Joint Hurricane Testbed Semi-Annual Report for Year 1
September 1, 2005-February 28, 2006

Project: *Continued Development of Tropical Cyclone Wind Probability Products*
Principal Investigator: John Knaff and Mark DeMaria
Affiliation: Knaff (CIRA) and DeMaria (NESDIS)
Project dates: September 2005-August 2007
TPC Points of Contact: Rick Knabb, Michelle Mainelli, Chris Sisko and Jim Gross

1. Background

This project is continue the development of the Monte Carlo wind probability program and assist with the implementation of new products that are derived from the output. A verification system for the probabilities will also be developed. At the request of TPC, a new task involving the evaluation of the probabilities associated with hurricane watches and warnings from the 2004 and 2005 hurricane landfalls was added. The timeline and deliverables for this project are listed below in the Appendix.

2. Accomplishments

M. DeMaria coordinated with Rick Knabb of TPC to provide feedback on a training session that was developed to help explain the new probabilities to NWS forecasters and other users of the new products. In addition, several cases from the 2004 and 2005 seasons were re-run using the most current version of the program for Pablo Santos from the Miami WFO, for the development of an experimental algorithm that utilizes the probability output. A web site was created at CIRA displaying the complete set of probabilities from all of the cases for Hurricanes Charley, Frances, Jeanne and Ivan from 2004 and Katrina and Rita from 2005
(see http://rammb.cira.colostate.edu/projects/tc_wind_prob). A short description of the MC program is also provided on the web site to assist with training.

Work has begun on the evaluation of the probabilities associated with hurricane warnings from the 2004 and 2005 seasons. Table 1 lists all the storms that had a warning issued for at least on time period. The probability program was adapted so that it provides probabilities directly at the same set of coastal breakpoints that are used to issue warnings. This set includes 195 points along the U.S. coastline from Brownsville, Texas to Eastport, Maine. The distance between these points is fairly irregular with spacing ranging from about 5 to 50 nmi. To provide more even coverage, the official breakpoints were supplemented by additional coastal points, so that the difference between points is no more than 15 nmi. The final set includes 342 coastal points. The MC model runs at the supplemented breakpoint set for all 14 storms in Table 1 were completed.

A program to match the points with a hurricane warning with the probability output has also been developed. Results show that for all the coastal points for which a warning was issued for these 14 storms, the average 5-day cumulative probability was 28%. This is consistent with previous analysis of the warning regions which suggests that when a warning is issued there is actually only about a 1 in 4 chance of the point experiencing hurricane winds. This data will be further analyzed to determine the distribution of probabilities and the values at the end points of the warning areas. This work may lead to a new application of the MC probability program, which would provide objective guidance for issuing hurricane watches and warnings.

Table 1. Atlantic Storms with at Least One Hurricane Warning

Storm Name	Year
Alex	2004
Charley	2004
Frances	2004
Gaston	2004
Ivan	2004
Jeanne	2004
Arlene	2005
Cindy	2005
Dennis	2005
Emily	2005
Katrina	2005
Ophelia	2005
Rita	2005
Wilma	2005

The 14 storm cases in Table 1 are also being used as a test dataset for the development of the verification program. There were 375 times when a warning was issued or hurricane winds were observed along the coast for at least one breakpoint for these 14 storms, which provides 128250 points (375 x 342) for development of the verification program.

Three methods were proposed for evaluating the probabilities, including a bias check, a Brier Skill Score, and a Relative Operating Characteristic (ROC) score. Progress on the three verification techniques is described below.

a) Bias

The bias program has already been developed, and compares the integrated probability to the total number of points with observed hurricane winds. Results for the 128250 point verification set showed that there were 5025 points with observed winds, and 4489 were expected from the summed probabilities. Thus, the bias is about 0.89. The slight low bias is probably due to the fact that many of the landfalling cases in Table 1 were much larger than the typical storm.

In response to feedback from TPC during a visit in December, 2005, the bias verification was expanded to include a stratification by probability categories. For this calculation, the points are divided into 10 categories according to the probabilities estimated by the MC program (0-10, 10-20, …, 90-100%). The average estimated probability in each group is calculated, and then compared to the percentage of points in each group that actually experienced hurricane winds. If the probability program was perfect, the estimated and observed probabilities would be the same.

Figure 1 shows the estimated and observed probabilities for each group. The correspondence is quite reasonable, given the limited sample size of the test verification set. The observed frequency and estimated probably agree to within ~20% in every category.

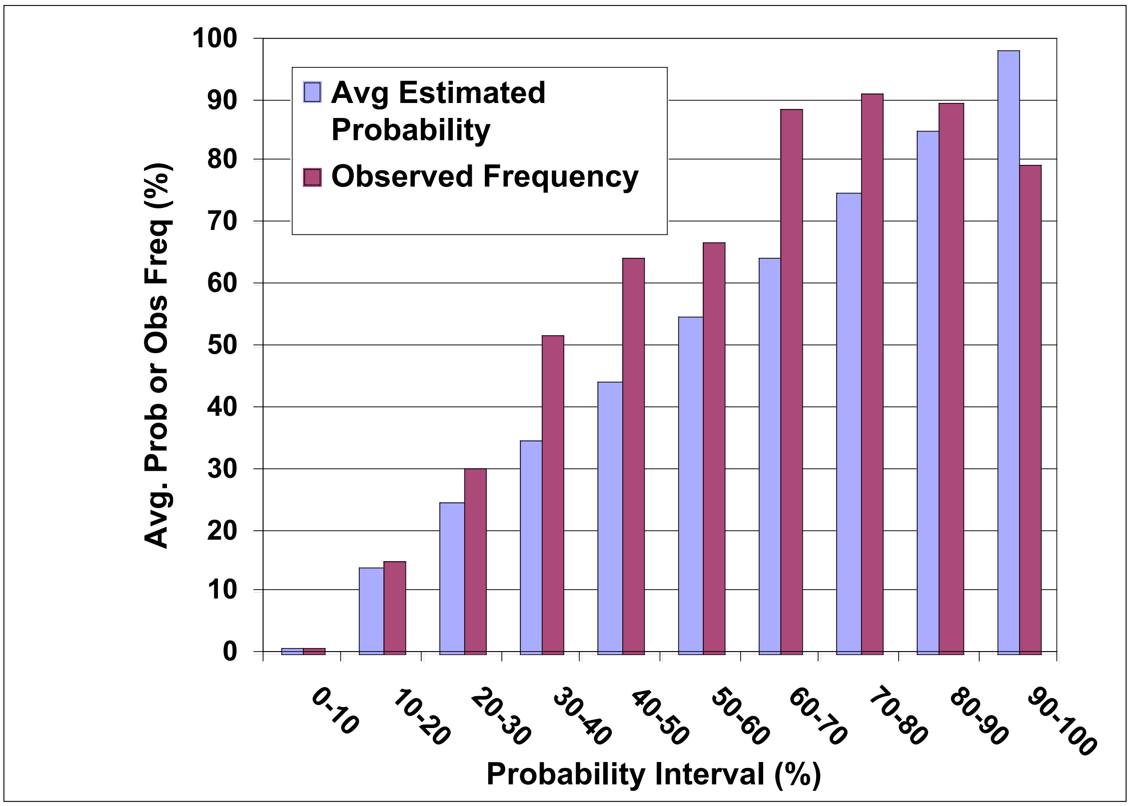

Figure 1. A comparison of the average estimated probability from the MC model and the frequency of occurrence of hurricane winds for 10% probability intervals of the estimated probability.

b) Brier skill score

The program for the Brier skill score first requires the development of a program for the Brier score. This is just the square of the difference of the estimated probability minus the verification (1 for a point with an occurrence of hurricane winds, 0 for a point with no hurricane winds), summed over all points at all times. If the probability forecast was perfect (100% at all points with hurricane winds and 0% at the points without hurricane winds), the Brier score would be zero. The Brier skill is the improvement of the Brier score of the probability, relative to some baseline. The baseline that will be used is what would be available without the probability, which is just the deterministic forecast. This will be converted to a probability by assigning 100% for all points that fall within the 64 kt radii along the forecast track, and 0% for points that fall outside the radius. The program for this calculate is under development.

c) Relative Operating Characteristic (ROC) score

Contingency tables can be constructed from the outcomes of events and warnings (i.e., a binary forecast system), where the issuance/non issuance of a warning is contingent on a threshold probability (e.g., one may put up warnings when the probabilities along the coast greater than or equal to 10%). Table 2 shows the organization of the contingency table, where *h* is the number of hits, *f* is the number of false alarms, *m* is the number of misses, and *c* is the number of correct rejections. For our project we are interested in the trade-off between two quantities that can be estimated from the values compiled in the contingency tables; the hit rate (***hr***) = $h/(h+m)$ and the false-alarm rate (***far***)=$f/(f+c)$. As the threshold probability for issuing warning varies (i.e., 10%, 20% …100%), ***hr*** and ***far*** also vary. By calculating ***hr*** and ***far*** over a range of different threshold probabilities for issuance of a warning for the sample of 128250 points, a curve (e.g. the ROC curve) can be constructed as shown in Fig. 2. The area under this curve is related to the skill of the probabilistic forecasts. If the curve is above the line where far is equal to hr, there is an indication of skill for that portion of the curve. The overall skill of a probabilistic forecast scheme can be estimated from the area above the ***hr***=***far*** line; ***skill*** = $2 X (A - 0.5)$, where A is the area under the curve. The ***skill*** will vary from 1.00 for perfect forecast to -1.0 for a perfectly bad forecast. For our test dataset, ***skill*** = 0.885 and the probability threshold where ***hr*** is maximized relative to ***far*** (i.e., the maximum likelihood ratio) is 60%. With the hurricane warning problem, however, it is likely more important to determine the threshold that maximize the ***hr*** suggesting a threshold closer to 10%. Such issues will be investigated during the rest of the project.

Table 2. Two-by-two contingency table for verification of a binary forecast system

	Forecasts		
Observation	**Warning (W)**	**No Warning (W')**	**Total**
Event (E)	h	m	E
Nonevent (E')	f	c	e'
Total	w	w'	N

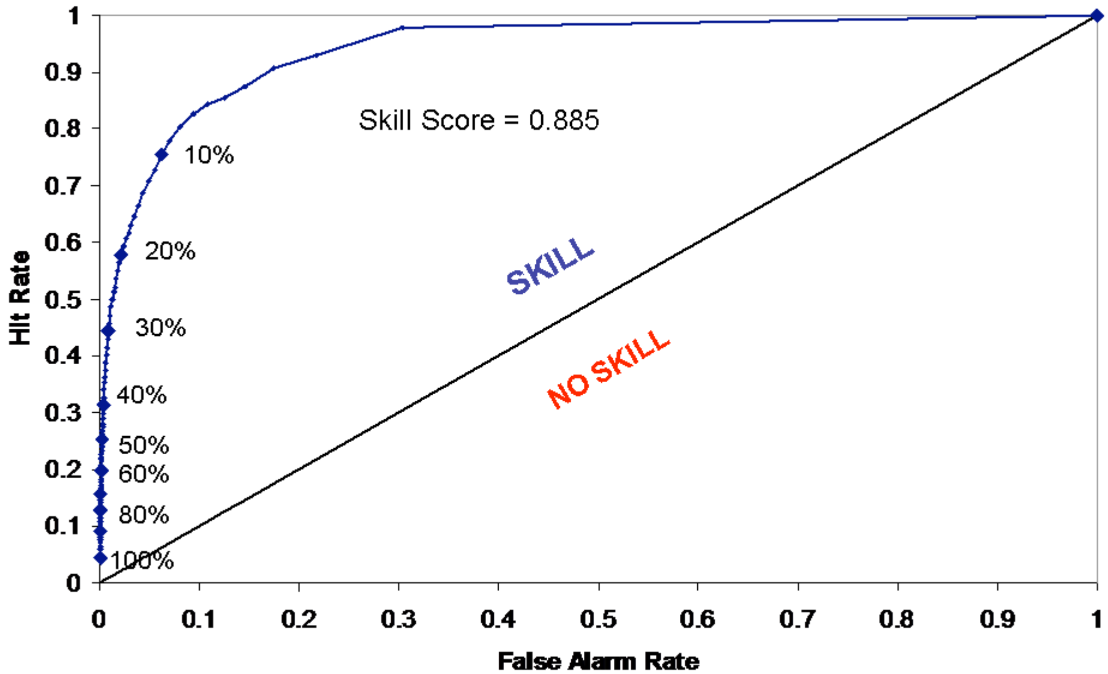

Figure 2. Hit rates vs. false-alarm rates associated with varying the threshold probabilities given by the 5-day cumulative probabilities from the MC model when warning were issued for the storms listed in Table 1.

3. Plans for the Next Six Months

The project will continue according to the schedule shown in the Appendix. The development of the verification program will continue using the 14 storm test set, and then run on the full 2005 season. Results from the evaluation of the probabilities at the hurricane warning locations, and additional progress on the verification program will be reported at the Interdepartmental Hurricane Conference in March of 2006.

Appendix

Year- one project timeline and deliverables:
Sep 2005 – Funding arrives at CIRA
Sep 2005 - Finalize training material in coordination with TPC
Oct 2005 - Begin verification code
Feb 2006 – Provide Year 1 semi-annual report
Mar 2006 - Identify (for the 2004 and 2005 seasons) the wind speed probabilities for the U.S. landfalling tropical storms and hurricanes at both the issuance of the watches and the warnings based upon breakpoints.
Mar 2006 – Perform evaluation of 2005 MC real-time runs

Mar 2006 – Present preliminary results at the IHC
Apr 2006 - Deliver verification code to TPC and provide training
May 2006 – Provide updated MC code to TPC for 2006 based upon 2005 verification
Aug 2006 – Provide Year 1 final report/proposal renewal