6.B3

VERIFICATION OF EXTRATROPICAL CYCLONES WITHIN NCEP FORECAST MODELS USING AN AUTOMATED TRACKING ALGORITHM

Brian A. Colle* and Michael Charles

Stony Brook University-SUNY, Institute for Terrestrial and Planetary Atmospheres

1. INTRODUCTION

Most recent predictability studies of major extratropical cyclone events have focused on individual forecast bust events, such as the January 2000 “surprise” East Coast snowstorm (Zhang et al. 2002). Since the mid-1990’s, there has been little recent objective cyclone verification of current NCEP operational models, such as the North American Mesoscale model (NAM) or Global Forecast System (GFS) model.

The goal of this work is to verify a full spectrum of cyclones in operational models in order to assess how their predictability varies across North America and adjacent oceans, as well as to investigate how model skill changes for different large-scale flow regimes. A large dataset of cyclone errors in NWP models can be used to quantify the evolution of forecast errors across North America. One important question is whether cyclone forecast and analysis errors over the Pacific degrade the cyclone forecasts over the eastern U.S. a few days later?

While ensembles have been shown to outperform deterministic models for several meteorological variables, and have been shown to improve cyclone forecasts on a case study basis (Buizza and Chessa 2002), there has been limited long-term assessment of ensemble predictability for extratropical cyclone strength and position. Froude et al. (2007) provided a quantitative assessment of the cyclone predictions in the ECMWF and NCEP GFS ensemble, but this study focused on the extended forecasts and was limited to 6 months of data. Our study compares the NCEP NAM and GFS for 5 cool seasons, and completes a 3-year verification of NCEP’s short-range ensemble forecast (SREF) system.

2. DATASETS AND METHODS

Cyclones were tracked for the 5 cool seasons (October through March) from 2002 to 2007. The GFS 120-h forecasts were available at the National Climatic Data Center (NCDC) at ~95 km resolution (213 grid) covering all of North America four times daily. During the first three cool seasons, the NAM was available at NCDC at about 40 km resolution (212 grid), covering the region surrounding the continental United States (CONUS) (Fig. 1), with 6-hourly forecasts out to 60 hours at 00 & 12 UTC, and out to 48 hours at 06 & 18 UTC. For the last 2 cool seasons (2005-2007) the NAM was available at NCDC at 12-km grid spacing 4 times daily out to hour 60 hour. All data was bi-linearly interpolated to 0.8 deg latitude longitude grid, which is close to the resolution of the available GFS grids.

The SREF data was available from NCEP from 2004-2007 at ~40-km grid spacing (212 grid) (Fig. 1). There were twice daily runs at 0900 & 2100 UTC out to 63 hours. The SREF consisted of 15 members for 2004-2006: 5 ETA members using Bett-Miller-Janic convection (EBM), 5 ETA members using Kain-Fritsch convection (EKF), and 5 RSM members using the Regional Spectral Model. For the 2006-2007 cool season, six Weather Research and Forecasting (WRF) members were also available, three using the Nonhydrostatic Mesoscale Model (NMM) core and the others using the Advanced Research WRF (ARW) core.

Originally, we hoped to use the North American Regional Reanalysis (NARR) for the observed cyclone locations and magnitudes. However, when comparing the NARR and NAM analyses with surface observations, both underpredicted the magnitude of cyclones by 2-4 mb on average (not shown), especially for deep cyclones over the Pacific. The GFS analysis had little bias and was more skillful than NARR and the NAM analyses, so the GFS was used to verify the cyclone central pressures. For the cyclone displacement verification, the average position of the GFS and NAM was used as the observed position.

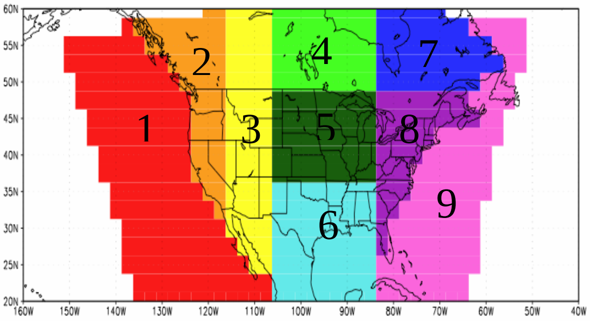

Figure 1. Regions 1-9 used to compare the results from the NAM, GFS, and SREF modeling systems.

**Corresponding author address:* Dr. Brian A. Colle, Marine Sciences Research Center, Stony Brook University/SUNY, Stony Brook, NY 11794-5000. email: brian.colle@stonybrook.edu

The cyclone tracking routine was developed at NCEP (Marchok 2002). A cyclone was identified and tracked by first locating the lowest sea-level pressure (SLP) value in a latitude-longitude grid. Next, a SLP gradient of 0.0015 mb km^{-1} was required in each direction outward at least 300 km from the SLP minimum. If this gradient was found, then the SLP minimum was checked to see if there was a 2 mb closed contour. If either of these tests failed, then the grid point was masked out and the grid point with the next lowest SLP was tested. To prevent any points near the cyclone from being identified in subsequent iterations, all grid points surrounding the cyclone, to the point in which the SLP gradient reversed, were masked out. All steps were repeated until either all grid points were masked out, or all remaining SLP values were greater than 0.5 standard deviations from the domain-averaged SLP. Once all forecast cyclones were located, they were paired with the closest observed cyclone within 800 km. When compared with NCEP analyses over a two-month period, it was found that this scheme was 95% accurate in identifying and pairing cyclone events.

A bootstrap method was used to test significance. In this method a data sample (timeseries of SLP errors in a given region and forecast hour) was resampled 1000 times. Resampling was done by randomly replacing one data value at a time from the sample until a pseudo-sample of the same size as the original sample is generated. The mean is calculated and saved, and the process is repeated until 1000 pseudo-sample means are calculated. From this, the 5% and 95% confidence intervals around the mean are obtained.

3. NCEP GFS AND NAM VERIFICATION

a. Short-term NAM and GFS forecasts

Figures 2 a,b show the NAM and GFS mean errors of cyclone central pressure at various forecast lead times during the 2002-2007 cool seasons. For the NAM (Fig. 2a), there is a 1-0-2.5 mb positive bias (underdeepened cyclones) over the Pacific Ocean, which is significant from zero at the 99% level. In contrast, there is a ~1 mb negative bias (overdeepened cyclones) over the eastern U.S. (regions 7-8). The GFS tended to develop a statistically significant negative bias of ~1 mb for regions 2-3 (western U.S.) (Fig. 2b). Many of the overdeepened cyclones in the GFS were found just inland of the coast of the Pacific Northwest and southwest British Columbia (not shown), which suggests that the GFS did not produce enough cyclolysis as cyclones interacted with the relatively smooth terrain in the model.

For both the NAM and GFS, the SLP mean absolute errors are larger in the Pacific (region 1) than the other regions, and are significant at the 99% level (Figs. 2c,d). The other regions are not statistically significantly different from each other. The NAM errors are about 0.5-1.0 mb larger than the GFS for most regions, which is significant at the 95% level.

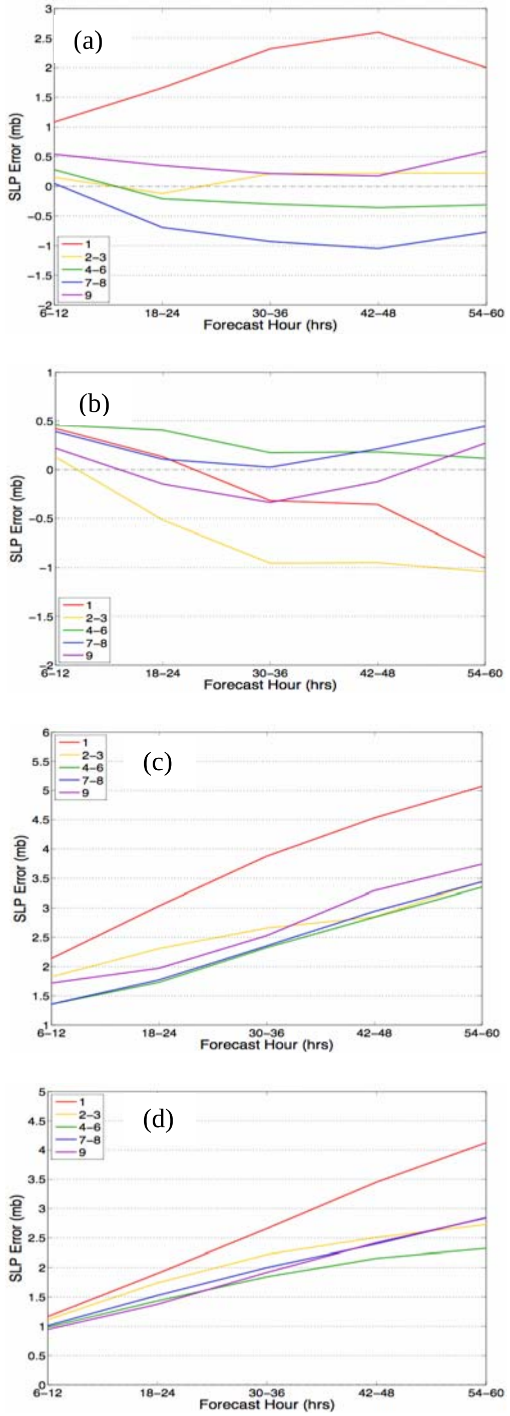

Fig. 2. Mean SLP error (in mb) versus forecast hour during 2002-2007 cool seasons for the (a) NAM and (b)GFS given the regions specified in Fig. 1. (c) and (d) Same as (a) and (b) except for the SLP mean absolute error.

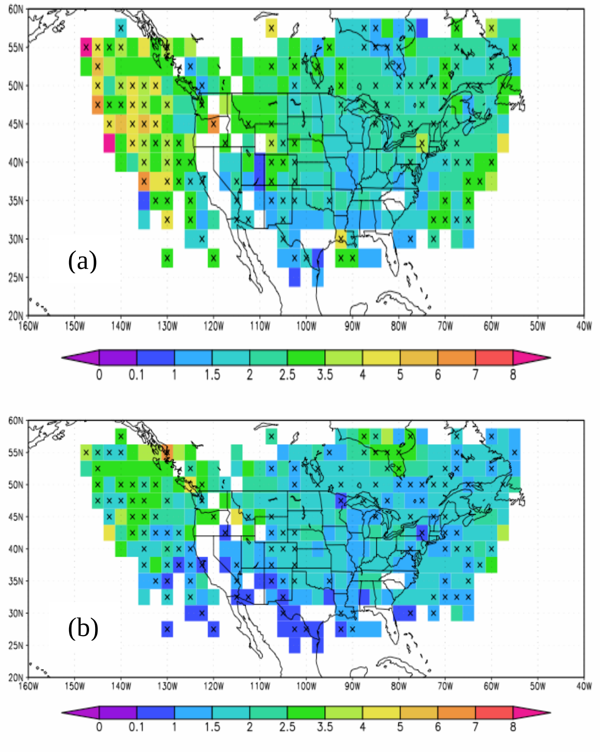

Figure 3. SLP mean absolute errors (shaded in mb on a 2.5 deg grid) for the (a) NAM and (b) GFS for the 2002-2007 cool seasons.

To better highlight those areas with the largest errors, Figure 3 shows the spatial errors across the regions on a 2.5 deg grid for the 18-36 h NAM and GFS forecasts. The largest errors in the NAM are located over the Pacific Ocean, northern Rockies, eastern Canada, the Northeast U.S., and the western Atlantic, with the Pacific errors clearly larger than the other regions. Those points where the GFS errors are significantly different than the NAM at the 95% level are shown with an X in Fig. 3. The GFS tends to be more skillful on average over much of the Pacific, parts of the intermountain West, Northeast U.S. and over the Gulf Stream. The NAM is only more skillful that than the GFS around the Hudson Bay region.

b. Mid-range GFS forecasts

The GFS was also verified out to day 5 (120 h). Figure 4 shows the errors for the relatively deep cyclones (> 1.5 standard deviations deeper than mean for each region). Region 1 (eastern Pacific) has larger errors than the other regions for the 30-60 h forecast. However, the region 9 (western Atlantic) errors become significantly larger than the Pacific (at the 90% level) at > 96 h. A large positive bias (cyclone underprediction) occurs in regions 8 and 9 in the day 3-5 forecast, which results in rapidly growing errors for the eastern U.S. and western Atlantic. By day 5, the absolute errors over the central U.S. and Canada (regions 4-6) are nearly half as large as the oceanic regions and eastern U.S.

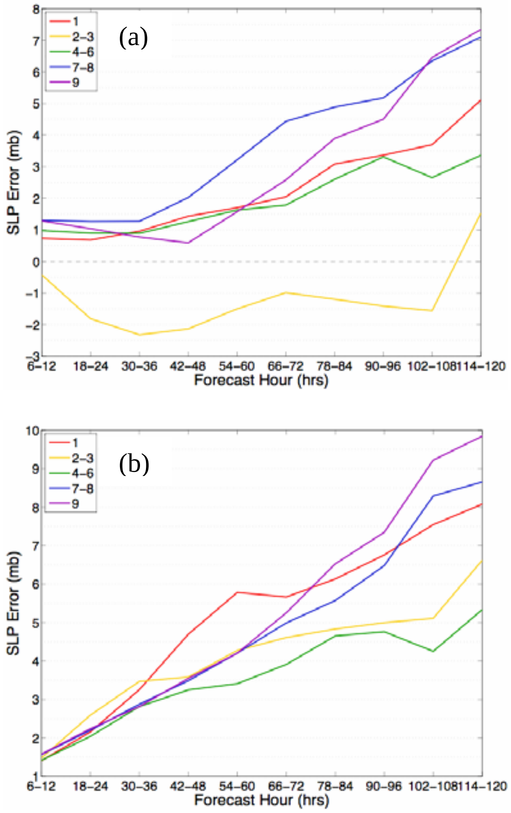

Figure 4. (a) Mean and (b) mean absolute SLP errors versus forecast hour for GFS cyclones deeper than 1.5 standard deviations from the mean cyclone intensity for each region individually for the 2002-2007 cool seasons.

To determine whether there was any relationship between the GFS cyclone errors over the eastern Pacific and the extended GFS forecasts over the eastern Atlantic, a time series was constructed in which the cyclone errors in the GFS over region 1 (West) between hours 6 and 18 were related to the errors later in the same forecast in regions 8 and 9 (East) (Fig. 5). The errors in the West were separated into large positive and negative SLP errors greater than 2 standard deviations from the mean as well as the small error events (less than 0.5 standard deviation from mean) and a random subset of ~50 events. For the positive error events over the West (Fig. 5a), the mean errors over the East by hours 78-96 were more positive by 3-4 mb, and the East mean absolute errors were comparable to the West by hour 66. The cyclone errors subsequently became more negative in the East and the West by hours 96 and 114, respectively. For the negative cyclone errors in the West (Fig. 5b), the errors in the East were again positive by 3-5 mb by hour 90, while the West errors reached another negative error peak at hour 84. The small errors cases over the West and the random events did not produce a positive bias in the East cyclones until after hour 102 (Figs 5c,d). Based on the different East responses for the different West error types, it seems there are relationships between the cyclone errors in the West and those in the East, which are worthy of further investigation.

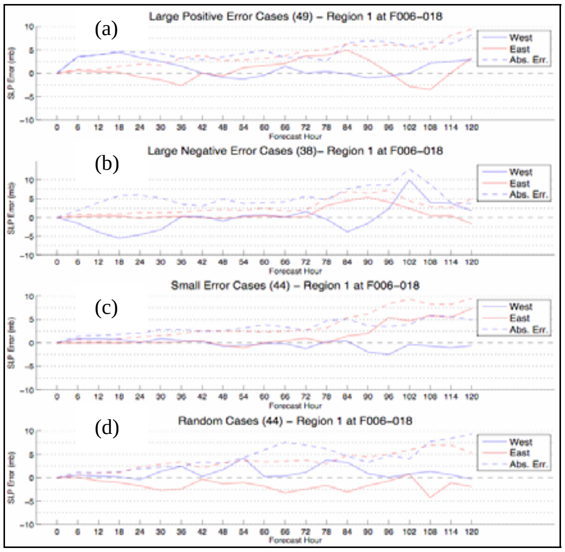

Figure 5. GFS cyclone errors (in mb) versus forecast hour for region 1 (West) and regions 8 and 9 (East) for those West cyclone errors between hour 6-18 that are (a) positive (2 std dev greater than mean), (b) negative (by 2 std dev), (c) small (within 0.5 std dev of mean), and a (d) a random set of ~50 error events.

c. Composite GFS error tracks

It was hypothesized that the various GFS errors, especially those well-defined large error events, may have favored tracks or large-scale flow patterns. Figure 6 shows the tracks for the GFS 48-h forecasts for the 40 events with the largest positive errors in regions 8 and 9 versus the 40 events with the largest negative error in those same regions. Many of the positive error cases at hour 48 (X locations) tend to cluster near the Gulf Stream east of the mid-Atlantic coast, where there is a cluster of tracks. Most of the absolute error growth (colored lines) leading to the positive errors occurs over the western Atlantic. In contrast, the negative error (X) locations at hour 48 tend to occur over a larger geographic region over the Eastern U.S., and a greater number of tracks are further north and west than the positive cases. A larger fraction of negative events occur over land, while the majority of positive cases are confined to the Atlantic region.

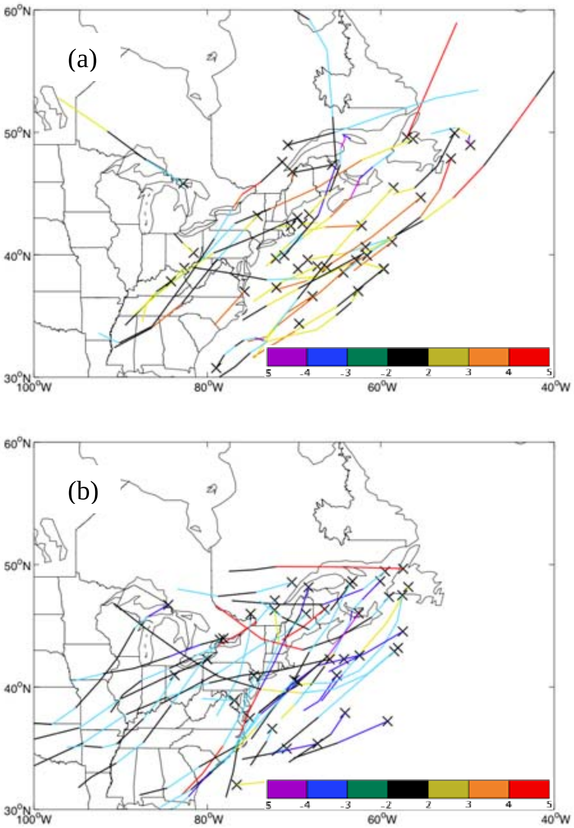

Figure 6. Cyclone tracks for regions 8 and 9 associated with the GFS cyclones at hour 48 (x locations) that have the 40 largest (a) positive and (b) negative pressure errors. The model pressure-error tendency is given by the colored lines in mb/6h.

4. NCEP SREF VERIFICATION

The NCEP SREF at 0900 and 2100 UTC was verified for the 2004-2007 cool seasons, and the errors were compared with the NCEP 1200 and 0000 UTC GFS and NAM (Figs. 7 and 8). Figure 7 shows the displacement errors for regions 1 (eastern Pacific) and 9 (western Atlantic). For the Pacific region (Fig 7a), the mean of each of the SREF components (RSM, EKF or EBM) are similar in skill to the operational NAM (Eta). The full SREF mean has smaller displacement errors than the operational NAM (significant at 99% level), while the operational GFS is more skillful than SREF before hour 48. The RSM is the best sub-ensemble for SREF, and it is comparable in skill to the full SREF. The SREF displacement errors are 25-50 km smaller over the western Atlantic than the Pacific between hours 18 and 48 (Fig. 7b). Over the western Atlantic, the full SREF is more skillful than the NAM after 21 h and is less skillful than the operational GFS.

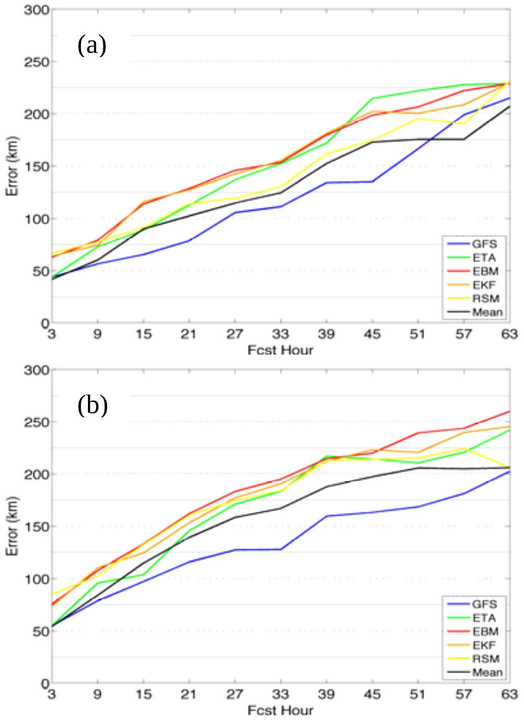

Figure 7. Cyclone displacement errors (in km) versus forecast hour for mean of the full SREF (black) and its components (EKF, EBM, and RSM) as well as the operational NAM (ETA) and GFS for (a) region 1 and (b) region 9.

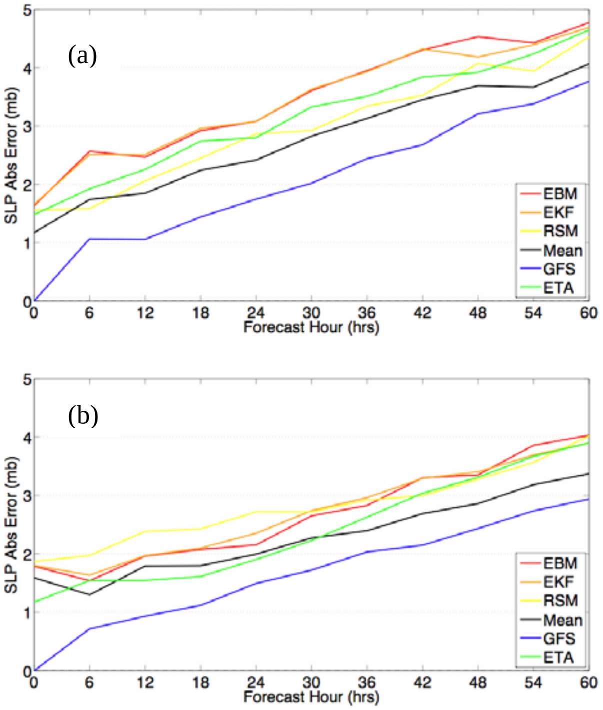

Figure 8. Same as Fig. 7 except for mean absolute error for the cyclone central pressure (in mb).

The mean absolute errors for cyclone central pressure for the full SREF are also slightly less than the operational NAM for regions 1 and 9 (Fig. 8), but the SREF was not as skillful as the operational GFS. The SREF errors were 0.5-1.0 mb larger than GFS in the eastern Pacific than eastern Atlantic. The WRF-NMM was included in the SREF for the 2006-2007 cool season, but this did not seem to improve the SREF results (not shown).

The rank histograms for the SREF were compared for cases in which the cyclone-matching algorithm identified and paired at least 12 SREF members with an observed cyclone. Even at hour 3 (Fig. 9), the ensemble is overdispersed over the eastern Pacific (region 1) and is biased towards weaker cyclones than observed for regions 8 and 9, as well as regions 4 and 5 (not shown). This suggests that even the SREF initialization did not properly cover the observed cyclone phase space. By hour 36 (not shown), these biases persist, with a greater tendency for overdispersion over the western Atlantic.

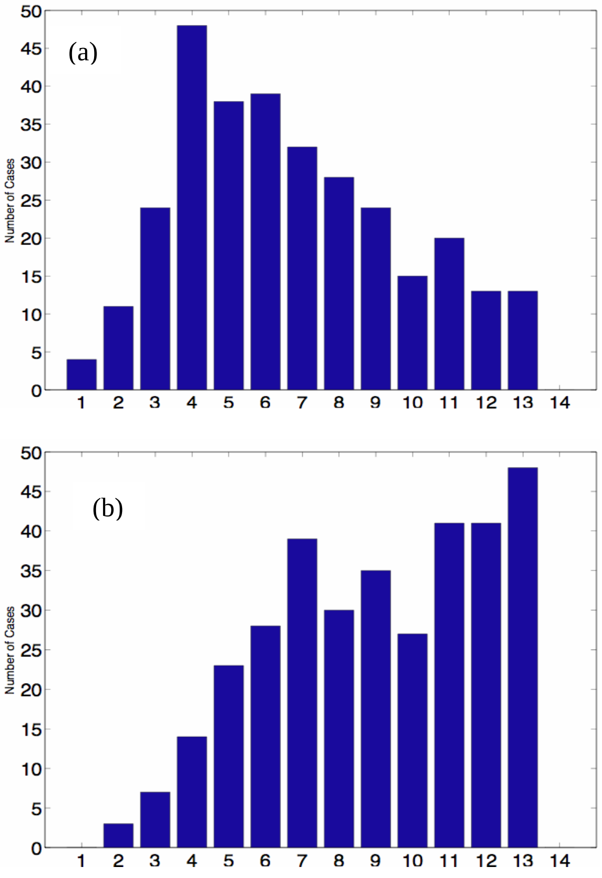

Figure 9. Rank histogram for the SREF at hour 3 for the (a) region 1 (eastern Pacific) and (b) regions 8 and 9 (eastern U.S. and western Atlantic). Only those cases with at least 12 SREF members having a cyclone match with the observed were utilized.

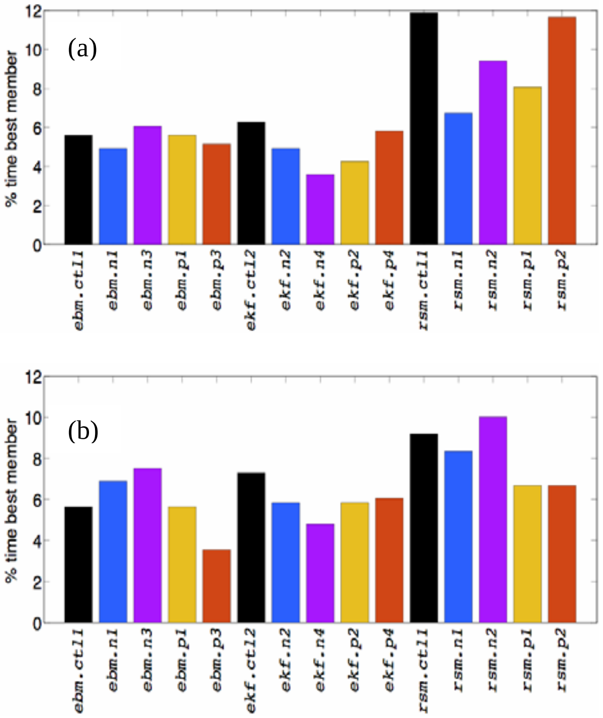

Figure 10. The percentage of time that each member of SREF is the best for central pressure for (a) region 1 and (b) region 8 (eastern U.S.) at hour 63.

The other characteristic with the SREF predications is that all the ensemble members are not equally skillful in forecasting the cyclone central pressure. Figure 10 shows the percentage time that each member is the best in forecasting the central pressure at hour 63. For both the eastern Pacific and eastern U.S., many of the RSM members are more skillful than the other SREF members on average.

5. CONCLUSIONS

This study is the first comprehensive verification of the NCEP extratropical cyclones around North America and adjacent oceans in several years, and the first to quantitatively evaluate the NCEP SREF cyclones. The results suggest that the NCEP GFS is more skillful than the NAM over many regions, especially over the eastern Pacific, where the NAM has a large positive error bias. The cyclone errors are larger over the Pacific than other regions for short (0-60 h) forecasts. However, the errors for relatively deep cyclones over the eastern U.S. are larger than the eastern Pacific by hour 72, since these cyclones are underpredicted in the GFS at the extended range over the eastern U.S. and western Atlantic. This suggests that there is still a need to improve > day 3 forecasts for major cyclones over the eastern U.S. We also show that the larger cyclone errors over the eastern Pacific associated with the largest errors (positive biases) over the eastern U.S. by hour 72-96.

The NCEP SREF was also verified for the 2004-2007 cool seasons. The SREF is more skillful than the operational NAM in many regions, but not the operational GFS. The SREF suffers from overdispersion and positive biases in many regions early in the forecast, which is hurting its performance. Also, not all members are equally skillful. Future work will identify those synoptic patterns associated with the best and worst SREF forecasts.

This research also suggests that there are synoptic flow regimes favoring certain cyclone biases. For example, cyclone underprediction in the day-2 GFS forecasts is associated more with cyclones moving east-northeastward over the Gulf Stream. These favored tracks for certain biases will also be investigated more closely in the future.

6. ACKNOWLEDGMENTS

We thank Tim Marchok for providing and helping to implement his cyclone tracking algorithm. We appreciate the assistance from Mr. Jeff McQueen and Dr. Jun Du in obtaining the SREF data, as well as support from UCAR-COMET on ensemble predictability (Grant S0238662).

7. REFERENCES

Buizza, R., and P. Chessa, 2002: Prediction of the U.S. storm of 24–26 Januar y 2000 with the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **130**, 1531–1551.

Froude L., L. Bengtsson, and K. Hodges, 2007: The Prediction of Extratropical Storm Tracks by the ECMWF and NCEP Ensemble Prediction Systems. *Mon. Wea. Rev.*, accepted.

Marchok, T.P., 2002: How the NCEP tropical cyclone tracker works. *Preprints, 25th Conference on Hurricanes and Tropical Meteorology, Amer. Meteor. Soc.*, 21-22.

Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the ‘‘surprise’’ snowstorm of 24 –25 Januar y 2000. *Mon. Wea. Rev.*, **130**, 1617–1632.