3A.1 Verification of 12 years of NOAA seasonal hurricane forecasts

Eric S. Blake and Richard J. Pasch
NOAA/NWS/NCEP/National Hurricane Center, Miami, FL

Gerald D. Bell
NOAA/NWS/NCEP/Climate Prediction Center, Camp Springs, MD

1. INTRODUCTION

Seasonal hurricane forecasts with varying lead times have been produced in the Atlantic basin since 1984 (Gray 1984). Partly as a result of the success of those early forecasts, many different research and operational groups have made seasonal hurricane forecasts and have expanded their use to other tropical cyclone basins. After the large forecast failure of the Gray forecast for the 1997 season, a movement began within the NOAA Climate Prediction Center (CPC) to issue seasonal hurricane predictions. The first seasonal hurricane outlook by NOAA was issued in August 1998, and NOAA has released seasonal outlooks in May and August for the Atlantic basin since that time. These NOAA/CPC outlooks are a collaborative effort with the National Hurricane Center and the Hurricane Research Division of the Atlantic Oceanographic and Meteorological Laboratory.

During the past twelve years, NOAA's seasonal hurricane outlooks have sometimes been accurate (e.g. the 2008 and 2009 forecasts) but have also been criticized as not being skillful, especially after unsuccessful forecasts (such as 2006). Although a verification of Gray's seasonal hurricane forecasts showed there was skill above climatology (Owens and Landsea 2003), no such work has been done to compare the NOAA outlooks to any benchmarks of skill or to those of any another forecast group.

2. METHODOLOGY AND DATA

In this study, NOAA forecast skill is compared to seasonal forecasts issued by the Colorado State University (CSU) Gray project group and the Tropical Storm Risk (TSR) group in England. NOAA's outlooks have always conveyed likely ranges of expected activity rather than an exact number, which is fairly typical of other forecast groups. To facilitate comparisons with the single-number forecasts, the mid-point of the NOAA range was used for verification purposes.

The skill benchmark used is a seasonal forecast comprised of the most recent 5-yr running mean of a given forecast parameter. This benchmark is the same one recommended by a World Meteorological Organization (WMO) panel on seasonal forecast accuracy standards (WMO 2008). The skill measures employed are the Mean Absolute Error (MAE) and Mean Squared Error (MSE).

These measures are used to compare the mid-May NOAA forecasts against the 1 June outlooks issued by CSU, TSR, and to the preceding 5-yr mean. In addition, the early August forecasts issued by NOAA were compared to the respective early August forecasts issued by CSU, TSR, and the 5-yr mean. Note that only activity that occurred after 1 August was evaluated in the August forecasts, and any activity that occurred before 1 August was deducted from the total seasonal activity forecasts.

The parameters examined are the numbers of tropical storms (including subtropical), hurricanes, and major hurricanes and ACE (Accumulated Cyclone Energy, the sum of the squares of the maximum wind speeds every six hours for (sub)tropical storms and hurricanes, [Bell et al 2000]). For the ACE forecast comparisons, the Net Tropical Cyclone (NTC) index used by CSU is substituted for the ACE index since CSU has only recently begun making ACE forecasts. This substitution is reasonable given that both indices measure the combined intensity and duration of tropical storms during the season and hurricanes and are correlated with each other at r = 0.98.

One other aspect of the verification is to compare the ranges of the NOAA outlooks to a WMO-suggested range. The WMO recommended range is the average 30-yr standard deviation of the 5-yr means for each hurricane activity parameter (WMO 2008). This metric is used to compare NOAA's and TSR's forecast spreads only (CSU only recently began issuing a range).

3. RESULTS

Figure 1 shows a graph of the MAE statistics of the 5-yr climatology and the May and August NOAA forecasts. All of the NOAA forecasts showed a lower mean absolute error than the 5-yr mean, though the improvements were

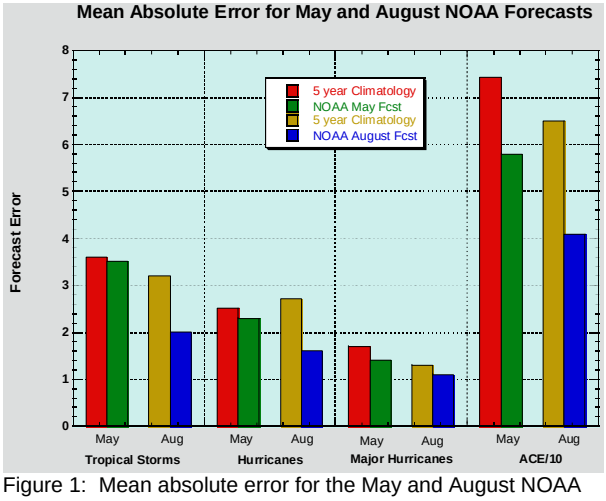

Figure 1: Mean absolute error for the May and August NOAA seasonal outlooks (1999-2009 for May, 1998-2009 for Aug).

Corresponding author address: Eric S. Blake, NHC, 11691 SW 17th Street, Miami, FL 33165; e-mail: Eric.S.Blake@noaa.gov.

small in May. Only the May forecasts of ACE had a statistically significant improvement (Student's t-test, >95% confidence) from the 5-yr mean ACE forecasts. The August forecasts, however, were quite skillful with the largest improvements noted in the predictions of ACE and number of hurricanes (which were both statistically significant at a 99% confidence level). Only the outlook for the number of major hurricanes was not statistically significant over the no-skill forecast. On average, the August outlooks reduce the error of tropical storm and hurricane counts by about one system per season, and improve the ACE forecast by about 25% of the median.

Another way to measure skill is to compare the MSE of the NOAA forecasts and a forecast based on the previous 5-yr mean. Figure 2 displays the improvements relative to the 5-yr mean from the May and August NOAA forecasts. Note that a skill of zero would represent no difference in the MSE of the 5-yr mean from the outlooks. In the NOAA May forecasts (blue bars), the largest skill values are seen in the hurricane and ACE predictions. NOAA's August outlooks, however, vastly outperformed the 5-yr mean forecast for all parameters, with improvements ranging from 32% to 62%. Similar to the MAE results, the May ACE (90% confidence) outlooks and the August tropical storms, hurricanes and ACE (95% confidence or greater) outlooks showed statistical significance.

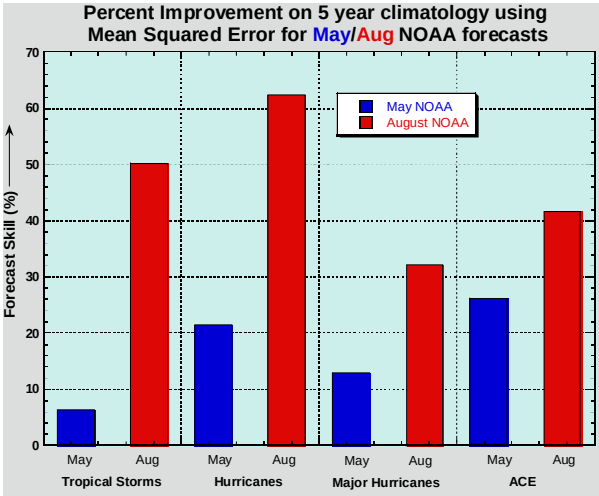

Figure 2: Forecast skill of the May and August NOAA seasonal hurricane forecast for all forecast parameters.

The next part of the verification was to compare NOAA forecasts to other well-known groups making seasonal hurricane predictions. Figure 3 shows the percent improvement over climatology for NOAA, CSU and TSR. The skill of the NOAA and CSU forecasts are comparable in May, with the TSR forecasts lagging somewhat behind. For the outlooks issued in August, NOAA shows more skill than either the CSU or the TSR forecasts, with the largest difference seen in the prediction of named storms.

There are several caveats regarding the forecast verification. Firstly, NOAA's outlooks reflect expected ranges of activity with roughly 70% probability, while the CSU and TSR outlooks are single number forecasts with error bars (usually 1 standard deviation) noted. Although

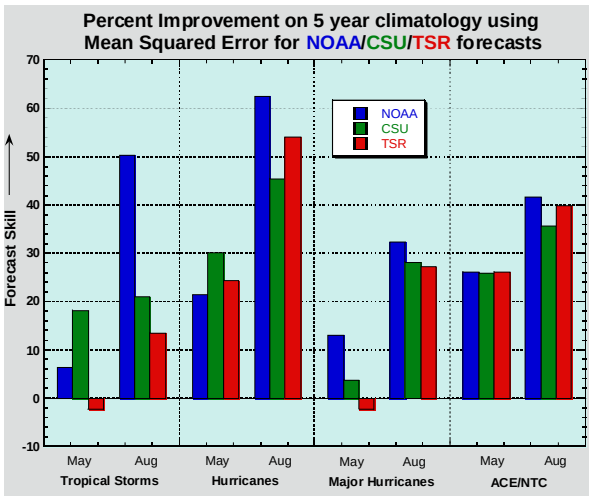

Figure 3: Comparison of May and August seasonal hurricane forecasts from NOAA, CSU and TSR from 1999-2009 (May) and 1998-2009 (August).

using the mid-point of the NOAA forecasts provides an easy comparison, it is not an exact comparison. The forecast tools and methodology of all of the forecast groups have changed over the past twelve years as well. Another potential issue is comparing ACE forecasts from TSR before 2003 with those of the other forecast groups. The first real-time TSR ACE forecasts were made in 2003, while the numbers used prior to that point were replicated real-time forecasts, using only forecast information from 1988 to the year before the forecast. A final point to consider is that the final forecasts are obtained in different ways from each group. TSR's forecast uses a statistical model with no human modification, but both CSU and NOAA have a more subjective approach. Those groups use a blend of statistical and analog models and NOAA also utilizes dynamical model predictions of tropical cyclones and derived tropical cyclone forecasts from global climate model fields to make its outlooks.

One part of the verification was to examine the ranges of the NOAA and TSR forecasts. Figure 4 shows a comparison of these forecasts to the WMO-recommended standard. Note that both the NOAA and TSR forecasts have a larger spread (defined as the maximum minus the minimum value) than the WMO standard. One of the problems with the WMO range standard is that it requires data from the past 30 years, which is mostly from an inactive era (1971-1994) in the Atlantic basin.

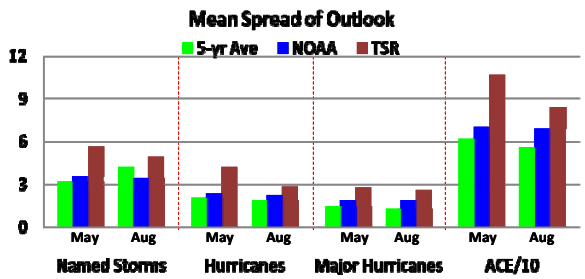

Figure 4: Mean spread of the seasonal outlooks from the WMO-recommended method (5-yr), NOAA and TSR.

Consequentially, the WMO-suggested ranges are much smaller than if a more representative average of the active era (1950-1969, 1995-present) is used, which has consistently shown a larger amount of variance than the inactive era.

Another notable aspect of Fig. 4 is that the ranges for the May and August NOAA forecasts are nearly the same. This was unexpected since, on average, there is more uncertainty involved in a May forecast than the August forecast, due to problems forecasting ENSO and Atlantic SSTs at longer-lead times. Since the May outlooks have not been verifying within the outlook ranges at the target rate of 70% (Fig. 5), one recommendation is to increase the size of the May outlooks to better reflect the higher uncertainty involved in the pre-season forecast.

Although the NOAA outlook ranges have only been slightly larger than the WMO benchmark, the percent of correct NOAA outlooks is significantly higher than the benchmark (Fig. 5). The only parameter that has shown the same skill as the WMO standard is the May named storms forecast. In the rest of the forecast categories, despite having only a slightly larger range than the benchmark, large increases in skill were observed, especially for the August forecasts of named storms and hurricanes, and in both the May and August outlooks of major hurricanes and ACE.

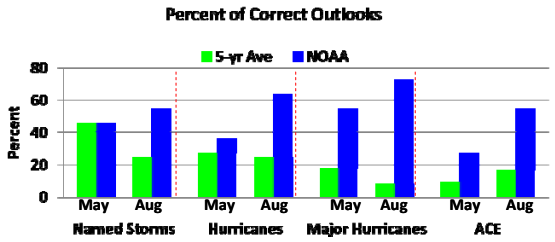

Figure 5: Percentage of correct outlooks for the 5-yr average and the NOAA outlooks in May and August.

Another way to visualize the forecast spread issue is shown in Fig. 6. NOAA forecasts for named storms (upper left) have been too low on average, but the verification for the other parameters have been closer to the middle of the forecast range, especially for hurricanes.

4. SUMMARY

In this study, NOAA seasonal hurricane outlooks are compared to forecasts issued by CSU, TSR and a benchmark of skill using the 5-yr running mean as a forecast. Overall, the May NOAA outlooks had slightly higher skill than the 5-yr mean and comparable skill to the other forecast groups. However, the August NOAA outlooks had considerably more skill than the 5-yr mean forecasts and generally were more accurate than forecasts issued by CSU or TSR, though those forecasts also had significant skill when compared to the 5-yr mean.

A verification of the spread of the NOAA and TSR outlooks was also conducted. It was found that the NOAA outlook ranges were generally smaller than the TSR ranges, with higher skill for the NOAA forecasts. However, both the NOAA and TSR forecasts had higher ranges than a WMO panel recommendation. It is worth noting than the WMO standard uses data from the last 30 years, which may not properly account for the multi-decadal cycle of higher and more variable activity that the Atlantic is currently experiencing. In addition, it was found that the May and August NOAA outlooks ranges are similar, which was not expected due to the higher uncertainties involved in the May outlook. Future NOAA outlooks will probably have larger May ranges to account for higher uncertainties at that time to meet the 70% target verification rate.

5. ACKNOWLEDGEMENTS

Data from CSU and TSR were kindly provided by Phillip Klotzbach and Adam Lea, respectively. Useful manuscript suggestions were made by Chris Landsea James Franklin and Todd Kimberlain.

6. REFERENCES

Bell, G. D., and co-authors, 2000: The 1999 North Atlantic Hurricane Season: A Climate Perspective. State of the Climate in 1999. *Bull. Amer. Meteor. Soc.*, **81**, S1-S68.

Gray, W.M., 1984. Atlantic Seasonal Hurricane Frequency. Part II, Forecasting its Variability. *Mon. Wea. Rev.*, **112**, 1669-1683.

Owens, B. F., and C. W. Landsea, 2003: Assessing the skill of operational Atlantic seasonal tropical cyclone forecasts. *Wea. Forecasting*, **18**,45-54.

WMO, 2008: Expert Meeting to Evaluate Skill of Tropical Cyclone Seasonal Forecasts, Boulder, Colorado, USA 24-25 April 2008,WWRP 2008-4.

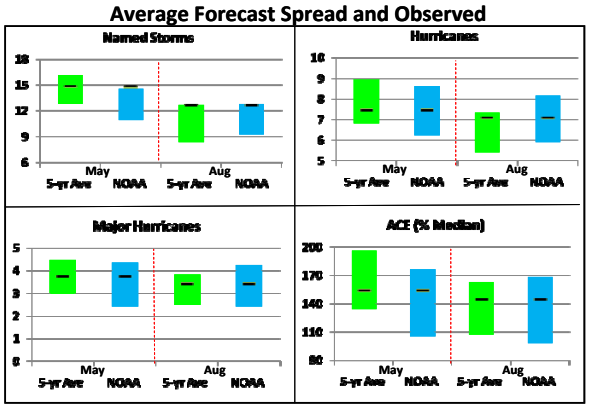

Figure 6: Average forecast spread (bars) and observed (horizontal black lines) values for the no-skill (green) and NOAA (blue) forecasts from May and August.