Summary of the 4Dvar Global Subjective Verification.

An important update was made to the Global forecast system in March 2005, which involved replacing the 3D variational method with the 4D variational method in the Global data assimilation system. This implementation was preceded by a parallel run which lasted three months, from November 27, 2004 until March 10, 2005. During this period, the meteorologists of the Analysis and Prognosis (A&P) division evaluated the performance of the new system (designated as GLB-4D in this paper) in comparison with that of the operational system (designated as GLB-3D in this paper).

The A&P meteorologists were also called upon to participate in the evaluation of the 4D-Var pre-parallel run, fed by archived data covering two two-month periods: the summer of 2004 and the winter of 2003-2004. The evaluation of these archived cases was not done during regular operational shifts. This document is intended to summarize the important points of this evaluation, both objective and subjective.

The daily evaluation consisted of comparing the 48, 72, 96 and 120 hour forecasts of the 500 mb geopotential height (GZ) and the mean sea level pressure (MSLP), with verifying analyses. The forecasts of the two systems were evaluated in comparison to their respective analyses. The evaluated area was subdivided into four regions; the Pacific, the central North America (NOAM), the Atlantic and the Arctic (see figure 1). The 24 hour cumulative precipitation (QPF) amounts were also evaluated for all the regions except the Arctic, a region where the observation network is insufficient to reach significant conclusions. Forecast periods covered were 24-48hr, 48-72hr, 72-96hr, and 96-120hr.

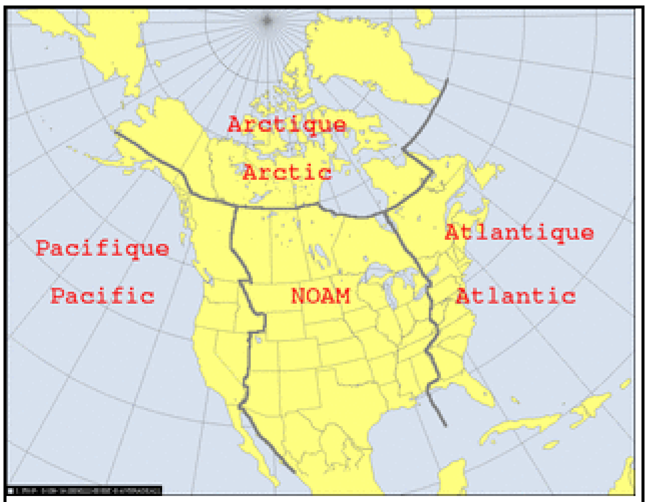


Figure 1 : Division of evaluated regions

In regards to the daily evaluation of GZ, MSLP or QPF, the meteorologists chose the letter 'O' when the operational GLB-3D assimilation system verified better for a given field, the letter 'P' when the GLB-4D assimilation system verified better and the letter 'E' when the two models verified in a similar manner. This evaluation was done for each region indicated in figure 1. The results of this exercise are summarized in the following tables and histograms.

GEM-GLOBAL

The following table shows the percentage of cases (from a total of 5839 evaluated cases) in favour of the GBL-3D, the GBL-4D or when the two systems were judged to be similar, for each of the three evaluated fields covering the particular region of interest. An improvement in forecasts was noted with the GBL-4D system for all three fields. However, these improvements were less significant for the QPF's.

Fields		Equivalent	GBL-3D	GBL-4D	
GZ 500mb		48.5%	21.2%	37.6%	
MSLP		51.7%	18.5%	29.8%	
QPF		60.6%	15.6%	23.9%	

Table 1 (winter 2004-2005)

Figure 2 displays the QPF evaluation for each of the three regions and evaluation period (24-48hr, 48-72hr, 72-96hr, 96-120hr). The histogram indicates the percentage of cases where the GBL-4D (yellow) and the GBL-3D (burgundy) were preferred for the QPF. The blue columns show the percentage of cases when the two models were judged to be equivalent. In the short term the GBL-4D and GBL-3D were rated equivalent in more than 70% of the cases except over the Pacific, where there were fewer equivalent cases. The most significant gains occur in the 24-72hr period over the Pacific region, whereas over the Atlantic region the most significant gains occur on days 4 and 5. Over the center of the continent (NOAM), the evaluation did not show any gain for day2, but some significant gains for the three other periods.

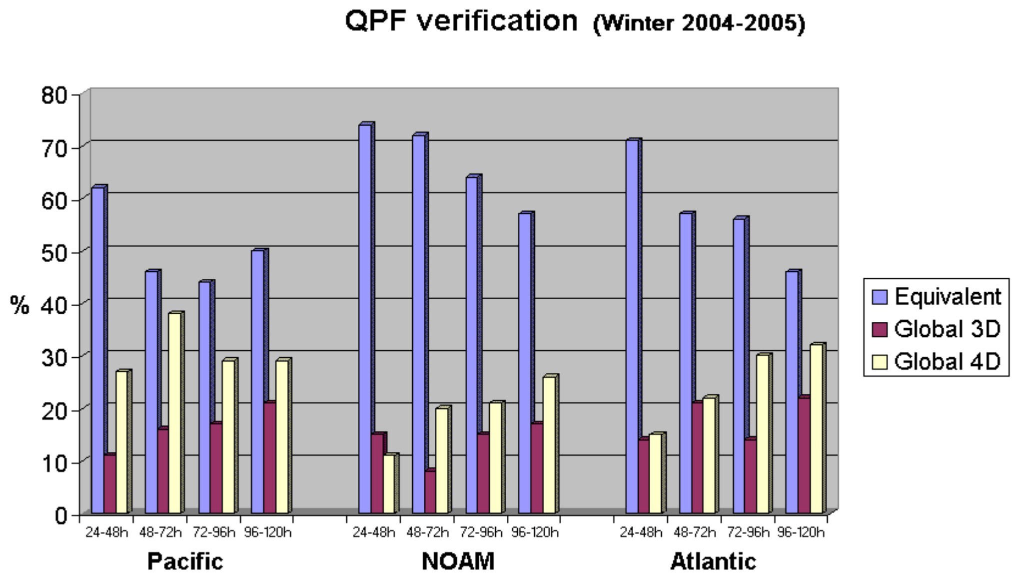


Figure 2: QPF evaluation per forecast period for each region during the winter of 2004-2005.

In summer, the gains for the QPF were minimal, with more than 73% of the forecasts judged to be equivalent. In situations where the systems were judged to be different, about 18% of these were in favour of GBL-4D and 9% in favor of GBL-3D. The differences between the two forecast systems occurred mainly over the eastern part of the continent, comprising more than 53% of the cases.

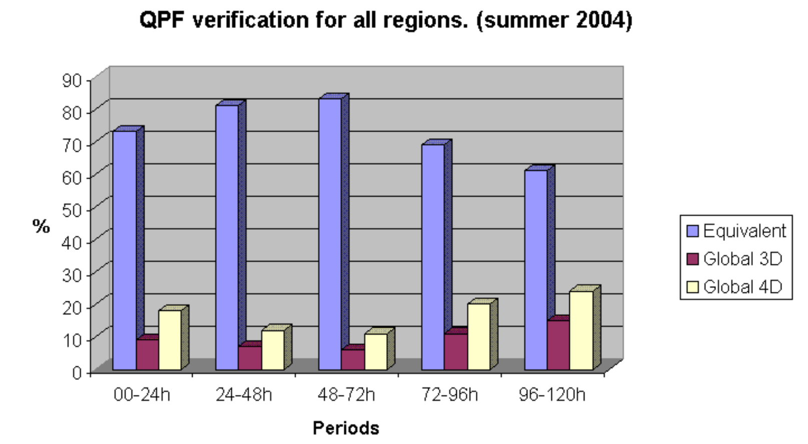


Figure 3: QPF evaluation per forecast period for all the regions during the summer of 2004

Figure 3 summarizes the evaluation done for the 97 summer runs over the period July 15 to September 15, 2004. Very little gain is noted in terms of QPF during the first 72 hours, when the GBL-3D and GBL-4D were also quite similar with the mass fields. During the summer, the gains in QPF were most noticeable on days 4 and 5, which were clearly related to a better verification of the mass fields.

With respect to the mass fields, overall, the gains were most noticeable for the GBL-4D across the Pacific, followed by the Prairies, then to a lesser extent the Atlantic and finally least of all over the Arctic. Figure 4 illustrates this behavior well, taking into account all three of the evaluated fields. Note that this evaluation is valid for the winter of 2004-2005. A similar exercise was performed for a reduced number of cases over the summer of 2004, during which time the Arctic was the region with the largest number of cases in which the GBL-4D verified better than the GBL-3D.

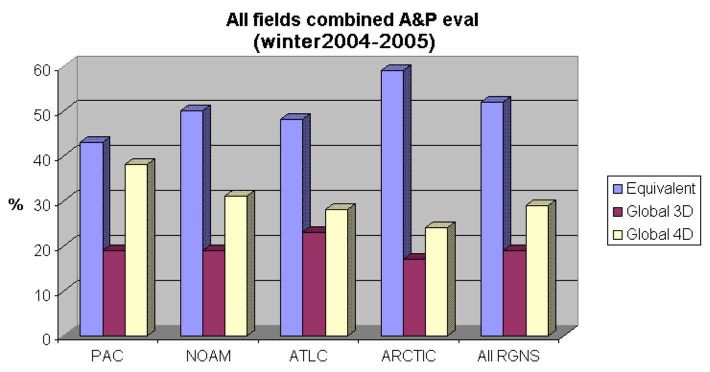


Figure 4: Summary of the evaluation for the 3 fields and all the forecast periods during the winter of 2004-2005.

An analysis of the A&P evaluation of the winter of 2004-2005 for GZ 500 and MSLP for different valid times highlights certain observations (figures 5 and 6).

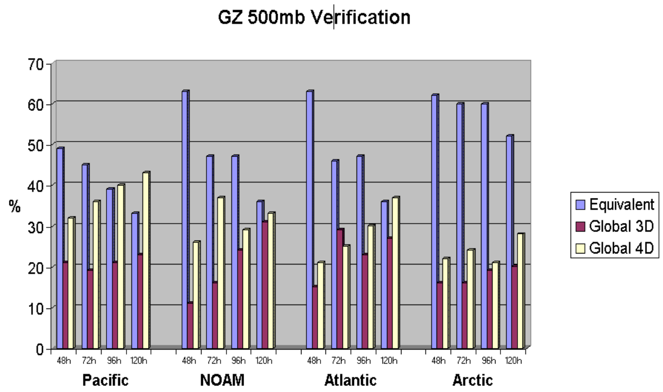


Figure 5: Evaluation of the 500mb geopotential height per forecast period for each of the 4 regions.

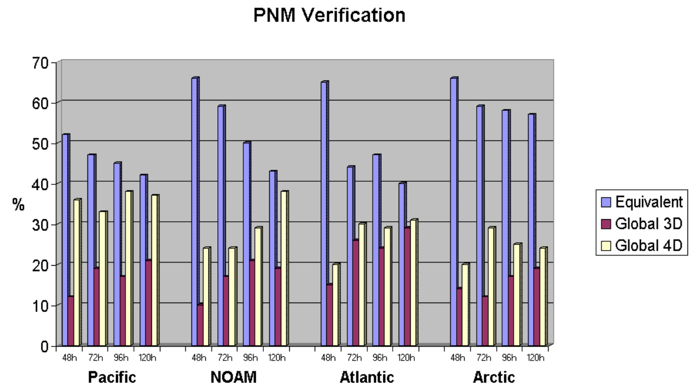


Figure 6: Evaluation of the mean sea level pressure per forecast period for each of the 4 regions.

For all regions, an improvement in the performance of the GBL-4D with increasing forecast time is noted, especially for the 500mb heights. The results of the evaluation for the Pacific region are clearly in favour of the GBL-4D; the largest differences between the two assimilation systems was for the MSLP field at 48hr. Over the center of the continent (NOAM), the GBL-4D verified the best at 72hr for GZ 500mb and at 120hr for the MSLP. This result is attributed to the trajectory of surface meteorological systems being better forecast in the medium term. For the Atlantic region, the gains were minimal for the GBL-4D. There are basically as many better performances for each of the two assimilation systems. However, when there was a significant difference between the two models (results not shown), the GBL-4D was better in 10 cases as compared to only 3 cases for the GBL-3D (out of a total of 137 cases).

The A&P meteorologists uncovered very few systematic gains or errors for the GBL-4D. Here as some of the observations noted during the GBL-4D evaluation.

1- The results of all the evaluations (as previously displayed) for the Arctic region show a slight gain in favour of the GBL-4D even though the forecasts were generally similar at all forecast times. This observation is true for winter, when blocking patterns are frequently the norm. However, the upper level circulation becomes much more zonal in summer, and results were correspondingly different during the summer of 2004. There were, unfortunately, no objective counts done during this period of the number of good and bad cases. Despite this, an improvement in the 4D version was noted, especially for the timing of meteorological systems.

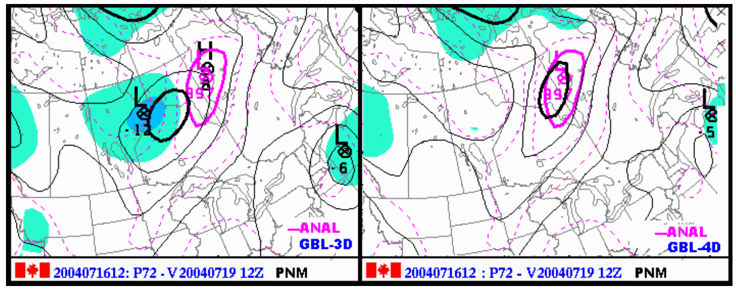


Figure 7: 72 hour forecast of the mean sea level pressure by the GLB-3D system (on the left) and by the GLB-4D system (on the right) represented by the black isobars; The verifying analysis is in magenta. The difference between the forecast and the analysis is colored at intervals of 4 mb .

Two typical cases are discussed here. The first case shows a 72 hour forecast (figure 7) in which the GBL-3D was too slow with a surface low pressure system analyzed over Hudson Bay. The GBL-4D was systematically faster for this depression and verified much better.

2- The second case (figure 8) shows an important improvement for the GBL-4D with respect to the speed and the amplitude of the 500mb trof over Nunavut and the northern Prairies. Many similar cases with smaller amplitudes were noted over the Arctic Ocean and the Beaufort sea during the summer of 2004. Note that the GBL-4D was worse with respect to the shortwave trough over the Atlantic region.

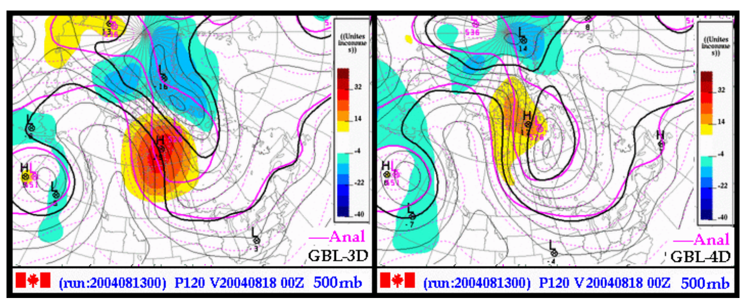


Figure 8: 120 hour forecast of the 500mb heights by the GLB-3D system (on the left) and by the GLB-4D system, (on the right) represented by the black isolines; the verifying analysis is in magenta. The difference between the forecast and the analysis is colored at intervals of 4 mb.

3- A problem was highlighted fairly early in the GBL-4D evaluation. Four situations over the Atlantic Ocean were noted wherein the GBL-4D over-developed low pressure systems originating from the Tropics, in particular from the Caribbean. There was a similar known behaviour in the GBL-3D in which storms were occasionally erroneously developed over the southwestern Atlantic Ocean and Gulf of Mexico. The frequency of these poor forecasts could be slightly higher in the GBL-4D system, mainly during hurricane season. An example of this (figure 9), shows a case from November 2004 where the GBL-4D system over-developed a surface low pressure system originating from the

Caribbean at a forecast period of 120 hours.

4- In general, the GBL-4D system forecast hurricanes slightly better, especially when the analysis had the benefit of dropsonde observations. The 4D-Var was particularly better with its forecast of Hurricane Karl during its extratropical transition. Figure 10 shows the significant improvement of the 4D-Var

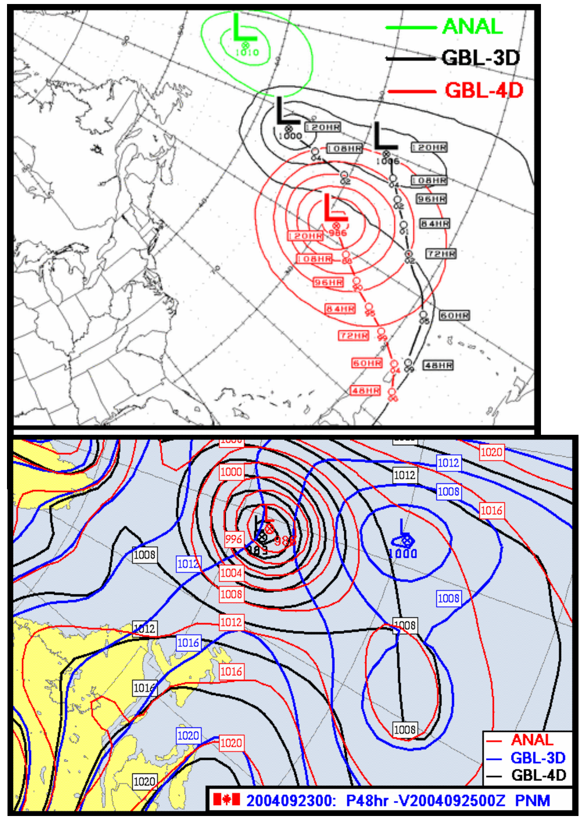


Figure 10: 48 hour forecast of the mean sea level pressure by the GLB-4D (in black), the GLB-3D (in blue). The verifying analysis is in red.

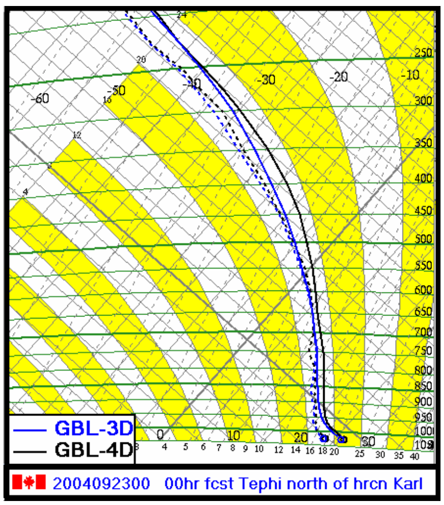


Figure 11: Temperature and humidity profiles at 00hr for the GLB-3D (blue) and the GLB 4D (black) to the north of Hurricane Karl.

system for a 48 hour forecast verifying September 9, 2004. The GBL-4D forecast the position of this storm very well whereas the GBL-3D was much too far south and not intense enough. Figure 11 shows an extreme case of a 00 hour forecast tephigram situated just north of Hurricane Karl. It was noted that tropospheric temperature profiles around tropical systems was generally warmer in the 4D-Var system. A similar behaviour was noted with Hurricane Frances, which had dropsonde data available for the analysis.

5- With regards to Pacific typhoons, the behaviour was different. In general, the GLB-4D improved the short term forecast, but deteriorated it in the medium range. The GLB-4D had a tendency to bring these systems too rapidly into the mid-latitude circulation, accelerating them eastward too rapidly. The final 120-hour position of surface low pressure systems originating from tropical systems often ended up much further east with the GBL-4D system as compared to the GBL-3D.

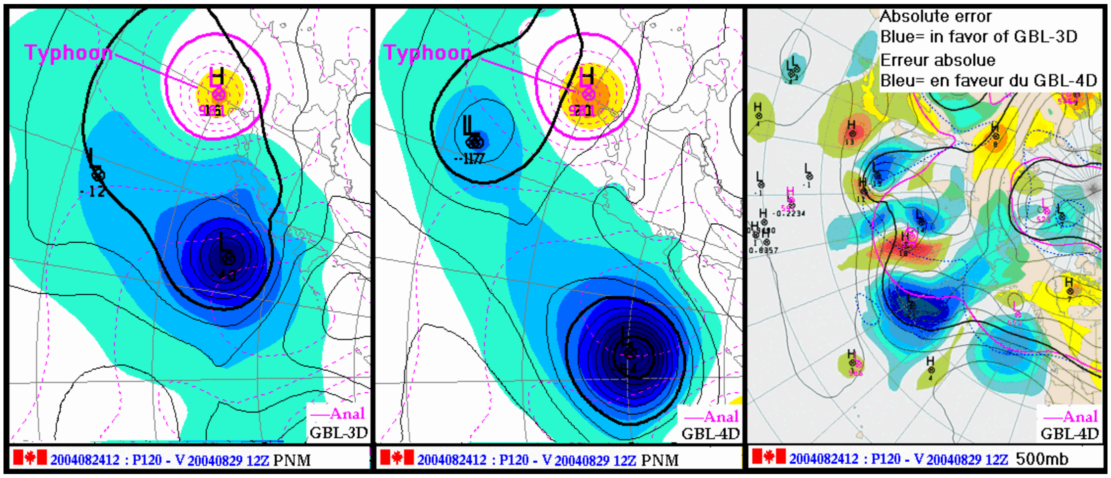
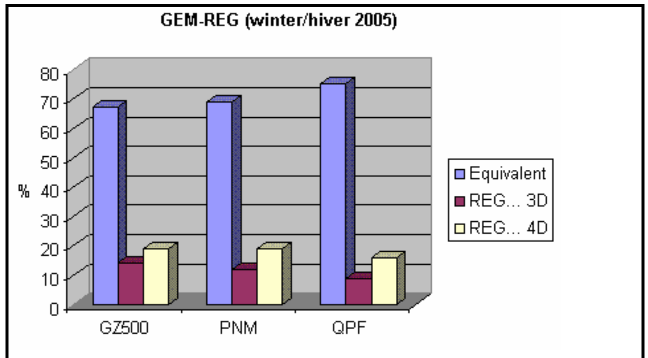


Figure 12: <u>Left and centre</u>: 120 hour forecast (in black) of the MSLP over the North Pacific by the GLB-3D (on the left) and by the GLB-4D (in the centre). The verifying analysis is in magenta, and the differences between the analyses and the forecast are coloured at every 4mb. <u>Right:</u> 120 hour forecast of the GZ at 500mb over the North Pacific by the GLB-4D (in black) and GBL-3D in magenta. The coloured zones represent the difference in absolute errors between the two systems. The GLB-3D has lower errors than the GLB-4D in the blue sector and vice-versa in the yellow/red sectors.

Figure 12 shows a case from the A and P evaluation period. The central image is a 120hr forecast of the GBL-4D indicating a more progressive storm position than the GBL-3D (left image). The coloured field shows differences between the analysis and the forecast. In this case, even though the typhoons are far away, there is an impact on North America in the medium to long range. The image on the right shows the absolute error of the GZ 500mb between the two forecast systems. The blue areas are in favour of the 3Dvar and the warm colour areas in favour of the 4Dvar. In this case, the shortwave troughs were forecast significantly out of phase all the way to just west of British Columbia.

<u>GEM-REG</u>

The GEM Regional forecast system continues to use a 3D-Var analysis at the end of a 12 hour assimilation cycle. However, this cycle is initiated every 12 hours by the global cycle. Since the 4D-Var system would eventually be implemented in the global cycle, a regional parallel run was conducted with the goal of evaluating the impact of the global cycle change on the regional system. A and P conducted the evaluation of this parallel run during the period January 20 to March 10, 2005. To simplify the text, the term *GEM-3D* will refer to the GEM regional model fed by an assimilation cycle initiated from a 3Dvar analysis, and *GEM-4D* for the system initiated from a 4D-Var analysis.



The evaluation process was similar to that used for the global system (see preceding section). The results showed a slight improvement for the GEM-4D system for the four regions (see figure 13). For all regions combined, the most significant gain, though modest, is with the MSLP. Over the Pacific and central North America regions (especially at day-2) the 4D-Var was significantly better than the 3D (figure 14). A gain is also noted over the Arctic at the end of day 2. With respect to the QPF's (see figure 15), it's the Pacific region which saw the largest gains over the 00-24hr period and over the Atlantic for the 24-48hr period. The gains were minimal over the continent. The gains were mainly attributed to corrections to the precipitation axis and the position of the precipitation envelope, thus a direct link to the forecast trajectory and speed of low pressure systems.

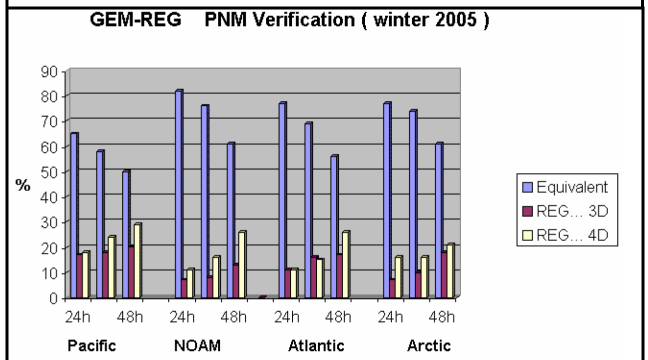


Figure 14: Evaluation of the impact of the GLB-4D, on the regional forecast system for the mean sea level pressure in

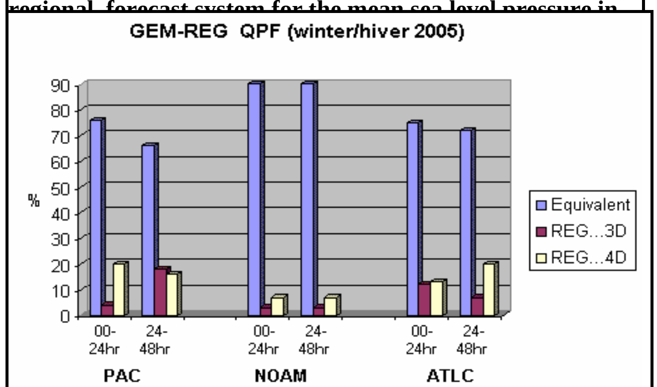


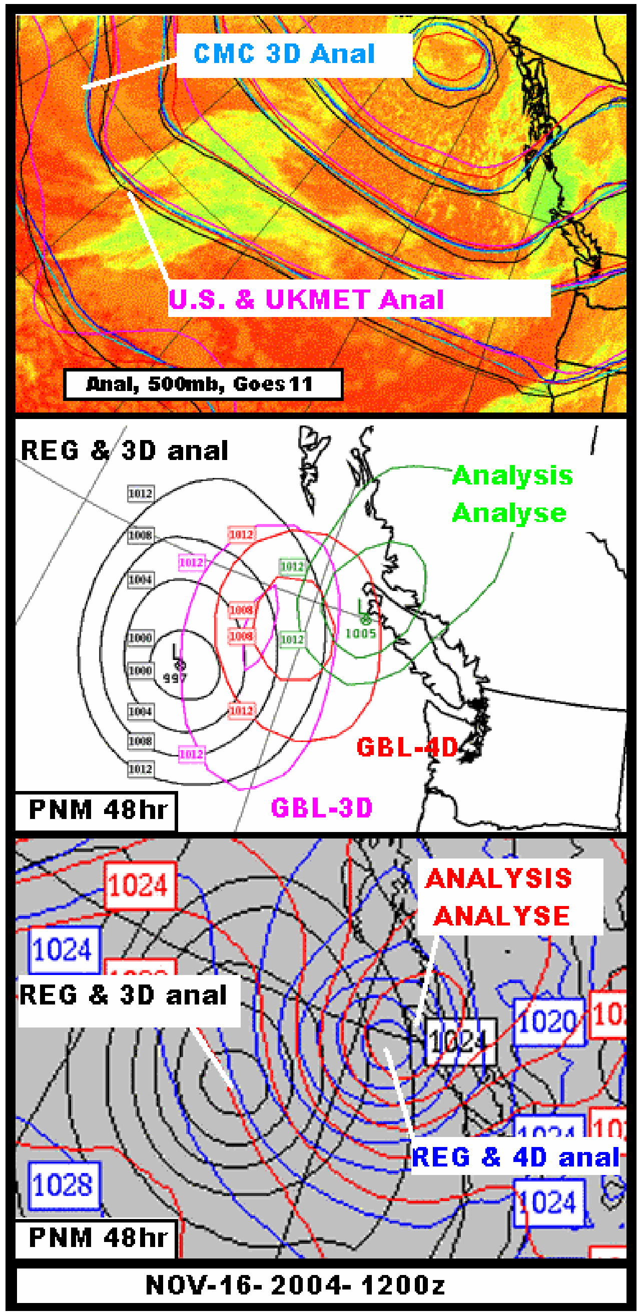
Figure 15: Evaluation of the impact of the GLB-4D, on the regional forecast system for the QPF in relation to the forecast period for several regions.

Generally, the impacts were relatively subtle during the regional evaluation, except for one interesting case. To the right, on the top image, a deficient analysis concerning the trough at 500 mb is evident. The pale and dark blue lines are the 3DVAR operational analyses of the regional and global systems. The heights in black and magenta represent the UKMET and NCEP analyses. We note that the CMC analysis is too far west with the upper level trough compared to the satellite imagery cloud signature.

The central image shows the 48hr forecasts of the MSLP against the analysis (green). The GEM regional forecast was the slowest among all the operational models available. The GEM global forecast the surface low a bit more progressive but was still too slow.

The 4D-Var analysis corrected the position of the upper trough, while on the same image (red), the position of the surface low forecast further east by the GLB-4D is evident.

The bottom image shows the effect of the 4Dvar analysis on the regional 3D-Var assimilation cycle. The regional forecast (blue isobar) is almost in agreement with the verifying analysis (in red).



Conclusions

It is clear that the 4D-Var implementation is a good step forward for the NWP system even though the handling of tropical systems remains a subject of concern.

In general, the 4D-Var analysis had a positive impact on the global model, especially for the Pacific region in the short to medium term. The improvements of the 4D-Var analysis become more muted for the mass fields as we approach the Atlantic region, even though we see an improvement in QPF on days 4 and 5.

The 4D-Var global cycle had an impact on the GEM-Regional model, especially for the MSLP at the 48 hr forecast period, across the continent (NOAM). QPF improvements were more frequent over the Pacific region. Elsewhere over Canada, many small differences were noted with precipitation axes and precipitation envelopes, in favour of the GEM-4D.