9C.5 TROPICAL CYCLONE INTENSITY CHANGE PREDICTABILITY ESTIMATES USING A STATISTICAL-DYNAMICAL MODEL

Mark DeMaria*
NOAA/NESDIS/StAR, Fort Collins, CO

1. INTRODUCTION

The recently established NOAA Hurricane Forecast Improvement Project (HFIP) has set a 5-year goal to reduce track and intensity forecast errors by 20%, and a 10-year goal for a 50% reduction. The 10-year goal corresponds to an average error reduction of 5% per year. Figure 1 shows the annual average 48 h Atlantic track and intensity errors of the National Hurricane Center (NHC) official forecasts from 1985 to 2009 and the corresponding linear trend lines. The track forecasts have been decreasing at an average rate of about 3.7% per year, so the 5% per year improvement proposed by HFIP represents an acceleration of the error reduction rate. However, the intensity errors have only decreased at about 0.6% per year. Thus, the HFIP goals represent an order of magnitude increase in the improvement rate.

In the 1970's and 80's the most accurate numerical track forecast models were statistical-dynamical (Neumann 1987, here after N87). This guidance included the NHC73 and NHC83 series of models, where the track was predicted statistically using input from climatology, persistence and the storm environment determined from a global atmospheric model. N87 also indicated that the reductions in the track forecast errors at that time were only about 0.5% per year and appeared to be leveling off.

N87 developed an interesting methodology to estimate how much improvement could be achieved in track forecasts from statistical models. For that purpose, the most accurate track forecast model at the time (NHC83) was run in a "perfect prog" mode, where verifying analyses replaced the forecast fields used to estimate the storm environment predictors and best track input replaced the operational estimates of the initial storm position and motion vector. Figure 2 shows the average NHC track errors from 1976-1985 (the baseline period used by N87) and the estimate of the how much improvement was possible. Also shown are the 5-year averages of the NHC track errors from 1986 through 2009. This figure shows that the roughly 50% improvements in track errors estimated by N87 were achieved at most forecast times by the late 1990s and surpassed in the 2000s. These improvements were not, however, achieved by better statistical-dynamical models, but were primarily due to better dynamical models starting with the operational implementation of the GFDL hurricane model in 1995 (Rappaport et al., 2009).

* *Corresponding author address*: Mark DeMaria, NOAA/NESDIS/STAR, CIRA/Colorado State University, 1375 Campus Delivery, Fort Collins, CO 80523-1375; e-mail: Mark. DeMaria@noaa.gov

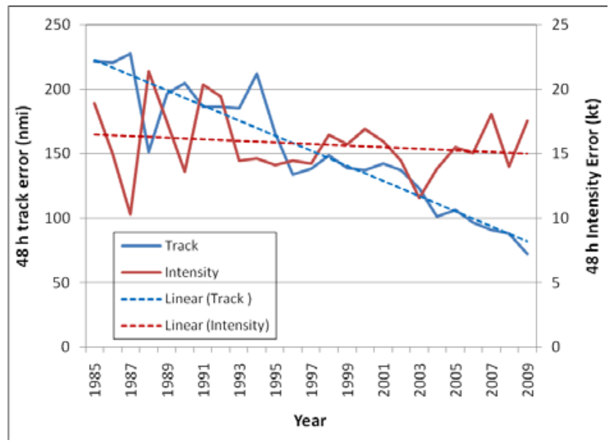

Figure 1. Time series of the annual average NHC official 48 h track and intensity forecast errors since 1985.The linear trend lines of each are also shown.

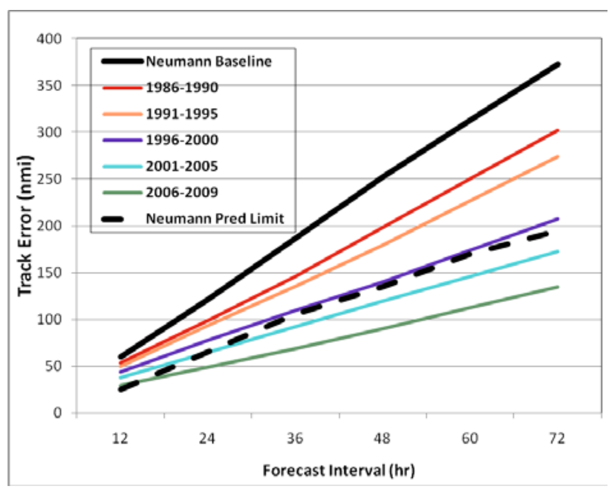

Figure 2. The average NHC track forecast errors for the 197-1985 baseline period and the statistical track model predictability limit estimated by Neumann (1987). Also shown are the 5-year average NHC track errors from 1986-2009.

The current situation for operational intensity forecasting is similar to that for track forecasting in the 1980s. The most accurate intensity prediction models are statistical-dynamical, and the rate of improvement is very small (Franklin 2010). To provide an estimate of how much intensity forecasts can be improved, the statistical-dynamical Logistic Growth Equation Model

(LGEM) is used in a “perfect prog” mode, similar to the N87 study with the statistical-dynamical NHC83 track model. LGEM (DeMaria 2009) has been operational since 2006 for the Atlantic, and eastern and central North Pacific, and uses a subset of the input for the Statistical Hurricane Intensity Prediction Scheme (SHIPS), which has been operational since 1991 (DeMaria et al., 2005). In terms of seasonal mean absolute error, LGEM was the most accurate Atlantic intensity forecast model in 2008 and 2009 (Franklin 2010). Recent changes to LGEM and SHIPS are described in section 3, and the intensity predictability results are presented in section 4.

3. THE SHIPS AND LGEM INTENSITY MODELS

SHIPS is a statistical-dynamical model that uses a multiple regression technique to predict changes in the maximum sustained surface winds at 6 hr intervals out to 120 h. Predictors include information from climatology, persistence, the atmosphere and the ocean. The atmospheric predictors are derived from forecasts of the NCEP global forecasting system (GFS) and GOES infrared imagery. The oceanic predictors are determined from Reynolds sea surface temperature analyses and oceanic heat content estimated from satellite altimetry measurements (Mainelli et al., 2009).

DeMaria et al. (2005) provides a detailed description of the predictors used in SHIPS through 2003. The predictors consist of two basic types, static, which are evaluated only at the beginning of the forecast period, and time-dependent, which are estimated at each forecast period relative to the NHC official forecast track. The regression coefficients for SHIPS are updated at the start of every hurricane season using a dependent sample from 1982 through the previous year. Major changes to SHIPS since 2004 are described below.

In 2004, two static predictors from GOES imagery and one time-dependent predictor, the oceanic heat content (OHC) from satellite altimetry along the storm track, were added. Because the developmental sample with satellite data was much smaller than the total sample, a correction step was applied to the SHIPS forecast without the satellite input. The OHC input was only available for the Atlantic version.

When the forecast track crosses land, the SHIPS forecast is adjusted using a climatological inland decay rate. In 2005, the decay rate was modified so that it is slower for storms where part of the circulation is still over water (DeMaria et al., 2006). This change removed a low bias in the SHIPS prediction. In 2006, the 250 hPa temperature was added as a predictor to account for cases where 200 hPa is above the tropopause and the 200 hPa temperature already included is not representative of a cold upper troposphere.

In 2007 a method for removing the storm circulation from the GFS forecast was added. This allowed the averaging area used to calculate the 850 to 200 hPa vertical shear to be reduced to circle of 500 km radius. Previously a 200 to 800 km annulus was used. As part of the vortex removal process, the time tendency of the 850 hPa tangential wind averaged from 0 to 600 km was added as predictor. For example for the 48 h SHIPS forecast, the difference between this parameter at 48 h and 0 h is used. With this predictor, the GFS model evolution of the storm circulation influences the SHIPS intensity forecast.

No changes were made to the predictors in 2008. In 2009, the GOES development dataset was expanded to include most cases back to 1983 and methods to fill in missing OHC and GOES data were developed. The OHC proxy is determined from a monthly OHC climatology modified by the current SST and the GOES proxy is a linear combination of other predictors. This allowed the satellite predictors to be included with all of the other input so the correction step was eliminated. Also in 2009, the OHC input became available for the east Pacific SHIPS.

Another change in 2009 was a new predictor related to the direction of the 850 to 200 hPa shear vector. The optimal direction of the shear vector was found to be a function of latitude so the shear direction predictor (SD) is given by

$$SD = f(\theta)\,|\beta-\beta_o| \qquad (1)$$

where θ is latitude, β is the direction of the shear vector, β_o is the optimal shear direction and $f(\theta)$ is the latitudinal scale factor given by

$$f(\theta) = a \qquad \theta \le \theta_1 \qquad (2a)$$

$$f(\theta) = a + 3(b-a)y^2 - 2(b-a)y^3 \qquad \theta_1 \le \theta \le \theta_2 \qquad (2b)$$

$$f(\theta) = b \qquad \theta_2 \le \theta \qquad (2c)$$

where a, b, θ_1 , and θ_2 are constants and y is the scaled latitude given by

$$y = (\theta - \theta_1)/(\theta_2 - \theta_1) . \qquad (3)$$

The latitudinal scaling factor *f* is equals the constant a for latitudes below θ_1, equals b for latitudes above θ_2 , and is a cubic polynomial between those two latitudes. The cubic polynomial was chosen to satisfy continuity across θ_1 and θ_2 and has a zero derivation at those two latitudes. For the Atlantic, the constants that maximize the variance explained of the observed intensity changes were found to be a=10^o, b=40^o, β_o =55^o, a=1.0, b=-0.75. At low latitudes the optimal shear direction is from the northeast. However, since b is negative, the optimal shear is from the southwest at high latitudes. This result suggests that the preferential locations for intensification are to the southeast of the center an upper-level ridge at low latitudes or southeast of an upper level trough at high latitudes. For the east Pacific, the constants were found to be a=10^o, b=30^o, β_o =70^o, a=1, b=-1.0, so SD has a qualitatively similar behavior to that in the Atlantic.

The vertical shear information in SHIPS is derived from the 850 and 200 hPa levels. However, the environmental wind can vary in the vertical in more complicated ways than can be captured by just two vertical levels. For this reason, a generalized shear parameter (GS) was developed that includes information from multiple levels. The generalized shear is given by

$$GS = 4/(P_2-P_1)\int_{P_1}^{P_2} [(u-u_b)^2 + (v-v_b)^2]^{1/2}\, dP \qquad (4)$$

where P is pressure, u and v are the horizontal wind components averaged over the horizontal circle with radius of 500 km, and u_b and v_b are the u and v values vertically averaged over the pressure layer from P_1 to P_2. The factor of 4 is included in (4) so that GS has the same value as the standard 850 to 200 hPa shear parameter for the special case where u and v vary linearly at the same rate as a function of pressure from P_1 to P_2. The integration in (4) is from 100 to 1000 hPa for use in SHIPS.

The GS has a fairly high correlation with the standard 850 to 200 hPa shear. To remove that effect, the GS and standard shear values for the developmental sample were fitted with a quadratic polynomial to provide the climatological relationship between the two. The new predictor is the deviation of GS from this standard relationship. Positive values of the GS deviation indicate that environmental wind profile has a nonlinear structure so that levels besides 850 and 200 hPa are contributing to the shear. As expected, the deviation GS parameter is negatively correlated with intensity changes. The generalized shear will likely be included in the 2010 version of SHIPS.

As described above, LGEM became operational in 2006 and was developed to relax some of the linear constraints of the SHIPS model. LGEM is governed by a first-order differential equation (the logistic growth equation) for the maximum sustained wind. This equation constrains the solution to lie between zero and the maximum potential intensity. The primary free parameter in LGEM is the growth rate, which is a function of a subset of the predictors used in SHIPS. When the track crosses land, the growth rate is replaced with same climatological decay rate used in SHIPS.

As described by DeMaria (2009), two methods were developed to estimate the growth rate. The first determines the growth rate from the best track data for the developmental sample, and fits that to the SHIPS predictors using a multiple regression method. A shortcoming of that method is that error in the growth rate is being minimized, rather than the error of the predicted maximum wind. An adjoint method was developed to determine the growth rate relationship that directly minimizes the predicted maximum wind, but that version has not been made operational.

The multiple regression version of LGEM has been run operationally since 2006 with the same input as for SHIPS, except that the satellite information was not included due to the smaller sample size. The satellite input was added to LGEM beginning in 2009, when the larger GOES data base and the proxy OHC and GOES variables became available when that input was missing. All the other new predictors described above for SHIPS such as the shear direction parameter and generalized shear were also added to LGEM. The 2010 version of LGEM is used in the predictability estimates described below.

4. INTENSITY CHANGE PREDICTABILITY

The starting point for the predictability estimation is the 2010 version of LGEM fitted to the 1982-2009 developmental sample. To determine the actual performance of LGEM with the most recent version it would be necessary to run the model with the same input that is available in real time. That includes the NHC forecast track and GFS forecast fields, rather than best track positions and GFS analysis fields. However, an archive of GFS 5-day forecasts is only available back to 2002. Therefore, the baseline period for this study is 2002-2009, which includes 135 Atlantic tropical cyclones. The LGEM coefficients used in this study are not independent, since the baseline period is a subset of the developmental sample. However, the developmental sample is sufficiently large and the forecast tracks and GFS fields are significantly different than the best track and GFS analysis fields, so that the results are very similar to what would be obtained from a version of LGEM developed from completely independent input.

To estimate the predictability limits of intensity forecasts, a procedure similar to that described in N87 is used, but with some intermediate steps. For this purpose, LGEM was run with four types of input: (1) NHC forecast tracks and GFS forecast fields; (2) NHC forecast tracks and GFS analysis fields; (3) Best track positions and GFS forecast fields; (4) Best track positions and GFS analysis fields. The fourth version is considered to be a lower bound on accuracy that is possible from a statistical-dynamical model. Versions (2) and (3) estimate the potential intensity error reductions from improving the prediction of the large-scale environment, and forecast track, respectively.

Figure 3 and Table 1 show the average forecast errors for the 2002-2009 sample for the four versions of LGEM, and the NHC official forecast. The forecasts were verified against the NHC best track intensities using the standard NHC verification criteria (homogeneous sample containing only those cyclones classified as tropical or subtropical). This figure shows that the NHC official forecast errors are smaller than the best version of LGEM at 12 to 36 h. Thus, there is little room for improvement in these short range forecasts. At the later times, the NHC official forecasts were larger than all the versions of LGEM, indicating that those can be improved by better statistical-dynamical models.

The errors for versions 3 and 4 of LGEM level off and actually decline a little for the longer range forecasts. This artifact is probably due to the reduction in sample sizes for the longer range forecasts. Generally speaking, without error growth in the track or large scale environmental predictors, LGEM has forecast errors of about 13 kt once the information from persistence and other static predictors has a lesser influence on the forecast.

Figure 4 shows the percent improvement of versions 2-4 of LGEM relative to the baseline model (version 1). This figure shows that most of the potential for improvement comes from reducing the track error (versions 3 and 4). Reducing the errors in the prediction of the large scale environment has only a small impact (version 2). Comparing the improvements for version 2 with the difference between versions 3 and 4 shows that there is a nonlinear interaction between the track and large scale environment errors, especially at the longer ranges. Replacing the GFS forecast fields with the verifying analyses but using the NHC forecast tracks actually increases the errors at the longer times. However, when the best tracks are used, replacing the GFS forecast fields with analysis fields improves the LGEM forecasts, especially at the longer times. This is probably due to the fact that the position of the storm where the shear and other predictors are calculated is more consistent with the GFS storm position when best tracks are used in combination with verifying analyses.

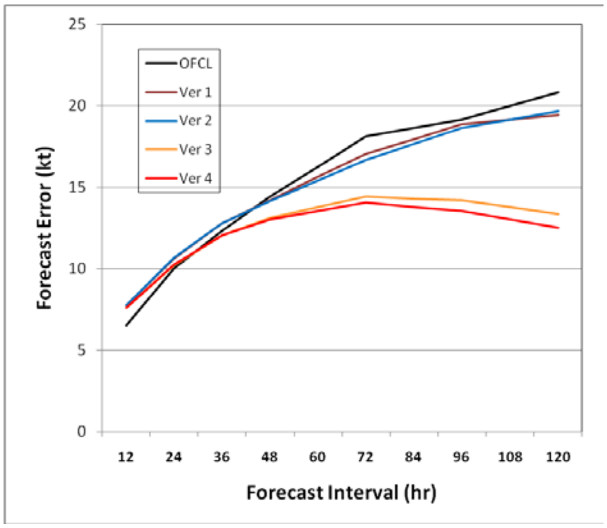

Figure 3. The average intensity forecast errors for the four versions of LGEM with varying input for the 2002-2009 sample. The NHC official intensity forecast errors are also shown.

Table 1. Average intensity errors (kt) to 120 hr from four versions of LGEM with various inputs for the 2002-2009 sample. The errors from the NHC official forecast (OFCL) and the sample size (N) at each forecast time are also shown.

Time	OFCL	Ver 1	Ver 2	Ver 3	Ver 4	N
12	6.5	7.8	7.8	7.6	7.6	2402
24	10.1	10.7	10.7	10.2	10.3	2159
36	12.3	12.8	12.8	12.0	12.1	1923
48	14.4	14.2	14.2	13.1	13.0	1709
72	18.1	17.1	16.7	14.4	14.1	1373
96	19.1	18.9	18.6	14.2	13.5	1076
120	20.8	19.4	19.7	13.3	12.5	859

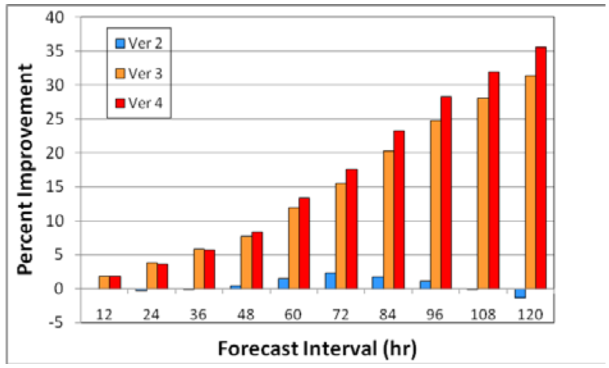

Figure 4. The percent improvement of versions 2-4 of LGEM relative to the baseline (version 1).

Returning to the initial discussion of the HFIP intensity forecast goals, the results in Fig. 4 show that these are going to very difficult to achieve for the shorter range forecasts, unless dynamical prediction systems can surpass statistical-dynamical intensity models, analogous to track forecasts in the 1990s. However, for the longer range forecast periods, at least the 5-year goal of 20% improvement might be achievable through statistical-dynamical models with improved track forecast input .

5. CONCLUDING REMARKS

This paper reviewed recent changes to the operational SHIPS and LGEM statistical-dynamical intensity forecast models and then used LGEM to estimate how much room there is for improvement. The predictability framework developed by Neumann (1987) for statistical-dynamical track models was used, where the model was run with “perfect prog” input. Results showed that there is much more room for improvement in the longer range forecasts (up to 35% at 120 h) than in the shorter ranges (about 8% at 48 h). Most of this improvement can be realized just by reducing track forecast errors, with a smaller contribution from reducing errors in the prediction of the large-scale environment.

It is extremely unlikely that the track errors can be reduced to zero. However, if the HFIP goal of a 50% reduction in track error is achieved, this would lead to a roughly 4 to 17% improvement in statistical-dynamical intensity errors for the 2 to 5 day forecasts, assuming a linear relationship between track and intensity errors.

There are other methods that have potential to improve the short range intensity forecasts beyond what is presented in this study. Additional information from lightning activity and satellite based total precipitable water estimates is showing promise for improving the SHIPS model (Knaff et al. 2010). Other data not utilized include microwave imagery, which is available from multiple satellites, and aircraft in situ data. In addition, the adjoint formulation of LGEM allows the complete intensification history of a storm up to the forecast time to influence the prediction, rather than just the previous 12 h intensity change as in the current version of SHIPS and LGEM. This might also provide some short range error reduction. Finally, using ensemble and model consensus approaches are showing promise for improving intensity forecasts (Sampson et al. 2008). As part of HFIP, methods to couple LGEM with a global model ensemble forecast system are being developed.

ACKNOWLEDGMENTS

The research was partially supported by NOAA Hurricane Forecast Improvement Project (HFIP). Many of the improvements in the SHIPS and LGEM models were supported by the Joint Hurricane Testbed. The views, opinions, and findings in this report are those of the authors, and should not be construed as an official NOAA and or U.S. Government position, policy, or decision.

REFERENCES

DeMaria, M., 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Mon. Wea. Rev*., **137**, 68-82.

_____, J.A. Knaff, and J. Kaplan, 2006: On the decay of tropical cyclone winds crossing narrow landmasses, *J. Appl. Meteor*., **45**, 491-499.

_____, M. Mainelli, L.K. Shay, J.A. Knaff, and J. Kaplan, 2005: Further Improvements in the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20**, 531-543.

Franklin, J.L., 2010: 2009 National Hurricane Center Forecast Verification Report. <Available from http://www.nhc.noaa.gov/verification/pdfs/Verification_2009.pdf>

Knaff, J.A., M. DeMaria, J. Kaplan, J. Dunion and R. DeMaria, 2010: Assessing the impact of total precipitable water and lightning on SHIPS forecasts. AMS 29th Conf. of Hurr. And Trop. Meteor., May 10-14, 2010, Tucson, AZ. <Available from http://ams.confex.com/ams/29Hurricanes/techprogram/paper_167919.htm>

Mainelli, M., M. DeMaria, L.K. Shay and G. Goni, 2008: Application of oceanic heat content estimation to operational forecasting of recent Atlantic category 5 hurricanes. *Wea. Forecasting*, **23**, 3-16.

Neumann, C.J., 1987: Prediction of tropical cyclone motion: Some practical aspects. Extended Abstracts, *17th Conference on Hurr. And Trop. Meteor*., April 7-10, 1987, Miami, FL., Amer. Meteor. Soc., 266-269.

Rappaport, E.N., J.L. Franklin, L.A. Avila, S.R. Baig, J.L. Beven, E.S. Blake, C.A. Burr, J.-G. Jiing, C.A. Junkins, R.D. Knabb, C.W. Landsea, M. Mainelli, M. Mayfield, C.J. McAdie, R.J. Pasch, C. Sisko, S.R. Stewart, and A.N. Tribble, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395-419.

Sampson, C.R., J.L. Franklin, J.A. Knaff, and M. DeMaria, 2008: Experiments with a Simple Tropical Cyclone Intensity Consensus. *Wea. Forecasting*, **23**, 304–312.