Improving the Validation and Prediction of Tropical Cyclone Rainfall

Joint Hurricane Testbed
Second Year Midterm Report
August 1, 2004 – January 31, 2005

Principal Investigators Timothy Marchok, NOAA/GFDL
Robert Rogers, NOAA/AOML/HRD
Robert Tuleya, SAIC at NCEP/EMC

TPC Contact Stacy Stewart

Project summary and timeline

Through funding from the Joint Hurricane Testbed (JHT), this project proposed to improve the validation and prediction of tropical cyclone rainfall. Improved validation of rainfall will enable the forecaster to identify errors and biases in the models, which can aid the forecaster in interpreting numerical guidance of rainfall and adjusting their forecasts accordingly. An accurate diagnosis of rainfall forecast errors requires a validation scheme that accurately measures the performance of the forecast system. However, there are many aspects of the rainfall field produced in tropical cyclones (TCs) that are not sufficiently addressed by standard validation techniques such as bias and threat scores. A tropical cyclone is a system that dynamically constrains convective development to storm-relative locations that persist for time periods from hours to days. A model's ability to reproduce the rainfall fields is thus dependent on its ability to capture these dynamical features (e.g., eyewall, rainband, and stratiform rain) and to accurately predict the track and intensity of the storm. Furthermore, a great deal of useful information can be obtained from considering the performance of the forecasts for the entire distribution of rainfall, not just peak rainfall amounts or point comparisons with specific rain gauges. This is particularly important when comparing models of varying resolution to observations based on comparatively small sampling areas such as radar data or rain gauges, since a spatially averaged field always has lower variability than point values (Tustison et al. 2001). Finally, most traditional validation schemes are run on fixed geographical domains. As done in Tuleya et al. (2005), limiting the validation domains to areas close to the storm track will narrow the focus of the validation to rainfall that is more directly linked with the storm, making the validation storm-specific.

The issues raised above highlight several aspects of a rainfall validation scheme that are desirable for TC QPF: 1) accounts for the varying abilities of the models to reproduce elements of the storm (e.g., structure, track, and intensity); 2) accounts for sampling size discrepancies; 3) performs comparisons of the entire rainfall distribution in addition to peak rainfall; 4) focuses on storm-related rainfall. Only then can objective evaluations of the TC QPF from various models be performed, allowing for the identification of biases in the models and ways to improve them.

To accomplish the goals stated above, several deliverables were proposed to be completed by the end of this 2-year project: development of new rainfall validation schemes that provide a baseline of comparison for different forecast systems; production of rainfall forecast error statistics for historic United States landfalling storms using traditional and new validation techniques for the operational GFDL, Eta and GFS models, and the benchmark Rainfall CLIPER

(R-CLIPER) model; and design of a new forecasting tool based on the R-CLIPER model that incorporates information related to vertical shear and storm track. With these goals in mind, the following tasks were proposed to be completed during the second year:

1) Continued development of new TC rainfall validation techniques;
2) Calculation of validation statistics to be presented to NHC;
3) Development of a version of R-CLIPER that accounts for vertical shear.

This document will report on the progress achieved up to this point.

Accomplishments to date

1) Summary of accomplishments

The following list summarizes the accomplishments during the first half of this year. Further explanations are provided in subsequent sections.

Continued development of new verification techniques
- Rain flux PDFs
- Track-relative Rain flux PDF in swaths surrounding the storm
- Storm-relative grid-shifted rainfall statistics (i.e., shifted ETS, shifted correlations)

Calculation of validation statistics
- Bias scores - all storms, 1998-2003 (GFDL, GFS, Eta, R-CLIPER)
- Equitable Threat Scores (ETS) - all storms, 1998-2003 (GFDL, GFS, Eta, R-CLIPER)
- Correlation coefficients – all storms, 1998-2003 (GFDL, GFS, Eta, R-CLIPER)
- Rain flux PDFs – all storms, 1998-2003 (GFDL, GFS, Eta, R-CLIPER)
- Track-relative Rain flux PDF in swaths surrounding the storm – all storms, 1998-2003 (GFDL, GFS, Eta, R-CLIPER)
- Storm-relative grid-shifted rainfall statistics for all storms 1998-2003 (GFDL, GFS, Eta, R-CLIPER)

Development of a version of R-CLIPER that accounts for vertical shear
- Acquisition of R-CLIPER code to prepare for modification
- Acquisition of satellite-based rainfall measurements stratified by storm intensity and vertical shear for quantification and incorporation into R-CLIPER

2) Continued development of new verification techniques

New validation techniques have been developed that account for the various issues discussed above (i.e., track, intensity, and structure errors, sampling size discrepancies, rain distribution vs. peak rainfall comparisons, and storm-related rainfall). These techniques are probability distribution function (PDF) comparisons, track-relative PDF analyses, track-relative mean rainfall swaths, and storm-relative ETS and pattern correlations analyses. The PDF comparisons show the distribution of rain flux as a function of rainfall threshold, where the rain flux is defined as the product of the rain rate in a given grid box times the areal coverage of the grid box. Using this technique can account for the differences in variability that arise due to averaging scale discrepancies (Tustison et al. 2001), though differences that arise in models due to the ability to resolve different features remain. Similar to the bias score, this metric is track-independent.

Another technique, track-relative PDF analyses, is used to compare rain flux distributions independent of track error and to isolate the ability of the various models to reproduce rain fields at radii normally associated with the eyewall, rainband, and stratiform regions. Figure 1 shows an example of the swaths over which the track-relative PDFs of storm-total rain flux are

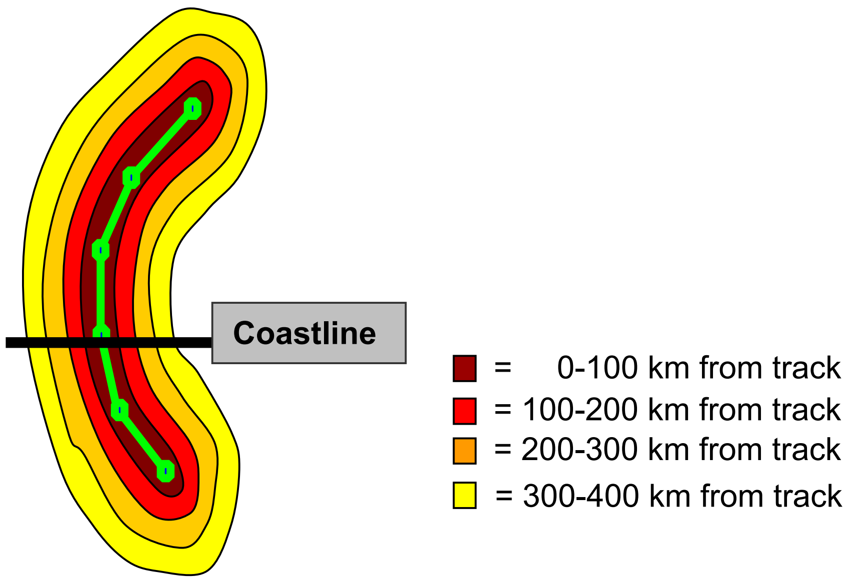


Figure 1. Schematic showing configuration of swaths used in track-relative PDF analyses.

calculated. As can be seen from this schematic, the storm total rain flux amounts are calculated in 100 km swaths. The innermost 100 km is dominated by rainfall produced in the eyewall region, while distances further from the track are likely a mixture of rainband and stratiform rain. A final technique, storm-relative analyses, is used to quantify the impact of track error on standard track-dependent validation algorithms such as the ETS and pattern correlations. Storm-relative analyses are calculated by shifting the 6-hour forecasted rainfall pattern by a distance equal to the difference in position of the forecasted vs. the observed storm location. These shifted rainfall analyses are then summed over the lifetime of the storm, producing storm-total shifted rainfall analyses. Figure 2 shows an example of a shifted Eta 6-hour rainfall field for Tropical Storm Bill of 2003. Figure 2a shows the Eta forecasted rainfall field and Fig. 2b shows the observed (Stage IV) field during the same time period. The white line in Fig. 2 denotes the best track position of Bill, while the red line is the Eta forecasted track. Figure 2c shows the resultant 6-hour Eta forecasted rainfall field after the field is shifted by an amount equivalent to the forecast track error. As can be seen from Fig. 2d, shifting the rainfall field results in a threefold increase in the correlations for this storm during the 24-30 h time period, indicating a significant contribution of track error in this case. These results suggest that a model's QPF

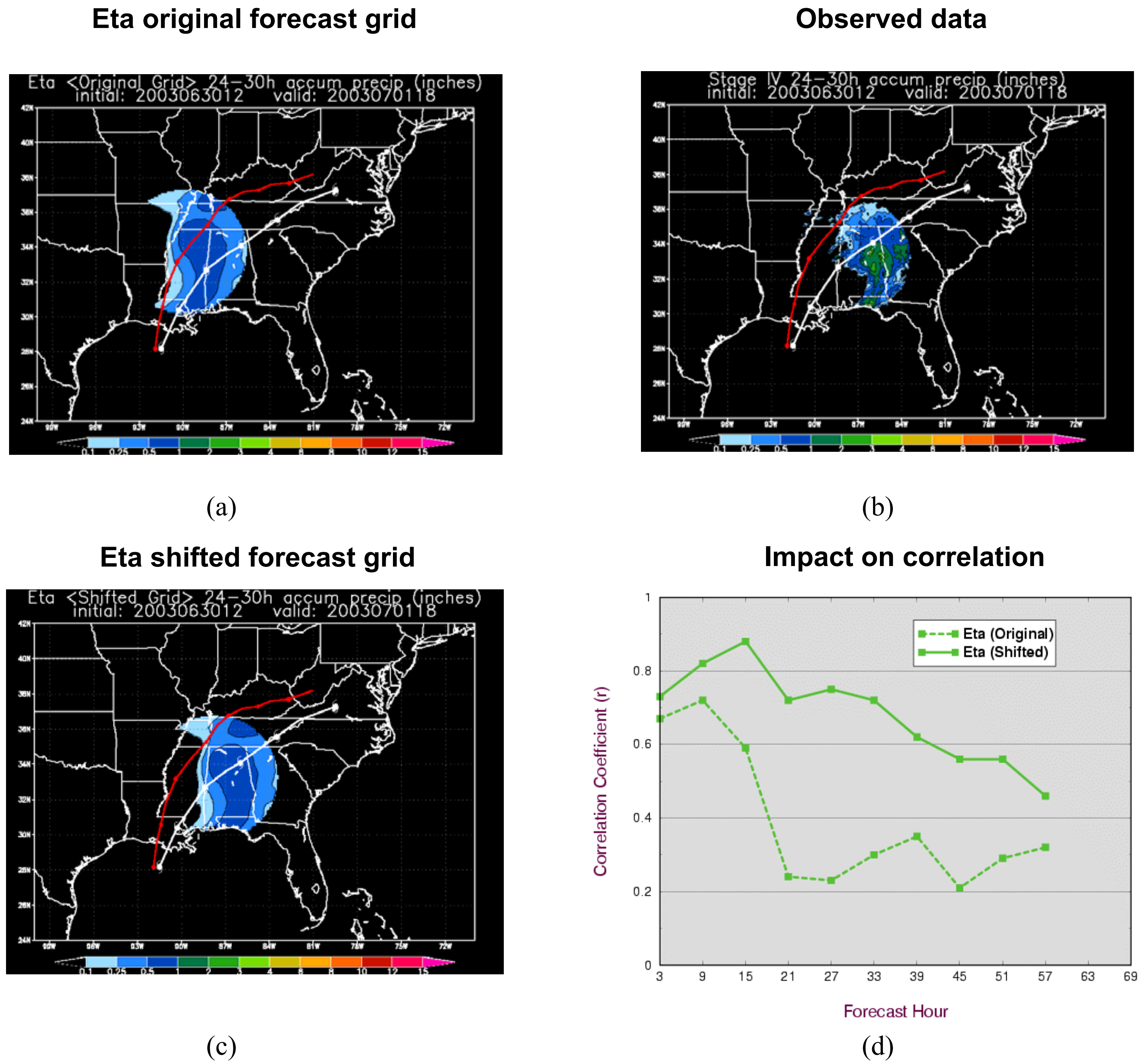


Figure 2. Example of storm-relative grid-shifted rainfall fields for Eta forecast of Tropical Storm Bill (2003) compared with Stage IV observed fields. (a) Original Eta 24-30 h forecast of rainfall; (b) Observed 24-30 h rainfall; (c) Shifted Eta 24-30 h forecasted rainfall; (d) Correlation coefficient as a function of time using original and shifted forecasts.

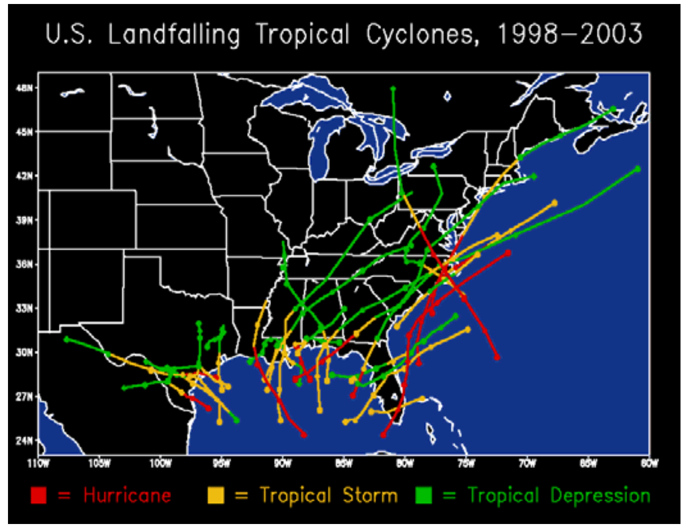
forecasts can provide valuable information (e.g., spatial distribution of the rainfall) even if that model's track forecasts are not perfect.

3) Calculation of validation statistics

The storms that have been included in the database so far are all U.S. landfalling storms between 1998 and 2003 (Figure 3) for a total of 28 storms. One forecast from each storm is included in the database. The cases that were selected had initial times that were always from

1998	1999	2000	2001	2002	2003
Bonnie 95	Bret 100	Gordon 55	Allison 45	Bertha 35	Bill 50
Charley 40	Dennis 60	Helene 65	Barry 60	Edouard 35	Claudette 75
Earl 70	Floyd 90		Gabrielle 60	Fay 50	Grace 35
Frances 45	Harvey 50			Hanna 45	Henri 30
Georges 90	Irene 70			Isidore 55	Isabel 90
Hermine 35				Kyle 35	
				Lili 85	

(a)



(b)

Figure 3. (a) Storms included in database; (b) Tracks of storms used.

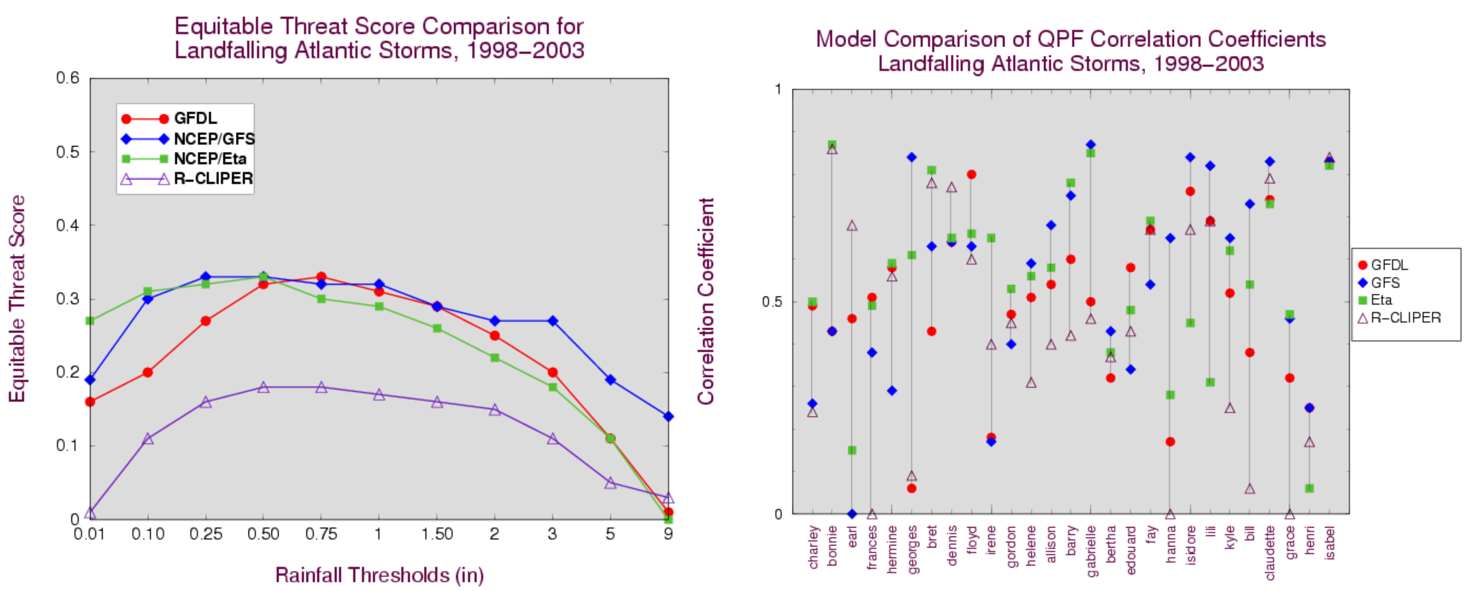
the last 12 UTC time within 24 hours of landfall, in order to coincide with the storm database used by Tuleya et al. (2005). Forecast and observed data for each storm were included in the database until advisories from the National Hurricane Center were no longer issued for them. Figure 3 shows that while the storms in this database took a variety of tracks over the Gulf and Atlantic Coast states, the majority (75%) of the storms made landfall along the Gulf Coast. The storm tracks passed over a variety of topographies with different translational speeds, and they span a wide spectrum of conditions that can produce many different rainfall distributions.

Rainfall observations were provided by Stage IV hourly 4-km gridded rainfall data (Stage II prior to 2002) provided by the Environmental Modeling Center (EMC) at the National Centers for Environmental Prediction. The thirteen regional RFCs perform quality control on these data, then send them to EMC where they are combined into a unified analysis. This data consists of multi-sensor (i.e., rain gauges, radar) rainfall maps covering the entire contiguous United States. It is available on an hourly basis for all times back to 1998.

The models used were the real-time configurations of the NCEP operational models used in forecasting hurricanes that can also output QPF data in real time: the GFDL hurricane model, the GFS, and the Eta model. The GFDL model is a nested, hydrostatic regional model whose current configuraiton runs at 1/6 degree grid length, the GFS is a global model run at 1/2 degree resolution, and the Eta is a limited-area model with a minimum grid length of 12 km. The R-CLIPER model is run for all storms, using the official TPC forecasted positions, to provide a benchmark against which the other models are judged. The R-CLIPER can be run at any resolution; in this case it is run at 4 km grid length.

All of the forecasts and observations use a land-sea mask, and Canada and Mexico are excluded from the analyses. For many of the validation statistics (e.g., bias score, ETS, correlation) only those areas within 600 km of the storm track are included. This restriction is to limit the inclusion of rainfall that is not directly related to the tropical cyclone, such as rainfall from a frontal boundary or mid-latitude cyclone well-removed from the TC. All of the statistics in this study are for storm total rainfall and include data up to a maximum cutoff forecast time of 72 hours within each forecast.

Validation statistics are separated into track-dependent and track-independent groupings. The track-dependent statistics (ETS and correlations) are presented in Fig. 4. Comparisons of the ETS for the different models (Fig. 4a) show that the GFS model generally performs the best across all rainfall amounts, including the highest amounts of 9 inches, while the R-CLIPER model does the worst across all rainfall amounts. The most significant difference between the GFS and the other models occurs at the low and high extremes of the rainfall distribution (i.e., < 0.25 inches and > 2 inches). The correlation statistics (Fig. 4b) show significant case-to-case variability in the performance of the models. In general, though, the GFS has the highest correlations while the GFDL and the R-CLIPER models have the lowest correlations.



(a) (b)

Figure 4. Track-dependent QPF statistics for all models and all U.S. landfalling storms between 1998 and 2003. (a) Equitable threat score; (b) Correlations

Track-independent statistics include the bias score, rain flux PDF score, track-relative rain flux PDF, and storm-relative grid-shifted rainfall. The bias scores for the various models are shown in Fig. 5. The GFS and the Eta models show a high bias for rainfall amounts up to 3 inches, above which point they show a low bias indicative of an inability to produce the extreme rainfall amounts. By contrast, the GFDL shows a comparable high bias to the other two dynamical models for the lower rainfall amounts, but above three inches the GFDL shows a very pronounced high bias. The R-CLIPER shows a small high bias for the low rainfall amounts and a small low bias for the high rainfall amounts.

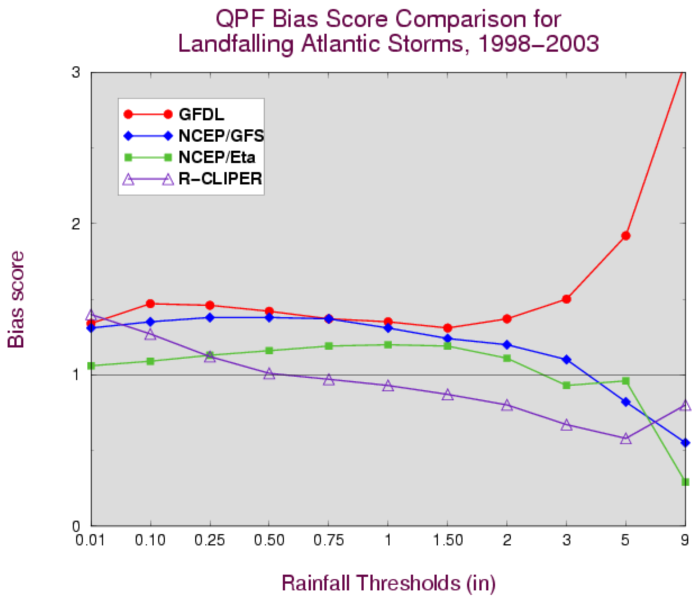


Figure 5. Bias scores for all models and all landfalling U.S. storms between 1998 and 2003.

A plot of the PDF of rain flux vs. rainfall amount for the different models is shown in Fig. 6. The observed rain fields show a log-normal distribution with a peak in the distribution (modal value) at about 4 inches. By contrast, the Eta model (Fig. 6a) has a modal value at around 2 inches, and the distribution of rain flux at the high rain rates is less than the observed distribution. The GFDL's modal value is at about 5 inches, and there is a tendency to underproduce rain flux for the intermediate rain amounts (2-4 inches) and overproduce rain flux at the higher rain amounts. The R-CLIPER model (Fig. 6b) shows a pronounced shift in the distribution toward lower rain amounts, with a modal value near 2 inches, higher rain flux for the light rain amounts, and reduced rain flux for the intermediate to heavy rain amounts. The GFS

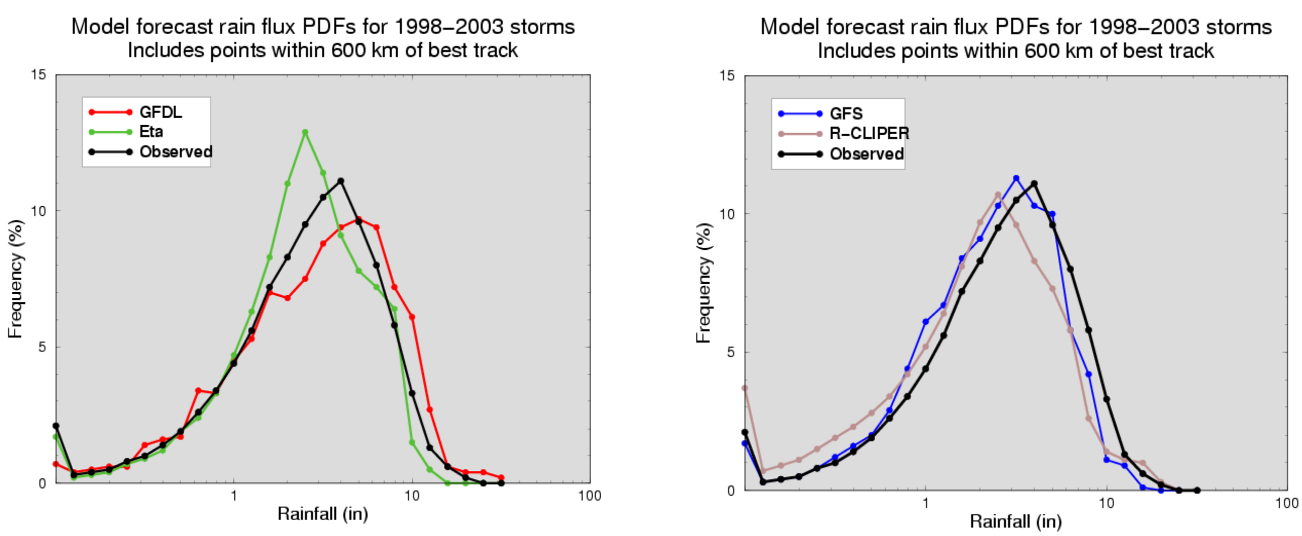


Figure 6. Plot of the PDF of rain flux (rain rate x grid box area) for each model and the Stage IV data for all U.S. landfalling storms from 1998-2003.

produces the best distribution of all, with a modal value close to the observed mode and a slight high bias for the light rain amounts and slight low bias for the heavy rain amounts.

Track-relative PDF analyses (Fig. 7) provide a means of comparing the models' abilities to reproduce features at different locations relative to the track of the storm, where features such as eyewall or stratiform rain would likely predominate. Figure 7a shows the PDF of rain flux for the GFDL, Eta, and observed rainfall fields within the 0-100 km band around the storm track, where eyewall (or eyewall remnants) would tend to predominate. The GFDL has a clear tendency to produce too much rain for the high to extreme rain amounts, while the Eta produces too much rain flux for the light to moderate rain amounts, suggesting that the GFDL may tend to overpredict eyewall rain while the Eta tends to underpredict eyewall rain. The GFS and R-CLIPER comparisons in this swath (Fig. 7b) show that the GFS produces a larger proportion of

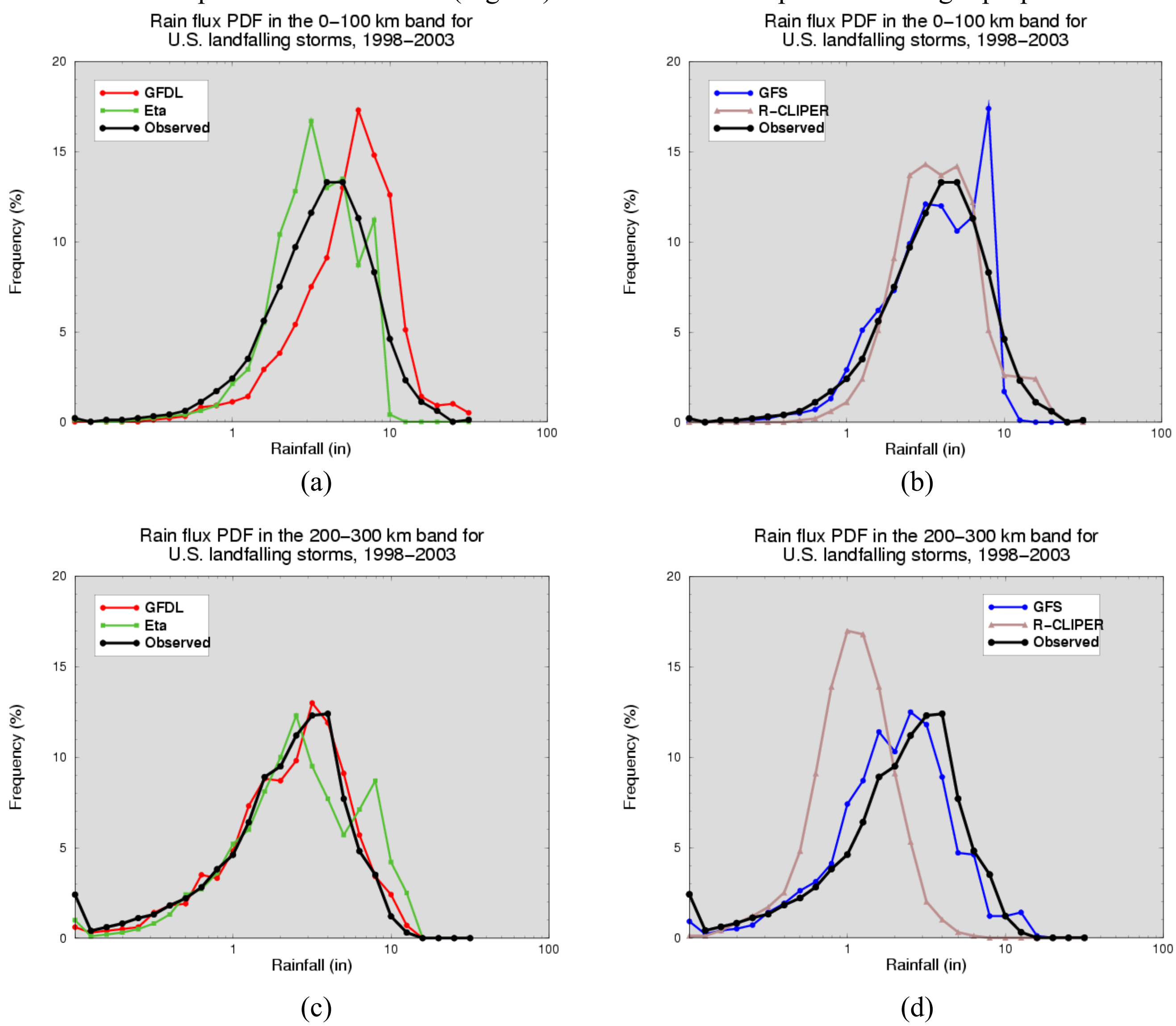


Figure 7. Rain flux PDFs in 100-km swaths relative to the tracks of the storms. (a) PDFs in the 0-100 km swath for the GFDL, Eta, and observed fields; (b) PDFs in the 0-100 km swath for the GFS, R-CLIPER, and observed fields; (c) PDFs in the 200-300 km swath for the GFDL, Eta, and observed fields; (d) PDFs in the 200-300 km swath for the GFS, R-CLIPER, and observed fields.

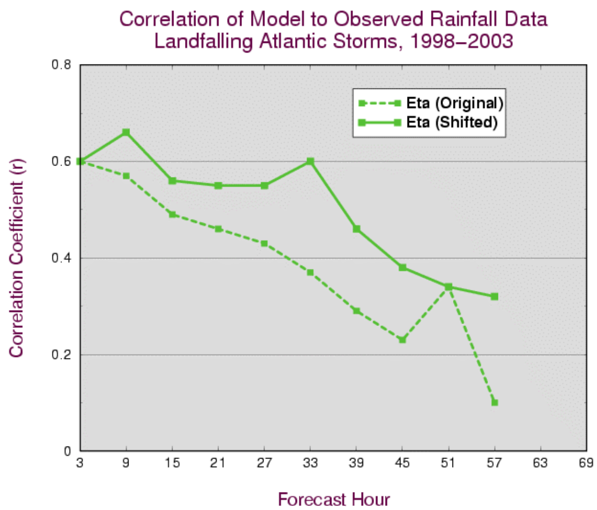
rain flux for moderate to heavy rain amounts (< 10 in), but the GFS is incapable of producing the extreme rain amounts (> 10-15 in). The R-CLIPER has the closest resemblance to the observed flux distributions, showing the ability to produce rain flux for light, moderate, heavy, and extreme rain amounts.

Figures 7c and d show the rain flux PDFs for the 200-300 km swath, where a mixture of rainband and stratiform rain would likely predominate. The flux PDF from the GFDL model (Fig. 7c) shows an excellent agreement with the observations at this distance, while the Eta model, while also showing an improvement over the 0-100 km swath, shows a two-peak structure in the frequency distribution at 2 and 9 inches, indicating that Eta produces too little rain flux for the moderate rain amounts and too much rain flux for the heavy amounts. The GFS (Fig. 7d) shows a slight tendency to overpredict rain flux for the light to moderate amounts and underpredict rain flux for the heavy rain amounts. The R-CLIPER is the most different from the 0-100 km results. Whereas the 0-100 km flux PDF agrees very well with the observations, the 200-300 km flux PDF is significantly skewed toward lighter rain rates. The R-CLIPER significantly overproduces rain flux for the light rain amounts and significantly underproduces rain flux for the moderate to heavy rain amounts.

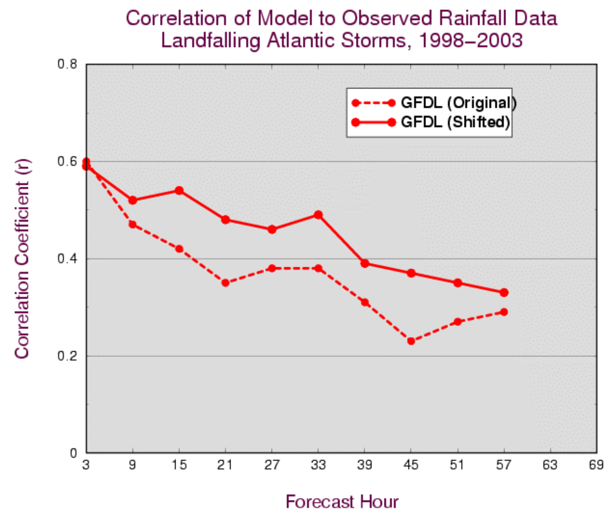
The impact of comparing storm-relative rainfall fields by shifting the rainfall fields an amount equal to the track error for a given model during a given time is shown in Fig. 8. This figure shows rainfall correlation patterns for different forecast hours for the various models, both for the original, unshifted rain fields and for the shifted, storm-relative rain fields. The amount of improvement realized by shifting the fields provides a qualitative assessment of the contribution of track error to the correlations. For the original, unshifted Eta fields (Fig. 8a) the correlations begin at 0.6 and fall to almost 0.2 by the 45-h forecast time. For the shifted Eta fields the correlations again begin at 0.6, but by 45 h the correlations are 0.4 – a doubling of the correlation at this time. The GFDL model (Fig. 8b) shows nearly comparable improvements in the correlations when the rain fields are shifted, while the GFS model (Fig. 8c) shows a much smaller improvement. This indicates that most of the error reflected in the GFS correlation fields is not due to track error, but rather is due to other aspects of the modeling system such as resolution and physics parameterization deficiencies.

4) <u>Development of a version of R-CLIPER that accounts for vertical shear</u>

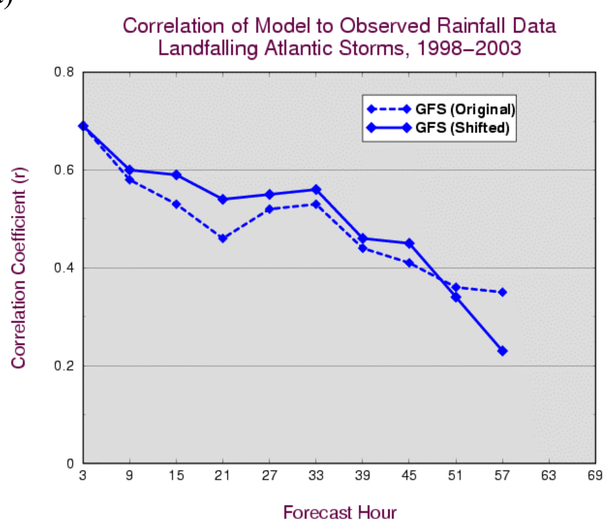
Another task that was proposed to be accomplished during the second year of this project was the development of a modified version of the R-CLIPER that would incorporate the impact of vertical shear on the rainfall distribution in landfalling tropical cyclones. The inclusion of this feature is based on many observational and modeling studies that have shown the development of azimuthal asymmetries in tropical cyclone rainfall as vertical shear increases (e.g., Black et al. 2002, Corbosiero and Molinari 2002, Rogers et al. 2003, Lonfat et al. 2004). The original proposal called for the quantification of azimuthal asymmetries as a function of vertical shear based on GFDL model output. However, recently work has been done (Lonfat et al. 2005) that has shown the ability to quantify these asymmetries based on thousands of overpasses of tropical cyclones over several years from the Tropical Rainfall Measuring Mission (TRMM) satellite. The strength of the azimuthal asymmetries as a function of vertical shear and storm intensity is shown in Fig. 9. As can be seen in this figure, a pronounced downshear left asymmetry is created when the vertical shear exceeds 5 m s^{-1} regardless of intensity. The shear values were obtained from the SHIPS database, which is the same source for the shear data that would be



(a)



(b)



(c)

Figure 8. Plots of correlation coefficients as a function of forecast time for (a) the Eta; (b) the GFDL; and (c) the GFS model. The dashed line in each figure represents the correlation coefficients from the original, unshifted rainfall fields, while the solid line represents the correlations based on rainfall fields shifted to reflect the track error in the model.

used for the R-CLIPER program. We are currently working to quantify the asymmetry as a function of shear for incorporation into the R-CLIPER program. We anticipate completing this task within the month and beginning test runs immediately thereafter. Once the technique is developed we will run this modified R-CLIPER program through the same validation techniques presented here to determine whether any skill has been added to the R-CLIPER.

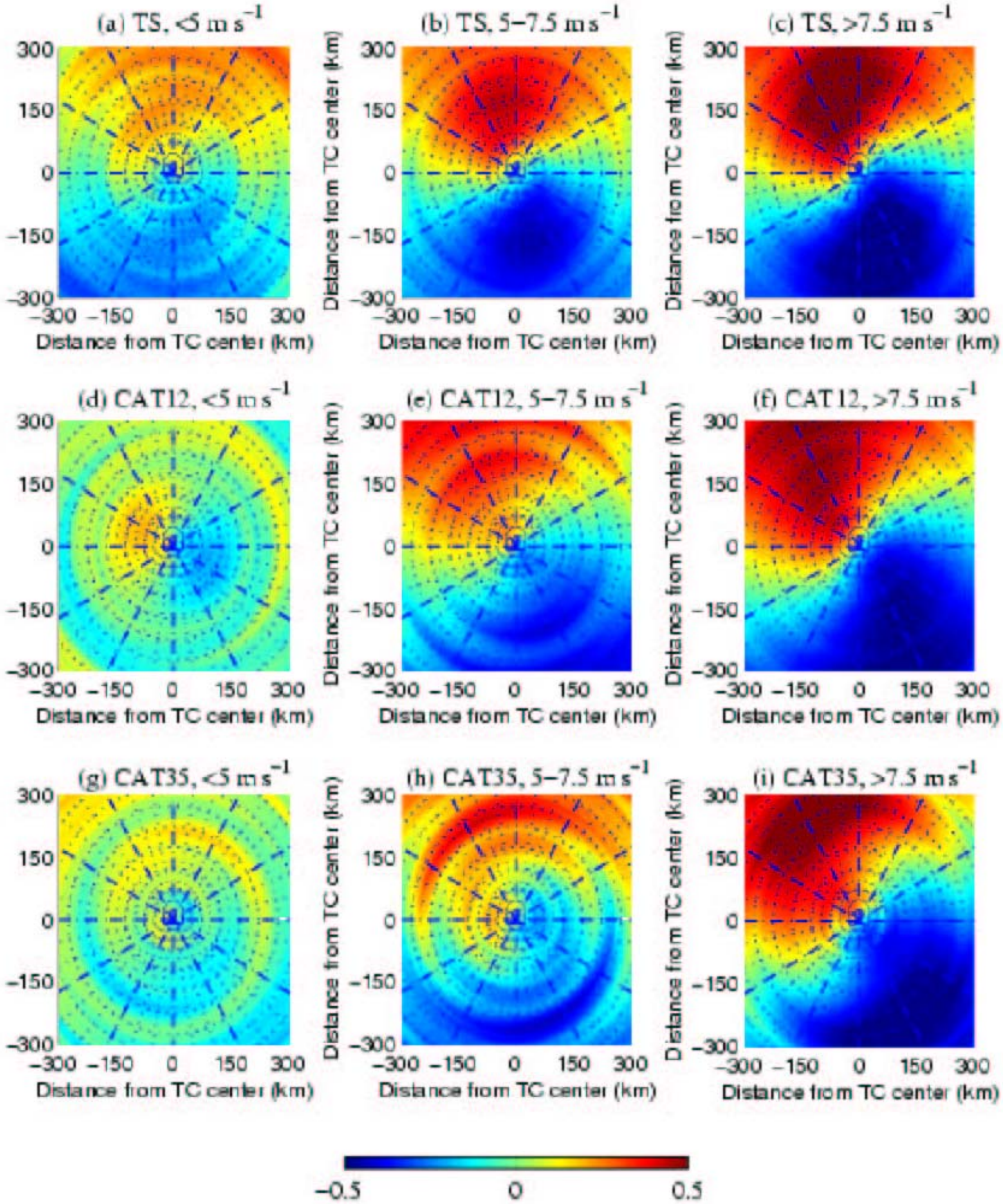


Figure 9. First-order rainfall asymmetry magnitude normalized by the azimuthally-averaged rainfall magnitude for weak shear (< 5 m s^{-1};a,d,e), moderate shear (5-7.5 m s^{-1}; b,e,h), and strong shear (> 7.5 m s^{-1}; c,f,i) for tropical storms, category 1-2 hurricanes, and category 3-5 hurricanes. Data taken from TRMM TMI rainfall estimates and TRMM PR data from 1998-2000. Taken from Lonfat et al. (2005).

Completion of second-year activities

The remainder of this year will see the completion of the project. Specifically, we will add the storms from 2004 into our validation database. In addition, another model, the GFDL model run with the NOAH Land-surface model, has been run and will be compared with the version of the GFDL we have used that runs with a standard slab model. Another task we will perform is to combine the metrics we have presented here into one (or more) unified rainfall validation parameters. This will facilitate comparisons among models in an efficient and effective manner. We will divide the metrics into two groups: one that is track-dependent (i.e., ETS and correlation) and one that is track-independent (i.e., bias score, rain flux PDF, track-relative flux PDF, storm-relative shifted-field PDF). Once this is completed we will provide our validation statistics and validation algorithms to NHC and EMC. The final task we will complete is the incorporation of the vertical shear into the R-CLIPER and the evaluation of this modification on the performance of the modified R-CLIPER. If skill is added, we will provide the code for adding vertical shear to the operational version of R-CLIPER to provide another product for predicting rainfall.

References

Black, M.L., J.F. Gamache, F.D. Marks, C.E. Samsury, and H.E. Willoughby, 2002: Eastern Pacific Hurricanes Jimena of 1991 and Olivia of 1994: The effect of vertical shear on structure and intensity. *Mon. Wea. Rev.*, **130**, 2291-2312.

Corbosiero, K.L., and J. Molinari, 2002: The effects of vertical wind shear on the distribution of convection in tropical cyclones. *Mon. Wea. Rev.*, **130**, 2110-2123.

Lonfat, M., F.D. Marks, Jr., and S.S. Chen, 2004: Precipitation distribution in tropical cyclones using the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager: A global perspective. *Mon. Wea. Rev.*, **132**, 1645-1660.

Lonfat, M., S.S. Chen, J.A. Knaff, and F.D. Marks, Jr., 2005: Effects of environmental vertical wind shear and storm motion on tropical cyclone rainfall asymmetries deduced from TRMM. Submitted to *J. Atmos. Sci.*

Rogers, R.F., S.S. Chen, J.E. Tenerelli, and H.E. Willoughby,2003: A numerical study of the impact of vertical shear on the distribution of rainfall in Hurricane Bonnie (1998). *Mon. Wea. Rev.*, **131**, 1577-1599.

Tuleya, R.E., M. DeMaria, and R. Kuligowski, 2005: Evaluation of GFDL model rainfall forecasts for U.S. landfalling tropical storms. Manuscript to be submitted to *Wea. And Forecasting.*

Tustison, B. D. Harris, and E. Foufoula-Georgiou, 2001: Scale issues in verification of precipitation forecasts. *J. Geophys. Res.*, **106**, 11,775-11,784.