15D.2

VERIFICATION OF THE NATIONAL WEATHER SERVICE TROPICAL CYCLONE INTENSITY PROBABILITIES AND FUTURE PLANS

Michael J. Brennan[1], Daniel P. Brown[1], Richard D. Knabb[2], and Mark DeMaria[3]

[1]NOAA/NWS/National Hurricane Center, Miami, Florida
[2]NOAA/NWS/Central Pacific Hurricane Center, Honolulu, Hawaii
[3]NOAA/NESDIS/CIRA, Fort Collins, Colorado

1. INTRODUCTION

The National Hurricane Center (NHC) began issuing intensity probabilities for tropical cyclones (TCs) in the 1990s. The original intensity probability product used the long-term mean NHC intensity forecast errors and the deterministic NHC intensity forecast to compute the likelihood that a tropical cyclone would fall within certain intensity categories (dissipated, tropical depression, tropical storm, hurricane) and the various categories of the Saffir-Simpson Hurricane Scale at specified times during the 72-h forecast period. However, these probabilities did not directly take into account land interaction.

In 2008, the NHC and the Central Pacific Hurricane Center (CPHC) began computing intensity probabilities from a set of 1,000 realizations, or alternate tracks and intensities, that vary around the official forecast based on a Monte Carlo sampling of historical errors in the NHC and CPHC track and intensity forecasts (DeMaria et al. 2009). The Monte Carlo technique accounts for land interaction in the 1,000 realizations and should therefore provide a more accurate assessment of the chances that the intensity of a tropical cyclone will fall within the various categories. This study presents a verification of the new intensity probabilities issued by NHC in the eastern North Pacific and Atlantic basins during the 2008 and 2009 hurricane seasons.

A discussion of planned future enhancements to the product will be provided in the conclusions.

2. METHODOLOGY

The intensity probabilities issued for all 2008-2009 Atlantic and East Pacific TCs were verified. Forecast probabilities at each lead time (e.g., 12 h, 24 h, 36 h, etc.) for each intensity category were cataloged. To increase sample size, the forecast probabilities were grouped into 10% bins (e.g., 0%, 10%, 20%, etc.). The status of the cyclone was then cataloged at the appropriate verifying time from the final NHC best track and the observed frequencies for each forecast probability bin, intensity category, and lead time were computed.

Since the methodology for computing the intensity probabilities does not make a distinction between TCs and other types of cyclones (extratropical, remnant low), the verification was performed regardless of whether or not the system was a TC at the verifying time. For example, a 20 kt remnant low would be counted in the tropical depression category, and a 50 kt extratropical cyclone would be counted as a tropical storm. Dissipated forecasts were verified when a best track point for that cyclone was not available at the verifying time. Probabilities issued with "special" advisories were not verified, as the original probabilities issued with the routine advisory package were retained and verified for that initial time.

Reliability diagrams for each forecast intensity category and lead time were constructed to evaluate the reliability of the probability forecasts.

Despite verifying all probability forecasts for two seasons in both basins, the sample sizes for some forecast probability bins remains quite small [sample sizes are provided on the reliability diagrams (Fig. 1-8)], particularly for the individual Saffir-Simpson Hurricane Scale categories. For that reason, only the results for the tropical storm and hurricane intensity categories are discussed here.

3. RELIABILITY OF THE 2008-09 FORECASTS

Before assessing the reliability of the intensity probability forecasts, the bias of the official NHC intensity forecasts during the same time period was examined, since the official forecast is the basis for the intensity probabilities. During 2007-08, the intensity forecast bias for Atlantic tropical cyclones was positive for each forecast period (12, 24, 36, 48, 72, 96, and 120 h), meaning that the official forecast over-predicted the verifying intensity. However, the bias was small (1.8 kt or less) at each verifying forecast period. For the eastern North Pacific, the NHC official intensity forecast bias was small (less than 0.3 kt) at 12 and 24 h. It was negative (-2.2 to -3.1 kt) at 48-96 hours, meaning that the official forecast under-predicted the verifying intensity. The bias at 120 kt was also negative, but less than 1 kt.

For the Atlantic, the probability forecasts of tropical storm intensity at lead times of 12, 24, 36, and 48 h, show relatively good reliability, as the observed frequency steadily increases with the forecast probability (Fig. 1). However, the technique appears to

---

Corresponding author address: Michael J. Brennan, NOAA/NWS/NHC, 11691 SW 17th Street, Miami, FL 33165; e-mail: Michael.J.Brennan@noaa.gov.

over-forecast the occurrence of tropical storms at 12 h in the 10% to 40% bins, and at 24-h in the 30% and 50% bins. At 12 and 24 h, the distribution of the forecast probabilities was weighted toward the tails of the distribution, with relatively few forecasts around the 50% probability bin. At longer lead times (72-120 h), the forecasts appear quite reliable for probabilities up to 50% (Fig. 2), with more forecasts found in the middle of the probability range.

For the East Pacific, the tropical storm forecast probabilities show a pronounced over-forecast from 10-50% at the 24, 36, and 48 h lead times compared to the observed frequency (Fig. 3). For example, forecast probabilities in the 40% bin have an observed frequency of around 15% at those lead times. At the 12-h lead time the forecasts appear to be much more reliable for probabilities of 30-50%. At the higher probably thresholds (> 50%), the forecasts between 12-48 h exhibit better reliability, however there is still a tendency for the observed frequency to be 5-10% less than the forecast probabilities for probabilities up through 80%.

For the longer term (72-120 h) forecasts in the East Pacific, the tendency for the technique to over-forecast the probability of the cyclone being a tropical storm continues (Fig. 4), despite the slight under-forecast bias in the official NHC intensity forecasts. The relatively flat curves on the reliability diagram suggest that the technique has a limited ability to differentiate between systems that will or will not be tropical storms at these lead times. For example, the observed frequency remains relatively constant, between 15 and 30%, as the forecast probability increases from 40 to 60%.

For the hurricane intensity category in the Atlantic, the probabilities at short ranges (12-48 h) generally show a steady increase of observed frequency with forecast probability, although there is a tendency for the technique to over forecast the probability of the cyclone being a hurricane at probabilities less than 50% (Fig. 5). Exceptions to this general increase in observed frequency occur near the middle of the probability distribution (50%) at 12 and 24 h, however at these lead times the number of forecasts in these probability bins is very small. For example, there were only two forecasts in the 50% bin for a lead time of 12 h and only seven forecasts in the 50% bin for the 24-h lead time.

At longer forecast lead times (72-120 h) there is a sizeable increase in observed frequency as forecast probabilities increase from 30 to 70% (Fig. 6). However, at the 10-40% probability range, the technique consistently over-forecasts the frequency of hurricane intensity. In the higher probability bins (60-90%) the forecast probability is too low compared to the observed frequency technique, although the sample size is smaller for these probability ranges. Overall, the technique is able to distinguish between Atlantic cyclones that will or will not be a hurricane in 3 to 5 days, but the technique is under-dispersive.

In the East Pacific, the results are much noisier for the hurricane category, even at the shorter lead times (Fig. 7). Overall there is some reliability, however, the small sample sizes at the mid-range of the probability distribution makes the results difficult to interpret. Interestingly, in contrast to the Atlantic, the short range hurricane probabilities in the East Pacific have a tendency to *under-forecast* the observed frequency in the lower half of the probability distribution. At the higher probability ranges, there is a tendency for the technique to over-forecast the observed frequency, especially at 36 and 48 h.

There is some signal of reliability for hurricane forecasts in the East Pacific at 72-h (Figure 8), but the technique appears to over-forecast the observed frequency at the higher probability thresholds. The small sample size at 96 and 120 h precludes a meaningful interpretation of the results.

Overall, the results from the two basins are quite different. In the Atlantic, the forecasts for tropical storms have greater reliability than in the East Pacific. The 12-48 h forecasts in the Atlantic, the forecast probabilities for hurricane intensity are too high (low) compared to the observed frequency at the low (high) probability thresholds. The opposite is true in the East Pacific. These results suggest that in the East Pacific the technique has more difficulty discerning which cyclones will be hurricanes in the first 48 h of the forecast period than in the Atlantic. At forecast lead times of 3-5 days, the hurricane forecast probabilities in the Atlantic show much greater reliability than those in the East Pacific, although the sample size for the East Pacific forecasts was small at the higher probability thresholds.

4. FUTURE APPLICATIONS OF THE MONTE CARLO WIND SPEED PROBABILITIES

Emergency planners often base coastal evacuation decisions on the risk of a particular hurricane category (on the Saffir-Simpson Hurricane Wind Scale) affecting their particular area. Unfortunately, in their present state, the intensity probabilities <u>cannot be used</u> to estimate the chances of a particular cyclone falling within the various categories <u>at landfall</u>. The intensity probabilities represent the chances of a cyclone having certain maximum winds at a <u>specific time</u>, not location. Since the probabilities account for land affects (e.g., weakening after landfall), the intensity probabilities cannot be used to infer the intensity of a cyclone at the time of landfall. For example, if a storm is forecast to be near land at a specific time, many of the 1,000 realizations will have tracks that have already intersected land, which results in weaker storms in those realizations. Therefore, the intensity probabilities tend to be spread nearly equally among the various intensity categories when a tropical cyclone is forecast to be near land. While these probabilities accurately represent the chances of a cyclone having a maximum intensity in the various stages (or categories) at the forecast time, they do not accurately reflect the chances of a storm being in a specific intensity category when the center approaches or crosses a coastline.

To obtain accurate landfall intensity probabilities, a planned enhancement will produce conditional probabilities using the tracks and intensities of only those realizations that cross the coast between pre-defined points. It is hoped that NHC forecasters will have the ability to define coastal segments using a Graphical User Interface (GUI) by the peak of the 2010 hurricane season. Once the segment is defined, the intensity of each realization at landfall can be used to create landfall intensity probabilities. It is planned that landfall timing distributions will also be produced. Figure 9 shows an example of the GUI and the graphical output.

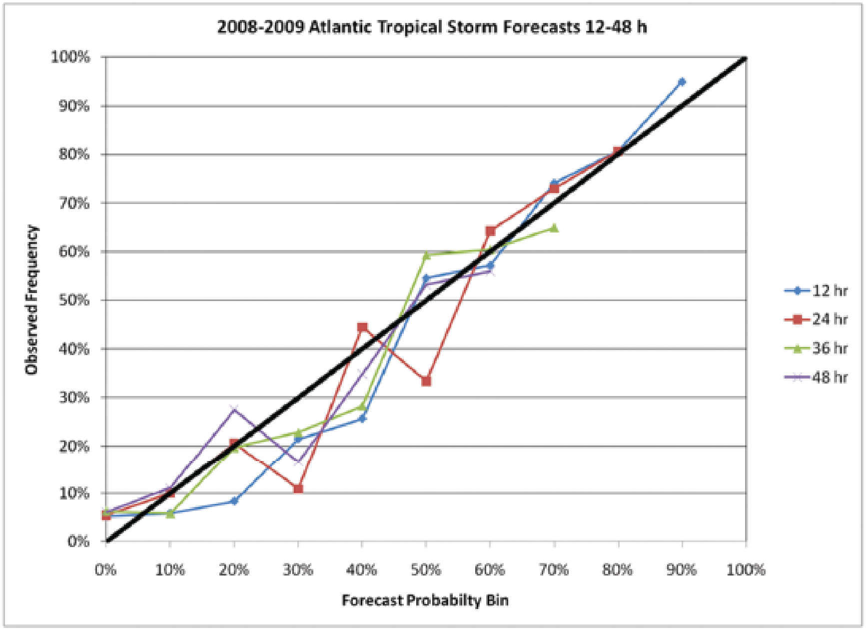
Another proposed utility of the technique is the timing distribution for the onset of 34-, 50-, and 64-kt winds. Since the onset of 34-kt winds makes outside preparedness activities difficult, most emergency managers plan to have these activities completed before the onset of tropical-storm-force winds. Currently, emergency planners only have the ability to graphically display the deterministic forecast of onset of the various wind thresholds in HURREVAC [a decision assistance tool for government emergency managers (http://www.hurrevac.com)]. The plan is to produce onset and ending ranges (in hours) that correspond to specific risk thresholds. An emergency planner could determine a confidence level such that by “x” hours 30 percent of the realizations would have produced 34-kt winds at their location. Therefore they would be 70% sure that tropical-storm-force winds would not have started by that time.

NHC is also exploring the use of the wind speed probabilities to provide line integral probabilities and objective tropical cyclone watch/warning guidance. The line integral probabilities provide the chance of 34-, 50-, or 64-kt winds occurring anywhere between specific locations, rather than at individual locations. For example, this would provide the chance of hurricane force occurring anywhere within a hurricane warning area. Work has begun on producing objective tropical cyclone watch/warning guidance (Schumacher et al. 2010), based on the chance of hurricane-force winds from the Monte Carlo probabilities at the end points of NHC hurricane warnings in the continental United States between 2004 and 2008.

5. REFERENCES

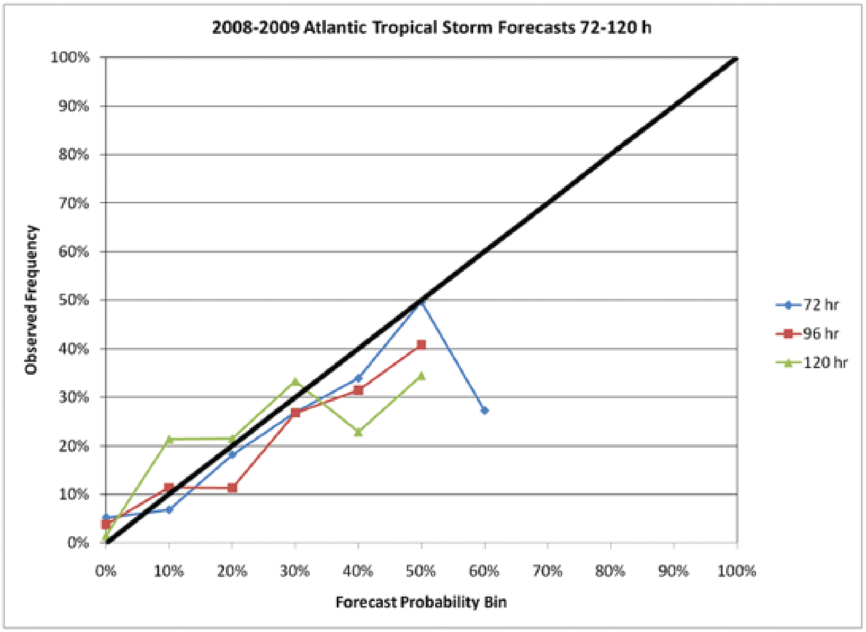
DeMaria, M., J. A. Knaff, R. Knabb, C. Lauer, C.R. Sampson and R.T. DeMaria, 2009: A new method for estimating tropical cyclone wind speed probabilities. *Wea. Forecasting,* **24**. 1573-1591.

Schumacher, A.B., M. DeMaria, J.A. Knaff, C.R Sampson, and D.P. Brown, 2010: Objective tropical cyclone warning guidance using Monte Carlo wind speed probabilities. *Preprints, 29th Conf. Hurr. Trop. Meteor.*, Tuscon, AZ., Amer. Meteor. Soc.



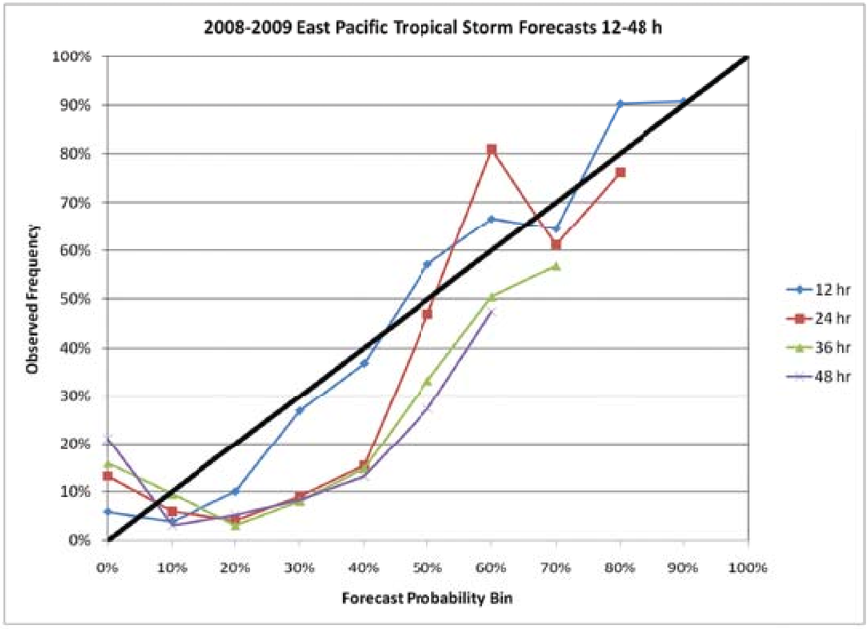
Number of Forecasts	95	34	36	28	39	11	36	54	62	138
	73	40	39	46	36	33	70	96	93	0
	48	52	56	57	46	54	96	111	111	0
	49	36	40	60	92	94	127	0	0	0

Figure 1. Reliability diagram of 2008-2009 Atlantic tropical storm forecasts at forecast lead times of 12 (blue), 24 (red), 36 (green), and 48 (purple) hours. The forecast probability is shown on the x-axis, and the observed frequency is shown on the y-axis. The thick black diagonal line indicates perfect reliability.



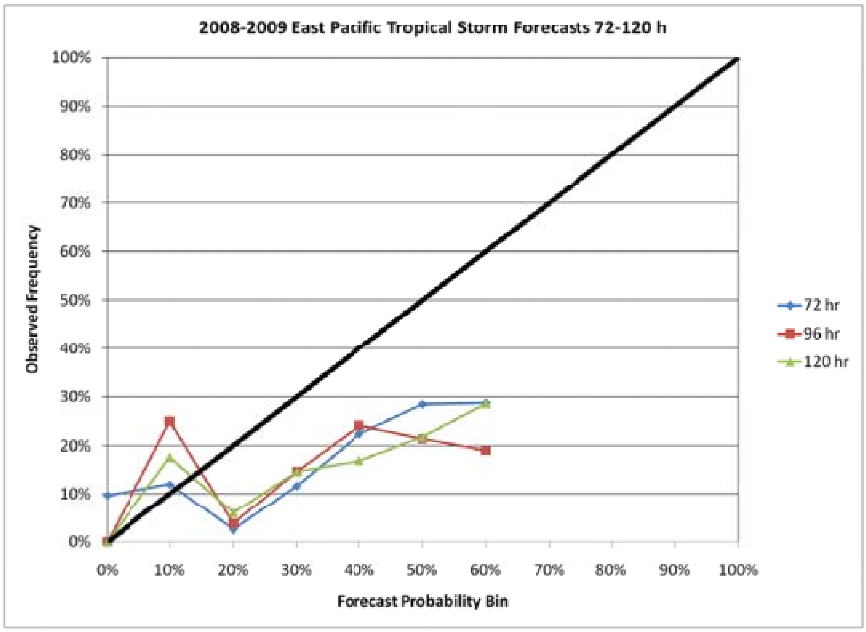
Number of Forecasts	38	44	44	67	112	157	11
	52	44	53	71	89	135	0
	63	28	79	39	144	58	0

Figure 2. As in Figure 1, except for 72 (blue), 96 (red), and 120 (green) hour forecasts.



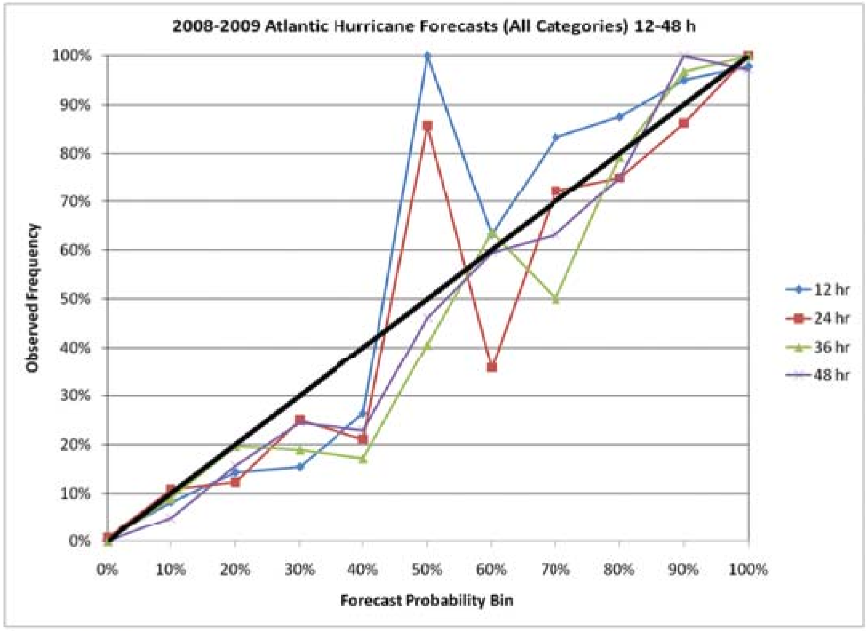
Number of Forecasts	85	77	60	52	38	7	51	31	72	129
	53	67	73	22	65	81	21	85	126	0
	44	42	64	74	54	60	93	146	0	0
	38	32	95	72	61	84	175	0	0	0

Figure 3. As in Figure 1, except for the East Pacific.



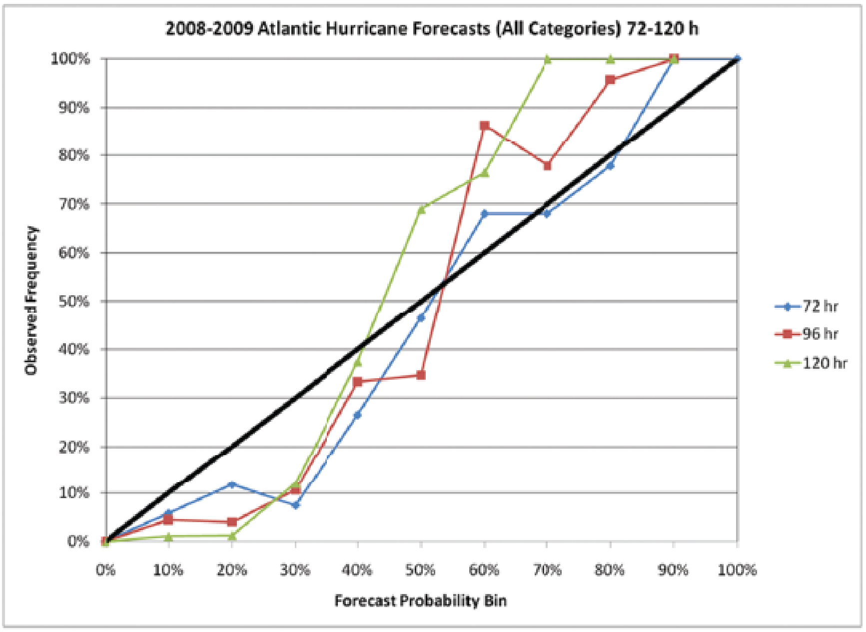
Number of Forecasts	31	25	116	94	67	88	94
	41	20	105	123	54	61	58
	44	17	148	75	59	46	21

Figure 4. As in Figure 2, except for the East Pacific.



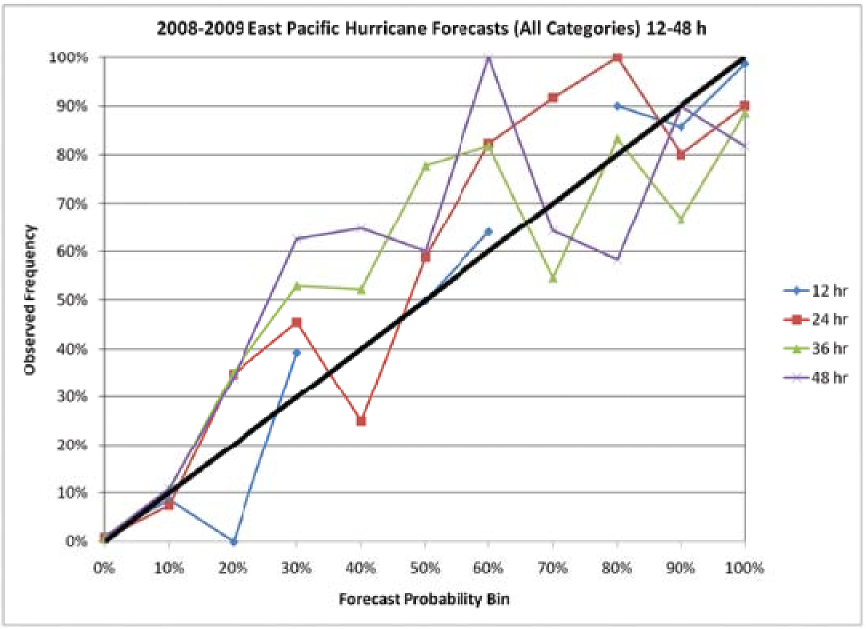
Number of Forecasts	224	73	42	13	19	2	19	12	16	20	90
	142	111	57	20	38	7	14	18	20	29	70
	123	100	56	37	35	32	11	20	29	31	46
	97	101	51	49	35	26	37	19	24	24	34

Figure 5. As in Figure 1, except for Atlantic hurricane forecasts.



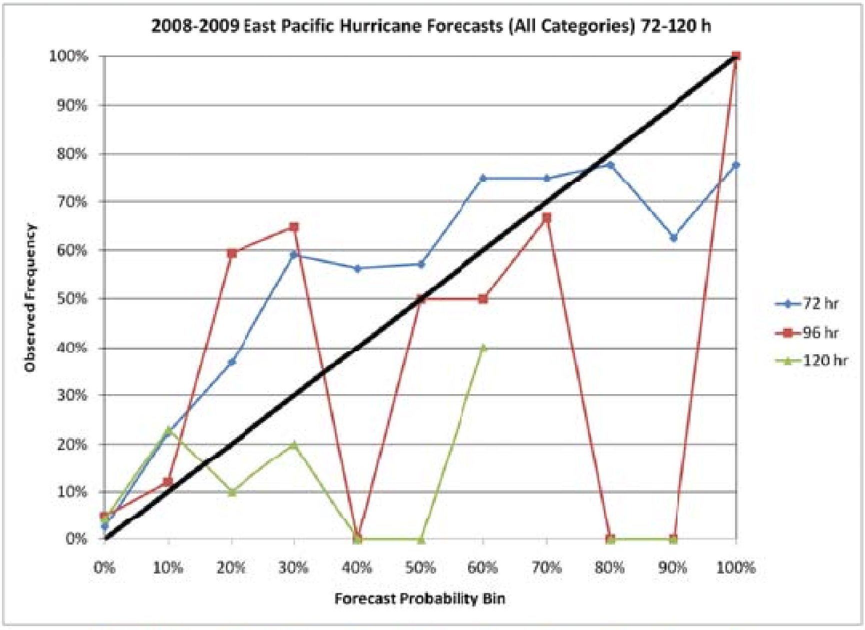
Number of Forecasts	77	102	68	40	49	28	25	25	27	22	10
	90	89	75	36	36	26	22	18	13	18	0
	85	90	78	32	32	29	17	10	16	3	0

Figure 6. As in Figure 2, except for Atlantic hurricane forecasts.



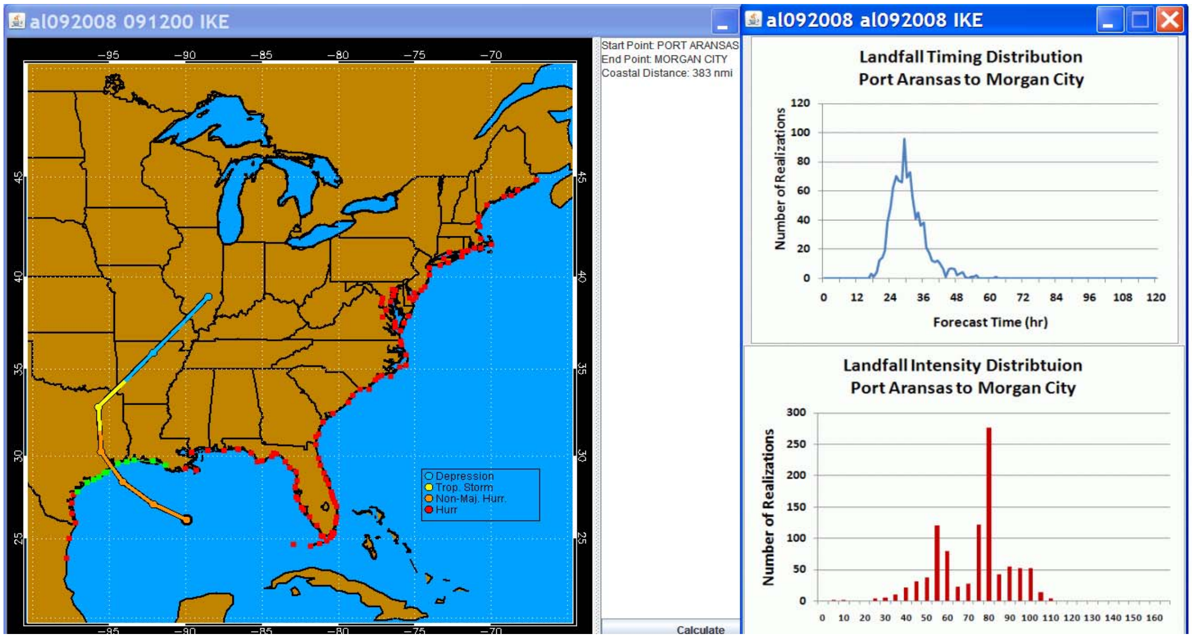
Number of Forecasts	339	82	5	28	0	2	25	0	20	28	73
	277	94	46	33	8	34	17	12	7	15	50
	238	112	74	34	23	18	11	11	12	9	35
	225	138	47	48	17	15	9	14	12	10	22

Figure 7. As in Figure 5, except for the East Pacific.



Number of Forecasts	275	112	38	22	16	14	8	4	9	8	9
	306	75	32	17	2	10	4	5	5	3	2
	326	52	10	5	4	5	5	0	2	1	0

Figure 8. As in Figure 6, except for the East Pacific.



Start Point: PORT ARANSAS
End Point: MORGAN CITY
Coastal Distance: 383 nmi

Figure 9. Example of the proposed GUI that NHC forecasters may use to define coastal segments to determine conditional intensity probabilities and landfall timing distributions.