P2.1 IMPROVED FORECAST SKILL USING PSEUDO-OBSERVATIONS IN THE NCEP GFS

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

1. INTRODUCTION

Numerical weather prediction (NWP) models use a wide array of conventional and non-conventional observations estimating the state of the Earth's environment for their initial conditions (ICs). Successful assimilation of observations involves sophisticated algorithms and techniques for quality control (QC) and analysis. Models that embody the physical laws governing the behaviour of the Earth's atmosphere, ocean and land surface, and computers with the power to run these models rapidly enough to make timely predictions are an essential element of an effective environmental analysis and prediction system. NCEP's Global Forecast System (GFS) model integrates forecast cycles every 6 hours generating a background (guess) for the next assimilation and analysis of observations. The analysis system in production at NCEP is the Grid-point Statistical Interpolation (GSI) (Wu et al. 2002) which uses the background and all available conventional and non-conventional observations to generate an optimal analysis including the global surface pressure, and 3 dimensional dependent variables of motion, mass and moisture.

On approximately a monthly basis, poor forecasts or "Skill Score Dropouts" plague GFS performance. Other national center forecasts, for example, the European Centre for Medium-range Weather Forecasts (ECMWF), often do not exhibit this loss in skill. We attempt to quantify the skill differences between the GFS and ECMWF forecast system when there are dropouts, and define area(s) at IC time that have an impact on 5-day forecasts. Our goal is to find differences that can lead to algorithms to detect and correct data QC, bias correction and analysis issues in ICs before the forecast begins. To do this one needs to construct experiments that will objectively compare results from these national center forecast systems. For the experiments, we use the ECMWF standard 15 level pressure, longitude/latitude, 1°x1° ICs converted to simulated or "pseudo" RAOB observations. To analyze low model forecast skill, we compare the operational GFS and ECMWF analyses as well as forecasts from these analyses. Treating the ECMWF gridded ICs as pseudo-observations, and using them as sole input into the GSI analysis, which then acts as a "grand interpolator", generating new ICs that inherit ECMWF analysis system characteristics are labelled as "ECM" runs described in more detail in section 3. From these ICs, GFS forecast experiments, close to operational configuration (at T382L64), are made for comparison with NCEP's operational forecasts (control) to detect differences between these systems in time and space for QC and other investigations. The GSI analysis and GFS forecast model are each complex systems which require varied input and settings. We also investigate controlling certain aspects of these systems such as the influence between different versions of the GSI over the last year as it has progressed with upgrades, and the method by which the model background (guess) is used.

Generally, the ECM results show improvement in GFS 5-day skill scores in practically all Southern Hemisphere (SH) and most Northern Hemisphere (NH) cases. In addition, application of GSI pseudo-observations derived from similar standard GFS (instead of ECMWF) post processed ICs produces forecasts similar to GFS production for typical and dropout cases. This provides a way to make comparisons between forecast systems and isolate differences in QC of observations from conventional and non-conventional observation sources as well as other analysis differences. ECM runs can be used to find the locations that are responsible for dropouts.

Regions that influence the sensitivity of forecast skill can be found by creating a hybrid IC from selectively using the ECMWF or GFS pseudo-observations as input data for the GSI analysis. A region is chosen where "patches" over special areas are substituted in the pseudo-observation file, e.g., areas where there is ambiguity in observation quality, perhaps from

* *Corresponding author address:* Jordan C. Alpert, DOC/NOAA/NWS/NCEP/EMC, 5200 Auth Rd. Rm 207, Camp Springs, MD 20746; e-mail: jordan.alpert@noaa.gov

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

areas of cloudiness and other observation contamination, or in latitude bands to isolate bias and quality control problems that alter 5-day forecasts. We show areas of meteorological potential action. The interplay between QC, the assimilation of various observations, the analysis constraints on the incremental changes from a previous forecast, and the character of the model guess as “memory” of past events, is studied for GFS dropout forecasts. Once the ECM analysis experiments are run, one can use the results interpolated to individual observation locations to analyze how the GSI and ECMWF analyses “draw” to these observations. The resulting differences in statistics can be used to discover quality control issues and create algorithms for complex QC, and implement real-time QC detection/correction schemes. This is also demonstrated in an accompanying report (Ballish et al. 2009).

2. DESCRIPTION

Differences in forecast skill between national weather centers can be due to differences in the observations ingested or from the use of observations and the QC employed. Other differences can be due to the way the analysis system assimilates and “draws” for the observations. The observation data cut-off for each analysis system cycle varies across national centers, 6-h for GFS and most national centers and 12-h for ECMWF. The cycling analysis system, as used by national centers, is a never ending cycle of influence from model forecasts and the models physics package used there in that begins the analysis process with a background guess in combination with the available observations. “The analysis background (guess), which comes from the forecast model, which came from the analysis, which comes from the previous background (guess), which comes from that previous analysis, etc”…makes the determination of analysis problems difficult as the influence from forecast model physics and analysis differences become a complex mix. To investigate this problem with the goal of improving the QC and evaluating the conventional and non-conventional observations, a collection of tools for case study examination using an analysis system and statistics are used. A suitable, but admittedly arbitrary criteria, to determine when “dropout” cases occur, that is, the GFS 5-day Anomaly Correlation (AC) 500 hPa forecast height score necessary to be considered a dropout, has the following criterion:

- At least two of the following criteria must be met:

    - a) ECMWF minus GFS AC > 15 AC points
    - b) GFS AC < 0.70
    - c) ECMWF AC < 0.75
    - d) Monthly avg. GFS AC score minus GFS forecast > 15
    - e) Monthly avg. ECMWF AC score minus ECMWF forecast > 15

The point system is the following: if ECMWF=0.95 and GFS=0.80, this represents a 15 point difference in 5-day AC score. The dropout criterion is the same for NH and SH. Using this criterion we list the IC dates for the last year in the form YYYYMMDDHH where HH is the cycle time at 00, 06, 12, or 18 Z in Table I. A number of these days are selected for experiments and comparison with ICs supplied from ECMWF analysis as indicated in Fig 2.

3. ECM EXPERIMENTS

An independent national center source of information is needed to compare with GSI analyses. Since ECMWF forecasts do not often exhibit the loss of skill that occasions the GFS forecasts, we chose it as a proxy for ground truth. That is, we chose fields from ECMWF to engineer GFS analyses suitable for comparisons and useful as a means to construct controlled experiments. Analyses that are derived from the ECMWF information by use of the GFS Gridded Statistical Interpolation (GSI) herein, will be called ECM analyses and the method to calculate them are schematically shown in Fig. 2.

The ECMWF operational medium range prediction model is spectral T799 with 91 vertical levels. The fields used for this study are from 15 standard pressure levels, interpolated, and post processed files on a 1°x1° equally spaced cylindrical projection longitude/latitude grid. Each file contains surface pressure, u-, v-components of the wind, temperature and relative humidity on 15 standard levels (including the surface pressure) 1000., 925., 850., 700., 500., 400., 300., 250., 200., 150., 100.,50., 20., and 10 hPa. The GFS operational model, by comparison, also is a spectral model with truncation T382 and uses a physics Gaussian grid of about 0.3 degrees with 64 vertical levels.

An orography grid from GFS operations is interpolated to the 1°x1° grid and a new surface pressure is constructed hydrostatically taking account of these elevations for pseudo-observations. Specific humidity and temperature are calculated from the given variables and converted to profiles with appropriate coding and headers to have them appear as profiles of “pseudo-observations” for the above given pressure levels at all the 360x180 grid points. To assimilate these pseudo-observations we include fixed and background guess fields from production archives necessary to run the analysis system. The GSI observation input is confined to the pseudo-observations so they are assimilated in conjunction with the background guess taken from a previously run production analysis for the date/time in question. Using 6-h forecast from the previous cycle or an analysis were found to give similar forecast skill for ECM runs. The resulting analysis has some influence from the GFS background guess and this is discussed in section 6. To decrease the GFS influence or memory stored in the background, the resulting GSI analysis from the pseudo-observations is recycled as the background guess, and the GSI is rerun using the same pseudo-observations as the only observations to produce an analysis that can be considered a proxy for the ECMWF analysis.

The resulting product is an analysis that can be considered as a “grand interpolation” for GFS ICs from the original ECMWF information. These analyses are used in conjunction with surface and fix fields from GFS production archives as ICs for experiments. The AC scores for ECM experiments (Fig. 3 top) used the Statistical Spectral Interpolation (SSI) system, the forerunner to the GSI which also gives similar results to the current GSI (Fig. 3 bottom). For a period JAN-MAR of 2007 for the SSI and DEC2007 for the GSI, we show AC skill scores in Fig. 3 for a series of GFS and ECMWF production and ECM 5-day forecasts. These runs confirm similarities between the ECMWF production and the ECM pseudo-observation runs in terms of AC skill. In addition, if we repeat this process for the GFS production output files with the same information (15 pressure levels, 1°x1° grid, surface pressure, u-,v-components of the wind, temperature, and relative humidity) as contained in the ECMWF files, then the results return forecast skill very similar to GFS operations. The conclusion drawn from this is the analysis system faithfully reproduces a proxy for GFS or ECMWF model analysis system using pseudo-observations as input to the analysis system. From Fig. 3, it is shown that significant improvement from using these ECMWF analysis as IC for 5-day forecasts in terms of the AC height score improving 3 points in the NH to 8 points in the SH for the period shown for the SSI. The GFS with the new GSI is improved for this period but the ECM runs skill falls between GFS and ECMWF operations. It should be noted that one difference between the hemispheres is the amount of Ocean area and therefore the amount of land based conventional observations especially RAOBs contributing to the analyses and forecasts. This could be the reason for the ECM runs consistently making greater improvement in the SH skill scores compared to the NH.

4. OVERLAY FOR A TYPICAL DROPOUT CASE

An example of a NH dropout is the 2007102212 GFS production IC (F00) which when integrated resulted in a 5-day forecast that verified with an AC skill score of 0.61, as shown on the banner at the top of Fig. 4. This particular case by definition is a dropout based on criterion a, b, and d earning this case a place on the dropout list (Table 1). The production ECMWF 5-day forecast had AC score of 0.87 and the ECM run was 0.89 (Fig. 4), both alleviating the dropout. Comparison of the 500 hPa geopotential of the GFS and ECMWF production (similar to Fig. 4c) IC show very little difference and confirms that a slight difference is sufficient at IC time to cause very different day 5 forecasts. What difference is present between these two national center model IC’s ranges as much as ± 20m in height with virtually all large differences located within a broad trough in the Central Pacific (within the box drawn in Fig. 4c as shown by the red color fill in that area). The height fields at other levels (not shown) give a similar result, and similar differences in temperature (not shown) are predominately from this same Pacific location. A number of wind maximums are present in both the ECMWF and GFS analyses (not shown but consider the height gradients) which completes a synoptic picture of a volatile broad trough with a number of short waves moving within. The red color fill area indicates higher heights for the GFS (Fig. 4c). This shows that the GFS IC difference is largely an amplitude problem and not from a phase error. The largest 5-day forecast error at 500 hPa, in this case, is found to the east of Greenland and is largely responsible for the low AC score as shown in Fig. 5c.

To test that the dropout originated from the IC differences in the Pacific region we integrate an “Overlay” (OVRLY) forecast experiment for 5-days with GFS pseudo-observations but with ECM pseudo-observations only over a prescribed area or “patch” in the Pacific as shown by the box in Fig. 4c. The pseudo-observations used for the GFS and the ECM overlay include all the dependent variables and surface pressure at each latitude. This “hybrid” set of pseudo-observations is used as the only observation input to the GSI analysis system. The GFS production analysis is used as the background guess and GFS production fixed fields such as albedo, snow, etc… are needed to start the analysis as described in Fig. 2, and the resulting analysis is used to make a 5-day forecast called the “OVRLY run”. The 5-day forecasted height and their difference to the verifying analysis for this OVRLY run (GFS IC with the ECM overlay substituted only over the patch area) at 500 hPa, is shown in Fig. 5b. The color fill in Fig. 5a compared with that in 5b shows much greater forecast error in the GFS production thus, the ECM values over the Pacific OVRLY patch area are sufficient to alleviate the dropout. The OVRLY skill score shown in the banner of Fig. 5 confirms this finding. For a 5-day forecast, the GFS forecast errors are largest east of Greenland extending across the 0 meridian but these errors are greatly reduced in the OVRLY experiment. The associated trough error in the Greenland area 5-day forecast can be traced back to the Pacific OVRLY region described above in Fig. 4c at IC time. The resulting analysis is a hybrid of the two national center analyses, but the information content of the dependant variables from the better scoring ECMWF analysis is placed only over the Pacific area in question as shown in Fig. 4c. The resulting 5-day OVRLY experiment forecast skill score is shown on the banner of Fig. 5 and is 0.90 confirming that the problem area is the outlined Pacific area. Smaller rectangles centered on the broad Pacific trough were studied with similar results; however when the OVRLY was moved to an area far away from the Pacific, the system reverted to the production GFS and the dropout re-occurred with similar loss of skill. This gives rise to the idea of possible areas of analysis/model sensitivity and we note the lack of conventional observations causing the analysis system to rely more on non-conventional observations as it does in the SH.

In the SH, the differences between the GFS and ECMWF ICs are seldom centered in a single area, as shown in Fig. 6 for a typical case, 2008030312 at 500 hPa. The differences shown in Fig. 6 color fill are aligned with the active areas of troughs and ridges, for example at 500 hPa (not shown) as in the NH except that the SH is more active. Experiments are run replacing GSI observational data with ECMWF pseudo-observations over two latitude bands: 20-60S and 60 – 90S as shown in Fig. 6. The skill score results for this SH case, 2008030312, are listed in Fig. 6 as well. The GFS operational run has a score of 0.59 classifying it as a dropout, compared to 0.85 for both the ECMWF operational run and the ECM run. When we apply the indicated overlays, each latitude band contributes information which improves the 5-day forecast, 0.84 from the southern mid-latitude band and 0.74 from the Antarctic latitude band. However, neither overlay area experiment returns the very high skill found in ECM forecast runs from global application of ECMWF pseudo-observations, but they do alleviate the dropout according to our criteria. The 5-day forecasts for these two overlay cases are shown in Fig. 7 with their respective skill scores. The forecast error 500 hPa map for the operational run is shown in Fig. 7a where there is significant error in the height, for example in the southern part of South America. The forecast error in either overlay is reduced in this area as shown in Fig. 7b, and greatly reduced elsewhere in the mid-latitude overlay experiment. Thus, the forecast error shown in Fig. 7b color fill is the smallest and the GFS minus the ECM overlay for the mid-latitude experiment (Fig. 7c color fill) is therefore similar to the GFS forecast error showing large forecast error but similar to Fig. 7a.

5. SOUTHERN HEMISPHERE ECM EXPERIMENTS

We have collected 10 SH dropout cases from the first half of 2008 and present the results of 5-day forecast runs for GFS production, ECMWF production, and ECM runs in Table 2. The GFS production has skill scores less than 0.7 in all cases while the ECMWF production is close to 0.8 or above, except for one case 2008031812. The ECM runs have significantly better skill by this measure than all of the GFS operations and by our criteria, alleviating all of the dropouts, and in the case mentioned, improved upon the ECMWF production forecast skill although this is not usually the case. Most of the time the ECM runs are a few AC points behind ECMWF probably because of the vertical resolution, 15 vertical levels instead of 91 used in ECMWF production, and the 1°x1° longitude/latitude grid compared to the higher resolution ECMWF physics grid from their corresponding T799 spectral truncation Gaussian grid of about ~0.2 degrees.

One issue is could improvement in the background guess for the GSI analysis result in better forecast skill? Using the given ECMWF pressure (GRIB) file as a background is not possible as the GSI normally uses 3-6-9-hour forecasts from the GFS model previous 6-hour cycle in an internal format with spectral and vertical coordinate representation for the background. The ECM runs can be used as palatable background guess input to the GSI which remove past memory of the cycling production GSI. Another issue is that the second GSI run used in creating the ECM analysis (see Fig. 2) will cause noise in the resulting analysis. We note that we always verify with the GFS operational analysis. Such noise usually dissipates in the first forecast day and does not affect a 5-day forecast. At this point we continue to investigate these details in order to obtain clean comparison tools to better examine assimilation and QC behavior.

To examine the influence of the set of production observations one may use the ECM analysis, as described in Fig. 2, as the background guess plus the full set of observations from production[†]. The analysis produced from this procedure can be integrated (T382L64) for the dropout cases and the result of these experiments are shown in Table 2 under the column ECMANLGES. Comparing the ECMANLGES with integrations from the ECM runs (ECM column in Table 2) show that adding the conventional and non-conventional observations, and their processing by the GSI, causes degradation in skill score. The GSI has had a number of upgrades and changes over the time the dropout examples have been compiled. These improvements include the addition of new observation types and calibration changes for observations such as satellite radiance bias corrections. These changes have improved the analysis which we rerun as a new control (the CNTRL column in Table2) to compare with the GFS column containing the GSI that was in force when the dropout occurred. The results show that the GSI improvements have resulted in skill improvements in most cases.

Finally we test the influence of replacing the GSI analyses previous 3, 6, & 9-hour forecasts (instead of a single 6-hour background guess) with ECM forecasts as background guess plus GDAS observations which is more like the GSI production. This means running a previous ECM run from 6-hours earlier than the IC time and using the resulting 3, 6, and 9-hour forecasts as background guess input to the GSI. The full set of observations is used for the analysis and the AC skill score is in the InterpECMGES column in Table 2. Comparing the InterpECMGES with the ECMANLGES shows using the above ECM derived background guess improves the GSI analysis and alleviates some dropouts in a few cases, as did the CNTRL and ECMANLGES experiments, but does not have the skill found in the ECM runs.

The implication that running the GSI with operational input data and an ECM background guess for some dropout cases decreased the skill of subsequent forecasts compared to ECM runs leads to the question of what is the cause of the degradation in forecast skill? One possible cause we can investigate is whether some observation type and/or observation contributed to the degradation. Work has been done in this area using adjoint methods (Zhu and Gelaro 2008), but it is instructive to run a set of experiments that include only one observation type to test the influence on the analysis and subsequent forecast skill. This experiment is done for the 2008020300 dropout case which originally had a skill score of 0.65 from GFS operations. A control run of the latest GSI system is rerun (slightly different from the earlier control “CNTRL”) described in Table 2 and had a skill score of 0.70. Conventional data including RAOBS, ships, buoys, aircraft, satellite cloud track winds, and other observation types were used, but not including the radiance observations, are called “PREPBUFR” runs after the file name that stores them. In Table 3, shown in blue, a radiance observation type can be removed as in “CNTRL NO AIRS” meaning all data is present except these satellite radiance observations. Or experiments with the conventional observations (PREPBUFR) present can add a single satellite radiance type which is shown in red. In Table 3, the blue section indicates that removing each satellite radiance observation contribution from AIRS, HIRS 2,3,4, AMSUA, AMSUB, MHS, or no satellite radiances

[†] Experiments shown use the production Global Data Assimilation System’s (GDAS) observational data employing a 6-h cycle which accumulates observations over 6 hours instead of the analysis used for the operational GFS which has observation data cutoff of abut 2 ½ hours.

present (only PREPBUFR present) had little effect in improving the 5-day forecast skill for this dropout case. When individual satellite radiance contributions are included one at a time, in addition to the conventional observations in PREPBUFR, the result in Table 3 (red) show 3 observation types which give significant positive improvement in skill score. These are AMSUB (Advanced Microwave Sounding Unit B) at 0.77, GPSRO (GPS Radio Occultation) at 0.79 and MHS (Microwave Humidity Sounder) at 0.78. Including these three radiance observation types in addition to the conventional observations alleviates the dropout and returns the ECM (and ECMWF) skill. The IC 500 hPa height and difference between the GFS operational run and the best three radiance observation types are shown in Fig. 8 for the above mentioned case of 2008020300. The total height fields are quite similar to the eye (Figs. 8a, b) but their difference, shown in Fig. 8c (left), indicate adding the radiance observation types causes a raising of the GFS heights (green areas) in a number of trough areas and lowering heights in ridges (red areas) along the active latitude band from 40 to 75S as well as a raising of heights in part of the Antarctic. The corresponding 5-day forecasts in Fig. 8a shows the GFS operational forecast error verifying against its own analysis with forecast errors ranging almost 500 m which is reduced by 4 fold in the forecast using the best three observation types. The forecast difference shown in Fig. 8c (right) therefore shows large differences because of the reduced forecast error. The three radiance observation types, AMSUB, GPSRO and MHS influence the specific humidity which is surprising as emphasis has been on mass and motion quality control issues. However, applying this to two other dropout cases (two so far) on the list in Table 2 did not give similar improvement.

6. APPLICATION OF THE EADY BAROCLINIC INDEX TO IDENTIFY SENSITIVITY OF ANALYSIS ERRORS

There is a connection between baroclinic disturbances, growing model forecast errors, and synoptic activity in mid-latitudes (Klinker and Ferranti, 2001). Accordingly, one measure of baroclinicity we have been applying is the Eady Baroclinic Instability (EBI) index, a rate with units of day^{-1}, representative of the growth rate of unstable disturbances (Hoskins and Valdes, 1990). The application of EBI to NH and SH forecast dropouts are demonstrated in a companion paper by Ballish et al. (2009).

Figs. 9a,b (bottom left panel) show the EBI for the ECM run with two GSI iterations, and bottom right panel shows the ECM with a single iteration. A second GSI iteration is found beneficial, in terms of improved AC skill score, to force the ECM analysis/forecast to be more similar to ECMWF analysis/forecast (see Fig. 2). The second application of the GSI causes EBI to be noisier and amplified compared to both the operational GSI (Fig. 9), and ECMWF (not shown). With the operational GSI, the EBI rates for the background (06 UTC forecast from Dec22, 2008) and the corresponding analysis (12 UTC, Dec 22, 2008) show very similar characteristics (Figs. 10a,b). The initial noise and amplification of the EBI for the second application of the GSI disappears completely in the forecast evolution cycle as evident in the 5-day 500 hPa AC scores. Figs. 11a,b show the total EBI at 500 hPa for the 24-h ECM forecast (bottom panel) from August 16, 2008 00 UTC (a SH dropout event) and the corresponding operational GFS forecast (top panel). The ECM EBI distribution (Fig. 11b) is much smoother, and the operational GFS forecast shows more pronounced baroclinicity in the SH. The noise found at IC and the forecasts soon after is not present by 24-h forecast, and has little or no influence on the model skill. A number of issues could be causing the noise in the ECM runs but it is probable that using an analysis as the background (guess) instead of a 6-h forecast from a previous cycle (as is done in GFS operations) is the cause. This is due to a digital filter that is employed in the GFS model beginning at 3-h forecasts so the 6-h forecasts have a filter applied while analysis and ICs do not. The 6-h and 18-h cycles from ECMWF model runs are needed to create a 6-h forecast for input as a background (guess) and this is now available. The enhanced baroclinicity of the GFS as measured by the EBI may indicate potential areas that can cause dropouts and could be an indicator in a dropout detection scheme compared to the more expensive and sophisticated adjoint techniques.

7. SUMMARY

The use of ECMWF analysis pressure files to generate “pseudo-observations” for input to the Gridded Statistical Interpolation (GSI) and subsequent GFS 5-day forecasts, yield results that have the character of the ECMWF model in terms of forecast error and skill. These are called “ECM” runs, where the GFS operational skill is improved in the NH and more so in the SH for every day cases as well as when the GFS model 5-day forecast has very low skill which we have termed "dropouts". Dropouts seen in the GFS model seem to occur once a month in the NH and more often in the SH. A climatology of NH and SH dropouts has been generated to describe the systematic differences when the model has forecasts of very low skill. The goal of this work is to diagnose problems in Quality Control and other analysis issues to implement operational improvements.

GFS runs from ECM analyses show dropouts can be alleviated in GFS forecasts. Running the operational GSI with an ECM derived background guess results in better forecast skill than the operational GFS but not as good as ECM runs. Running the operational GSI after removing select observation types offers a systematic approach for assessing the impact of different observation types. Work continues to analyze what is the optimal fit of the analysis to observation types and to determine an implementable algorithm for improved quality control, bias correction, and analysis weighting of observations.

Application of diagnostic tools such as EBI are found to be useful in identifying the location of sensitive regions from which GSI analysis errors grow and further degrade GFS forecast skill. The second application of the GSI causes the EBI to be noisier compared to the single application of the GSI for an ECM run. This could be an artifact of the digital filter which is under investigation. The noise from the EBI calculations is transient in nature and does not affect the forecast. 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

Ballish, B. A., J. Alpert, D. Carlis, and V. K. Kumar, 2009: What Causes NCEP GFS Model Forecast Skill "Dropouts", 13th AMS IOAS-AOLS, 2.1

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.

Klinker, E. and L. Ferranti, 2001: Forecasting system performance in summer 1999. Part 1: Diagnostics related to the forecast performance during spring and summer 1999. ECMWF Tech. Memo. 321. Available from ECMWF, Shinfield Park, Reading UK.

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

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.

Table 1. Dates of Northern Hemisphere (NH) and Southern Hemisphere (SH) skill score dropout cases specified by initial condition date.

Dropout Table

Northern Hem. (NH)	Southern Hem. (SH)
2007102112	2007092912
2007102200	2007100212
2007102212	2007100412
2007102312	2007100612
2007111112	2007100700
2007122012	2007101300
2008012100	2007101400
2008021712	2007102000
2008030112	2007111912
2008030400	2007121612
2008030412	2007122000
2008060400	2007122012
2008060500	2008011100
2008060600	2008011112
2008062500	2008011212
2008070200	2008020100
2008070212	2008020112
2008070300	2008020300
2008070412	2008021512
2008070600	2008021700
2008070700	2008022000
2008070712	2008030112
2008070812	2008030212
2008071000	2008030300
2008071900	2008030312
2008092300	2008030912
2008092312	2008031012
2008100400	2008031212
2008100412	2008031300
2008101012	2008031412
2008101100	2008031800
2008101112	2008031812
2008101200	2008032012
2008101212	2008040900
2008101300	2008042500
2008102100	2008042512
	2008042600
	2008050900
	2008051000
	2008051512
	2008052200
	2008052212
	2008061212
	2008062500
	2008062512
	2008072500
	2008080500
	2008081500
	2008081600
	2008081912
	2008090212
	2008090300
	2008100912
	2008101212
	2008101300
	2008101912
	2008102112
	2008102200
	2008110600
	2008110612
	2008110700
	2008110900
	2008110912

Updated December 1, 2008

Table 2. SH 5-day anomaly correlation scores for experiments performed on dropout cases.

Initialization Date	GFS	ECMWF	ECM	ECMANLGES	CNTRL	InterpECMGES
2008011100	0.68	0.83	0.80	0.76	0.74	0.82
2008011212	0.69	0.89	0.86	0.83	0.82	0.77
2008020300	0.65	0.83	0.83	0.82	0.69	0.63
2008030312	0.59	0.85	0.85	0.77	0.69	0.78
2008031800	0.59	0.79	0.75	0.56	0.69	0.70
2008031812	0.66	0.63	0.76	0.66	0.78	0.75
2008042512	0.67	0.80	0.72	0.72	0.63	0.69
2008042600	0.61	0.91	0.89	0.68	0.65	0.72
2008052200	0.60	0.87	0.84	0.83	0.73	0.70
2008062512	0.66	0.87	0.77	0.70	0.72	0.77

Table 3. SH 5-day anomaly correlation scores. The Q4 FY08 GSI is used for all experiments. Tests in blue involve a GSI analysis without different satellite types in the analysis process (i.e. NO AIRS suggests that all other satellite data is available except AIRS). The CNTRL (red) uses all available data and uses a 6-hr cut-off for the assimilation cycle; however the CNTRL W/PREPBUFR uses only the conventional data of the prepbufr file to create an analysis. Other experiments involve the prepbufr file plus one satellite type to create an analysis.

Initialization date: 2008020300 Operational GFS SH AC=0.65	SH AC Score
CNTRL NO AIRS	0.67
CNTRL NO HIRS(2,3,4)	0.67
CNTRL NO AMSUA	0.73
CNTRL NO AMSUB	0.67
CNTRL NO MHS	0.66
CNTRL NO RADIANCE	0.72
CNTRL	0.70
CNTRL NO SATELLITE DATA (EXCEPT SSMI & TRMM)	0.70
CNTRL W/PREPBUFR	0.68
CNTRL W/PREPBUFR + AMSUA	0.70
CNTRL W/PREPBUFR + AMSUB	0.77
CNTRL W/PREPBUFR + HIRS2	0.68
CNTRL W/PREPBUFR + HIRS3	0.72
CNTRL W/PREPBUFR + HIRS4	0.67
CNTRL W/PREPBUFR + AIRS	0.75
CNTRL W/PREPBUFR + GPSRO	0.79
CNTRL W/PREPBUFR + MSU	0.68
CNTRL W/PREPBUFR + MHS	0.78
CNTRL W/PREPBUFR + SNDR	0.68
CNTRL W/PREPBUFR + GPSRO,AMSUB,MHS	0.87

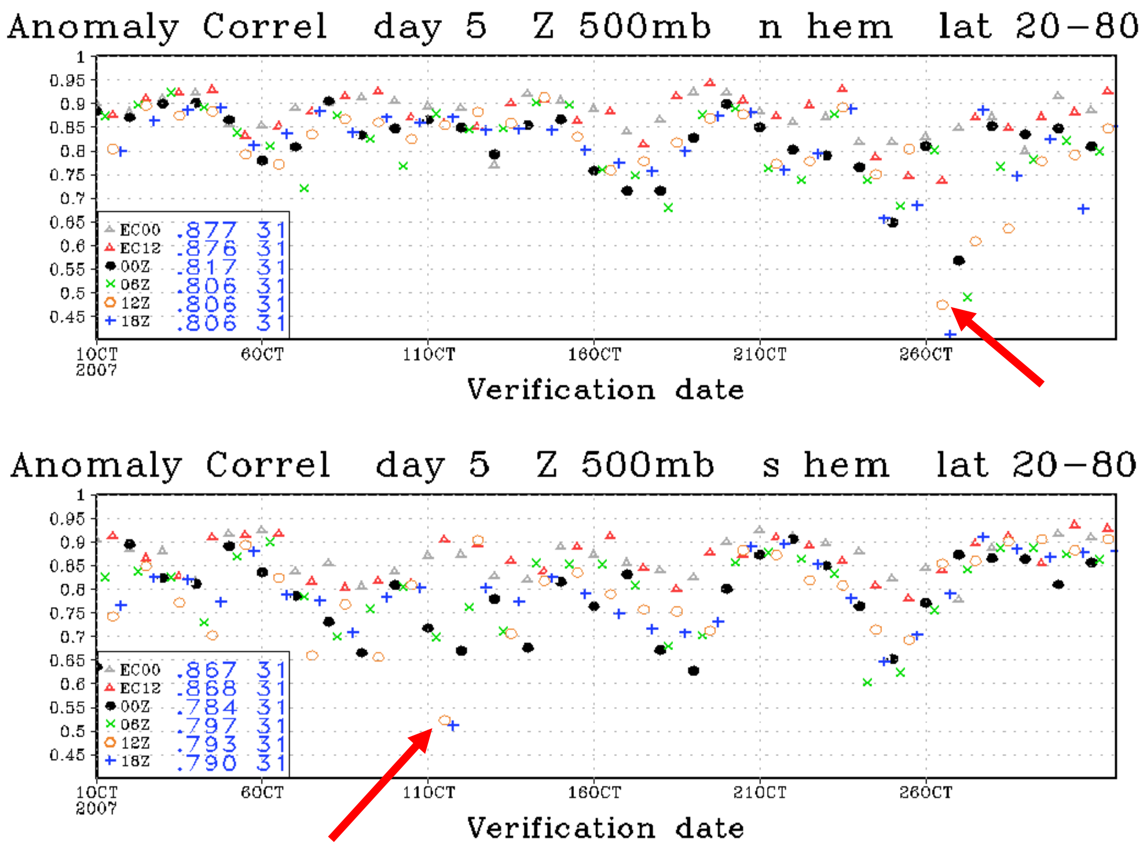


Figure 1. 5-day forecast 500 hPa anomaly correlation skill score for 20-80 North (top) and 20-80 South (bottom) for the GFS at 00, 06, 12 and 18 Z cycles and ECMWF 00 and 12 Z cycles during October 2007. Red arrows point to skill score dropouts.

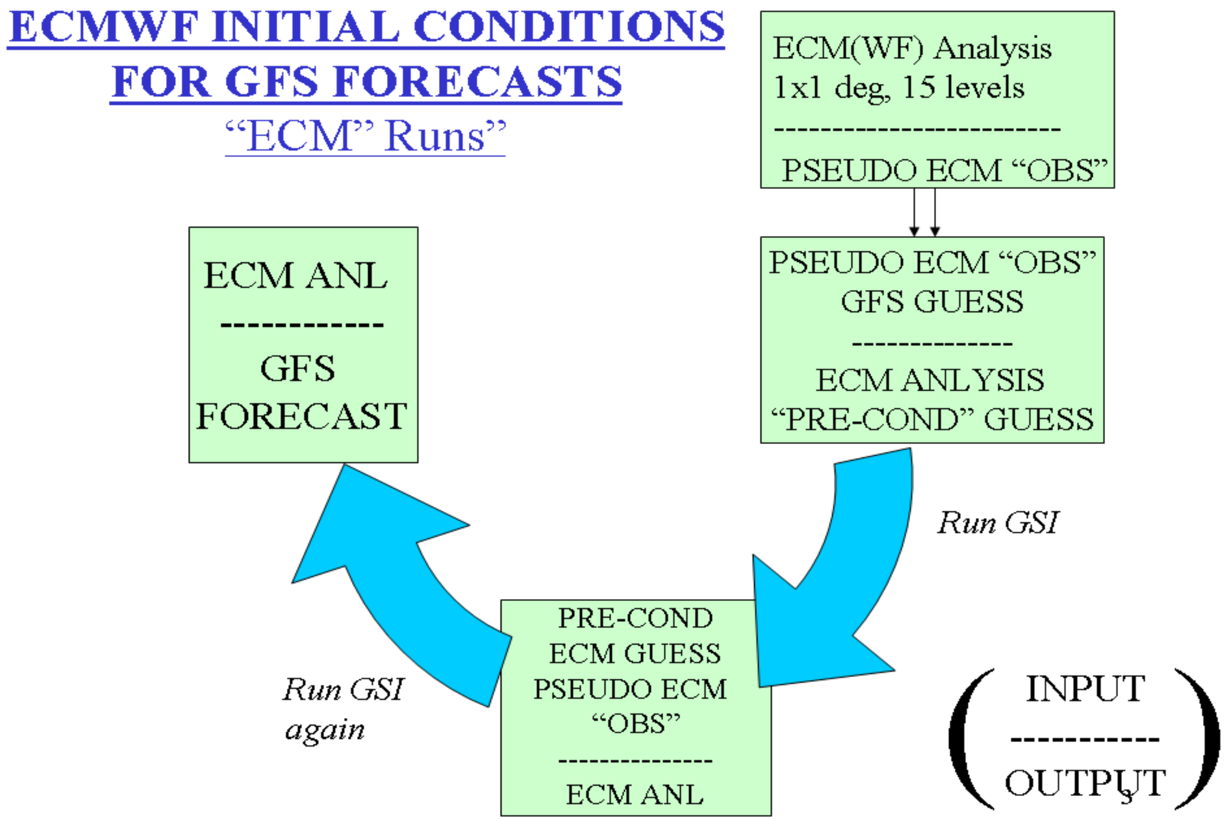


Figure 2. Schematic representation of an ECM run using the GSI/GFS system and ECMWF pressure grib analysis.

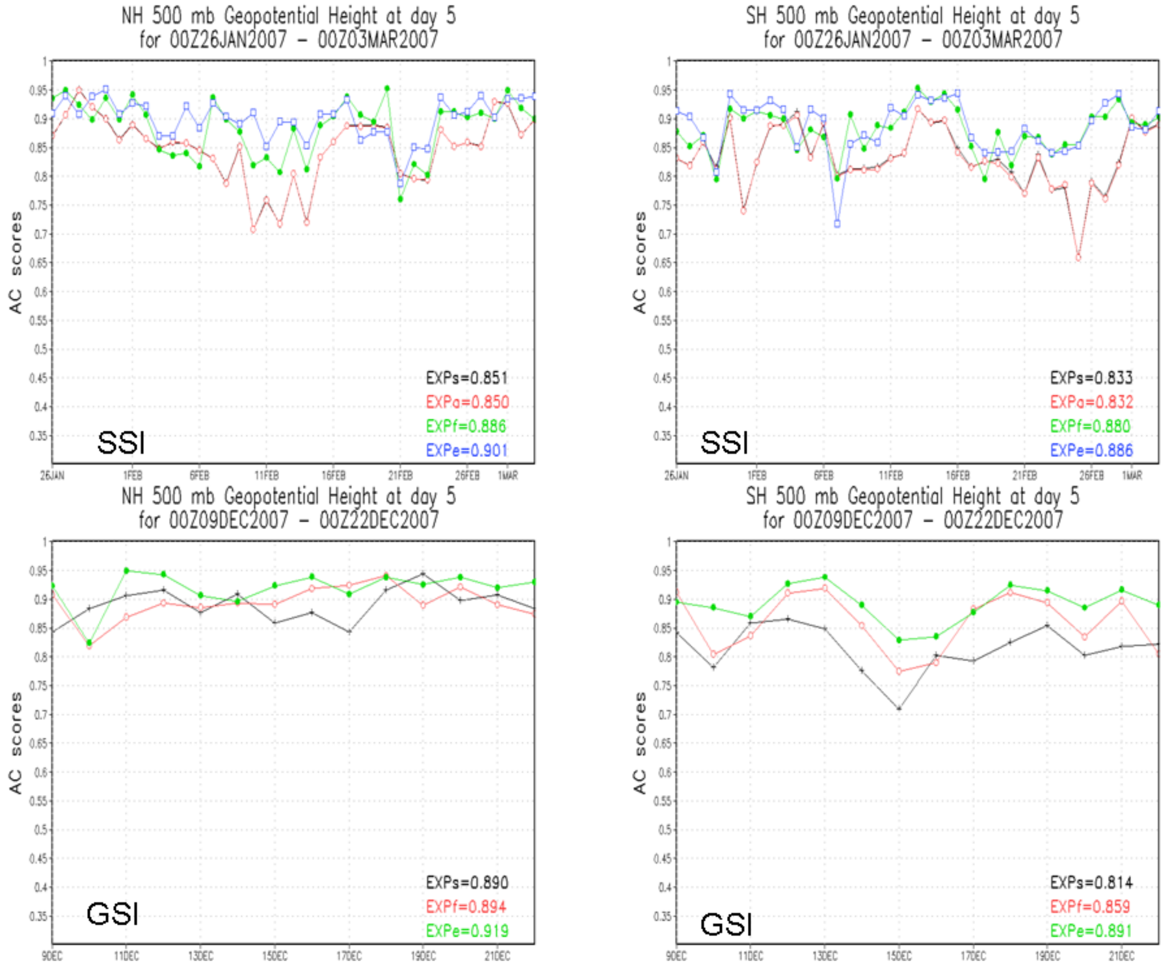


Figure 3. NH (left) and SH (right) 5-day anomaly correlation skill scores at 500 hPA for SSI (top where EXPs are the GFS operations, EXPa are the GFS run from the cycling analysis, EXPf are the ECM runs, and EXPe are ECMWF operations) from 26JAN2007 to 3MAR2007, and for GSI (lower where EXPs are the GFS operations, EXPf are the ECM runs, and EXPe are the ECMWF operations) from 9DEC2007 through 22DEC2007.

Lat/Lon Box	IC Date	GFS	ECMWF	ECM	OVRLY
20N→ 80N 150E→ 230E	2007102212	0.61	0.87	0.89	0.90

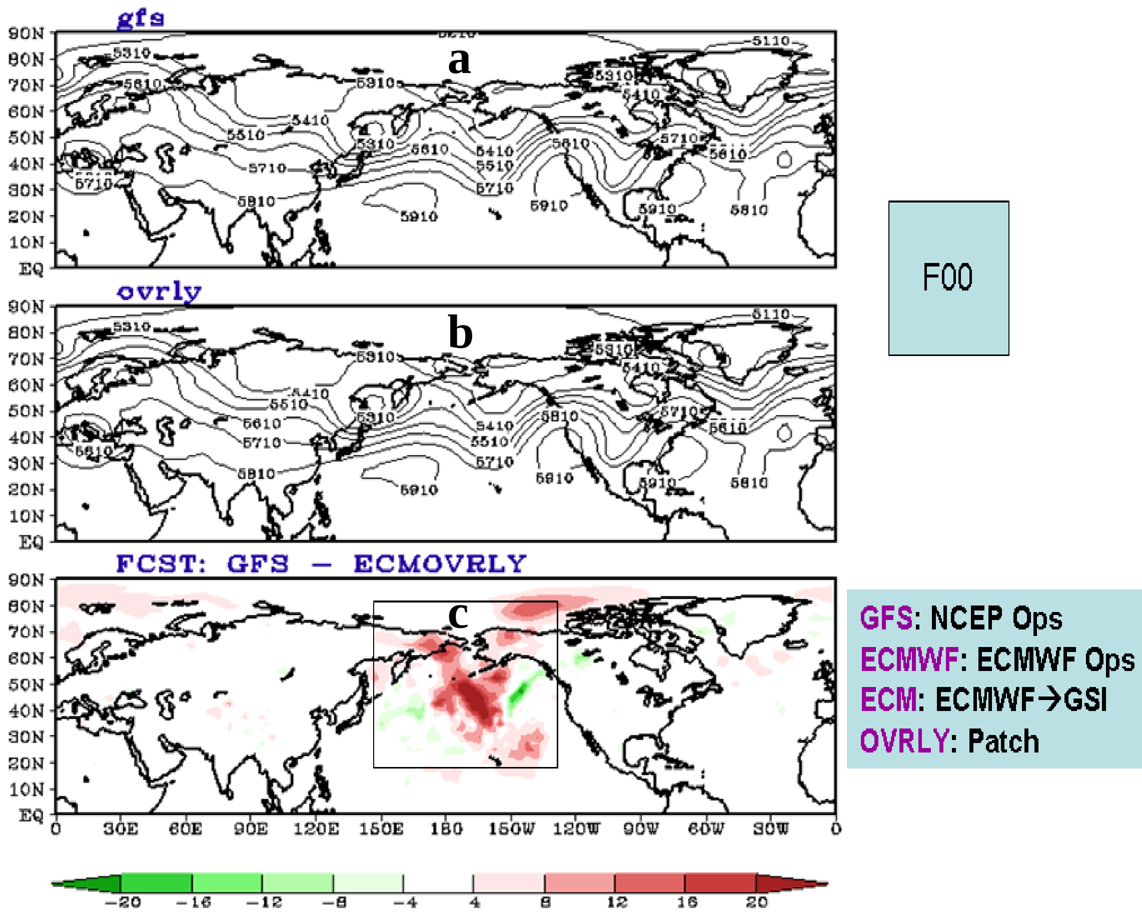
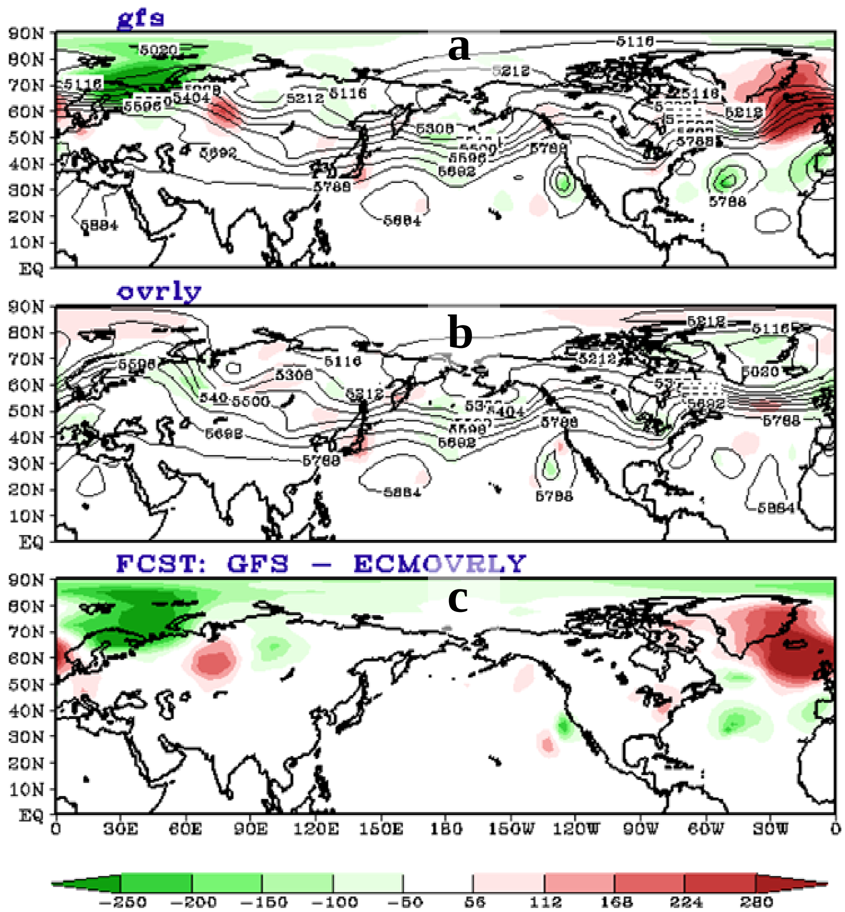


Figure 4. Central Pacific "overlay/patch" (OVRLY) run for a dropout case initialized on 2007102212. Patch is represented by box in (c). Top banner shows NH 5-day AC skill scores for the GFS, ECMWF, ECM, and OVRLY model runs. Graphical maps represent the 500 hPa heights at F00, the initial condition, for the GFS (a) and OVRLY or ECMOVRLY (b) runs. The forecast difference (GFS – OVRLY) map of the two F00 forecasts is shown in (c).

Lat/Lon Box	IC Date	GFS	ECMWF	ECM	OVRLY
20N→ 80N 150E→ 230E	2007102212	0.61	0.87	0.89	0.90



F120

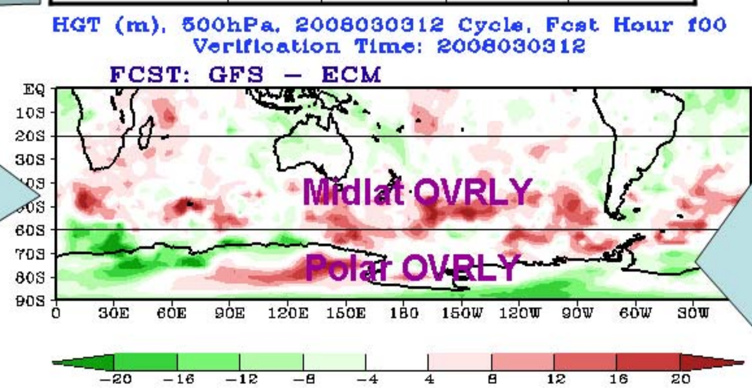
GFS: NCEP Ops
ECMWF: ECMWF Ops
ECM: ECMWF→GSI
OVRLY: Patch

Figure 5. Banner is same as Fig. 4. The NH 5-day F120 forecast of 500 hPa heights (contours) for the GFS (a) and OVRLY (b) runs. Color fill in (a & b) represent the forecast minus verifying analysis differences for both model runs. The F120 GFS minus OVRLY difference is color filled in (c).

Southern Hemisphere Ovrly
Polar vs. Midlat Expt
Case date: 12Z March 3, 2008

IC Date	GFS	ECMWF	ECM	OVRLY
2008030312	0.59	0.85	0.85	0.74

IC Date	GFS	ECMWF	ECM	OVRLY
2008030312	0.59	0.85	0.85	0.84



9

Figure 6. Mid-latitude and Polar OVRLY table and differences at initial condition (F00) time. The latitude bands indicate the two overlay/patch areas. Tables show the AC skill scores for the 20080303 12Z SH dropout experiments.

Lat/Lon Box	IC Date	GFS	ECMWF	ECM	OVRLY	Lat/Lon Box	IC Date	GFS	ECMWF	ECM	OVRLY
60S→ 90S 0→ 360E	2008030312	0.59	0.85	0.85	0.74	20S→ 60S 0→ 360E	2008030312	0.59	0.85	0.85	0.84

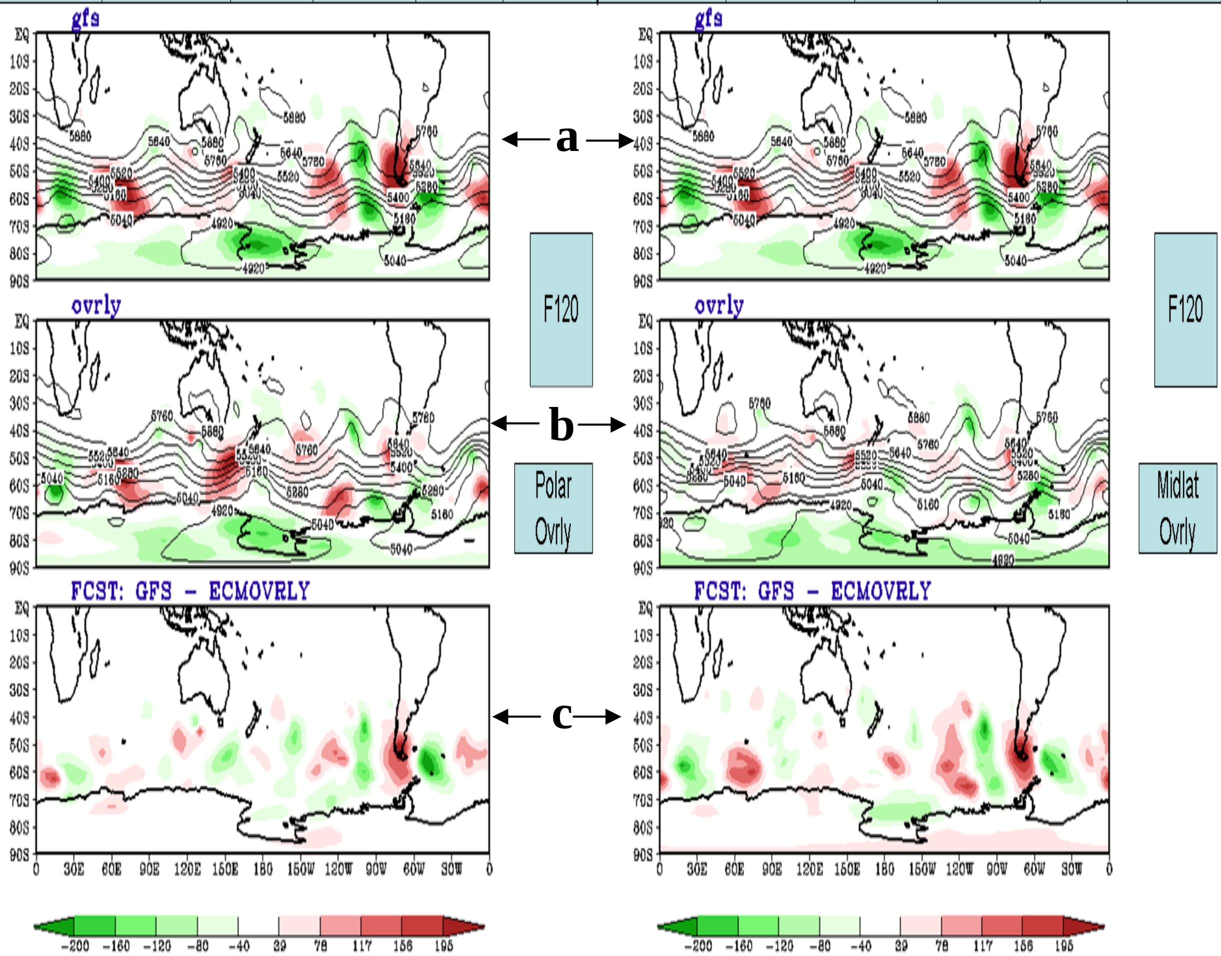


Figure 7. Mid-latitude (right) and Polar (left) OVRLY experiments of 5-day forecast (500 hPa) height contours and color fill as in Fig. 5 but for the 20080303 12Z dropout case: a) GFS operational 5-day forecast (note Left and Right maps are the same) b) overlay experiments, and c) GFS operations minus the respective ECMOVRLY run at F120.

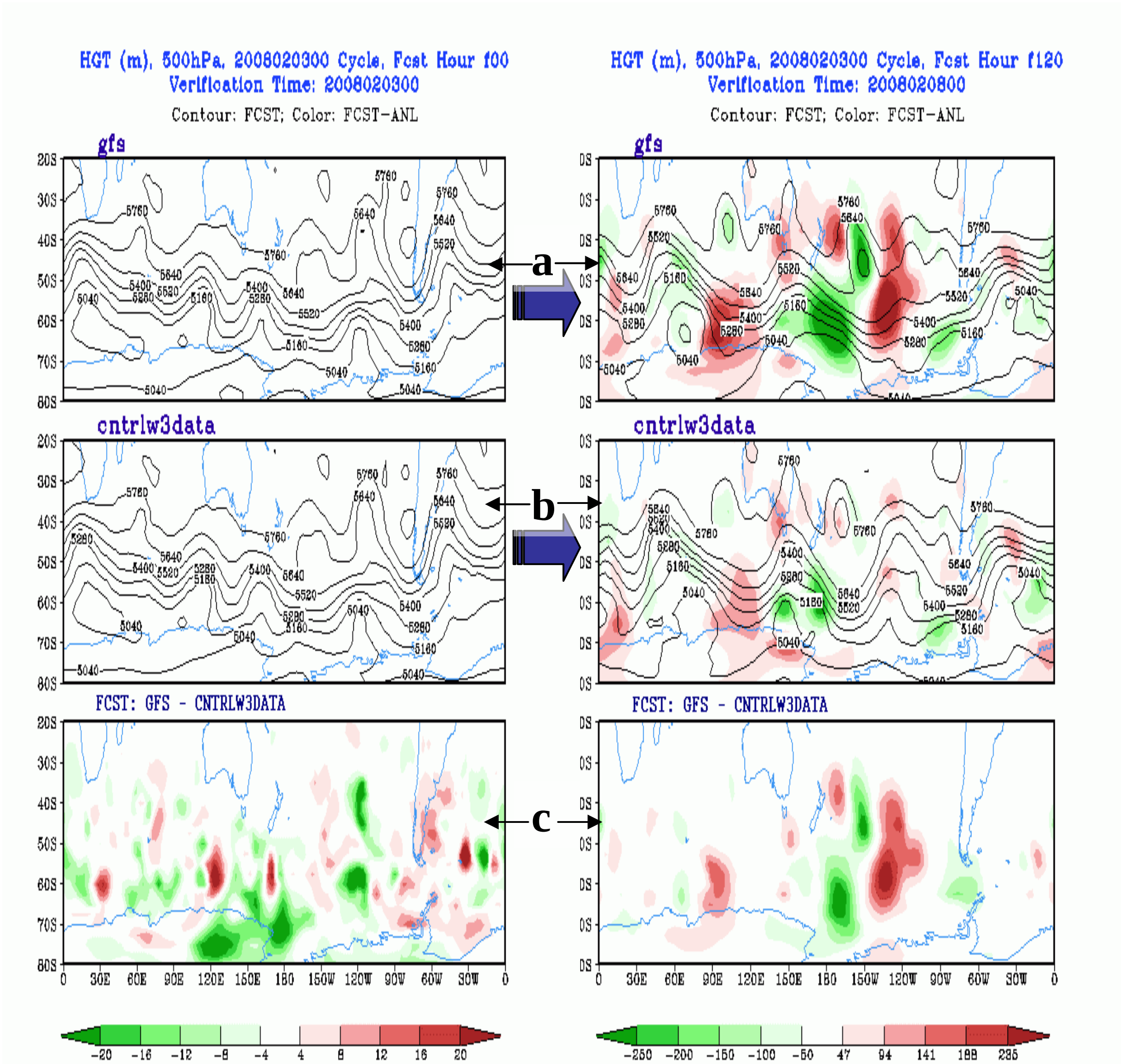


Figure 8. Southern hemisphere dropout case (20080203 00Z) of 500 hPa height contours at F00 (left) and F120 (right) for the GFS (a) and CNTRLW3DATA (b) experiments. CNTRLW3DATA uses GDAS conventional observations (PREPBUFR format) plus AMSUB, MHS, and GPSRO satellite radiance data to produce a GSI analysis. Forecasts errors are shown in (c) at F00 and F120 as color fill. The color fill in (a) and (b) is the same as Fig. 5.

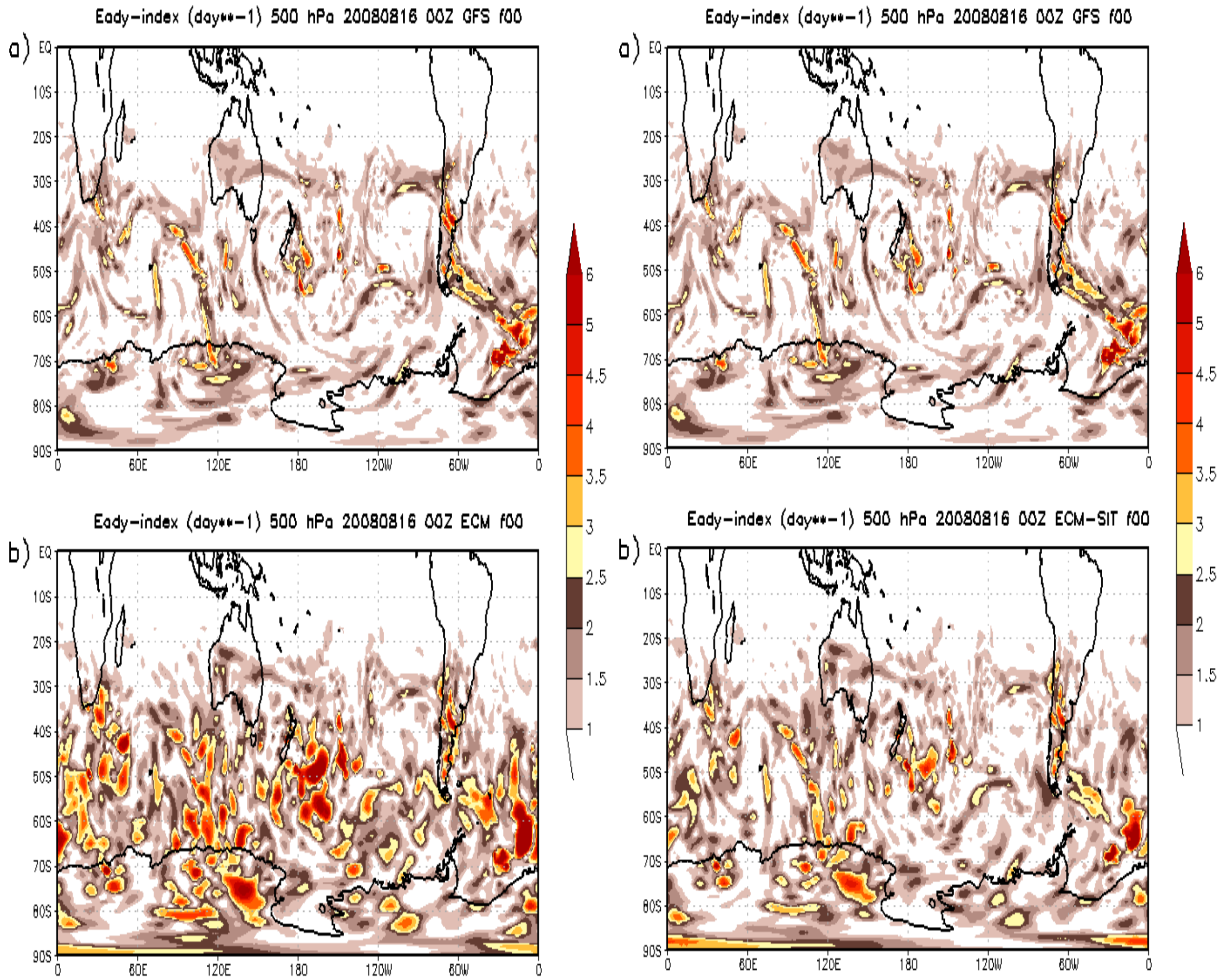


Figure 9. The total Eady Baroclinic Index (EBI) (day $^{-1}$) at 500 hPa for the F00 forecast from the August 16, 2008, 00 UTC (2008081600) initial conditions for the a) GFS model (top panels) and b) ECM model with two iterations of the GSI (bottom left panel) and ECM with a single application of GSI (bottom right panel).

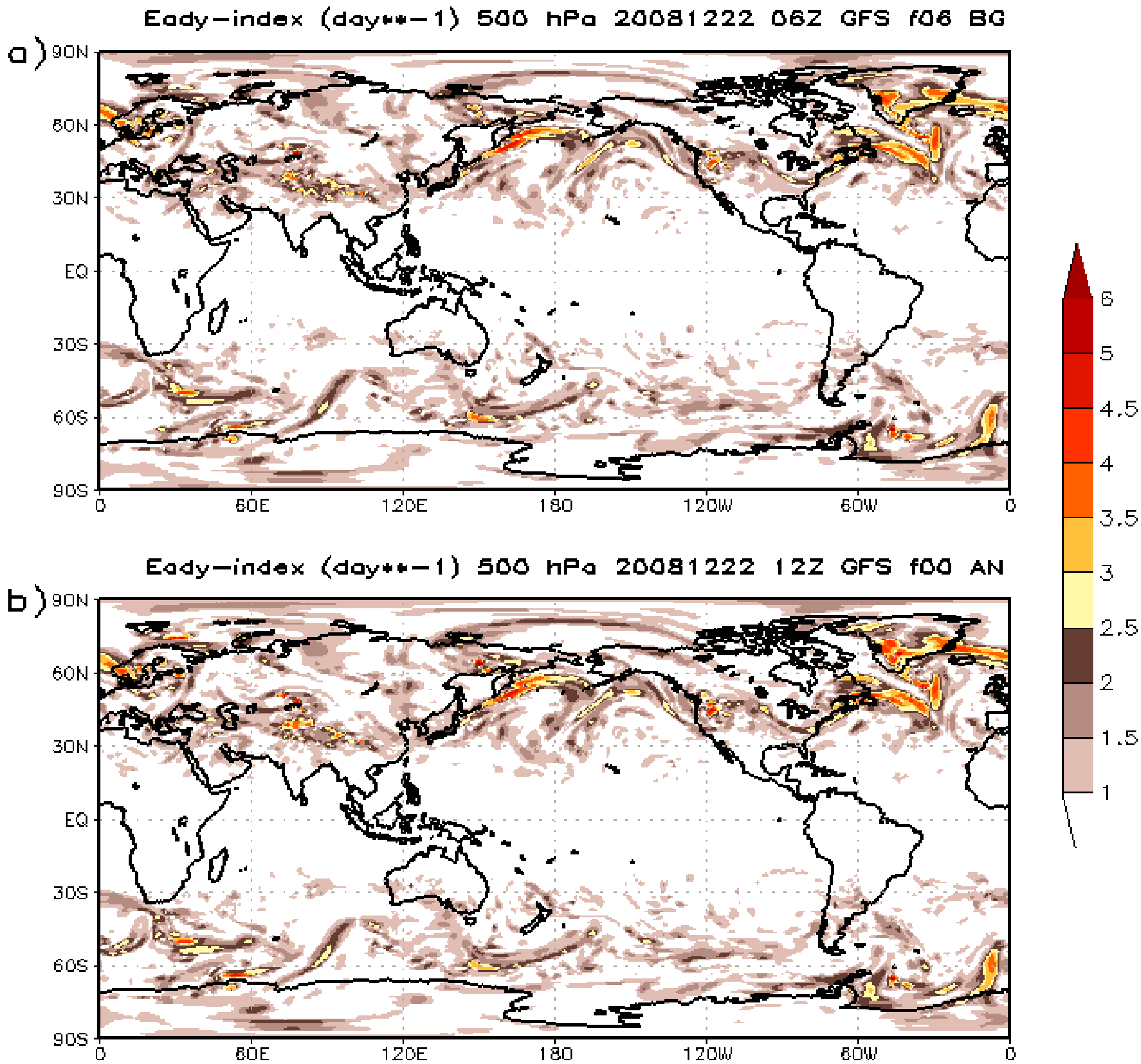


Figure 10. The total Eady Baroclinic Index (EBI) (day $^{-1}$) at 500 hPa for the a) 6-h forecast from December 22, 2008 06 UTC (BG) b) corresponding analysis valid at 12 UTC December 22, 2008.

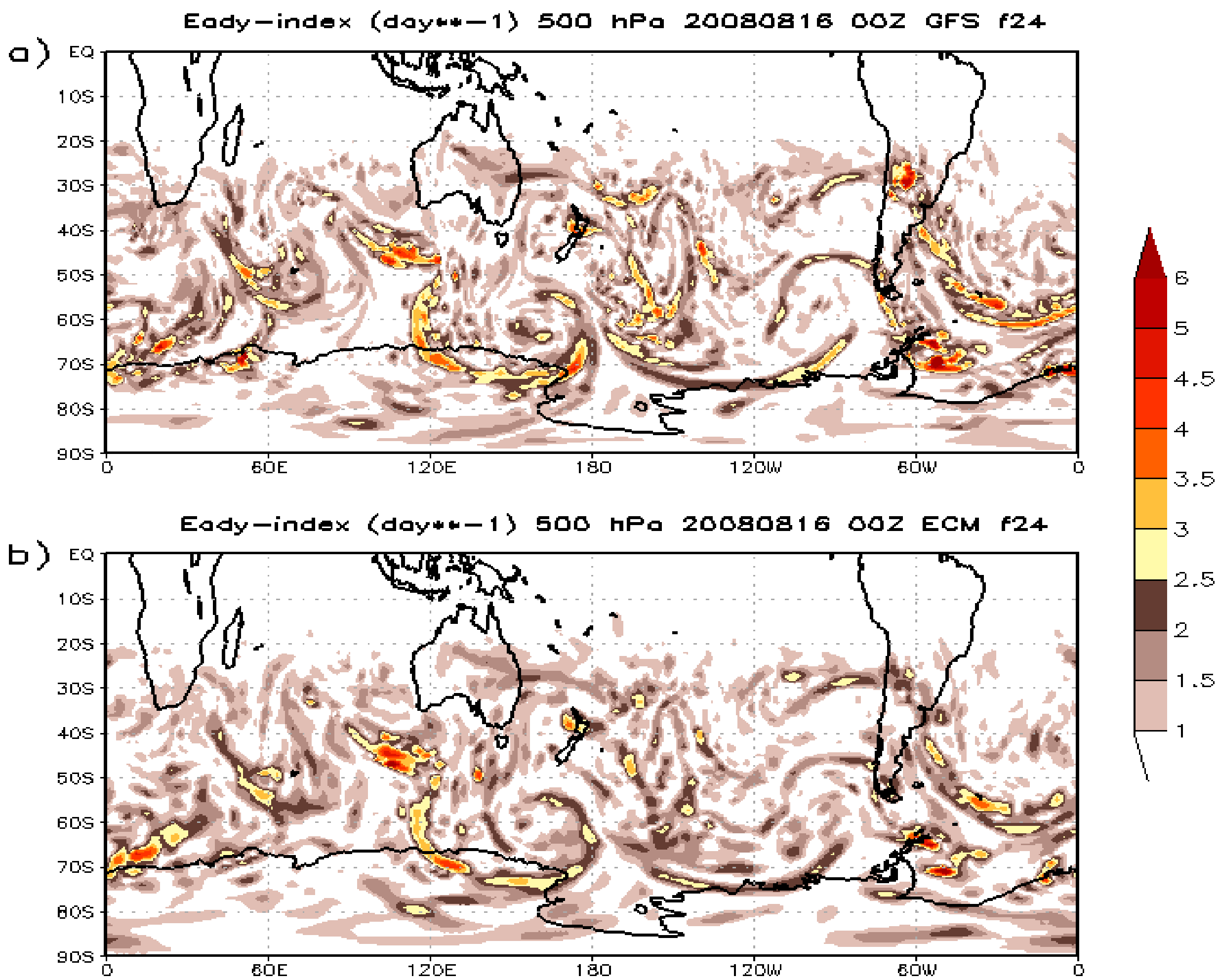


Figure 11. The total Eady Baroclinic Index (EBI) (day $^{-1}$) at 500 hPa for the 24-h forecast from the August 16, 2008, 00 UTC initial conditions for a) GFS (top panel) and b) ECM models (bottom panel).