NOAA Joint Hurricane Testbed (JHT)
Final Report

Date: January 6, 2014
Project title: *Improvements in Statistical Tropical Cyclone Forecast Models*
Principal Investigators: Renate Brummer, Mark DeMaria
Affiliation: Brummer (CIRA/CSU), DeMaria (NOAA/NESDIS)
Progress report dates: Aug 2011-Dec.2013

1. Long-Term Objectives and Specific Plans to Achieve Them

Although considerable effort is being made to improve dynamical tropical cyclone forecast models, statistical-dynamical models have generally provided the most accurate intensity predictions over the last few years. Four improvements to statistical-dynamical tropical cyclone forecast models were proposed. These included: (1) Improving the method to estimate the intensity growth rate in LGEM so the forecasts can be extended to seven days; (2) Developing special versions of SHIPS and LGEM for the Gulf of Mexico region; (3) Improving the databases used to develop SHIPS and LGEM through the use of the NCEP's new Climate Forecast System Reanalysis (CFSR); and (4) Developing an extended range climatology and persistence (CLIPER) model for track and intensity.

2. ACCOMPLISHMENTS

The accomplishments on the four main topics are summarized in Section 2a, including suggestions from the developers on which of the new capabilities might be transitioned to operations. Supporting details are provided in Section 2b.

a. Summary and Recommendations

(1) 7-Day LGEM

The persistence component was separated from the LGEM formulation so the model could be fitted to the full life cycle of each storm with a much smaller number of coefficients (separate coefficients are not needed for each forecast internal). Because the coefficients are not time dependent, the model can then be run to any length. Atlantic and East Pacific versions were developed from 2000-2011 data and then run out to 7 days on independent cases from 2012-2013 using operational input. The new version of the model is referred to as LGEM7. As shown below in Section b, the results were somewhat mixed. The model was generally well behaved and had skill relative to the new extended range baseline model at most forecast times out to 7 days in the east Pacific. In the Atlantic, LGEM7 only had skill at 7 days. The 2012-2013 sample sizes were fairly small at the extended ranges and the Atlantic sample was somewhat atypical, with most the intensification occurring at higher latitudes.

Because of the complexity of the new fitting technique for LGEM7, the test model only included the top 5 of the 20 LGEM predictors. The predictors related to baroclinic interactions, such product of product of shear with latitude, and the vortex in the GFS model were not included, which may explain the more negative results for the Atlantic version. Also, for days 1-5, LGEM7 had larger errors than LGEM, especially in the Atlantic.

Developer Suggestion: Defer transition decision. Redevelop LGEM7 with the 2012-2013 cases added, and test again during the 2014 season. Add some of the predictors from LGEM that were not included in the initial LGEM7 tests.

(2) The Gulf-specific version of LGEM.

The Gulf-specific version of LGEM uses the same formulation as LGEM7, but was developed from cases that were initially in the Gulf of Mexico region. The coefficients for the Gulf of Mexico version were different than those from the Atlantic basin version. In particular, the Oceanic Heat Content predictor was more important for the Gulf of Mexico. The Gulf of Mexico version was included in the LGEM7 runs for the 2012-2013 independent tests, but the sample of Gulf cases was too small to make a meaningful comparison.

Developer Suggestion: Defer transition decision. Redevelop the Gulf of Mexico version along with LGEM7 and test again in 2014. It should also be tested on the retrospective runs that are usually performed prior to the new SHIPS model upgrades. Those runs include cases from the previous 5 seasons, which would provide a more representative sample for evaluation.

(3) SHIPS and LGEM from the new Climate Forecast System Reanalysis

The new CFSR fields from 1979 to 2009 were obtained from NOAA/ESRL and combined with operational GFS analysis fields from 2010-2012. Versions of the Atlantic SHIPS and LGEM models were developed from a 1 deg version of the CFSR fields for comparison with the 2013 operational model that was developed from the old SHIPS database (2 deg operational GFS analysis fields from 2000-2012 and 2.5 deg NCEP reanalysis fields from 1982-1999). The results with the dependent data showed a considerable improvement in the variance of the intensity changes explained by the model. The CFSR versions of SHIPS and LGEM were then tested with real time input for cases from 2008-2012. Unfortunately, the tests with the operational input showed a statistically significant degradation for the both SHIPS and LGEM for the runs with the coefficients developed from the CFSR data.

Developer Suggestion: Do not implement SHIPS and LGEM developed from the CFSR data set. The CFSR data can be used to fill some holes in the SHIPS development data set.

(4) Extended range baseline models

Extended range baseline models were developed in the first year of the project and run operationally in 2012 and 2013 as part of the SHIPS model processing. The new baseline model uses a trajectory approach for track and intensity so it is referred to as Trajectory CLIPER (TCLP). Retrospective runs were also performed for a 10 year sample. For the 10 year sample, the mean track and intensity errors were within a few percent of those from the current baseline model OCD5 out to five days. Also, the TCLP errors provided just as good of a measure of annual forecast difficulty as OCD5, as indicated by correlations with the NHC Official track and intensity errors. The similarity of OCD5 and TCLP was confirmed in the real time runs during 2012 and 2013. The advantage of TCLP is that it can be run to any forecast length. The real time runs extend to 10 days.

Developer Suggestion: Accept TCLP as a new operational model. No additional action would be required since TCLP is already part of the operational SHIPS model.

b. Supporting Details

(1) 7-Day LGEM

The 7-Day LGEM was developed following the method outlined in the proposal. The primary unknown in LGEM is the growth rate. In the operational version, the growth rate at each 6 hr forecast time from 0 to 120 h is fitted using a least squares procedure. Some predictors are time dependent, such as those that come from the GFS forecast fields, and some are only available at the initial time, such as the GOES predictors and the growth rate during the12 hr before the initialization time. The 2013 operational LGEM model has 19 predictors, and the coefficients are different at each 6 hr time interval. Thus, there are 21*19 = 399 prediction coefficients. The length of the model run is limited by the maximum time period that the coefficients are available (currently 5 days). If the model was extended beyond 5 days using this same method, the sample sizes would be become very small at the longer ranges, since they would only include tropical cyclones that lasted that long.

An alternate version was developed for the extended range forecasts, where the growth rate is assumed to be a linear function of various predictors, but the coefficients are not time dependent. The coefficients are determined by minimizing the error between the forecasted and observed intensity over the full life cycle of each cyclone in the developmental data set. Once these coefficients are determined, the model can be run to any forecast time. Fitting the coefficients to minimize the prediction errors over the entire life cycle of each cyclone is more involved because the error measure is a nonlinear function of the coefficients. However, this problem can be solved using the method outlined in DeMaria (2009), MWR, which is also utilized in some data assimilation procedures. Basically, the model equations are appended to the error function as constraints with Lagrange multipliers. The integration of the adjoint of the LGEM model provides the gradient of the error function with respect to the prediction coefficients. Then, a steepest descent method can be used to find the coefficients that minimize the prediction error. This is an iterative procedure where the model equation are integrated forward in time to update the error function, the adjoint equations are integrated backwards to update the gradient, and then the coefficients are adjusted using the gradient descent. The iteration is repeated until convergence. Typically, about 2000 iterations were needed for convergence.

Because the minimization procedure is more involved than a simple least-squares method, only five predictors were included in the new LGEM model. These were chosen based on the analysis of the gradient from the adjoint model, which provides an estimate of the importance of the predictors on the forecasts. These included the vertical shear, the average of the 200 and 250 hPa temperature, the oceanic heat content, the difference between the translational motion direction and the shear direction, and a Gaussian function of the storm speed that enhances the sensitivity to very slow storm motion. The shear-motion angle difference is a simplification of the shear direction predictor in the operational LGEM and SHIPS models. Also, because the new CFSR fields were not found to be useful for the operational version of LGEM, only the operational analyses were used for the fitting. These were available back to 2000. However, because only 5 coefficients are determined instead of 399, it was not necessary to include as large of a development sample as the operational LGEM, which uses data back to 1982. The fitting was performed separately for the Atlantic and combined East/Central Pacific data using data from 2000-2011, leaving 2012 and 2013 available for independent testing.

Once the coefficients are determined, persistence is included through a separate minimization procedure, where the final growth rate is a weighted average of the growth rate from the 12 h period up to the initialization time, and the value from the fit to the full life cycle of each storm. The weight on the persistence growth rate decays exponentially with the forecast time, where the e-folding time was chosen to minimize the forecast error. The e-folding time was found to be 12 h for the Atlantic and 14 h for the east/central Pacific. The new version of LGEM is referred to as LGEM7.

Figure 1 shows the normalized LGEM7 growth rate coefficients for the Atlantic and East/Central Pacific sample. A negative value means that predictor reduces the growth rate as that predictor becomes larger. The shear is the dominant predictor, with the OHC, shear-motion angle and storm speed predictors of secondary importance. The upper level temperature has the least influence on the prediction.

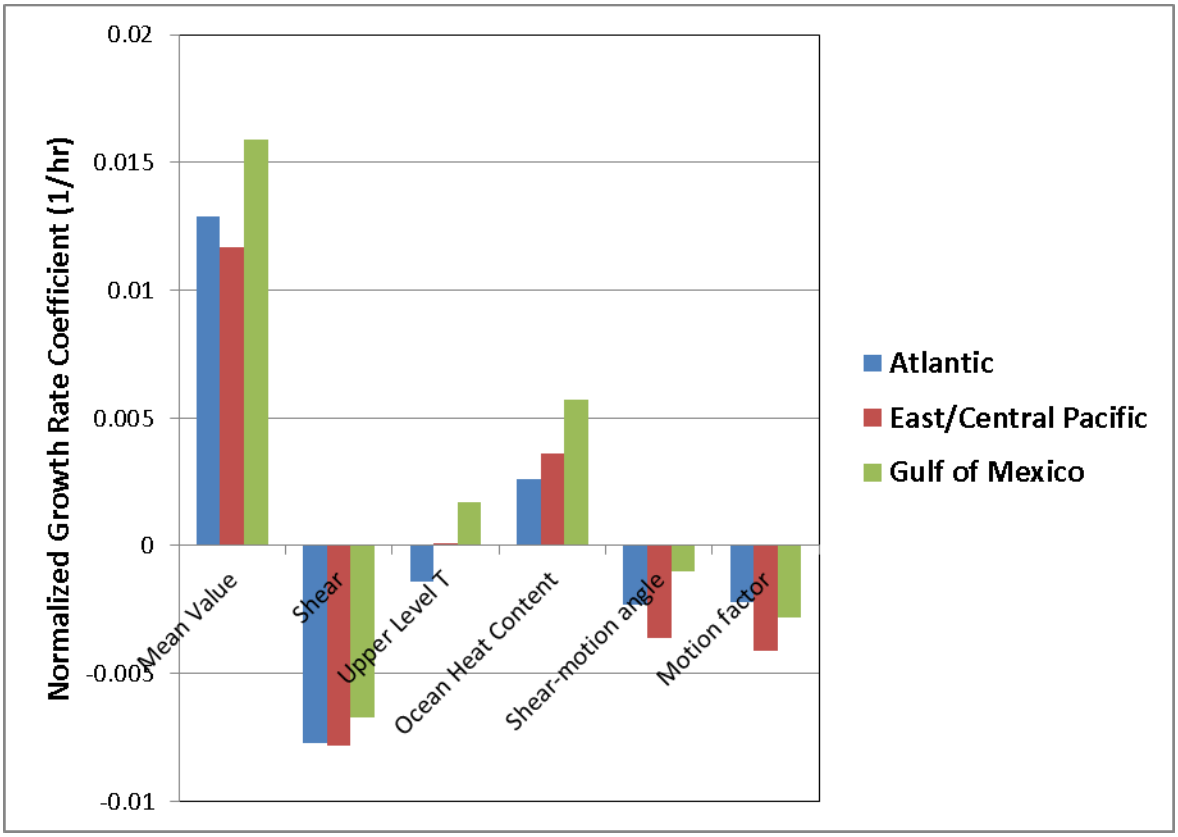

Figure 1. Normalized coefficients for the LGEM7 growth rate for the three versions of the model.

Although the LGEM7 was not ready in time for real time tests, it was re-run for all Atlantic and east Pacific cases from 2012 and 2013 using only information that was available in real time. The track forecasts are from the NHC experimental 7 day predictions. This represents a fully independent test since operational data was used as input and the developmental sample only included 2000-2011.

Figure 2 shows the errors of LGEM7 and the TCLP model and Fig. 3 shows the percent improvement of LGEM7 over TCLP (skill). LGEM7 had smaller errors than TCLP in both basins at 7 days (positive skill), although the number of cases with a 7 day verification was fairly small (85 in for the Atlantic and 24 for the east Pacific). However, in the Atlantic, the LGEM7 errors were larger than those of TCLP at most other forecast times. In the East Pacific, the LGEM7 errors were smaller than TCLP at most forecast times. These results suggest that LGEM7 potentially has some long range forecast skill, but the sample sizes were too small to make a firm conclusion. Additional testing on independent cases is needed to resolve this issue.

The LGEM7 errors were also compared to the LGEM errors out to 5 days as shown in Fig. 4. For the east Pacific, the LGEM7 errors 8% or less larger than the LGEM errors. However, for the Atlantic, the LGEM7 errors were up to about 22% larger than those from LGEM at the longer ranges. This suggests that the 5 coefficient version of LGEM7 is not capturing all of the predictive information in the operational LGEM. Many of the factors that were not included were related to high latitude cyclones such as the shear times latitude and the average tangential wind from the GFS model forecast. That may explain why the degradation was worse for the Atlantic than the East Pacific. The number of predictors in LGEM7 was restricted to gain experience with the more complex fitting procedure. The results in Fig. 4 suggest that additional work should be done to add back some of the LGEM predictors not currently included in LGEM7.

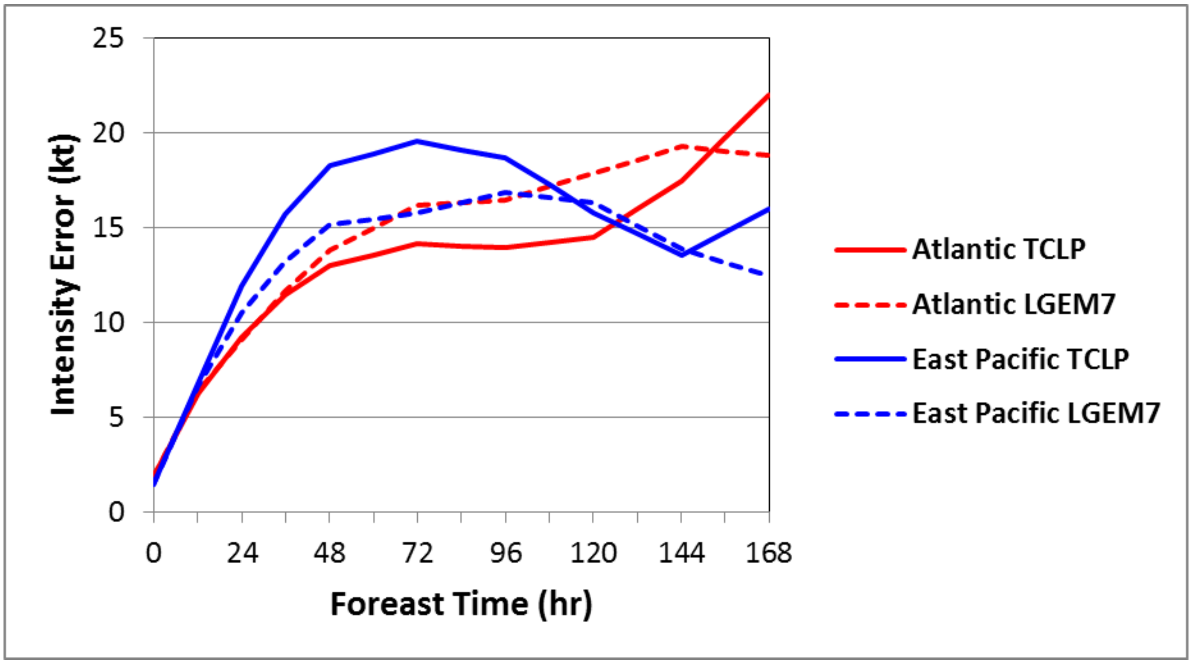

Figure 2. The average intensity errors from the extended range LGEM7 model and the trajectory CLIPER model (TCLP) for the 2012-2013 Atlantic and East Pacific cases. The Atlantic (East Pacific) sample includes 561 (555) cases with a 12 h verification and 85 (24) cases with a 168 h verification.

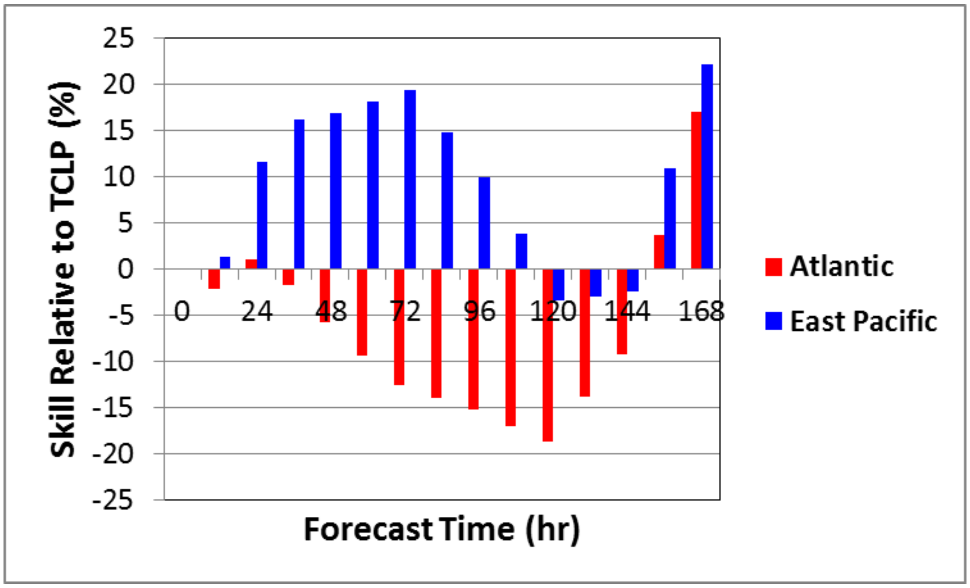

Figure 3. The skill of LGEM7 relative to TCLP.

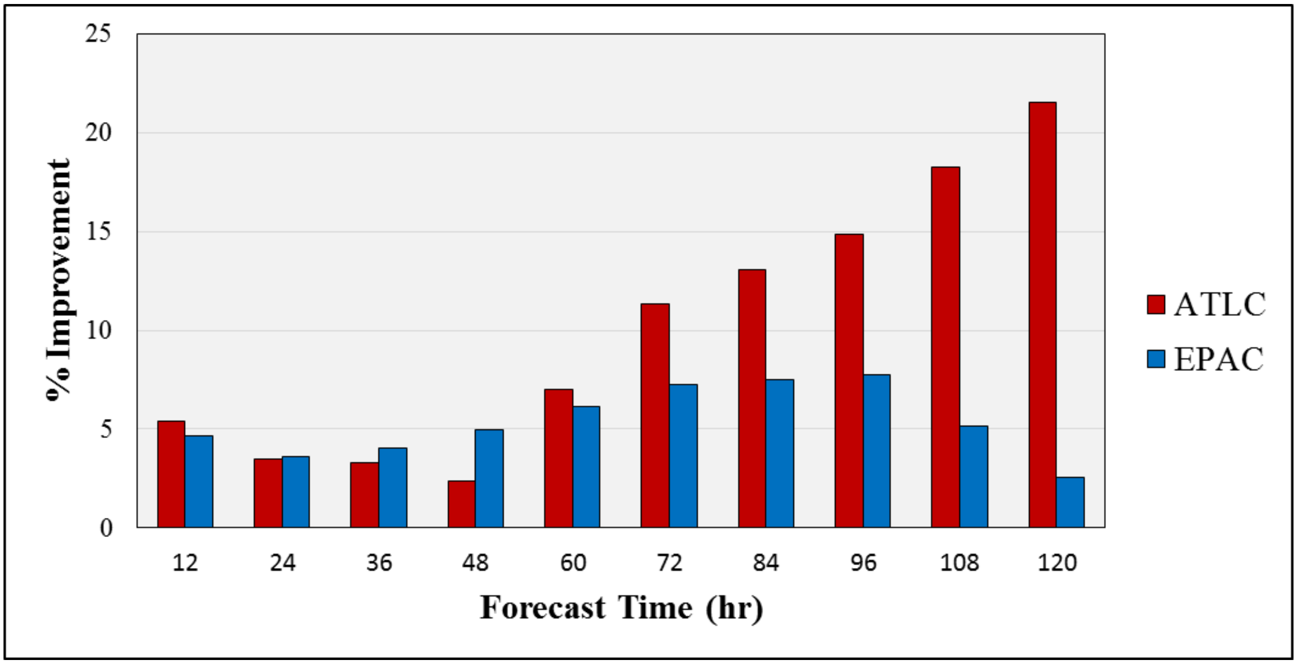

Figure 4. The improvement (percent error reduction) of the operational LGEM model compared with the experimental LGEM7 model for the 2012-2013 Atlantic and east Pacific forecasts. Positive improvement indicates that LGEM was better than LGEM7.

(2) The Gulf-specific version of LGEM.

As expected, the Gulf sample from 2000-2012 was significantly smaller than the total Atlantic sample. Because of the small sample size, especially at the longer ranges, the operational SHIPS and LGEM models were not fit to the Gulf data. Instead, the LGEM7 model was used, since it combines all forecast times together in the fit, and so can provide stable estimates of the coefficients with a smaller developmental data set. Figure 1 shows the LGEM7 coefficients for the Gulf model. The values are somewhat different than for the total Atlantic sample. In particular, the OHC coefficient is larger for the Gulf model.

The Gulf version of LGEM7 was run for all the Gulf TCs (defined as cases north of 17.5^{o}N and west of 81^{o}W) in the 2012-2013 independent data test. The mean errors from the Gulf version were within a few percent of those from the Atlantic basin version, and the sample size was not very large (only 22 cases with at least a 72 h verification). Also, the 2012-2013 cases, with mostly tropical storms and only two short-lived category one hurricanes, were not very representative of the longer term Gulf sample. Thus, a larger data sample will be needed to evaluate the Gulf version of LGEM7.

(3) SHIPS and LGEM from the new Climate Forecast System Reanalysis (CFSR)

The details of the evaluation of the SHIPS and LGEM models developed from the CFSR were described in the previous progress report for this project (available from http://www.nhc.noaa.gov/jht/current_projects.php). Basically, the dependent results showed significant improvements in the variance explained from the model fits with the CSFR data base compared to the operational model database. The model development uses the perfect prog approach, where best tracks and analysis fields are used for the fit. However, these results did not carry over when the models were run with operational input, which includes forecast tracks and model fields.

(4) Extended range baseline models

The new baseline models were developed using a trajectory approach as described in the original proposal. For track, the forecast is determined by integrating $dx/dt = c_x$ and $dy/dt = c_y$ where c_x and c_y come from the climatological storm motion fields modified by a beta-drift correction and persistence. For intensity, the prediction equation for LGEM is used where the growth rate and maximum potential intensity come from climatological values. The growth rate is modified to include a persistence factor. Because the model uses climatological and persistence input, it is referred to as the trajectory CLIPER model (TCLP). Because of the trajectory approach, TCLP can be run to any forecast length.

The primary purposes of the baseline models are to provide a measure of inter-annual variability in forecast difficulty and to use as method for evaluating forecast skill of more general models. A model should have errors smaller than those from the baseline to have skill. NHC currently uses the OCD5 model to evaluate forecast skill. OCD5 uses the traditional CLIPER model for track, and the decay version of the SHIFOR model, where track comes from the CLIPER forecast.

The TCLP model was run in real time in 2012 and 2013 as part of the operational SHIPS model on the NCEP super computer. Figure 5 shows the track and intensity errors for the

NHC Official, OCD5 and TCLP forecasts. The track errors from OCD5 and TCLP are nearly identical out to 5 days. This provides confidence that the TCLP model can be used as a baseline beyond 5 days. The OCD5 and TCLP intensity errors show greater differences out to 5 days, although the errors are still less than about 10%. A comparison with a 10 year sample of retrospective runs (described below) showed that the intensity errors from TCLP ranged from 8% larger to 5% smaller then those from OCD5 at all forecast times out to 5 days.

The procedure for evaluating the utility of TCLP as a measure of forecast difficulty was described in detail in the presentation by DeMaria et al. at the 2013 Inter-Departmental Hurricane Conference (available from the http://www.ofcm.gov/ihc13/67IHC-Linking-File.htm). The 10 year sample of retrospective runs was used to correlate the annual TCLP errors with those from the NHC official forecasts. Results showed that TCLP-NHC Official error correlations were comparable to those from OCD5 for intensity and better than those of OCD5 for track.

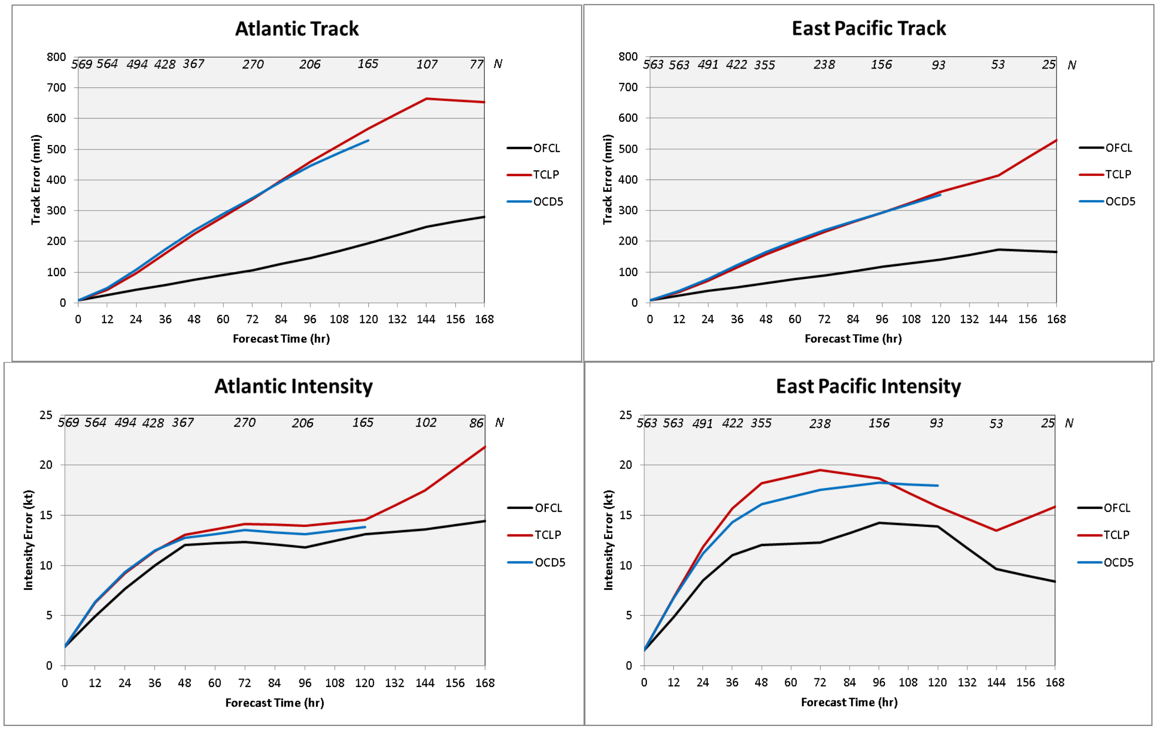

Figure 5. The average track and intensity errors for the 2012-2013 real time Atlantic and east Pacific NHC Official and baseline OCD5 and TCLP forecasts. The NHC and TCLP forecasts extend to 7 days. The sample sizes are shown along the top of each plot. All samples are homogeneous and use NHC's usual verification rules.

3. FOLLOW UP PLANS

The LGEM7 model (Atlantic, East/Central Pacific, and Gulf of Mexico-specific versions) has already been added to a parallel version of the operational SHIPS processing script. The LGEM7 will be included in the operational runs during the 2014 season for additional testing. Depending on the NHC recommendation, the forecasts from this new model might be added to the ATCF, or could just be saved on the WCOSS system for later evaluation without making them available to the hurricane specialists. The new baseline TCLP forecasts have put into the ATCF in real time since 2012, so that will continue unless NHC decides not to implement that model. The small degradation in the TCLP intensity forecasts relative to OCD5 at some forecast times will be re-evaluated following the 2014 season. Some minor retuning of TCLP might be needed to address that issue.