2.1 WHAT CAUSES NCEP GFS MODEL FORECAST SKILL "DROPOUTS"?

Bradley A. Ballish*, Jordan C. Alpert, DaNa L. Carlis, and V. Krishna Kumar[1]
NOAA/NWS/NCEP, Camp Springs, Maryland 20746

1. INTRODUCTION

Numerical Weather Prediction (NWP) models have improved remarkably in the past decade and progress is continually being made year to year in all disciplines of NWP particularly in the field of data assimilation. Regardless of the steady progress, most NWP models are blighted with the occasional forecast busts, referred here as dropouts which influence deeply the overall skill of the 5-day forecasts. In a companion paper, Alpert et al. (2009), we define the model dropouts in terms of 5-day 500 hPa anomaly correlation (AC) scores, and show that the NCEP Global Forecast System (GFS) can significantly reduce the number of dropouts when using analyses from the European Centre for Medium-Range Weather Forecasts (ECMWF) to initialize the GFS (ECM runs). The primary motivation of this study is to understand the genesis of these dropouts with a view to alleviate the forecast dropouts as they produce poor quality forecast guidance and are responsible for the reduction of NCEP GFS 5-day forecast skill.

2. BASIS OF FORECAST DIFFERENCES

There are numerous reasons why different forecast models will have forecast differences. There are differences in whole classes of observational data being used in analyses, differences in data time windows and cutoff times, data bias corrections, data Quality Control (QC), as well as analysis and forecast model differences. Even though the GFS performs well in these ECM runs, it is still possible that deficiencies in the GFS model could be harming the performance of the NCEP Gridpoint Statistical Interpolation (GSI) analysis, (Wu et al. 2002). For example, a temperature bias in the model may have small impact on the 5-day AC score but could cause incorrect perturbations in temperatures and winds in the assimilation system. Differences in two analyses at one point are due to all of the above listed factors accumulated nonlinearly over a long time as each analysis has large impact from the background, which has memory of all these differences. Forecast differences are due not only to forecast model differences but also the analysis differences accumulated for many reasons. Large errors in the model 5-day forecast in one area can be due to errors in the analysis at different locations as well as model errors. Since the analyses and forecast models are very complicated in addition to the above reasons, it can be difficult to find the cause of a forecast dropout.

3. ANALYSIS OF FORECAST DIVERGENCE AND MODEL DROPOUTS

One simple method of analyzing large forecast errors is to track the location of large errors during the 5-day forecast as a function of forecast time. For example, if there is a 400m height error at 500 hPa southeast of Greenland, where the error is the difference in the 5-day forecast and the verifying analysis, then we can examine how this error grew with time and location. Readers can use our website:

http://www.emc.ncep.noaa.gov/gmb/dcarlis/directory/dropout_cases/

to track forecast errors for many GFS dropout cases. The error in 5-day forecasts can be much bigger than the analysis uncertainty, but for analyzing short-range forecast errors, it is very important to have analyses from different centers to check forecast errors. For example, if the ECMWF has a small forecast error compared to the above GFS forecast error near Greenland, then the ECMWF analysis at hour zero in areas important for this case can be considered more reliable than that of the GSI. Then how the GSI and ECMWF analyses differed at hour zero as well as how the GFS forecasts differ from the ECMWF analyses in time can help find problem areas in the analysis that appear to have led to the 5-day forecast error.

* Corresponding author address: Dr. Bradley Ballish, NOAA/NCEP/PMB, 5200 Auth Rd. Room 307, Camp Springs, MD 20746; email: Bradley.ballish@noaa.gov

[1] Also at Perot Systems Government Solutions (PSGS), 8270 Willow Oaks Corporate Drive, Fairfax, VA 22031

This sort of analysis can be useful, but it can be complicated to analyze a dropout case. For example, an analysis error around jet-level at hour zero in the Pacific can lead to large 500 hPa height errors at 48 hours over the USA. A separate analysis error in moisture at hour zero in the tropics can then feed into the weather system over the USA at roughly 48 hours making the 500 hPa forecast error at 5 days even larger. In addition to the two above analysis errors, there could be other minor errors in the analysis as well as GFS errors to complicate analysis of the final errors in the 5-day forecast. Although the above analysis can be difficult and is not fruitful in all dropout cases, it often indicates that dropouts in the Northern Hemisphere (NH) for the GFS often originate in the Pacific. GFS dropouts in the Southern Hemisphere (SH) often originate from roughly 40 to 60 degrees south. For both of these areas, there are few radiosondes or other high quality conventional observational data to help both the analysis and estimates of analysis error.

It would be useful to have an adjoint or tangent linear version of the model forecast, such as in (Errico and Raeder, 1999), to estimate the analysis error based on the observed 5-day forecast error. However, the uncertainty in the verifying analysis coupled with large error in using such models well beyond their expected usefulness of about one day makes this option non-robust. It would be useful to have a combined adjoint of the analysis and forecast model to estimate whether observed data helped or hurt the forecasts through short ranges such as one day (Zhu and Gelaro 2008). Unfortunately, such adjoint tools are not currently available at NCEP.

However, it is useful to check if various observed data may have caused analysis error in select areas that are suspected to be leading to large forecast errors. We can compare the GSI analysis and background in such areas with the ECMWF analysis and observational data and then try to estimate if either error in the background or observational data resulted in large analysis differences. Although many cases were found where clearly wrong observational data passed QC and had negative analysis impact, so far, no cases have been found where QC error led to a forecast dropout. It is possible that QC errors some time in the past before the start of the analysis that led to a dropout could be important, but this sort of problem is more difficult to analyze. As mentioned earlier, most dropouts seem to originate in areas with no high quality conventional observational data. Such areas often have satellite radiance, satellite wind and surface data. These data seldom have large egregious one time impact due to both QC protection and because they have smaller analysis weights than high quality conventional data. Never the less, such data do have systematic impacts that are important but are difficult to analyze.

Since some analysis errors lead to forecast errors that get smaller with time, but some errors lead to forecast errors that amplify with time, diagnostics are needed to check on model forecast sensitivity to analysis errors. Although adjoint model sensitivity error estimates are not perfect and are reliable to about one day, Rolf Langland's FNMOC website:

http://www.nrlmry.navy.mil/adap-bin/tcs_adap.cgi

is useful to estimate sensitive areas for growth in forecast error (click on "Global Domain NAVDAS" and select desired 00Z case). Although this website has estimates in sensitivity for the FNMOC model, it is helpful to use it to look for areas where GFS analysis differences may lead to large forecast errors. For example, Fig. 1 shows this sensitivity for 00Z 21 October 2007. Note the high sensitivity in the Pacific. For the next few days, there were areas of high sensitivity down wind of this initial time. It is no wonder that the GFS had low 5-day AC scores for this period, as analysis errors in the Pacific then moved into areas where the analysis and GFS may have high sensitivity down wind.

For cases where suspect observational data may be harming the analysis and leading to a bad forecast, such data can be deleted and the GSI analysis and the GFS rerun to see if the forecast skill improves. So far, no examples have been found where deleting select observational data has prevented a dropout, but such cases are still looked for. For cases where the GSI analysis is suspect in an area, the GSI analysis can be rerun with ECM pseudo observations (fake radiosonde reports derived from ECMWF analyses) overlayed in the suspect area (Alpert et al. 2009) to confirm that this was a problem area. Often such experiments result in better forecast skill but not usually as good as full global ECM runs.

4. INVESTIGATION OF SYSTEMATIC GSI VERSUS ECMWF ANALYSIS HEIGHT AND TEMPERATURE DIFFERENCES

Since egregious QC errors have not been found to be leading to dropouts, systematic differences between the GSI and ECMWF analyses are now investigated. One systematic analysis difference is the tendency for the GSI analysis to be approximately 10 meters higher than that of the ECMWF at 200 hPa for roughly 40 to 60 degrees south and in the north Pacific. See Fig. 2 for a NH example and Fig. 3 for a SH example. Note in both of these cases, there is a predominance of red indicating that the GSI heights are higher than the heights from the ECMWF. For other examples of these height differences for dropout cases, see

http://www.emc.ncep.noaa.gov/gmb/dcarlis/directory/dropout_cases/

Since the GSI has heights that are systematically higher than that of the ECMWF analysis, experiments were performed using an ECM analysis run to provide a new background for the GSI, where the background has a height bias similar to that of the ECMWF. Using these backgrounds as input along with the full suite of operational data, the GSI was rerun for ten different cases and composite analysis minus background (ANMBG) were constructed as shown in Fig. 4. The top portion of the figure shows zonal average cross sections of ANMBG temperatures for the control and InterpECMGES experiments. The control is a rerun of the operational GSI. The InterpECMGES experiment is the same as the control except the background was from a 6-hour previous ECM run. The average heights from the ECM derived background are similar to the appropriate ECMWF analysis (not shown). The zonal averaged analysis temperature changes in the control run are small compared to the InterpECMGES experiment, which has relatively large temperature changes at roughly 60 degrees south around 900 hPa. This later experiment shows mostly higher heights in the analysis compared to the background around 50 to 80 degrees south. Additional work suggests that large numbers of automated aircraft temperatures with relatively warm biases as in Ballish and Kumar (2008) is partly to blame for the GSI analysis being too warm. The resulting warm bias in the analysis may impact the GSI use of satellite radiances and add to these height differences.

Rolf Langland of FNMOC has produced Hovmoller diagrams showing systematic height differences between the GFS and ECMWF analyses. See Fig. 5 showing 500 hPa height differences, GFS – ECMWF, from 35-65 degrees north from October to December 2007. He reports that other centers show similar height differences compared to ECMWF analyses. This is an important finding needing further analysis. The height differences show patterns suggestive of differences in use of satellite radiances. Langland points out that the ECMWF is the only center using a time window of +/- 6 hours rather than the more common +/- 3 hours. Bias correction differences in satellite radiances could also be a factor.

5. INVESTIGATION OF SYSTEMATIC GSI VERSUS ECMWF ANALYSIS FITS TO OBSERVATIONS

Since the ECM analyses are designed to be similar to ECMWF analyses and because the ECM analyses can be interpolated to observations with exactly the same methodologies as the GSI analyses, various statistics comparing these analysis fits to observations were generated. Here the analysis is converted from spectral coefficients to the model's Gaussian grid and then interpolated horizontally and then vertically to the observations. For these statistics, non-satellite radiance data were used for GDAS data that passed QC for all 12Z runs in April 2008. Similar statistics were found using roughly the same number of cases for various model run times in October 2007 (not shown). The statistics were for many classes of data for temperature, surface pressure, relative humidity and winds in 100 hPa thick layers. Since these interpolations do not include time interpolation of the analysis to the observations, only observations with in +/- 1.5 hours from the run time were used. The statistics were binned in different categories of how different the two analyses differed, and were made for different regions of the globe. Note that the ECM analyses are not identical to those of the ECMWF, and for this study, it was not known what data was available or passed QC at the ECMWF; however, these statistics are very informative.

In Fig. 6 is a comparison of whether the GSI or ECM analysis draws more closely towards various observational wind types in the pressure range of 300 to 200 hPa for all 12Z runs in April 2008. For the 10 data types on the right side of Fig. 6, the GSI draws more closely than the same for the

ECM analyses. These summaries based on percentages of better draw are consistent with RMS winds differences of the analyses versus observations (not shown). For Canadian AMDAR (CANDAR) and European satellite winds (EUSATW), the ECM analyses draw more closely probably because the GSI is monitoring these data types and not yet assimilating them in the analysis. For the vertically averaged azimuth Doppler radar data (VADW), that do not report pressure, the interpolation of the analysis to the observations used in this analysis uses pressure derived from the data's altitude along with a standard atmosphere. The GSI uses pressure derived from the height pressure relation of the background and analysis to determine pressure, which has been shown to be more accurate.

Figure 7 shows that the GSI draws more closely to radiosonde winds than that of the ECM analyses for 12Z April 2008 except for 300 to 200 hPa. The relative increase in the draw of the ECM versus GSI analyses at jet-level could in part be due to the ECMWF analyses using thinning of aircraft data (Cardinali et al. 2003). In Fig. 8, the GSI draws much more closely than that of the ECM analyses when the two analyses have vector wind differences in the category of 5-10 m/sec (outliers). This analysis draw difference is smaller near jet-level. If the GSI were to be found to perform better by drawing less for wind outliers, this could be accomplished by tuning the variational data QC.

Figure 9 shows that the GSI and ECM analyses have similar fits to radiosonde temperatures for the same time period as above, but with the ECM analyses fitting more closely 400 hPa and above. In Fig. 10, for analyses fits to the above temperatures when the two analyses differ by 2-4 degrees (outliers), the GSI fits more closely from about 900 to 400 hPa. Otherwise, the ECM analyses fits more closely. Again, tuning of the variational data QC could perhaps make these data fits more similar along with improving forecast performance.

Figure 11 shows that the GSI draws more closely for Antarctic radiosonde temperatures and winds than that of the ECM analyses below about 700 hPa for the same period as above. The Antarctic region has radiosondes only along the coastline, except at the South Pole, and there are few other high quality observations. This lack of a network with good data coverage coupled with energetic and difficult to predict weather at low levels, could mean that we do not have enough data coverage to successfully analyze the Antarctic atmosphere in too much small scale detail at lower levels. It is possible that the GSI would result in better quality forecasts if it drew less for the radiosonde data at lower levels. This supposition is contrary to the opinion that we need to draw more closely for such data in data sparse areas.

Figure 12 shows the speed biases in m/sec of GSI and ECM analyses versus four different types of satellite wind observations that passed QC in GDAS analyses for 12Z April 2008. At first glance, the GSI appears to have better biases by roughly .5 m/sec compared to the ECM analyses. However, this bias difference is likely due to the ECMWF data QC either rejecting or giving lower weights to satellite winds with speeds appreciably slower than the background, see the ECMWF report available at:

http://www.ecmwf.int/research/ifsdocs/CY31r1.

These bias differences imply that the ECM analysis speeds are faster than that of the GSI by roughly .5 m/sec at satellite wind locations. This needs further investigation, as these speed differences over large areas could have significant impact.

6. SPATIAL DISTRIBUTION OF FORECAST ERRORS AND EADY BAROCLINICITY INDEX

One rationale behind developing dynamical tools to assist the forecast busts/dropout analysis is to understand and predict the location of sensitive regions from which the GSI analysis errors grow disproportionately and consequently degrade forecast skill of the GFS. These sensitive regions result because of dynamical reasons emanating from flow dependent characteristics and cause forecast amplification of analysis errors. Since the differences in GSI and ECMWF analyses are an estimate of analysis error, the geographical distributions of the root mean square error (RMSE) differences between GFS and ECM are constructed at each grid point for NH and SH dropouts. Fig. 13 represents the spatial RMSE difference between GFS and ECM runs for the forecast range of 24 hours verifying on GFS (operational) analysis for a SH dropout initial condition (IC) dated 00Z 16 August 2008. Large error differences displayed as shades of red

(shades of blue) colors corresponding to positive (negative) values, meaning that GFS errors are larger (smaller) than ECM errors. The evolution and propagation characteristics of large forecast error differences are significant. Both models display equal areas of red and blue colors between 30S to 90S with some deeper red colored areas, implying the higher forecast errors for the GFS compared to ECM, indicating that the sensitive regions lie primarily in the southern belt of mid-latitudes. Vector wind differences between GFS and ECM displayed as arrows are superimposed in Fig. 13. Large wind differences are observed in a few areas near the Antarctic. In order to delineate the origin of analysis and forecast error differences between two models, it is paramount to understand the critical differences between their respective thermodynamic and dynamical characteristics. This is explored computing the baroclinic instability growth of the transient eddy activity of the GFS and ECM basic flows.

The growth of transient waves in the mid-latitude westerlies in the presence of vertical shear originates from baroclinic instability mechanism discovered by Charney (1947) and Eady (1949). The maximum growth rate of the most unstable mode provided by the Eady's model, i.e., the measure of the Eady Baroclinicity Index (EBI) as shown by Hoskins and Valdes, 1990, is given by

$$\sigma_{BI} = 0.31\, f\, |(-gp/RT)\, \partial \mathbf{V}/\, \partial p|\, N^{-1},$$

where f is the Coriolis parameter, **V** is the total vector wind, N is the Brunt Väisällä frequency and all other parameters have their usual meaning. EBI is proportional to the vertical wind shear and the static stability of the basic flow. The Eady index is computed using the three dimensional analysis and forecast fields of GFS, ECM as well as the ECMWF models to show potential action areas or volatility to propagate IC errors into forecast differences. Figs. 14a,b show respectively the total EBI at 800 hPa in units of per day (day^{-1}) for the 24-h GFS forecast (top panel) and the corresponding ECM model forecast (bottom panel) from 00Z 16 August 2008 ICs. It is quite evident that the GFS model shows more pronounced baroclinicity in the SH compared to the ECM model and it is quite likely that the larger RMSE as shown in Fig. 13 for the GFS originates due to enhanced unstable conditions in the SH mid-latitude band. The increase in baroclinicity of the GFS in a SH mid-latitude band, the differences in this case, may potentially cause a dropout in the 5-day forecast skill.

Figs. 15a,b show the total EBI for a NH dropout case on 12Z 21 October 2008 at 500 hPa for the F00 GFS forecast (top panel) and the corresponding ECMWF (bottom panel) model. It is clear that the greatest baroclinic potential lies in the eastern part of the broad Pacific trough, the differences in this case, cause a dropout in the 5-day forecast. The adjoint sensitivity of 24h forecast error to IC for the same case (Fig. 1) shows similar sensitive areas. Large differences are along the trough line with dipole structures (not shown) indicating differences in position (phase), and large potential for these Rossby wave details. The GFS model shows more pronounced baroclinicity compared to ECMWF operations. The differences are as much as 20% of the total index. Using this index to find potential baroclinic areas, and intersection with differences between background guess and analysis, shows promise to form the basis for an automated real-time dropout detection system.

7. CONCLUSIONS AND PLANS FOR FUTURE WORK

Since the GFS has better forecast skill in the ECM runs, which are initialized using the ECMWF analysis as input, than with the operational GFS running from the GSI analysis, attempts have been made to find systematic differences in the GSI and ECMWF analyses. One systematic difference is for the GSI analysis to have higher heights. This difference maybe due to the impact of overly warm aircraft temperature data in the GSI and possible resulting impact on satellite radiance bias corrections. It is also found that the GSI draws more for most wind observations, especially wind observations with moderate analysis differences. The ECM runs draw more for radiosonde temperatures than the GSI in the upper troposphere and above, but less so for temperature outliers.

Further work will involve tuning the variational data QC to try to optimize the GSI fits to observations, possibly more like that of the ECMWF analysis. Bias correcting of aircraft temperatures and tuning of the satellite radiance bias corrections could give heights more similar to that of the ECMWF analysis as well as better forecast skill.

Diagnostic tools such as EBI are useful in identifying the location of sensitive regions from which GSI analysis errors grow and subsequently degrade GFS forecast skill. The RMSE differences between GFS and ECM both verified against GFS analyses indicate errors concentrated in the mid-latitude baroclinic zones for the NH and SH dropouts, and these areas correspondingly show larger instability rates as diagnosed from the EBI. The EBI calculations show that the GFS model possesses more pronounced baroclinicity when compared to the ECM model and the ECMWF operations. Additional diagnostic tools for the dropout analysis are currently being developed.

Acknowledgements. The authors would like to thank Louis Uccellini and Steve Lord for their leadership and support of the dropout team. We are grateful to Rolf Langland for providing his plots of NCEP versus ECMWF height differences along with his ideas on this problem. Rolf also provided his website showing FNMOC model sensitivity estimates. Special thanks to our colleagues Fanglin Yang and Russ Treadon for their diagnostics and GSI support.

8. REFERENCES

Alpert, J. C., D. L. Carlis, B. A. Ballish, and V. K. Kumar, 2009: Improved forecast skill using pseudo-observations in the NCEP GFS. 13th AMS IOAS-AOLS, P2.1.

Ballish,B. and V. K. Kumar, 2008: Systematic differences in aircraft and radiosonde temperatures implications for NWP and climate studies. *Bull. Amer. Metoe. Soc*., **89,** 1689-1708**.**

Cardinali, C., L. Isaksen, and E. Anderson, 2003: Use and impact of automated aircraft data in a global 4DVAR data assimilation system. *Mon. Wea. Rev*., **131**, 1865-1877.

Charney, J.G., 1947: The dynamics of long waves in a baroclinic current. *J. Meteor*., **4**, 125-162.

Eady, E.T., 1949: Long waves and cyclone waves. Tellus, 1, 33-52.

Errico, R. M. and K. D. Raeder: An examination of the accuracy of the linearization of a mesoscale model with moist physics. *Q. J. R. Meteorol. Soc*. **125**: 169-195.

Hoskins, B.J., and P.J. Valdes 1990: On the existence of storm-tracks. J. Atmos. Sci., 47, 1854-1864.

Zhu, Y., and R. Gelaro, 2008: Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interploation (GSI) analysis system. *Mon. Wea. Rev*., **136:** 335-351.

Wu, W., R. J. Purser, and D. F. Parrish, 2002: Three dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev*., **130:** 2905-2916.

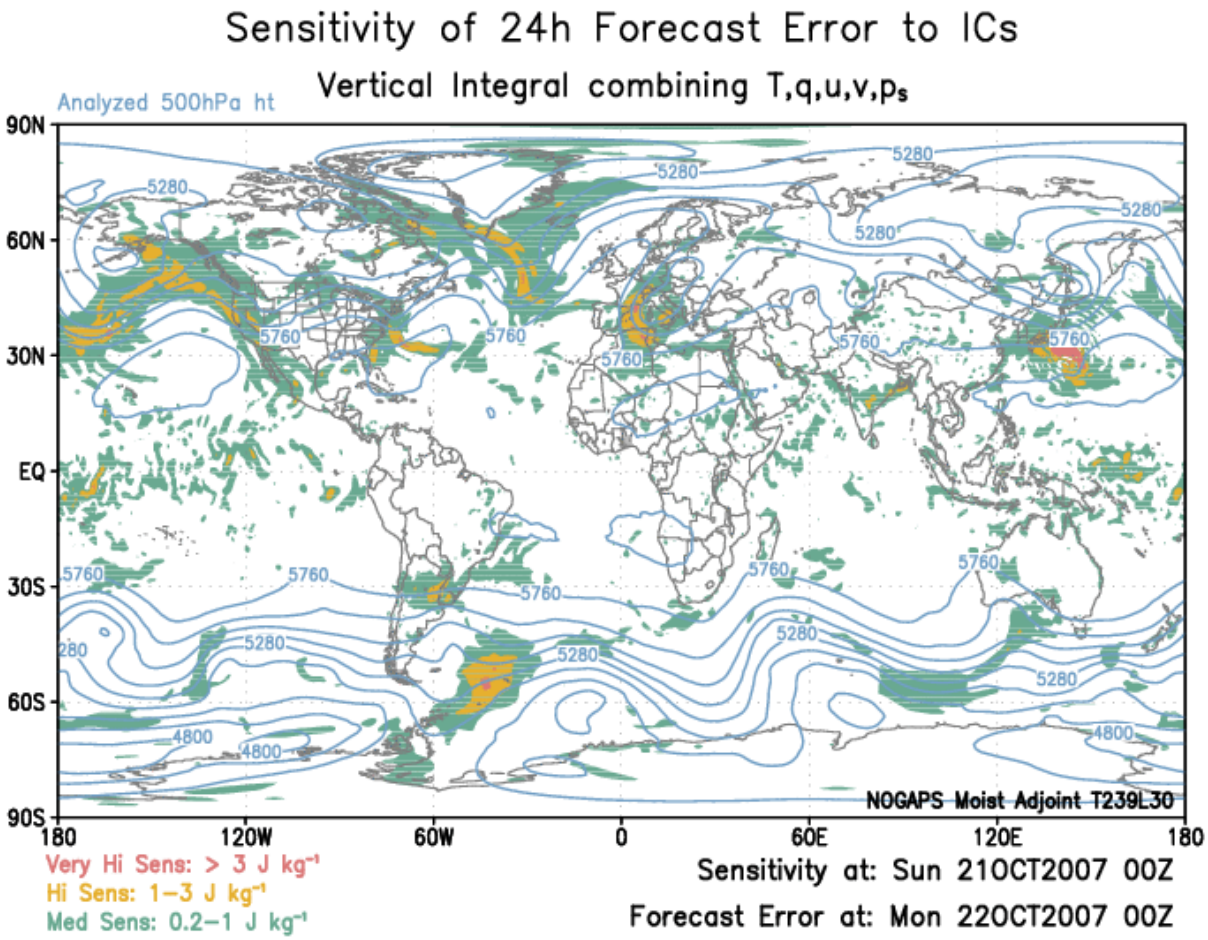


Figure 1. FNMOC model sensitivity error estimates from 00Z 21 October 2007.

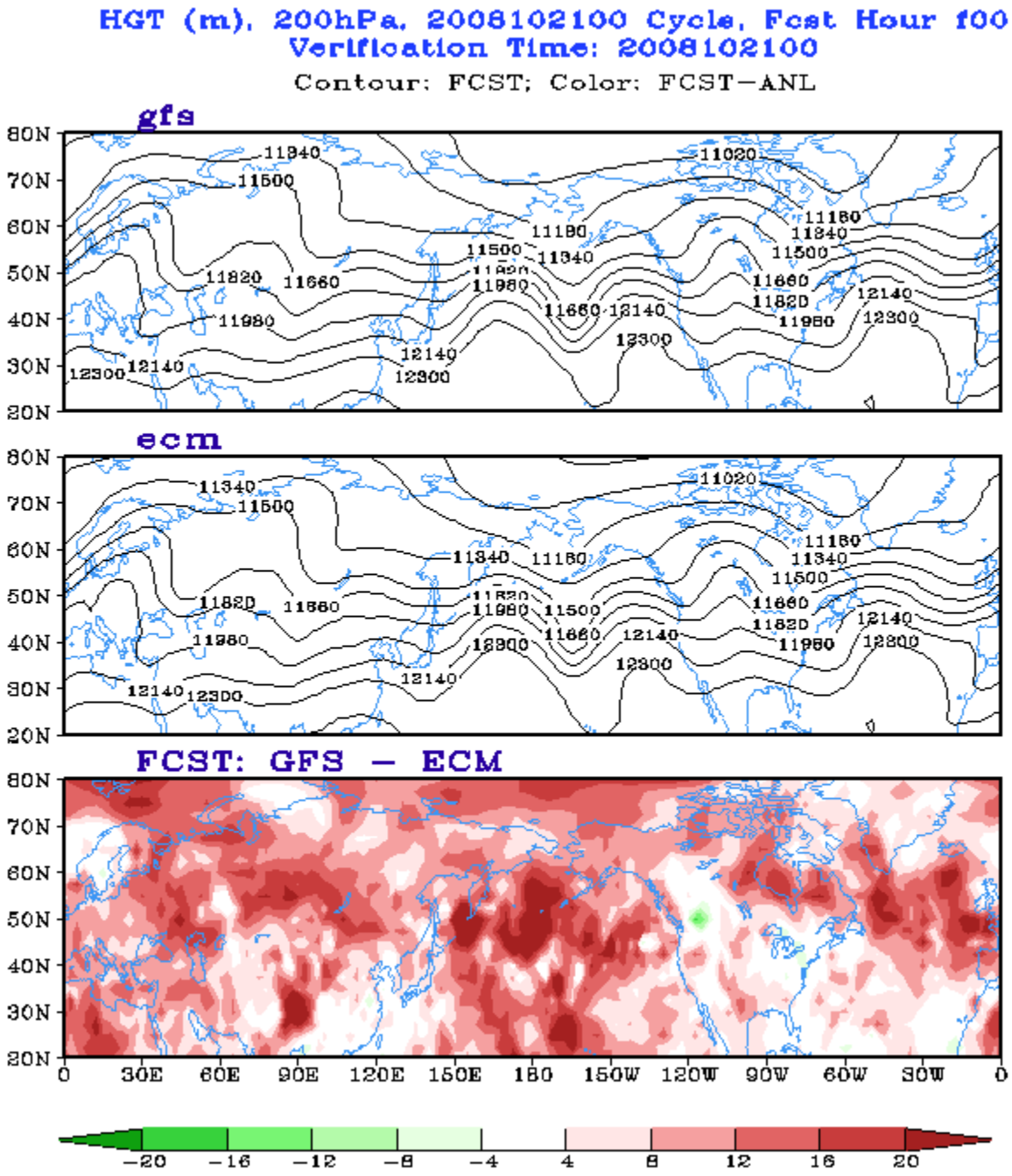


Figure 2. 200 hPa heights GFS versus ECMWF Northern Hemisphere 00Z 21 October 2007.

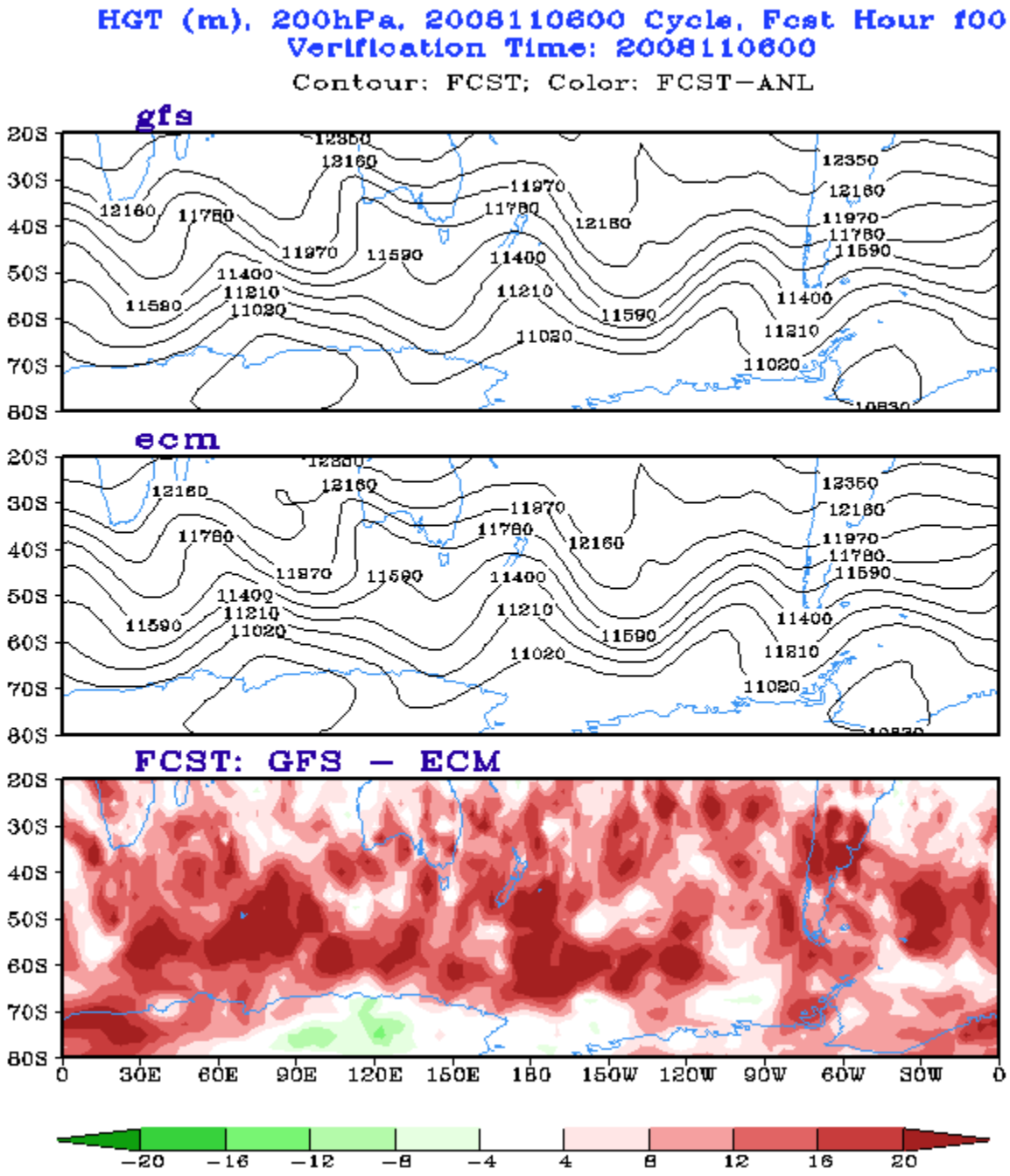


Figure 3. 200 hPa heights GFS versus ECMWF Southern Hemisphere 00Z 6 November 2008.

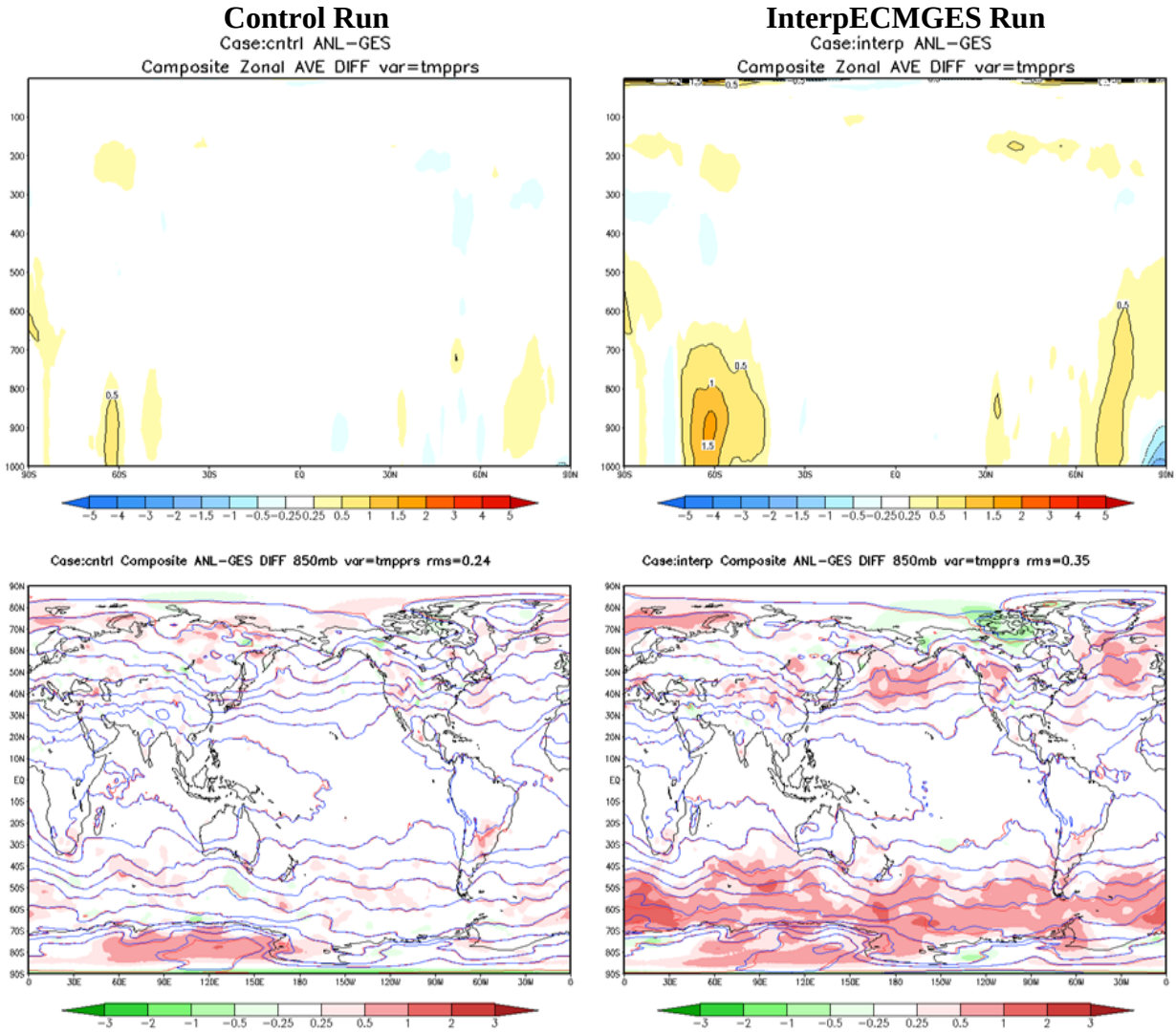
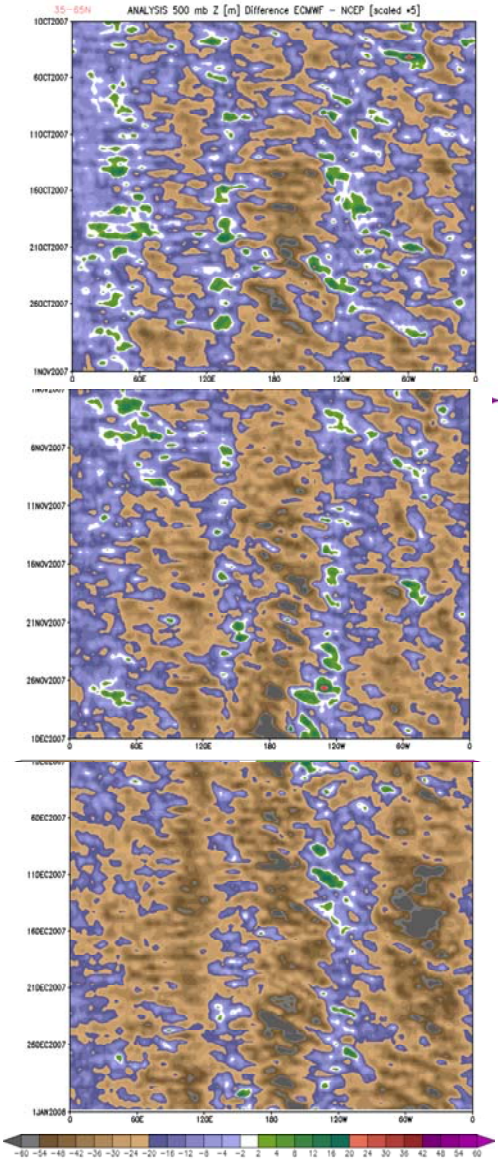


Figure 4. Differences of analysis minus background, control (left) and InterpECMGES (right). Temperature cross section differences (top) of zonal averages and 850 hPa height differences (bottom) for 5-day composite average.



Rolf Langland (NRL Monterey) shows systemic height differences between all models and ECMWF (shown is ECMWF-NCEP).

Cause may be the difference in satellite window coverage (under study):

ECMWF (12-h) vs. others (6-h).

Plots at left show height difference plots of time (October to December 2007) vs. longitude, averaged over 35-65N latitudes.

The range of the bias is ±12 m

Figure 5. Hovmoller diagrams of 500 hPa height differences (GSI - ECMWF) courtesy of Rolf Langland.

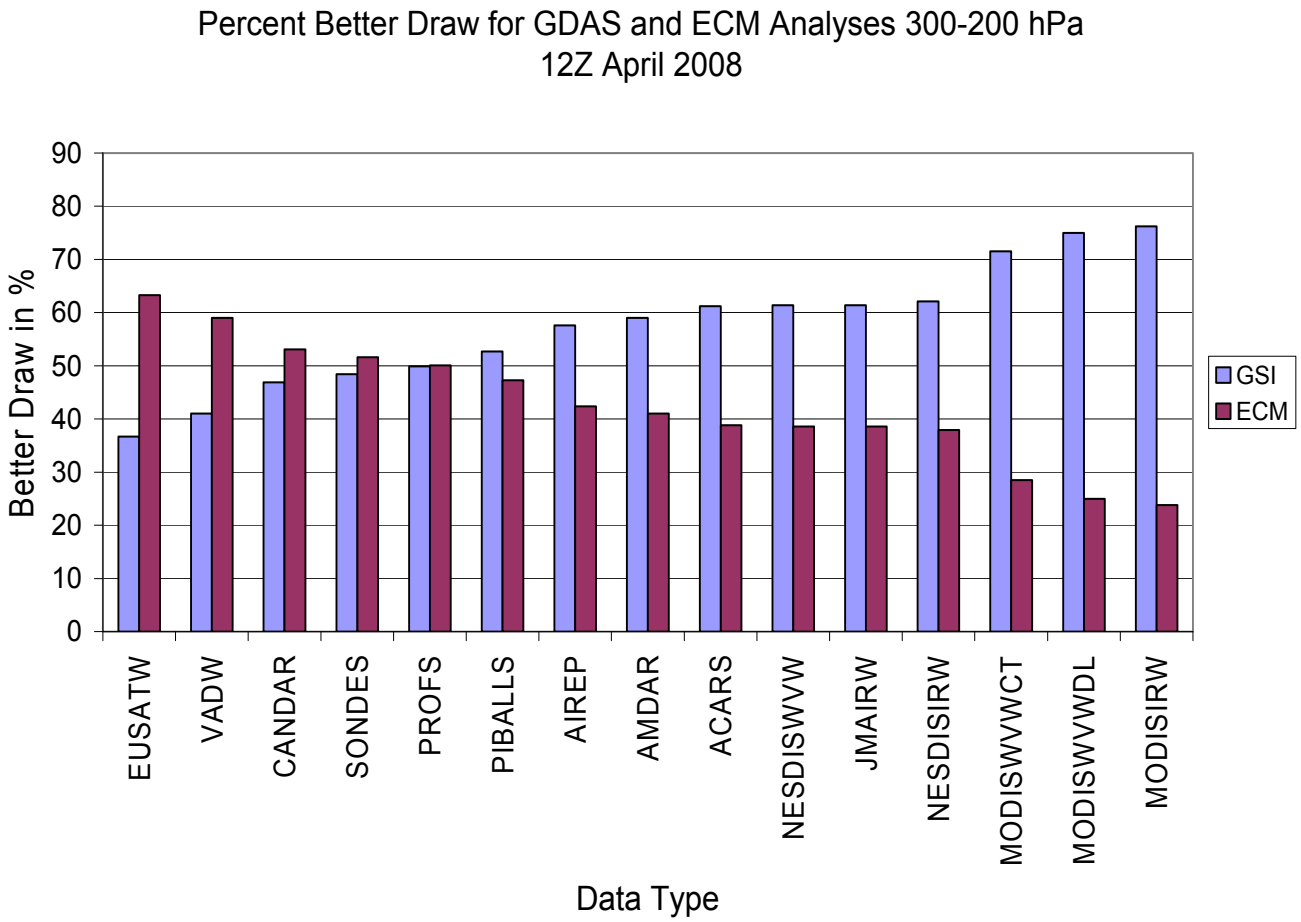


Figure 6. Comparison showing frequency of better draws to GDAS observations of the GSI and ECM analyses, 300 to 200 hPa, 12Z April 2008. See Table 1 for description of data types.

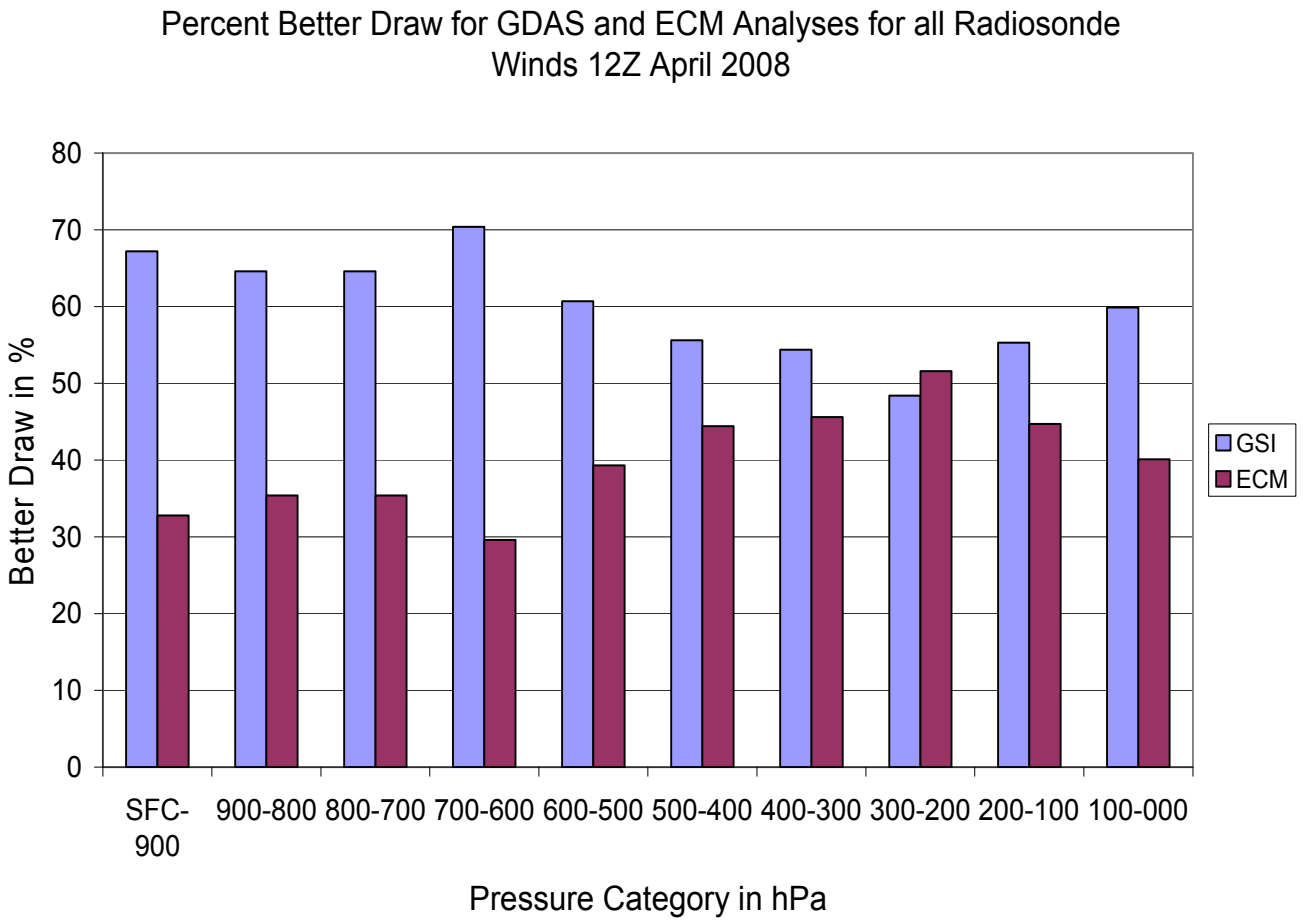


Figure 7. Comparison showing frequency of better draws to all GDAS radiosonde wind observations passing QC for the GSI and ECM analyses.

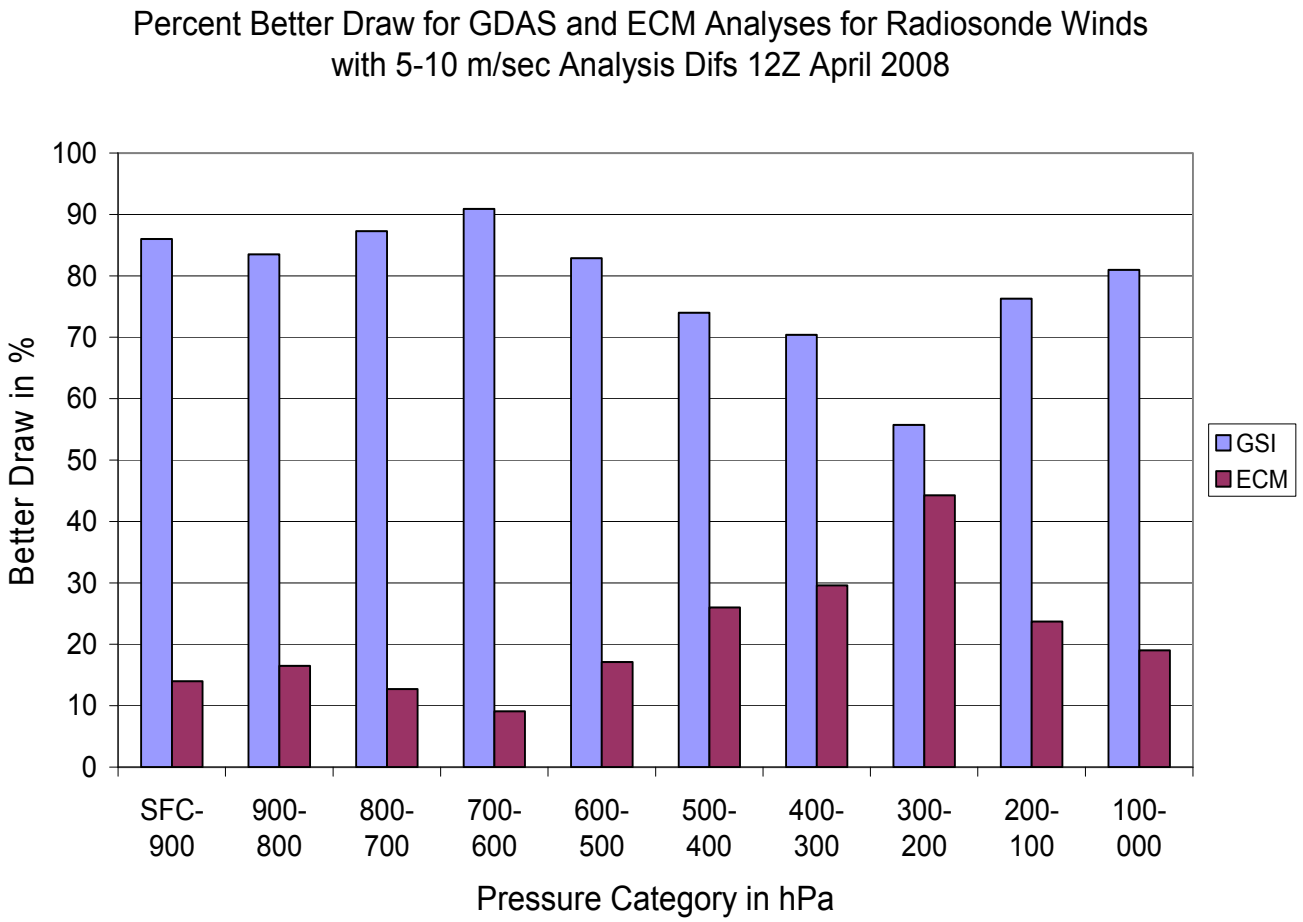


Figure 8. Comparison showing frequency of better draws to GDAS radiosonde wind observations passing QC where the two analyses have vector differences of 5-10 m/sec for the GSI and ECM analyses.

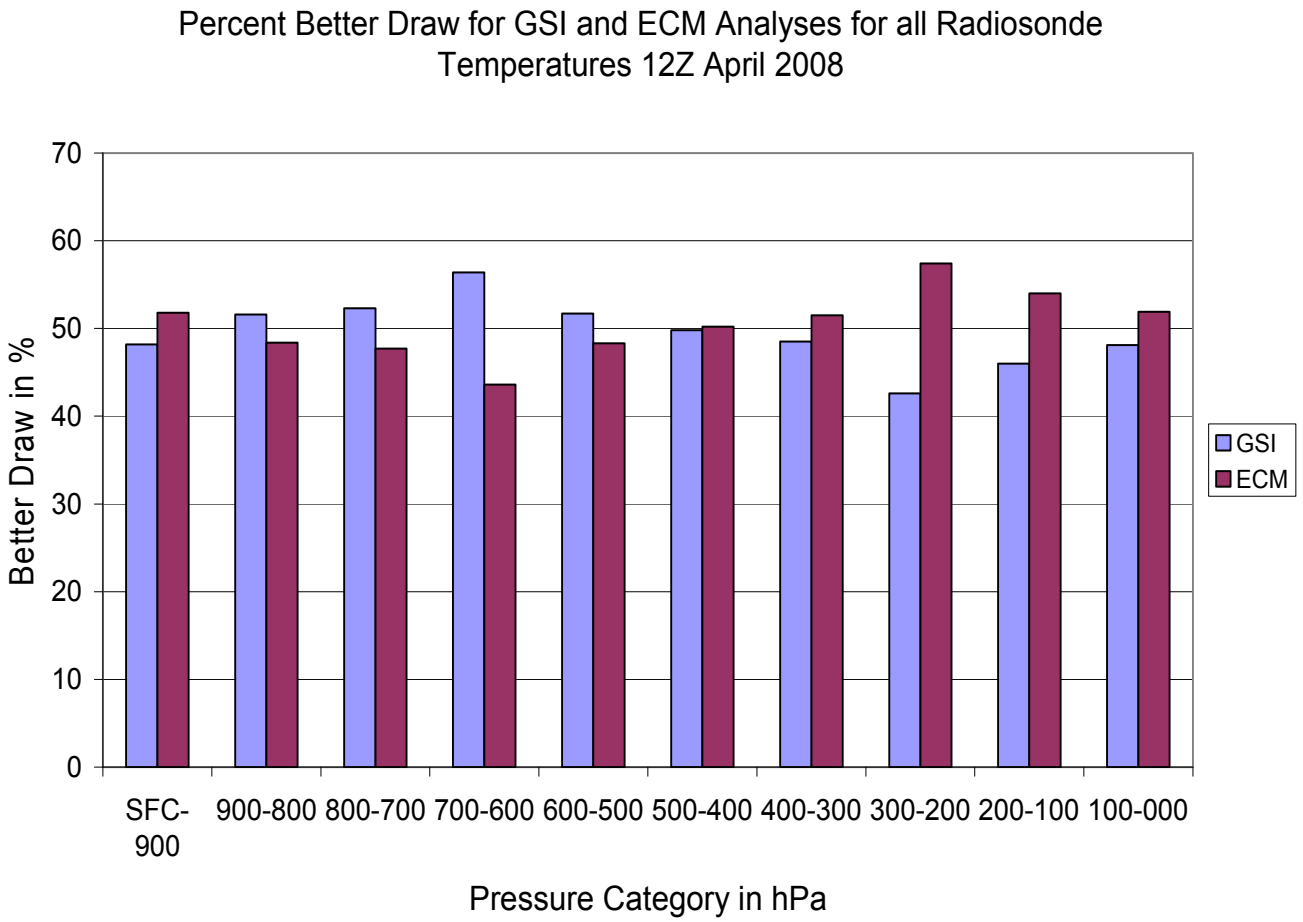


Figure 9. Comparison showing frequency of better draws to all GDAS radiosonde temperature observations passing QC for the GSI and ECM analyses.

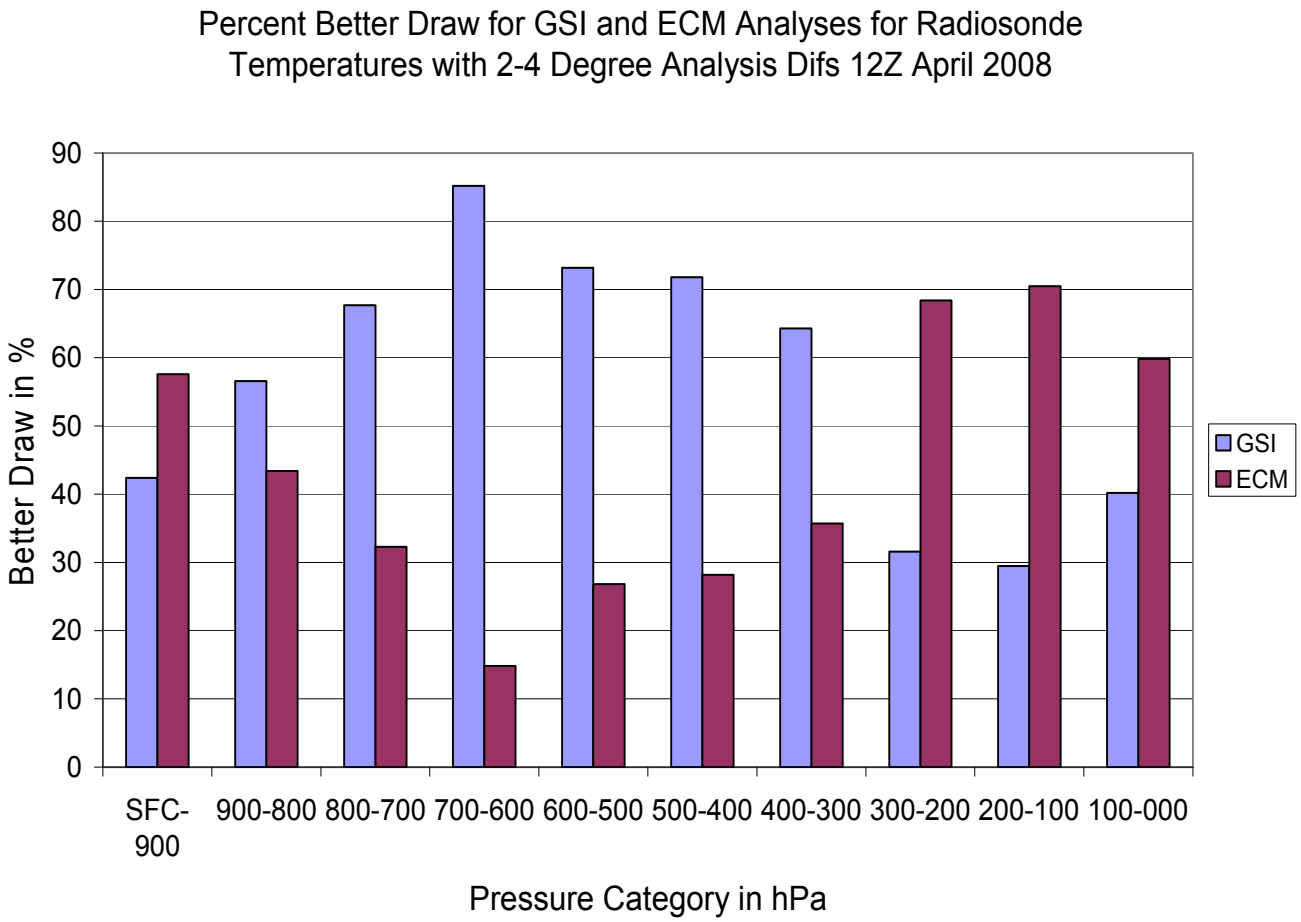


Figure 10. Comparison showing frequency of better draws to GDAS radiosonde temperature observations passing QC where the two analyses have temperature differences of 2-4 degrees C for the GSI and ECM analyses.

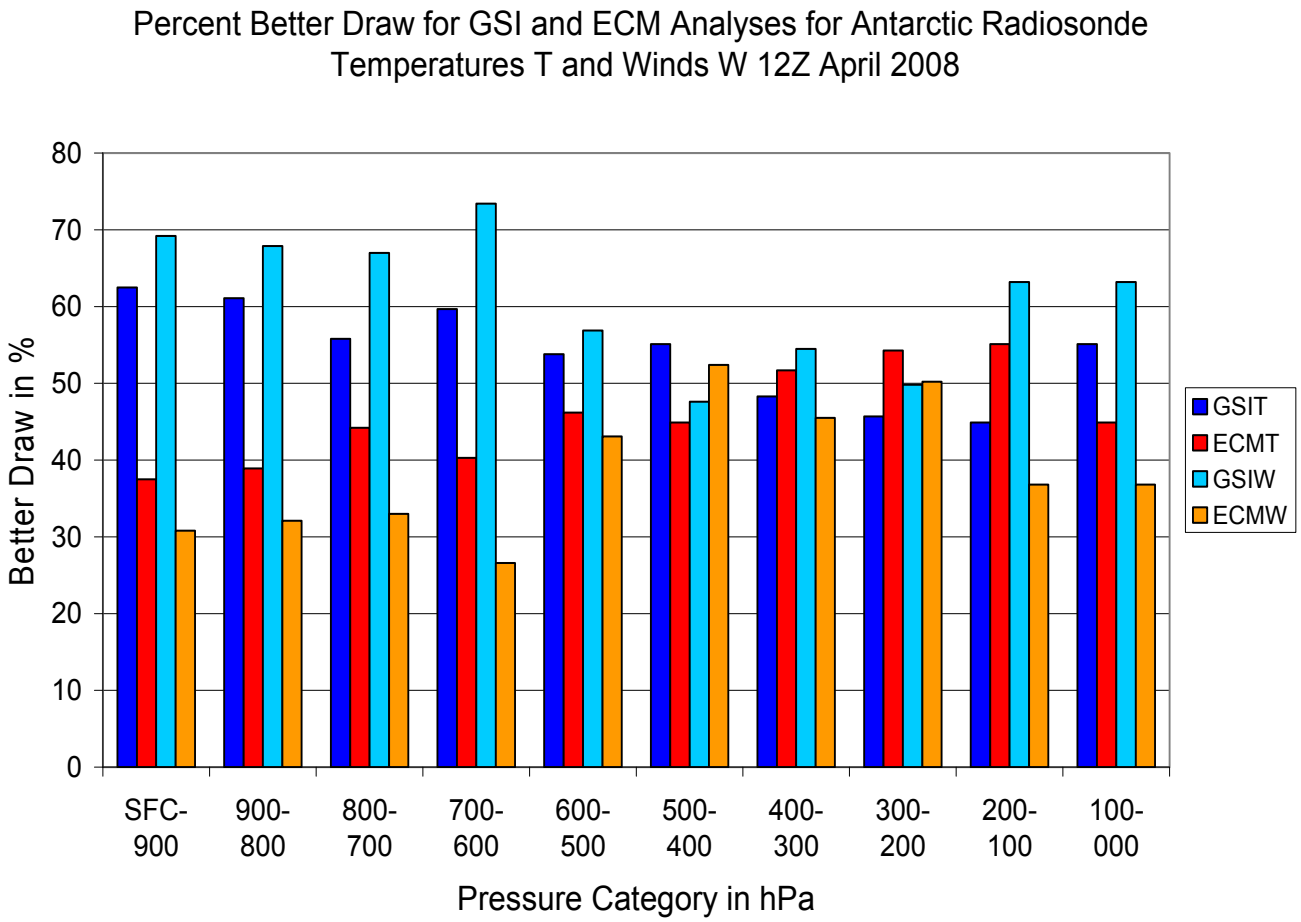


Figure 11. Comparison showing frequency of better draws to GDAS Antarctic radiosonde temperature and wind observations passing QC for the GSI and ECM analyses.

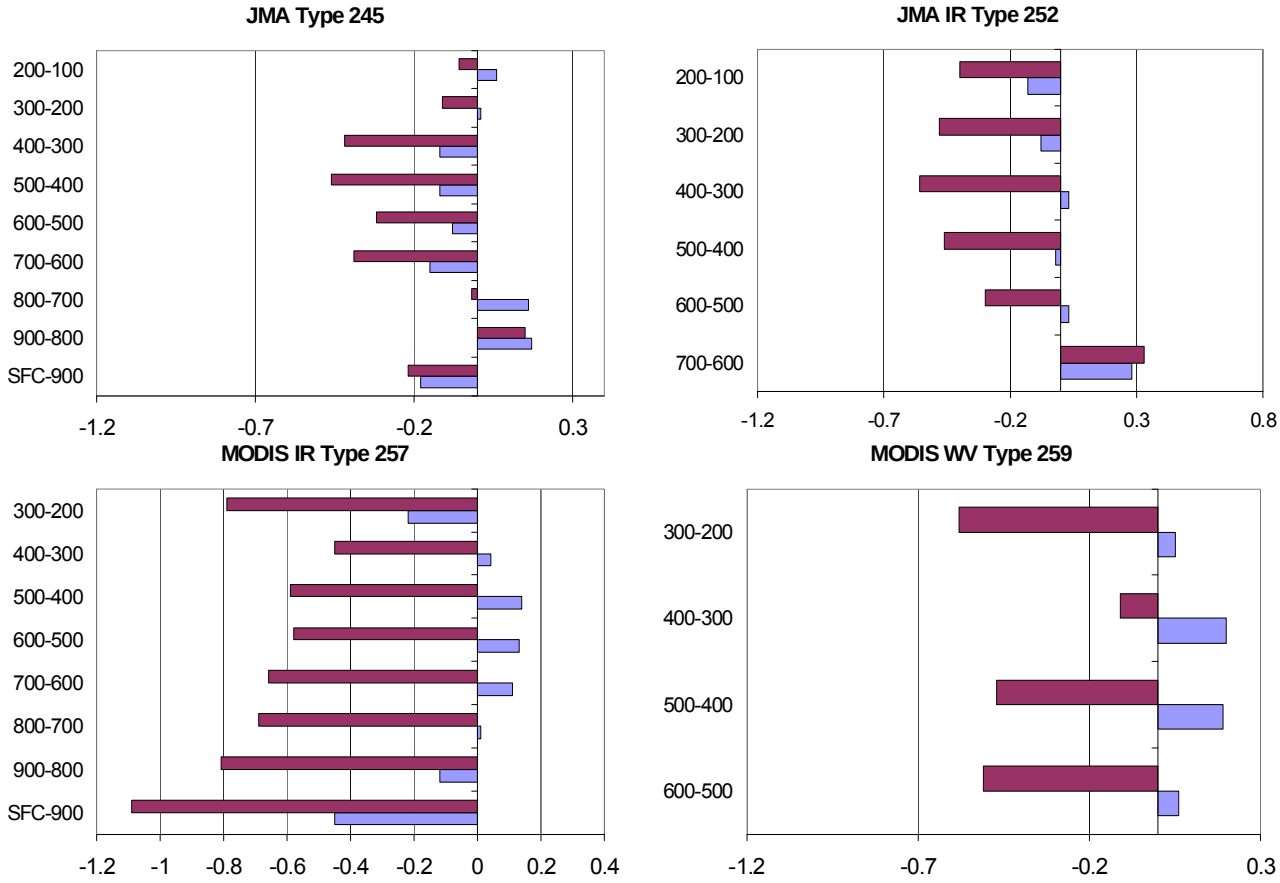


Figure 12. Comparison of speed biases in m/sec (observation minus analysis) for GSI (blue) and ECM (red) analyses for 12Z April 2008 for 4 different satellite wind types.

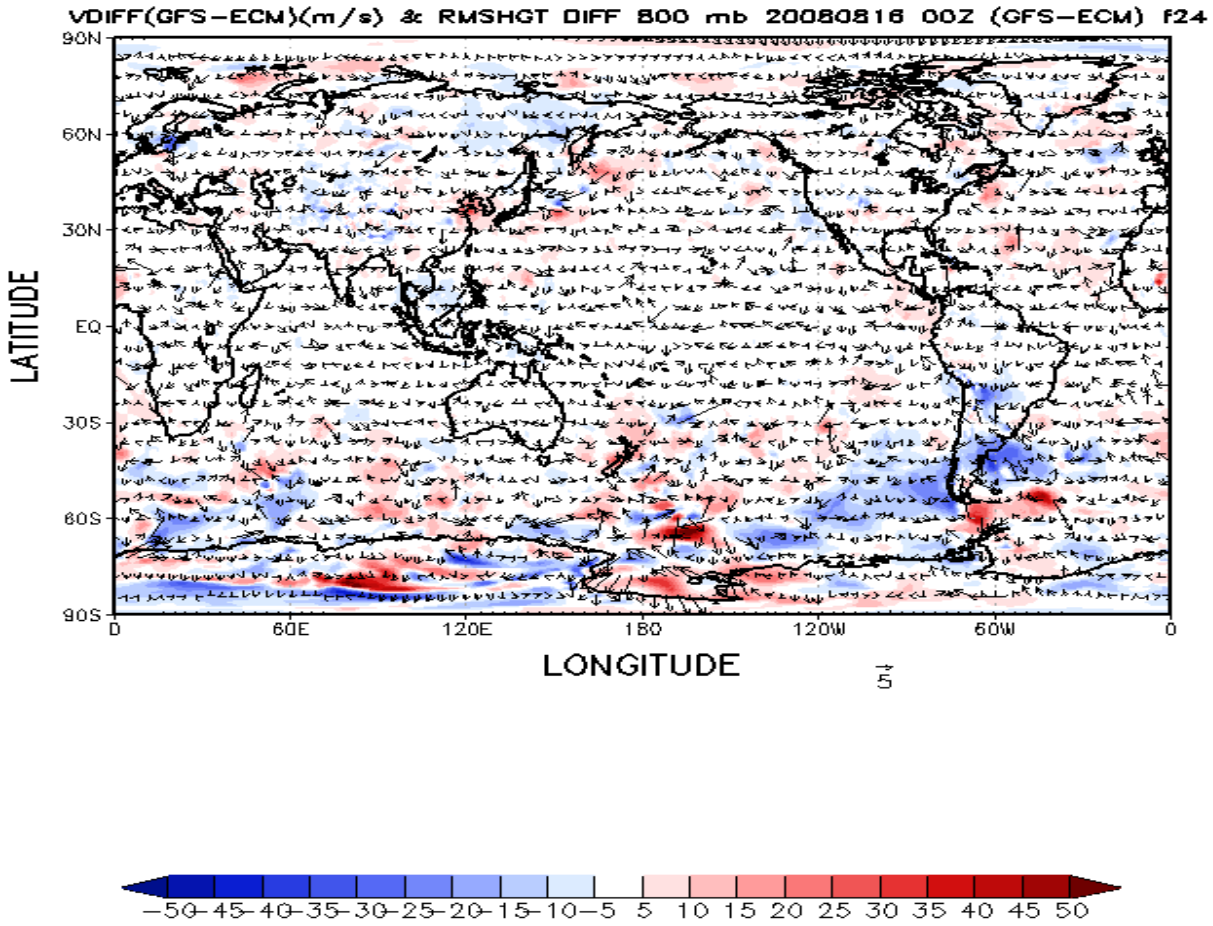


Figure 13. Geopotential height (m) root mean square 24-h forecast error differences between GFS and ECM runs at 800 hPa for 00Z 16 August 2008 . Vector wind differences (GFS – ECM) (m s^{-1}) are shown as arrows.

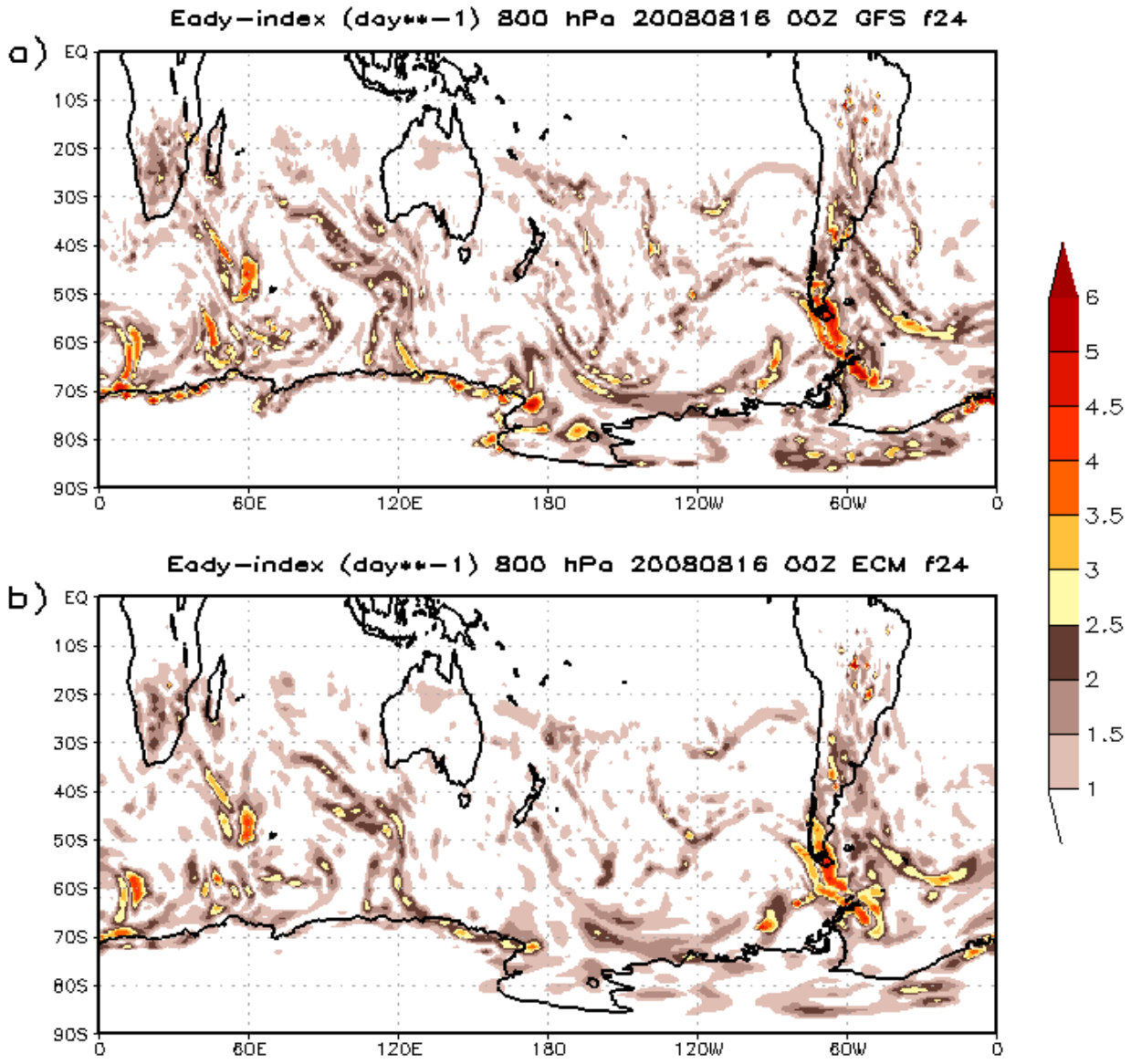


Figure 14. The total Eady Baroclinicity Index (EBI) (day $^{-1}$) at 800hPa for the 24-h forecast from the 00Z 16 August 2008 initial conditions for a) GFS (top panel) and b) ECM (bottom panel) runs.

Ho

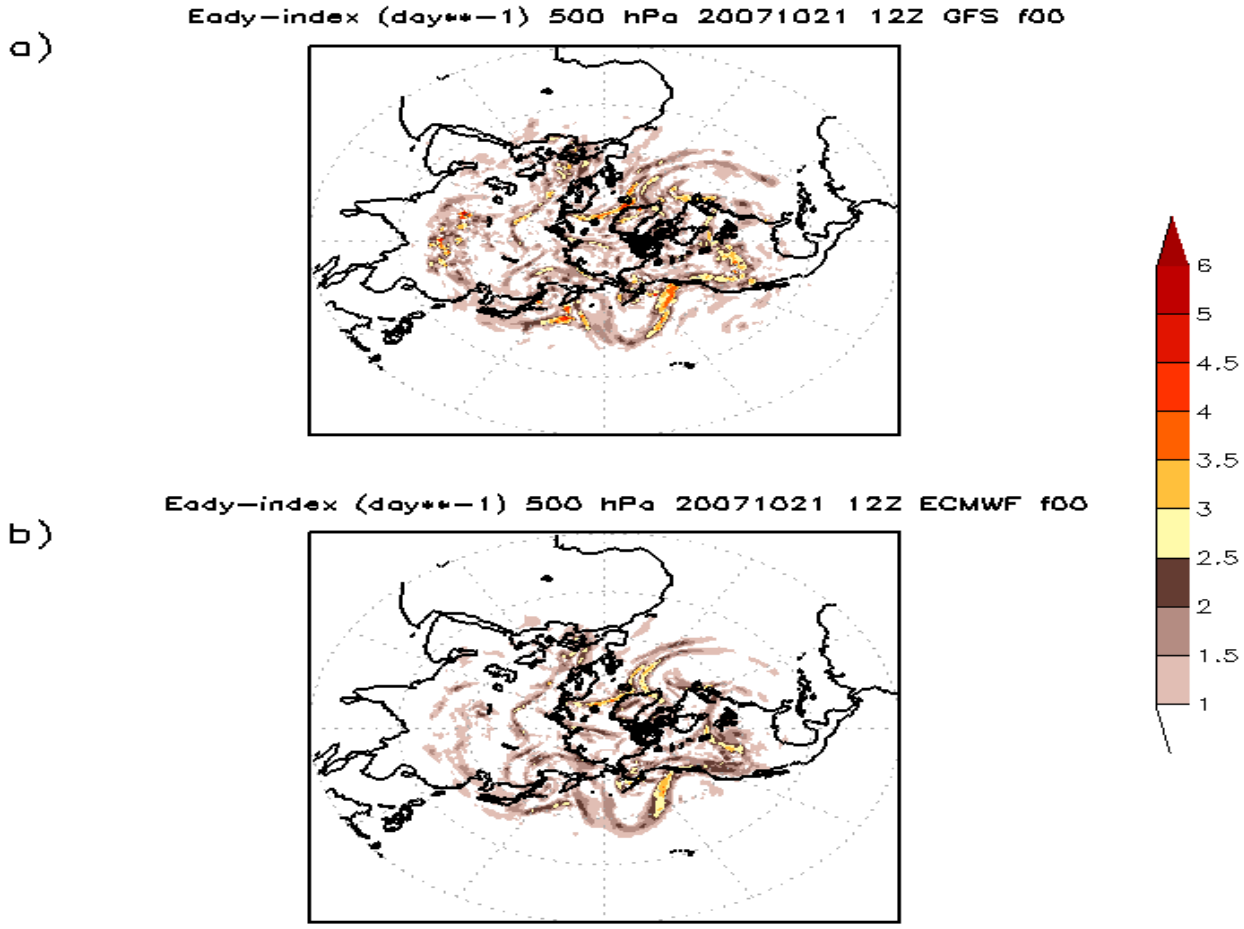


Figure 15. The total Eady Baroclinicity Index (EBI) (day $^{-1}$) at 500hPa for the F00 forecast from the 12Z 21 October 2007 initial conditions for a) GFS (top panel) and b) ECMWF (bottom panel) runs.