seen that in this region the benefits due to SATEM data extend vertically throughout the atmosphere and are largest around 300 hPa. Note also the large impact on the winds around the jet level induced by the SATEM thickness data alone. These results are generally consistent with the results of other recent studies (e.g., Kelly et al. 1993; Reale et al. 1993).

Results for August 1994, a period during Northern Hemisphere summer and the last complete month of the parallel run, are now presented. During this period, the radiation correction scheme for radiosonde data was in use with the revised analysis. With regard to the verifications against radiosonde data of the analyses and 6-h forecasts for this period (Figs. 6, 7, and 8), it should be noted that while the revised system is verified against *corrected* radiosonde data, the previous system is verified against *uncorrected* radiosonde data, that is, each system is verified against the radiosonde data that it uses in an internally consistent manner. However, during the period of the parallel run, there were no Southern Hemisphere stations using radiosonde equipment subject to radiation correction at the CMC, and therefore in the Southern Hemisphere winds *and* heights from both assimilation systems were being verified against the same (uncorrected) radiosonde data.

Figure 6 presents vertical profiles of rms differences of O-A and O-F for radiosonde observations for the month of August 1994. The format is the same as that used in Fig. 4. It can be seen that there has been a very

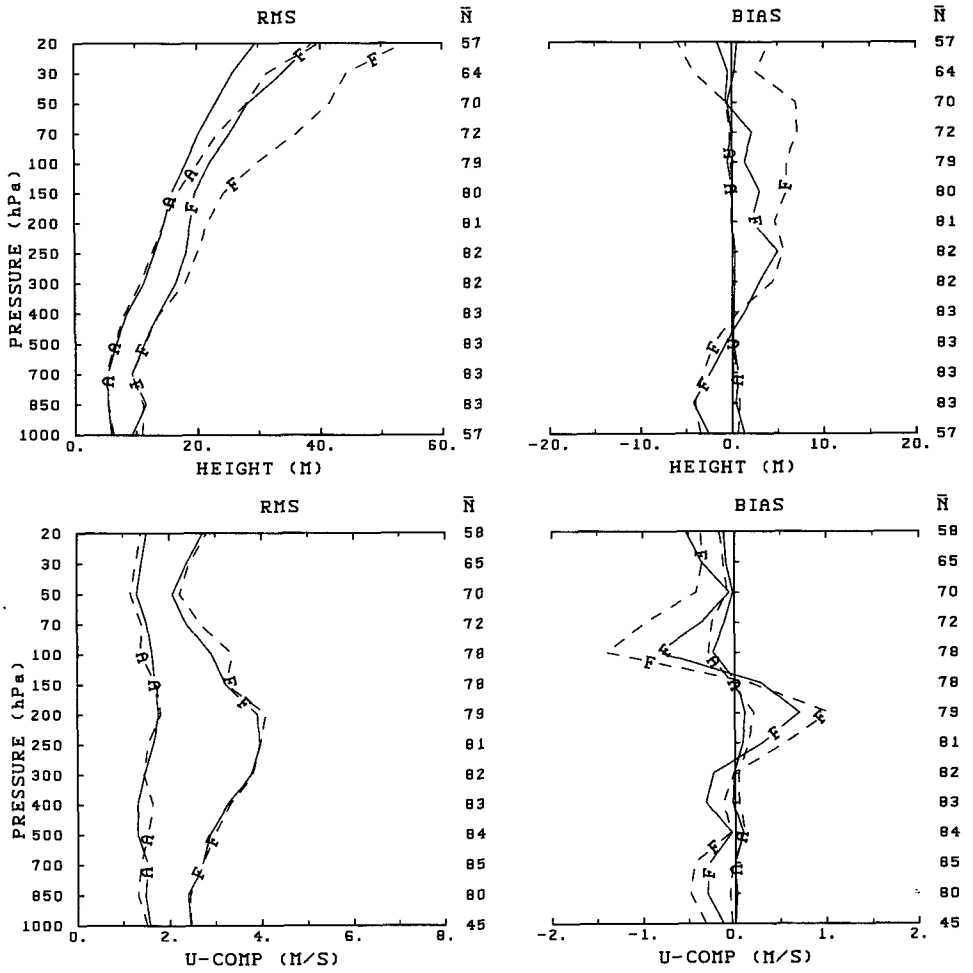


FIG. 8. As in Figs. 4 and 6 but from the regional spinup cycle for North American (30°–60°N) radiosonde observations at 0000 UTC for 1–15 September 1994. Left-hand panels are for total rms differences and right-hand panels are for the bias component. Upper panels are for height (m) and lower panels are for zonal wind component (m $s^{-1}$).

significant improvement in the analyses and 6-h forecasts in the Southern Hemisphere (the winter hemisphere) and at 1000 hPa in all regions. Since no radiosonde data was being corrected in the Southern Hemisphere, the improvements there are strictly due to the revised analysis algorithm. The improvement in the upper atmosphere of the Northern Hemisphere and the Tropics shows that there is much greater consistency between the analyses and 6-h forecasts of the revised system and the (corrected) radiosonde observations than was the case previously. This is due to a combination of the revised analysis algorithm and the corrections to the verifying radiosonde observations.

The impact of the radiation correction scheme can be seen in Fig. 7, which focuses on western North America at 1200 UTC (approximately 6 h before noon, local time). Both the total rms differences and bias component are shown. It can be seen that with the previous analysis the bias profile of the 6-h forecast is dominated by a significant negative bias with amplitude increasing upward. There are two causes for this negative bias of O-F. First, the reported heights at this time may be too low due to infrared radiative cooling of the instrument. A further and larger effect is due to solar heating of the instrument 12 h earlier in midafternoon. This heating can produce reported heights which are too high, yielding a subsequent forecast 12 h later, which is also too high. By contrast, it can be seen that the revised analysis system, which incorporates the radiation correction scheme for radiosonde data, yields a very small bias in the 6-h forecast.

Following an extended examination of the revised analysis in the context of the global assimilation system, attention turned to the regional spinup cycle, which uses the same unified analysis algorithm. To evaluate the impact of the revised regional analysis, a parallel regional spinup cycle was performed once per

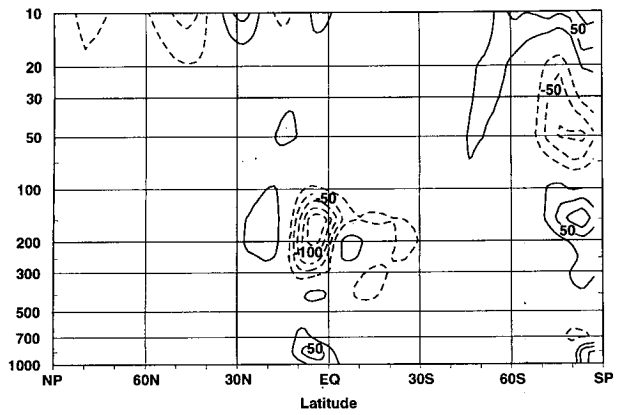


FIG. 9. The zonally and temporally averaged differences (new CMC minus old CMC) for the meridional wind (cm $s^{-1}$) for August 1994. The contour interval is 25 with the zero contour suppressed.

day to produce a final regional analysis at 0000 UTC. Figure 8 shows vertical profiles of rms differences of O-A and O-F for radiosonde observations for the period 1–15 September 1994. The rms results (left-hand panels) are seen to be consistent with the corresponding global system results for the Northern Hemisphere shown in Fig. 6. As shown in the right-hand panels, the biases have been reduced substantially. The 6-h zonal wind bias has been reduced but shows the same structure as before.

### *b. Global general circulation from the analyses*

In order to examine changes in the large-scale circulation and its statistics resulting from the new analysis, a lower-resolution subset of the analyses was archived. The 240 × 120 point Gaussian grids holding the global analyses were transformed to spectral coefficients with a triangular resolution at 32 waves. The statistics that follow were obtained by synthesis of these spectral coefficients on a 96 × 48 point Gaussian grid.

The new version of the analysis system addresses some of the difficulties identified in MCLHC. The new assimilation system produces a stronger Hadley circulation as revealed by Fig. 9, which displays the zonally and temporally averaged differences in the meridional wind produced by the revised and previous systems for August 1994. In equatorial regions the new analyses exhibit substantially increased low-level northerly flow and upper-tropospheric southerly flow. Since these changes reinforce the Hadley circulation itself (not shown), the new analysis system produces a significantly more intense Hadley circulation.

The comparison in MCLHC of the zonal wind between the old CMC analyses and those of the NMC for July showed that the CMC analyses exhibited weaker zonal winds in the southern polar regions than did the NMC. Subsequent comparison (not shown) indicated agreement between the NMC and the ECMWF analyses. This suggests that the old CMC analysis of the southern polar zonal winds was too weak. Figure 10 displays the differences between the new and the old CMC analysis for the zonal wind. The noticeable strengthening of the zonal wind over Antarctica is likely an improvement in the CMC analyses.

The differences for the zonally and temporally averaged temperature field are displayed in Fig. 11. The colder temperatures at about 300 hPa in the southern polar regions are consistent with the zonal wind increases shown in Fig. 10. The new analyses tend to exhibit a slightly cooler tropopause where operational analyses tend to be too warm. The colder temperatures above 20 hPa are likely due to the revised SATEM treatment. In any case, the model resolution above 50 hPa is very coarse, and these results are likely to change when the top of the model is raised and its resolution in this region improved.

Figure 12 displays the temporally averaged and vertically integrated global energy cycles (Lorenz 1955) for the new and the old CMC analyses. The new analyses are more energetic as evidenced by increased eddy available potential energy (EAPE) to eddy kinetic energy (EKE) conversion and higher levels of kinetic energy. There is a particularly striking increase in the zonal available potential energy (ZAPE) to zonal kinetic energy (ZKE) conversion. This term has a negative contribution from the Ferrel cells and a positive contribution from the Hadley circulation. The higher values of this conversion reflect the improved analysis of the Hadley circulation.

Analysis changes can introduce discontinuities in statistics computed from operational analyses. Table 4 displays the hemispheric rms differences between the old and new CMC assimilation systems. These differences are about two-thirds the rms differences between the CMC and the NMC analyses given in MCLHC. The magnitude of these differences underlines the major impact of this analysis change.

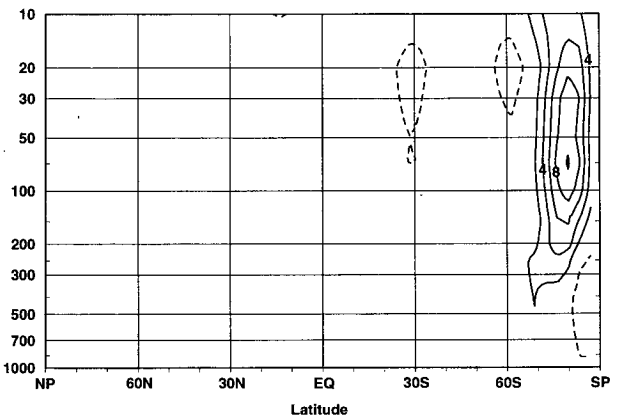


FIG. 10. As in Fig. 9 but for the zonal wind (m $s^{-1}$) and a contour interval of 2.

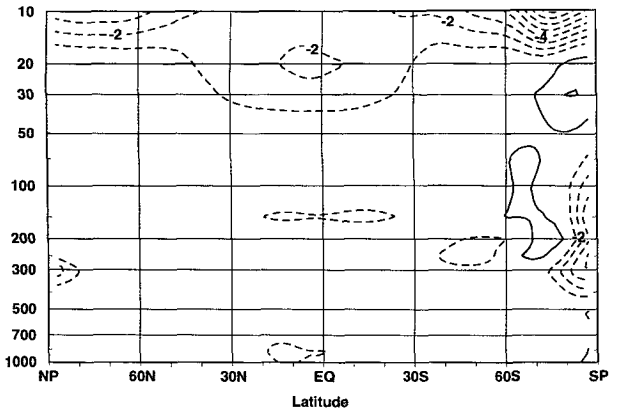


FIG. 11. As in Fig. 9 but for temperature (K) and a contour interval of 1.

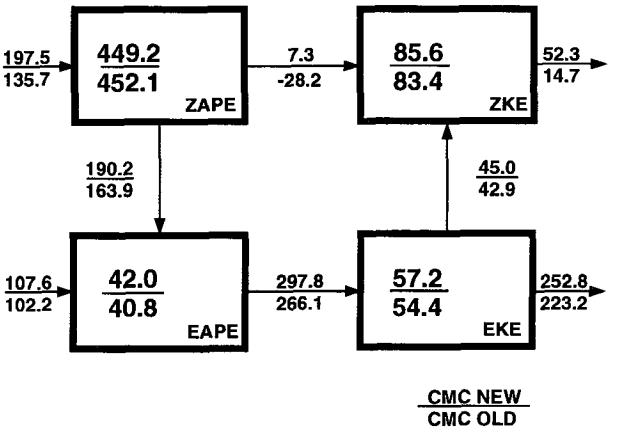


FIG. 12. Lorenz energy cycle for the old CMC and the new CMC analyses for August 1994. The values for the new analysis system are plotted above the values for the old system. The energies are in units of $10^4$ J m$^{-2}$ and the conversions are in units of $10^{-2}$ W m$^{-2}$.

### c. *Comments and evaluation of the forecasts beyond 6 h*

During the parallel run, medium-range global forecasts were produced and compared with the operational forecasts. In general, the results of this evaluation were consistent with the results of sections 3a and 3b, that is, it was found that the forecasts were generally of equivalent quality in the Northern Hemisphere but superior in the Tropics and especially so in the Southern Hemisphere. Pacific disturbances approaching the west coast of North America were usually slightly better forecast and, on occasion, significantly better forecast, as compared to the operational system.

The global forecasts from the revised analysis system consistently exhibited a significantly more rapid spinup of global precipitation as compared to the operational system. This is illustrated in Fig. 13, which shows the globally averaged rain rate as a function of time for forecasts with the operational model from the parallel and then-operational analyses for 0000 UTC 20 August 1994. An examination of the rain-rate spatial distribution (not shown) for both forecasts indicated that the largest differences occurred about the intertropical convergence zone (ITCZ). These differences appeared to be related to the stronger Hadley circulation noted above. Significant differences were also observed as far north and south of the equator as 40°N and 40°S, presumably due to improved static stability in the lowest layers. The fact that with the revised analysis the rain rate reaches the long-term mean in 6 h, whereas with the older analysis it took over 12 h to do so, is very significant in the context of an intermittent data assimilation system operating in a 6-h cycle.

## 4. Summary and conclusions

To provide appropriate initial conditions to its global and regional forecast models, a global data assimilation system and a regional spinup cycle have been in operation at the Canadian Meteorological Centre (CMC) for several years. These both utilize the same multivariate, 3D, statistical interpolation procedure, configured in each case, to produce analyses on the appropriate domain. A revised analysis component, with major modifications as listed in the abstract, was recently implemented after being tested on separate summer and

TABLE 4. Hemispheric root-mean-square differences between the new and the old CMC analyses for August 1994 for the zonal wind $u$, the meridional wind $v$, the geopotential $\phi$, the temperature $T$, and the specific humidity $q$.

| Level (hPa) | $u$ (m s$^{-1}$) | $v$ (m s$^{-1}$) | $\phi$ (m$^2$ s$^{-2}$) | $T$ (K) | $q$ (g kg$^{-1}$) |
|---|---|---|---|---|---|
| Northern Hemisphere | | | | | |
| 50 | 1.6 | 1.2 | 230 | 0.7 | |
| 200 | 3.1 | 2.9 | 119 | 1.1 | |
| 500 | 2.3 | 1.9 | 77 | 0.8 | 0.44 |
| 850 | 1.4 | 1.2 | 41 | 1.2 | 1.22 |
| Southern Hemisphere | | | | | |
| 50 | 3.4 | 2.6 | 463 | 0.9 | |
| 200 | 4.9 | 4.6 | 349 | 1.8 | |
| 500 | 3.4 | 3.2 | 268 | 1.2 | 0.30 |
| 850 | 2.5 | 2.3 | 230 | 1.6 | 1.06 |

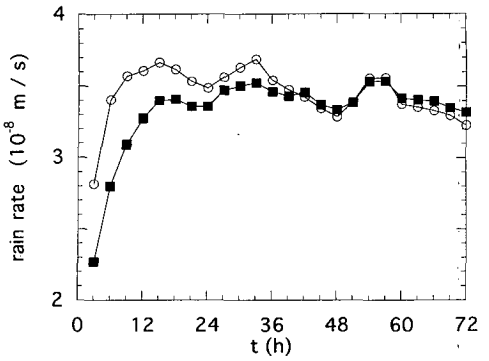


FIG. 13. Global rain rate for 72-h forecasts from 0000 UTC 20 August 1994. The solid squares and open circles are from the forecasts from the previous and revised analysis systems, respectively. (Note that $10^{-8}$ m s$^{-1}$ = 0.864 mm day$^{-1}$.)

winter periods. The fact that the revised analysis continues to proceed point by point allows correlation widths to vary with height as observed and as calculated from a recent set of innovations. This can be viewed as an intermediate step toward the use of a fully nonseparable correlation model.

The revised analysis has improved the assimilation of single-level reports and has consistently improved performance at 1000 hPa. Overall, the impact is moderately positive in the Northern Hemisphere and is more markedly positive in the Tropics and especially in the Southern Hemisphere. In that region, where the gains have been very substantial, it has been shown that the SATEMs have a large beneficial impact.

The revised system was evaluated against its predecessor during a parallel run in the summer of 1994. The rms and biases as verified against radiosonde data clearly show improvement in the analyses and 6-h forecasts, which persist beyond 6 h. As indicated by a set of global circulation statistics, the Hadley circulation is significantly strengthened, resulting in a much improved spinup of globally averaged precipitation. In addition, and consistent with the more energetic overall circulation, the revised analysis also exhibited stronger zonal winds about Antarctica.

The large impact of this analysis change is illustrated by the magnitude of the hemispheric rms differences between the old and new CMC assimilation systems for August 1994. These differences are about two-thirds the rms differences between the CMC and the NMC analyses given in Mitchell et al. (1993). It is also important to note that the changes to the CMC assimilation system did not involve changes to the assimilating model. If, as is often done, the model had been changed along with the analysis techniques, even larger rms differences would likely have resulted. For purposes of climate and other diagnostic studies, this reinforces the need for "reanalysis" projects in which long series of data are assimilated using a fixed data assimilation system.

Having addressed the problem that the previous analysis had in areas of dense single-level data, further improvements can now be undertaken. These can be expected to benefit both the global and regional data assimilation systems but, being aimed at improving analysis resolution, will perhaps have their largest impact in the regional system. Thus, additional data types can now be assimilated, including surface aviation (SA) reports (at present only SM reports are used), and ARINC (Aeronautical Radio, Inc.) Communication Addressing and Reporting System (ACARS) data. As indicated in section 2c, it would also be desirable to increase the number of analysis levels by using significant-level data from radiosondes. Most of the preliminary work for this has already been done in the context of the preparation of analyses for a mesoscale-model intercomparison project (Chouinard et al. 1994; Chouinard and Staniforth 1995).

In the Northern Hemisphere, the revisions to the statistics reflect the measured reduction in 6-h forecast errors. However, in the Southern Hemisphere, the use of the sparse radiosonde network led to problems in estimating the forecast-error statistics. In any case, forecast model resolution is continuously improving, and the classical approach of determining forecast-error correlation structures and variances based on innovations calculated from radiosonde observations is limited by the effective resolution of the network. One alternative approach for estimating forecast-error covariances consists in looking at the differences between 24- and 48-h forecasts valid at the same time (Derber et al. 1991; Rabier and McNally 1993). System simulation, using perturbed observations and variants of a data assimilation system, is another approach to forecast-error estimation (Houtekamer et al. 1996). These two methods, not directly related to any network distribution, are presently being considered for revising the forecast-error statistics for the CMC regional and global data assimilation systems as well as a future Canadian global 3D variational analysis.

*Acknowledgments.* As with all major data assimilation system changes, the development and implementation of the revised 3D analysis would not have been possible without the participation of many individuals. Particular thanks are due to the following members of Direction de la recherche en météorologie (DRM) and the Canadian Meteorological Centre (CMC): Roop Lalbeharry and Christopher Grassotti, for their contributions to the revised SATEM treatment; Michel Valin, for his valuable suggestions for optimizing the computer code; and Jean-Guy Desmarais, Richard Jones, Gilles Verner, and the other members of CMC's Analyses and Prognoses, Implementation, and Monitoring and Quality Control Sections.

We would also like to thank Peter Houtekamer, Andrew Staniforth and the anonymous referees for their thoughtful reviews of the manuscript; Luc Fillion for his comments and interest; André Plante who, while

monitoring the parallel run, first noted the more rapid spinup of global precipitation; Claude Girard for discussions about the globally averaged precipitation; and Yves Chartier and Mario Lépine for their help with the figures.

## REFERENCES


Anderson, D., A. Hollingsworth, S. Uppala, and P. Woiceshyn, 1987: A study of the feasibility of using sea and wind information from the ERS-1 satellite. Part I: Wind scatterometer data. ECMWF contract report to the European Space Agency, 125 pp. [Available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England.]

Andersson, E., J. Pailleux, J.-N. Thépaut, J. R. Eyre, A. P. McNally, G. A. Kelly, and P. Courtier, 1994: Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **120,** 627–653.

Bartello, P., and H. L. Mitchell, 1992: A continuous three-dimensional model of short-range forecast error covariances. *Tellus,* **44A,** 217–235.

Chouinard, C., and A. Staniforth, 1995: Deriving significant-level geopotentials from radiosonde reports. *Mon. Wea. Rev.*, **123,** 222–229.

——, J. Mailhot, H. L. Mitchell, A. Staniforth, and R. Hogue, 1994: The Canadian regional data assimilation system: Operational and research applications. *Mon. Wea. Rev.*, **122,** 1306–1325.

Cohn, S. E., and D. F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119,** 1757–1785.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Derber, J. C., D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Wea. Forecasting,* **6,** 538–547.

DiMego, G. J., 1988: The National Meteorological Center regional analysis system. *Mon. Wea. Rev.*, **116,** 977–1000.

ECMWF Research Department, 1992: ECMWF Data Assimilation Scientific Documentation. 3d ed., 94 pp. [Available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England.]

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.

Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Adv. Geophys.,* **33,** 141–266.

Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus,* **38A,** 111–136.

——, D. B. Shaw, P. Lönnberg, L. Illari, K. Arpe, and A. J. Simmons, 1986: Monitoring of observation and analysis quality by a data-assimilation system. *Mon. Wea. Rev.*, **114,** 861–879.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124,** 1225–1242.

Kelly, G., and J. Pailleux, 1988: Use of satellite vertical sounder data in the ECMWF analysis system. ECMWF Tech. Memo. No. 143, 46 pp. [Available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England.]

——, ——, F. Rabier, and J.-N. Thépaut, 1993: Observing system experiments made with the ECMWF system. *Some Impact Studies Carried out with Various Observing Systems,* COSNA Scientific Evaluation Group. WMO World Weather Watch Tech. Rep. No. 16, WMO/TD No. 594.

Lönnberg, P., and A. Hollingsworth, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus,* **38A,** 137–161.

Lorenz, E. N., 1955: Available potential energy and the maintenance of the general circulation. *Tellus,* **7,** 157–167.

Mailhot, J., R. Sarrazin, B. Bilodeau, N. Brunet, A. Méthot, G. Pellerin, C. Chouinard, L. Garand, C. Girard, and R. Hogue, 1995: Changes to the Canadian regional forecast system: Description and evaluation of the 50-km version. *Atmos.–Ocean,* **33,** 55–80.

McMillin, L., M. Uddstrom, and A. Coletti, 1992: A procedure for correcting radiosonde reports for radiation errors. *J. Atmos. Oceanic Technol.,* **9,** 801–811.

Mitchell, H. L., C. Charette, C. Chouinard, and B. Brasnett, 1990: Revised interpolation statistics for the Canadian data assimilation procedure: Their derivation and application. *Mon. Wea. Rev.,* **118,** 1591–1614.

——, ——, S. J. Lambert, J. Hallé, and C. Chouinard, 1993: The Canadian global data-assimilation system: Description and evaluation. *Mon. Wea. Rev.,* **121,** 1467–1492.

Rabier, F., and T. McNally, 1993: Evaluation of forecast error covariance matrix. ECMWF Tech. Memo. No. 195, 36 pp. [Available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England.]

Reale, A., M. Chalfant, R. Wagoner, T. Gardner, and L. W. Casey, 1993: TOVS operational sounding upgrades: 1989–1993. NOAA Tech. Rep. NESDIS, 76 pp. [Available from NOAA/NESDIS, Washington, DC 20233.]

Ritchie, H., and C. Beaudoin, 1994: Approximations and sensitivity experiments with a baroclinic semi-Lagrangian spectral model. *Mon. Wea. Rev.,* **122,** 2391–2399.

Rood, R. B., J. E. Nielsen, R. S. Stolarski, A. R. Douglass, J. A. Kaye, and D. J. Allen, 1992: Episodic total ozone minima and associated effects on heterogeneous chemistry and lower stratospheric transport. *J. Geophys. Res.,* **97,** 7979–7996.

Rutherford, I. D., 1976: An operational three-dimensional multivariate statistical objective analysis scheme. *Proc. JOC Study Group Conf. on Four-Dimensional Data Assimilation,* GARP Rep. No. 11, WMO, Geneva, 98–121. [Available as No. 1 of the series of Notes Scientifiques et Techniques de DRPN, Dorval, Québec, Canada, H9P 1J3.]

Schubert, S. D., R. B. Rood, and J. Pfaendtner, 1993: An assimilated dataset for earth science applications. *Bull. Amer. Meteor. Soc.,* **74,** 2331–2342.

Shaw, D. B., P. Lönnberg, A. Hollingsworth, and P. Undén, 1987: Data assimilation: The 1984/85 revisions of the ECMWF mass and wind analysis. *Quart. J. Roy. Meteor. Soc.,* **113,** 533–566.

Thépaut, J.-N., D. Vasiljevic, P. Courtier, and J. Pailleux, 1993: Variational assimilation of conventional meteorological observations with a multilevel primitive-equation model. *Quart. J. Roy. Meteor. Soc.,* **119,** 153–186.

Turtiainen, H., 1993: Radiation correction for RS80 radiosondes refined. *Väisälä News,* **131,** 18–19. [Available from Väisälä Oy, PL 26, FIN-00421, Helsinki, Finland.]