7.4

PROBABILISTIC GUIDANCE FOR HURRICANE STORM SURGE

Arthur A Taylor* and Bob Glahn
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA
Silver Spring, Maryland

1. INTRODUCTION

Hurricanes are one of the world's most destructive natural forces. Storm surge, the water pushed onto shore by the winds swirling around the hurricane, can wipe out entire communities in a matter of hours. Prior to the advent of numerical modeling of storm surge in the late 1960's, this was the largest cause of loss of life from a hurricane. Even today, as hurricane Katrina drastically demonstrated, we're still susceptible to storm surge, and it is critical that the National Weather Service (NWS) provide guidance on storm surge to emergency managers and the public so they can be prepared and react responsibly.

The National Hurricane Center (NHC) forecasters produce forecasts of storm surge as part of their advisory package. They base their storm surge forecasts on the results of the Sea Lake and Overland Surges from Hurricanes (SLOSH) model (Jelesnianski et al. 1992). The SLOSH model was designed to be used operationally, so the various inputs had to be available, assumed or parameterized. The key inputs which can't be prepared before the storm threatens are the wind and pressure fields used to describe the hurricane over time. To allow forecasters to provide those fields and run the model, SLOSH was designed with a simplified, parametric wind and pressure model which requires only readily available information. Specifically, SLOSH requires the track of the storm, the radius of maximum winds (Rmax) over time, and the pressure difference between the center of the storm and the ambient (or peripheral) pressure (DelP) over time. Since these parameters are either in the NHC's official advisory or can be computed from it, NHC forecasters can produce a storm surge forecast based on the current NHC hurricane forecast.

Because emergency managers and evacuation planners require guidance on potential storm surge flooding well in advance of a hurricane, NHC also provides composites of hypothetical storms grouped by Saffir-Simpson category, forward speed and direction. Each composite is generated by joining the results of running several dozen hypothetical storms through the SLOSH model. The hypothetical storms used in one composite are identical except that their positions have been shifted by a uniform amount. The composite is formed by determining the maximum value a given area attains at any time during any of the runs. While these composites are invaluable tools for evacuation planning, they are not directly associated with the current hurricane, and can result in overestimates of storm surge.

The accuracy of NHC's storm surge forecasts is directly related to the accuracy of the SLOSH model and the accuracy of the input parameters provided to it. Jelesnianski et al. (1992 p. 63) found that when the hurricane's track, intensity and size were estimated as well as possible after the event, SLOSH was correct to within approximately 20% of high water marks. This is particularly good, considering high water marks often vary by 20% for locations that are less than a mile apart (Jelesnianski et al. 1984). Unfortunately the errors in the hurricane input provided to SLOSH cause storm surge errors which are much larger than 20%. For example, in Figs. 1 and 2, the storm surge forecast for hurricane Ivan 2004 made 12 hours before landfall, predicted a surge of 10-12 feet for Mobile, Alabama. Due to errors in the predicted position and size of the

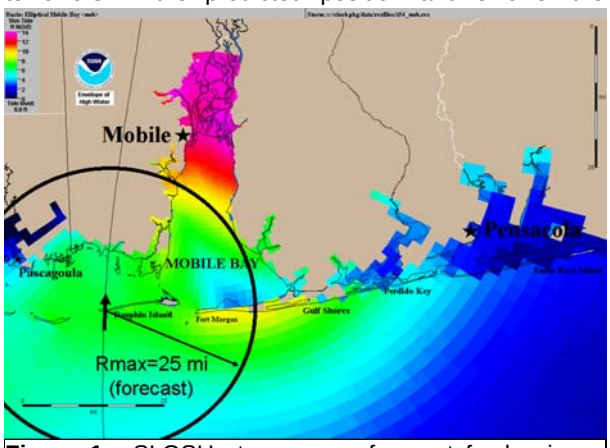

Figure 1. SLOSH storm surge forecast for hurricane Ivan 2004 made 12 hours before landfall.

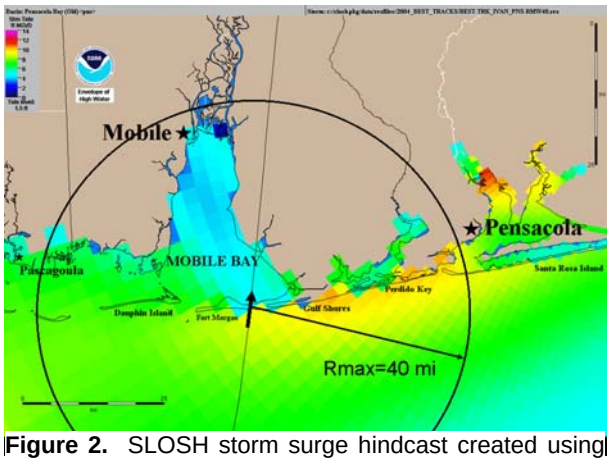

Figure 2. SLOSH storm surge hindcast created using the best available information for hurricane Ivan 2004.

*Corresponding author address: Arthur Taylor, 1325 East-West Highway, Station 10422, Silver Spring, MD 20910-3283; e-mail: arthur.taylor@noaa.gov

hurricane, only 3-5 feet actually occurred. Similarly Pensacola Bay, Florida was forecast to have only 2-5 feet, but actually experienced 7-11 feet.

Since the errors in the input provided to SLOSH far exceed the error in the model, a probabilistic approach for the inputs seemed desirable. The probabilistic hurricane storm surge (P-surge) model was developed to calculate the probability of storm surge from an ensemble of forecasts. The approach is, instead of using a single run of the model based on the current NHC hurricane forecast, to use an ensemble of hypothetical storms based on both the current NHC hurricane forecast and combinations of error distributions derived from historic NHC advisories, to estimate the probability of storm surge.

2. METHOD

The methodology of P-surge is to use NHC's official advisory to create a set of hypothetical storms, each of which has a specific chance of occurring. The set of hypothetical storms is created by permuting the hurricane's position, size, and intensity based on past errors of the advisories. Since each hypothetical storm represents a specific combination of errors in location, size, and intensity, the chance of that combination occurring is assigned to that hypothetical storm. P-surge then joins the results of running the hypothetical storms through SLOSH, along with their associated chance, in a probabilistic manner.

As mentioned earlier, SLOSH's inputs are the location of the storm, Rmax, and DelP; these can be derived from NHC's advisory. The position of the storm, both current and forecast, is available directly from the advisory. However, the advisory doesn't specify Rmax, and only provides the current pressure, from which DelP can be derived. Fortunately, SLOSH's parametric wind model relates Rmax, DelP, and the maximum velocity of wind (Vmax) in such a way that given any two, the third can be calculated. Vmax, both current and forecast, is available from the advisory, thus the current Rmax can be computed from the current DelP. This still leaves determining a forecast DelP and Rmax when neither is provided. To resolve this, we hold Rmax constant and use the SLOSH wind model to calculate the DelP. Thus the P-surge model can automatically compute a single hypothetical storm from the current NHC advisory.

In order to create an ensemble of hypothetical storms, the P-surge model also needs the error distributions associated with NHC's forecast. These are provided by analyzing the average errors of NHC's hurricane forecasts over the last few years, and estimating what they might be in the near future. The specific average errors that we estimate are the cross-track (perpendicular to the motion of the storm), along-track, and intensity errors. The P-surge model then combines them with an assumption that the errors follow a normal distribution to compute three permutations of speed (along-track error) and intensity. For the cross-track direction enough storms are created so that the storms cover 90% of the normal distribution. In addition, they are spaced so that the distance between storms at the 48-h forecast is the same as the 48-h forecast Rmax. This relationship determines the sampling of the normal distribution used in the cross-track dimension and is applied to all forecast projections (Fig. 3). The intent is to make sure that there is a good sampling of the cross-track dimension, since surge is highly dependent on that particular parameter.

The size of the storm also needs to be treated probabilistically. To do so, the size errors need to be determined. However, Rmax is not provided directly from the advisory, so error statistics have to be derived by another method. Since the P-surge model "forecasts" Rmax by holding Rmax constant, the size error statistics are based on the errors in this method. To get matching "observations", we use the matching advisory's DelP and Vmax to estimate the Rmax. Finally, we can't assume a normal distribution to the Rmax errors because a small storm can't get significantly smaller. So the P-surge model uses a different Rmax error distribution depending on the initial Rmax.

Once all the error distributions are established, the P-surge model creates one hypothetical storm for each category of error, and assigns that hypothetical storm a weight which is the product of the probability of each error. For example the P-surge model currently uses three categories weighted 30%, 40%, and 30%, for each of speed, size and intensity. So the fast, medium, and slow storms have weights of 30%, 40%, and 30%, respectively. The large, medium, small and the strong, medium, weak have similar weights. So, excluding the cross-track error dimension, the slow-large-weak storm has a weight of 0.3 * 0.3 * 0.3 = 2.7%, while the slow-medium-weak storm has a weight of 0.3 * 0.4 * 0.3 = 3.6%. Including the cross-track dimension further subdivides the 2.7% and 3.6%.

After the set of hypothetical storms is created, the storms are run and the maximum surge value a particular area attains at any time due to each hypothetical storm is calculated. The actual areas used vary, but typically cover an area between 500 m and 1 km across. The maximum surge values for all the areas are assigned the weight associated with the hypothetical storm which caused them. The surge values and associated weights from all the hypothetical storms are then combined to create probabilistic storm surge.

3. RELIABILITY

A probability forecast should be reliable. For example, if a forecast probability of 20% of storm surge exceeding 5 feet is made numerous times, then on 20% of those occasions storm surge should actually exceed 5 feet. The reliability of a probabilistic forecast system can be shown in a reliability diagram. This is done by plotting the ratio of occurrence, which is the number of times something occurs divided by the total number of

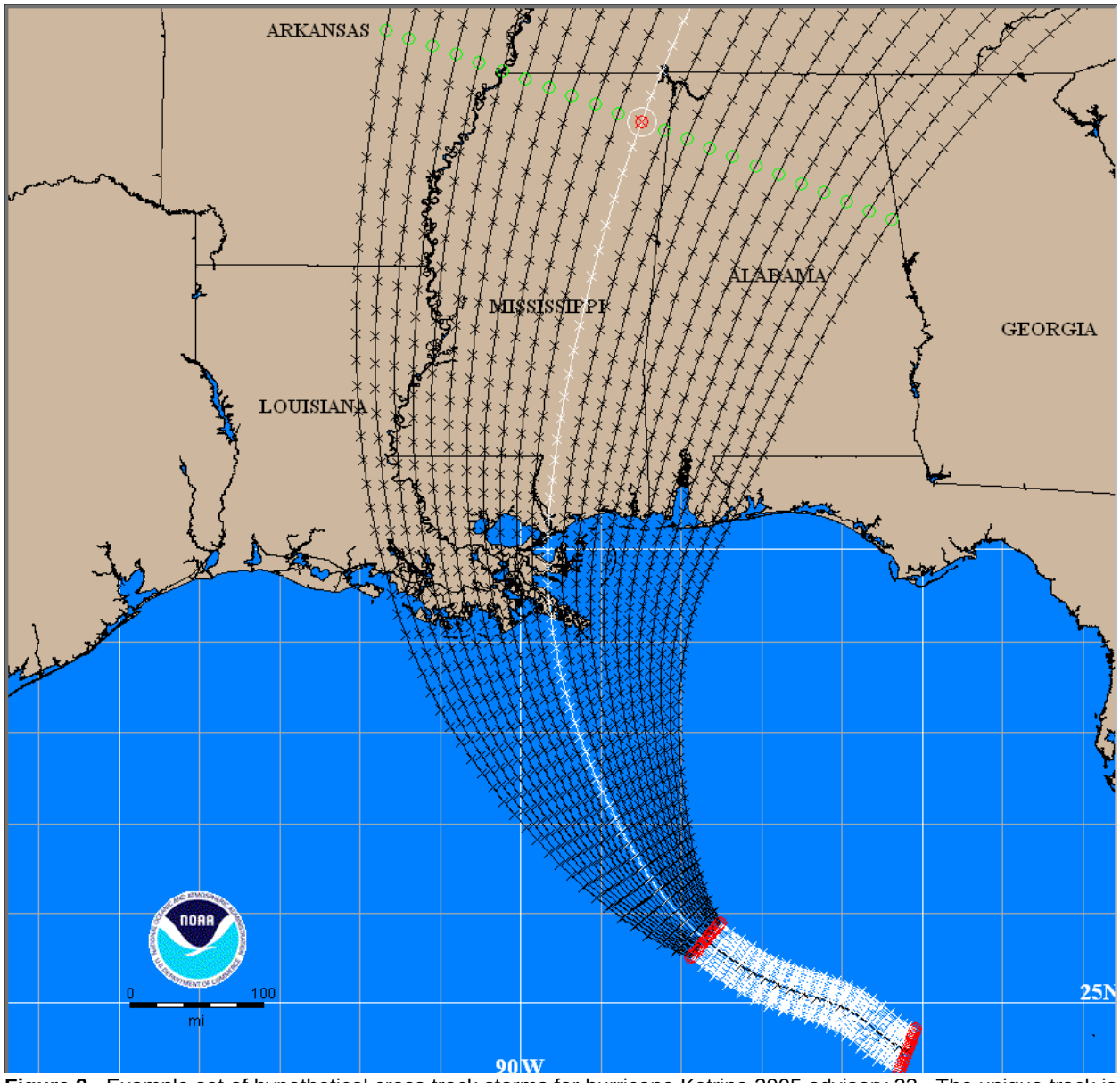

Figure 3. Example set of hypothetical cross track storms for hurricane Katrina 2005 advisory 23. The unique track is the unmodified NHC forecast track. The other tracks cover 90% of the normal distribution, and are sampled in such a way that the distance between storms at the 48-h forecast is the same as the 48-h Rmax forecast. The green circles are the 48-h forecast locations. The white portion is the 20 hours of hindcast, while the black is the forecast.

times it was forecast, against the forecast probability. Ideally the line on the reliability diagram approaches the 45 degree line.

After developing the P-surge model, we needed to determine whether or not it is reliable. If it isn't reliable, then we could calibrate the results by establishing rules which translated the model results to probabilities. For example, if it turned out that over all the times the model forecast a 25% chance of exceeding 5 feet it actually occurred only 20% of the time, then calibration could correct the 25% to 20%.

The problem with determining whether the P-surge model is reliable is that a hurricane making landfall in the United States is a rare event. In addition there are few actual observations of storm surge per hurricane, and those tend to be clustered around where the significant storm surge occurred rather than uniformly across the area. For all the hurricanes which made landfall between 1998 and 2005, NHC was only able to attain 340 excellent storm surge observations, the majority of which were split across seven storms. Unfortunately 340 observations isn't a large enough sample to make an assessment of reliability.

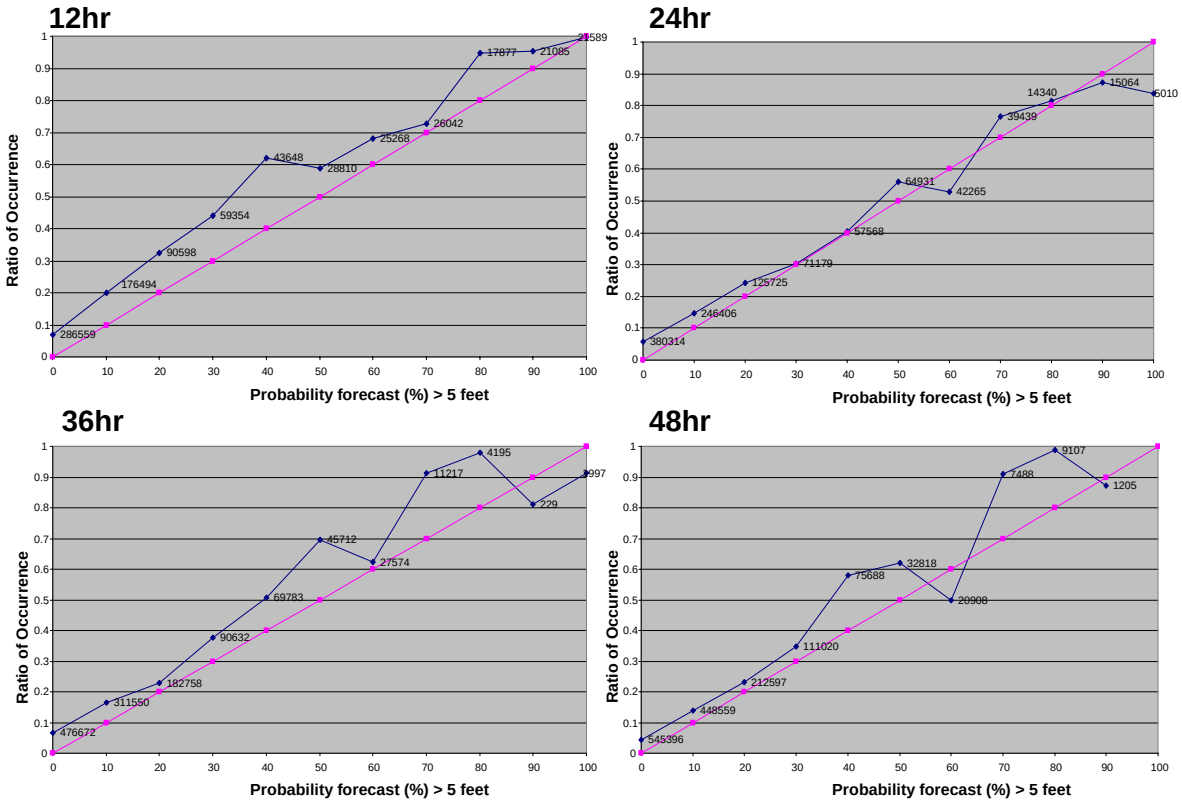

Figure 4. Reliability diagrams for forecasts of greater than 5 feet of storm surge for projections of 12, 24, 36 and 48 hours. The numbers printed on the graph are the number of sample points which went into calculating that value.

So we still needed a method to determine reliability and whether calibration was necessary. Since actual observations weren't available in sufficient quantities, we resorted to using SLOSH hindcasts. NHC has historically created SLOSH hindcast runs for major hurricanes to assess how the model performed. To do so, NHC used the best historical information as input to the SLOSH model. As mentioned earlier, given accurate input, SLOSH model results are within 20% of high water marks, which means the hindcasts are generally within 20% of actual observations. The advantage of using hindcasts is that there are "observations" throughout the area of interest, including where there was little or no surge. The disadvantage was that the hindcast fields we used were created by the same numerical storm surge model as was used in the P-surge model. While this is unfortunate, the fact that the hindcast fields are generally within 20% of actual observations somewhat justifies the process. Figs. 4, 5 and 6 show the reliability diagrams for storm surge greater than 5, 7 and 10 feet, respectively, created by comparing the forecasts with the hindcasts.

The method of using hindcasts, instead of observations, creates numerous sample points per hurricane, but those sample points are highly correlated. For example, if a storm caused flooding in excess of X feet at a particular location, then the locations nearby are also likely to exceed X feet. So while the number of sample points is large, the diagrams were really derived from approximately a dozen actual storm surge events. The plotted values, especially for the higher surges and longer forecast projections are quite erratic. However, this seems to be caused by too few independent sample points and there is no justification for calibration.

4. PRODUCTS

There are two types of products created from the P-surge model. One is the probability of storm surge greater than X feet (Fig. 7). This is calculated for a particular area by adding the weights of the hypothetical storms whose associated surge values are greater than X. For example, if five hypothetical storms have weights of 0.1, 0.2, 0.4, 0.2, and 0.1 and the first two exceeded X feet in a particular area, then the probability of exceeding X feet there is 0.1 + 0.2 = 30%.

The other product is the surge value in feet which is exceeded by Y% (typically 10) of the hypothetical storms (Fig. 8). This is calculated, for a given area, by sorting the surge values from the hypothetical storms, then adding, from the highest surge value downward, the weights associated with the hypothetical storm which caused each surge value. This stops when the sum of weights is $\geq$ Y%. The surge value associated

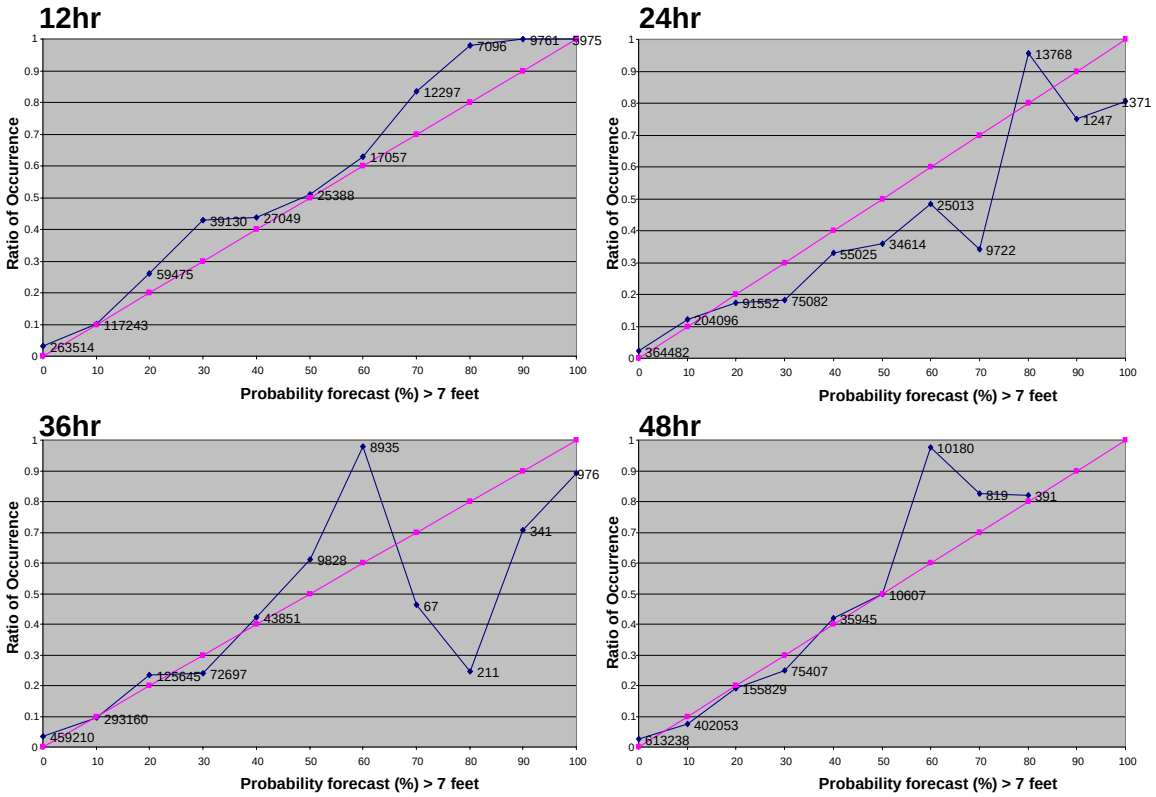

Figure 5. The same as for figure 4, except for forecasts greater than 7 feet.

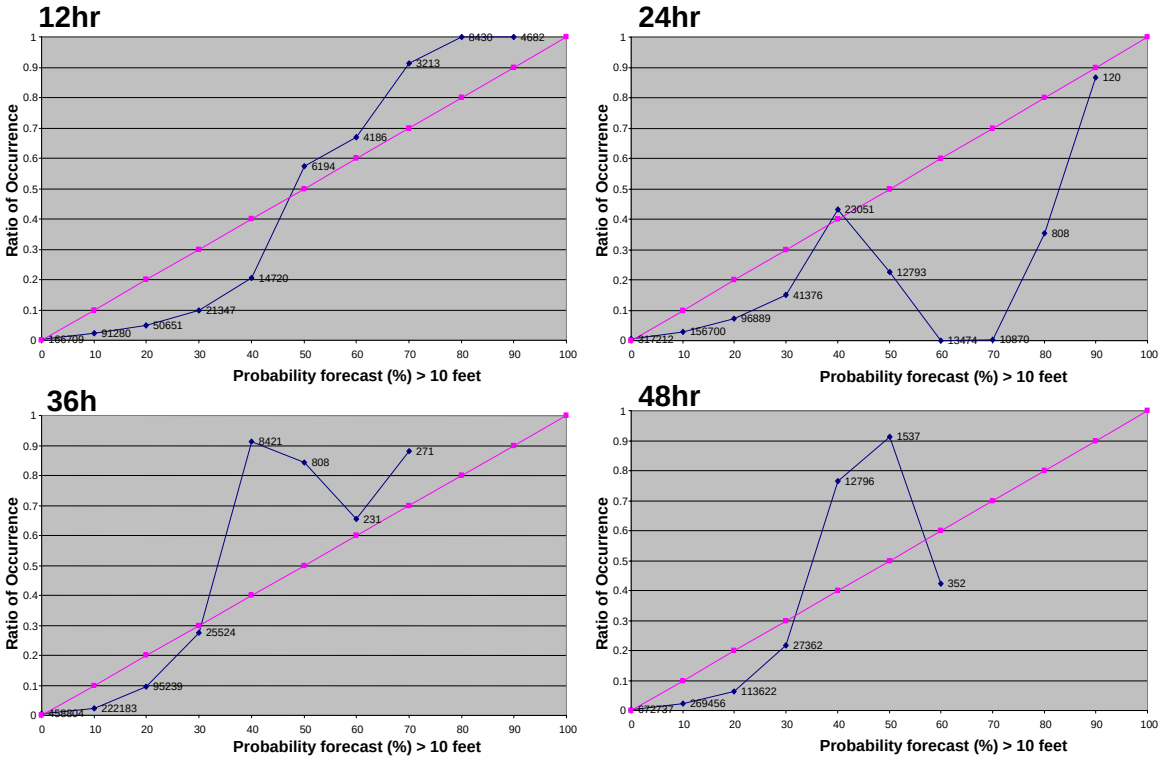

Figure 6. The same as for figure 4, except for forecasts greater than 10 feet.

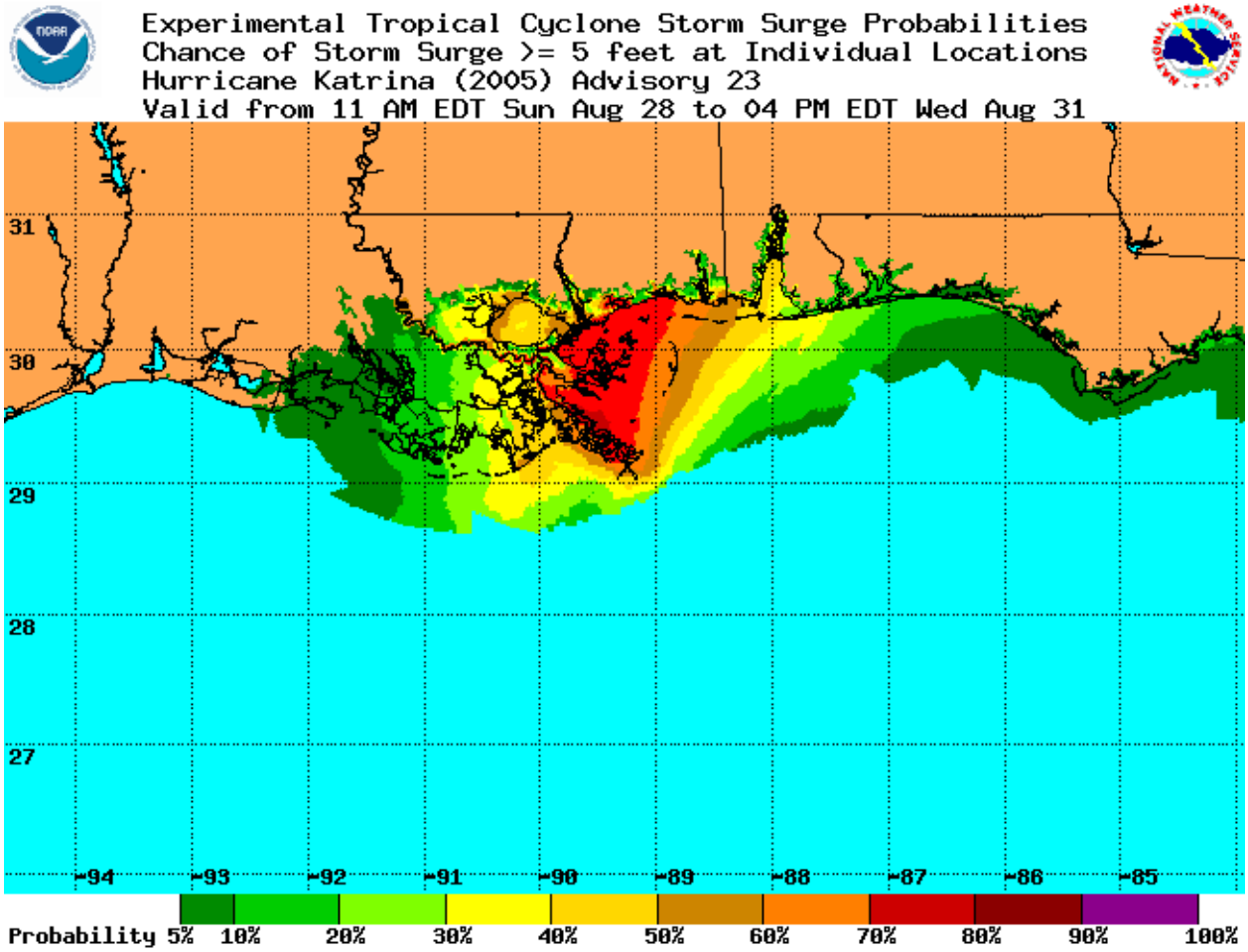

Figure 7. Example of one product from P-surge. It shows the probability of ≥ 5 feet of storm surge for hurricane Katrina 2005, based on the information available in NHC's advisory 23.

with the last weight added is the value exceeded by Y% of the ensemble of hypothetical storms. For example, suppose five hypothetical storms have surge values of 3, 6, 5, 2, and 4 feet and respective weights of 0.1, 0.2, 0.4, 0.2, and 0.1. After creating ordered pairs and sorting by the surge height we have: (6, 0.2), (5, 0.4), (4, 0.1), (3, 0.1) and (2, 0.2). If Y is chosen to be 10, then the height exceeded by 10% of the ensemble of hypothetical storms would be 6 feet, since 6, with its associated weight of 0.2, represents the top 20% of the ensemble's surge values. If Y were 60 instead of 10, then the height exceeded by 60% of the ensemble of hypothetical storms would be 4 feet, since 6, with its associated weight of 0.2, and 5, with its associated weight of 0.4, are both greater than 4 feet, making 60% of the ensemble's surge values greater than 4 feet.

SLOSH has been developed to run on grids specific to particular "basins". P-surge runs over the 38 basins which cover the Atlantic and Gulf of Mexico coasts of the United States; Puerto Rico; Bahamas; and the Virgin Islands. However a hurricane may affect more than one basin, so multiple basins are involved in a P-surge run.

For a particular hurricane, the results from the affected basins are merged onto a common grid which is consistent with the National Digital Forecast Database's (NDFD) (Glahn and Ruth 2003) continental United States (CONUS) grid, except that the resolution is eight times the resolution of the NDFD's CONUS grid. The result of merging overlapping domains can occasionally be seen in the final product and is the explanation for the "fan" shapes in Fig. 8. The merged results are stored in the World Meteorological Organization's (WMO) General Regularly-Distributed Information in Binary form (GRIB) edition 2 format (WMO 2001), and for active storms, are available at: http://weather.noaa.gov/pub/SL.us008001/ST.expr/DF.gr2/DC.ndgd/GT.slosh/AR.conus/. The results in GRIB format, along with images of them, for the active storm and a sampling of historic storms, are available at: http://weather.noaa.gov/mdl/psurge.

5. CURRENT STATUS AND PLANS

The Meteorological Development Laboratory and NHC have been working together to develop an official NHC web site for P-surge products anticipated to be

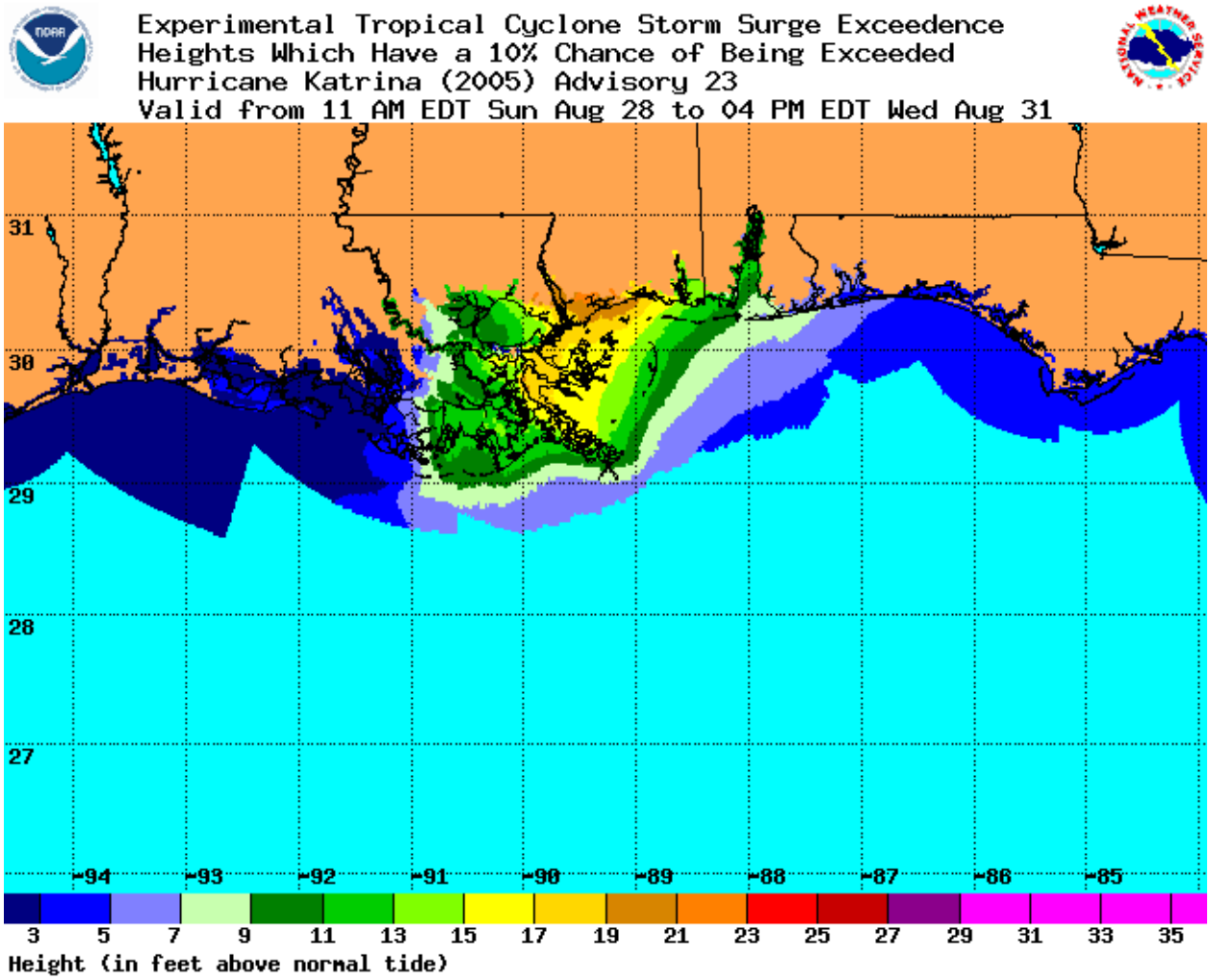

Figure 8. Example of the other product available from P-surge. It shows the values which 10% of the ensemble of hypothetical storms generated for hurricane Katrina 2005 exceeded, based on the information available in NHC's advisory 23.

active during the 2008 hurricane season. The P-surge products are available beginning when NHC issues a hurricane watch or warning for the United States, and are available as close to the advisory release time as possible.

The P-surge model was "experimental" during the 2007 season. The model is running in the National Centers for Environmental Prediction's job stream, the probabilistic forecasts are in the National Digital Guidance Database, an adjunct to the NDFD, and the data will soon be available to the NWS forecast offices. A decision will be made soon as to whether P-surge will move into its "operational" stage beginning in June of 2008. Meanwhile we continue to develop training material and update the error statistics used in the calculations.

6. SUMMARY

NHC has been providing two sets of guidance on hurricane storm surge to emergency managers and the public. One deals specifically with evacuation planning and is not specific to a current hurricane, while the other forecasts storm surge based on the current NHC forecast. As demonstrated with hurricane Ivan in 2004, the guidance based directly on a single run of the current NHC forecast can change drastically between NHC forecast release times. This is because numerical storm surge models are dependent upon an accurate forecast of the hurricane's track and surface wind structure, and today's best hurricane forecasts still have considerable uncertainty.

The probabilistic hurricane storm surge (P-surge) model uses the historic errors in the official NHC hurricane forecast to forecast the probability of storm surge from an active hurricane. The intent of P-surge is to provide information on both what could happen and how likely it is to occur. P-surge, as NHC's third set of storm surge guidance, complements the other two. It provides storm specific probabilities of potential storm surge which may enhance the non-storm specific guidance provided to emergency managers for evacuation

planning. In addition, it takes into account NHC's hurricane forecast errors, which allows it to be run earlier than is reasonable for a single run of the NHC forecast. Finally, when it is reasonable to create a forecast based on a single run of NHC's forecast, the P-surge guidance augments the single run by quantifying the uncertainty of the single run guidance.

7. REFERENCES

Glahn, H. R. and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, 84, 195-201.

Jelesnianski, C. P., J. Chen, W. A. Shaffer, and A. J. Gilad, 1984: SLOSH- a hurricane storm surge forecasting model. *Preprints Oceans 84,* Washington, D.C., Marine Technology Society and IEEE/Oceanic Engineering Society, 314-317.

Jelesnianski, C. P., J. Chen, and W. A. Shaffer, 1992: SLOSH: Sea, Lake, and Overland Surges from Hurricanes. *NOAA Technical Report*, NWS 48, U.S. Department of Commerce, 71 pp.

WMO, 2001: Manual on Codes, FM 92-XII GRIB - General Regularly-Distributed Information in Binary form, Vol. 1.2, WMO No. 306, WMO, 76 pp. [Current version 4, 7/11/2007, 79pp. available online at: http://www.wmo.ch/pages/prog/www/WMOCodes/GRIB.html]