Joint Hurricane Testbed Year 1 Report

1 September 2013 – 31 August 2014

A Probabilistic TC Genesis Forecast Tool Utilizing an Ensemble of Global Models

Principal Investigator: Robert Hart, Florida State University

Co-Principal Investigator: Henry Fuelberg, Florida State University

Co-Investigators: Daniel Halperin and Joshua Cossuth, Florida State University

Project Overview:

The ultimate goal of this JHT project is to create a disturbance-specific statistical TC genesis guidance product to aid the National Hurricane Center's Hurricane Specialists Unit in their decision making process when issuing Tropical Weather Outlooks. The guidance product will provide separate genesis probabilities for the 0-48 h and 0-120 h time periods. It will cover NHC's entire area of responsibility.

Accomplishments:

1. *Determined eastern North Pacific (EPAC) criteria thresholds for genesis and verified historical forecasts.*

The first task of this project was to extend the methodology of Halperin et al. (2013) to the EPAC basin in order to cover NHC's entire area of responsibility. That is, to determine calibrated threshold values for model-TC criteria and to verify the model genesis forecasts against the Best-Track (BT). Accordingly, Fig. 1 shows a performance diagram of 2007-2013 average model performance over the North Atlantic (NATL) and EPAC basins. Results show that on average over this 7 year period the models perform better over the EPAC than the NATL, due mainly to an increased probability of detection (POD). As in the NATL, the best ranking model varies from season to season (Fig. 2). It is interesting that although the POD and False Alarm Rate (FAR) differ from model to model (consistent with the results in Halperin et al. 2013), the critical success index (CSI) remains relatively constant among the models. Such approximate "conservation" of CSI may warrant further inspection beyond the work proposed here as it may suggest physical insight into the limits of predictability of genesis forecasting (especially over the NATL).

2. *Verified 2013 genesis forecasts*

Verification of 2013 genesis forecasts was completed. Figure 3 shows how the models' performance during 2013 compares with the 2007-2012 average. The models performed worse during 2013 compared to the 2007-2012 average over the NATL. The lower POD values may be a result of the 2013 TCs resulting from a genesis pathway (McTaggart-Cowan et al. 2008, 2013) that the models have greater difficulty predicting. There also were many TCs during 2013 that did not intensify past the tropical storm stage and were fairly short-lived. It is possible that the models do not handle marginal TCs well. In contrast, over the EPAC, the models performed better during 2013 than the 2007-2012 average. Perhaps 2013 EPAC TCs resulted from genesis pathways that are well handled by the models. The EPAC also contained stronger, longer-lived TCs compared to the NATL. We will investigate if there are statistically significant differences in peak intensity and TC duration between detected TCs and missed TCs in the models.

3. Regression model development: composite studies

The statistical guidance product will provide genesis probabilities for real-time model-indicated TCs based on the model's historical ability to predict genesis. Since a number of physically relevant variables (i.e., predictors) contribute to TC genesis—a dichotomous outcome variable—multiple logistic regression will be used. One of the first steps of logistic regression model development is the testing and selection of predictors. To help identify predictors that would discriminate well between the genesis (i.e., hit) and no genesis (i.e., false alarm) outcomes, composites of relevant variables were created for each outcome. The goal is to see whether there are structural differences between the genesis and no genesis events in the model forecast fields. If differences exist, they can be tested for inclusion in the regression models as categorical predictors. Thus far, we have found few notable structural differences between outcome types. The most obvious example is in the 925 hPa wind field (Fig. 4). For the genesis outcome, the location of maximum wind speed is north of the center, but for the no genesis outcome, it is north and east of the center. This difference does not appear to be a result of forward motion (speed or direction) alone, since the mean forward motion vector for each of the outcome variables is between 270 and 300°. In addition to structural differences, the values of physically relevant model output variables are analyzed next.

4. Regression model development: predictor testing and selection

Predictor testing and selection starts with relevant TC and environmental variables which are readily output in the model forecast fields (a non-comprehensive list is given in Table 1). Perturbations from the environmental average and time tendency terms also were tested for some variables. Given the basin-to-basin differences in model performance (Fig. 1), and the differing genesis pathways between the basins, predictors were tested separately for the NATL and EPAC. The predictors were tested for significance using backward elimination combined with a multiple fractional polynomial analysis, which checked for non-linear relationships between the predictor and the outcome variable (Sauerbrei et al. 2006; Hosmer et al. 2013). The historical cases were split into a developmental set, which comprised a random 95% of the events, and a verification set, which comprised the remaining 5%. A logistic regression model was fit using the developmental set, and the significant predictors were recorded. This process was repeated for 20 iterations. Each time, a different set of events was used as the verification set. Thus, each event was used once in the verification set. This process of cross validation (e.g., Wilks 2006) revealed that several predictors were significant during all iterations, and those were used as our initial predictor set. The predictor set was refined further based on the results of out-of-sample verification tests. This process was completed separately for the CMC, GFS, and UKM models.

5. Regression model development: assessing the goodness of fit

With the significant predictors identified, logistic regression models based on each global model's historical forecasts were created for the NATL and EPAC. Each model was fit with 2004-2010 events as the developmental set, while the 2011-2013 events were reserved as the out-of-sample verification set. For comparison, regression models using the same predictors, but only 2013 events as the out-of-sample verification set also were tested. The corresponding reliability diagrams are available at http://moe.met.fsu.edu/modelgen/histver.php. An example is given in Fig. 5. "Perfect" reliability is indicated by the orange diagonal line. Results are shown using 10% bins (blue line, consistent with NHC's presentation of their seasonal verification). For the remainder of this section, when referring to a model, we are discussing the *regression model based on the global model's historical forecasts* (e.g., "GFS" refers to the regression model based on the GFS historical forecasts).

The NATL 0-48 h regression models with the 2011-2013 events as verification perform fairly well. Reliability decreases in the forecast bins that contain few events (e.g., GFS forecast probabilities ≥ 60%, CMC forecast probabilities of 70-80%). The same reliability diagrams, except using only 2013 events as verification, look notably worse. Although overprediction may indeed have been a problem in some forecast probability bins (e.g., CMC forecast probabilities of 30-40%), the generally small sample size for all models precluded statistically significant or physically meaningful conclusions.

The NATL 0-120 h regression models with the 2011-2013 events as the verification set also perform well. However, each model struggled a bit with at least one forecast bin in the 50-70% range. Much like the 0-48 h reliability diagrams with only 2013 events as verification, the 0-120 h reliability diagrams for 2013 indicate some overprediction, but, again, the small sample size may be exaggerating the problem.

Due to recent active seasons and the greater average POD of genesis over the EPAC, the results for this basin are not impacted by the sample size issues seen for the NATL. The EPAC 0-48 h regression models with the 2011-2013 events as verification exhibit mixed results. The GFS underpredicts genesis for all forecast bins; the CMC overpredicts genesis for the 40-70% forecast probability range; and the UKM is closest to "perfect" reliability, with generally slight underprediction for forecast probabilities ≥ 40%. Results using only 2013 as the verification also are mixed. The GFS was more reliable during 2013 compared to the three-year average. The UKM overpredicted genesis for the lower half of the forecast probability bins. The CMC overpredicted in the 40-50% forecast range, much like the three-year average.

The EPAC 0-120 h regression models with 2011-2013 events as the verification set are mixed. The GFS again underpredicts at all forecast probabilities, more so than the 0-48 h regression model. The UKM underpredicts at forecast bins ≤ 60%, but is quite reliable in the 70-90% range. The CMC is most reliable, with only slight deviations from the "perfect" reliability line. When considering only 2013 as verification, the GFS is noticeably more reliable, except in the 30-40% range. The UKM diagram is a bit noisy, but generally underpredicts genesis. Again, the CMC is quite reliable, with slight underprediction of forecast probabilities ≥ 60%.

Overall, fairly large deviations from the longer-term mean reliability can occur, especially during seasons with relatively few forecast genesis events. Also, when more than one model predicts the same genesis event, consensus forecasts indicate a higher genesis probability. During seasons where the individual models are overpredicting genesis, the multi-model consensus probabilities will exacerbate the problem.

6. Quasi-operational products tested during 2014

The simplest way to output the quasi-operational products for viewing by the Hurricane Specialists Unit is via a locally-hosted website. The URL is http://moe.met.fsu.edu/modelgen and the following products are currently available on the site:

a. Overview of each basin that shows the location and categorical 0-48 and 0-120 h genesis probability of each model-indicated TC as well as the models available in the current initialization cycle (graphic).
b. 0-48 h and 0-120 h genesis probabilities for each model-indicated TC (graphic and text). This is available for each individual model and for the multi-model consensus.
c. Model-indicated tracks for each model-indicated TC out to 144 h (graphic and text).
d. The values of the criteria for defining a TC—including whether the values exceed the threshold required—for each 6 h forecast interval (text).
e. The values of each predictor used in the regression equations (text).
f. A history of the forecast genesis time, location, and probabilities for each model-indicated TC (text).
g. Real-time season-to-date verification of each regression model (reliability diagrams and geographical plots).
h. Historical verification of each regression model using 2011-2013 and 2013 only as the verification set (reliability diagrams).
i. An archive of all images.
j. A brief description of each product.

7. Began case studies of recent TCs

In an attempt to diagnose which aspects of the global model forecasts are failing, we plan to conduct a number of case studies of recent TCs. A few cases already have been completed (e.g., Bret 2011, Sean 2011, Debby 2012, Isaac 2012, Michael 2012). However, since this research task is still in its early stages, we cannot yet provide meaningful results or conclusions.

Comments on quasi-operational testing during 2014:

It has been difficult to evaluate the real-time products for the NATL given the lack of storms so far this season. The active EPAC, however, has provided an excellent opportunity to test the real-time guidance. The season-to-date verification products – located at http://moe.met.fsu.edu/modelgen/seasonver.php -- show some promise regarding the usefulness of the guidance probabilities.

We generally are pleased with the timely generation of the guidance products. In most cases, products are available 30-75 min prior to the synoptic/TWO issuance time (Table 2). Our biggest operational challenge thus far has been accounting for the recent operational upgrade to the UKM global model. The increased resolution and file size initially caused the UKM-based guidance products to be issued after the synoptic/TWO time. After consulting directly with

Julian Heming of UKMO, some modifications at both his end and at FSU made the data transfer and TC tracking code more efficient. As a result, the UKM-based products are once again usually available at least 30 min prior to the synoptic/TWO time. Despite the major changes to the UKM global model, we have not made any changes to the UKM-based regression models. We view this mid-season global model upgrade as an opportunity to test the robustness of the regression models.

Summary

Overall, we are slightly ahead of schedule on our original proposed plan and are pleased with how the work is proceeding. Regardless of the future activity (or lack thereof) in the NATL this season, the active EPAC has provided ample opportunity to test the effectiveness of the regression models and the timely generation of the guidance products.

References

Halperin, D. J., H. E. Fuelberg, R. E. Hart, J. H. Cossuth, P. Sura, and R. J. Pasch, 2013: An evaluation of tropical cyclone genesis forecasts from global numerical models. *Wea. Forecasting*, **28**, 1423-1445.

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant, 2013: Applied logistic regression. Wiley. com.

McTaggart-Cowan, R., G. D. Deane, L. F. Bosart, C. A. Davis, and T. J. Galarneau Jr, 2008: Climatology of tropical cyclogenesis in the North Atlantic (1948-2004). *Monthly Weather Review*, **136**, 1284-1304.

McTaggart-Cowan, R., T. J. Galarneau Jr, L. F. Bosart, R. W. Moore, and O. Martius, 2013: A global climatology of baroclinically inuenced tropical cyclogenesis. *Monthly Weather Review*, **141**, 1963-1989.

Sauerbrei, W., A. Meier-Hirmer, C. Benner, and P. Royston, 2006: Multivariable regression model building by using fractional polynomials: Description of SAS, STATA, and R programs. *Computational Statistics and Data Analysis*, **50**, 3464-3485.

Wilks, D. S., 2006: Statistical methods in the atmospheric sciences, Vol. 100. Access Online via Elsevier.

Table 1. Initial predictor pool (not comprehensive).

• forecast hour	• latitude	• longitude
• thickness (250-850 mb)	• mslp	• shear (200-850 mb)
• relative vorticity (850, 700 mb)	• year	• temperature (sfc, trop)
• 925 mb wind speed	• PWAT	• lapse rate (1000-700 mb)
• relative humidity (600, 700 mb)	• Julian day	• | season peak – Julian day |
• Okubo-Weiss (850, 500 mb)	• CAPE	• % land cover
• divergence (850, 200 mb)	• CIN	• sfc latent heat net flux
• Q vector convergence (200-400 mb)	• ENSO index	• thickness asymmetry
• maximum potential intensity (MPI)	• MJO phase	• genesis in another model

*Perturbation from environmental average and time tendencies to be tested for some variables

Table 2. A typical timeline showing when global model output is received locally and when the corresponding guidance products are available (all times given in UTC). These times occasionally change due to data transfer delays. The global model initialization time is given in parentheses. NHC's TWO issuance times are 0000, 0600, 1200, and 1800 UTC.

Model	Data received locally (cycle)	Guidance products available (cycle)
	NATL	
CMC	0445 (00); 1645 (12)	0501 (00); 1701 (12)
GFS	0435 (00); 1035 (06); 1635 (12); 2235 (18)	0459 (00); 1059 (06); 1659 (12); 2259 (18)
UKM	0430 (00); 1650 (12)	0501 (00); 1721 (12)
CONSENSUS		0505 (00); 1725 (12)
	EPAC	
CMC	0445 (00); 1645 (12)	0459 (00); 1659 (12)
GFS	0435 (00); 1035 (06); 1635 (12); 2235 (18)	0452 (00); 1052 (06); 1652 (12); 2252 (18)
UKM	0430 (00) ; 1650 (12)	0439 (00); 1659 (12)
CONSENSUS		0500 (00); 1700 (12)

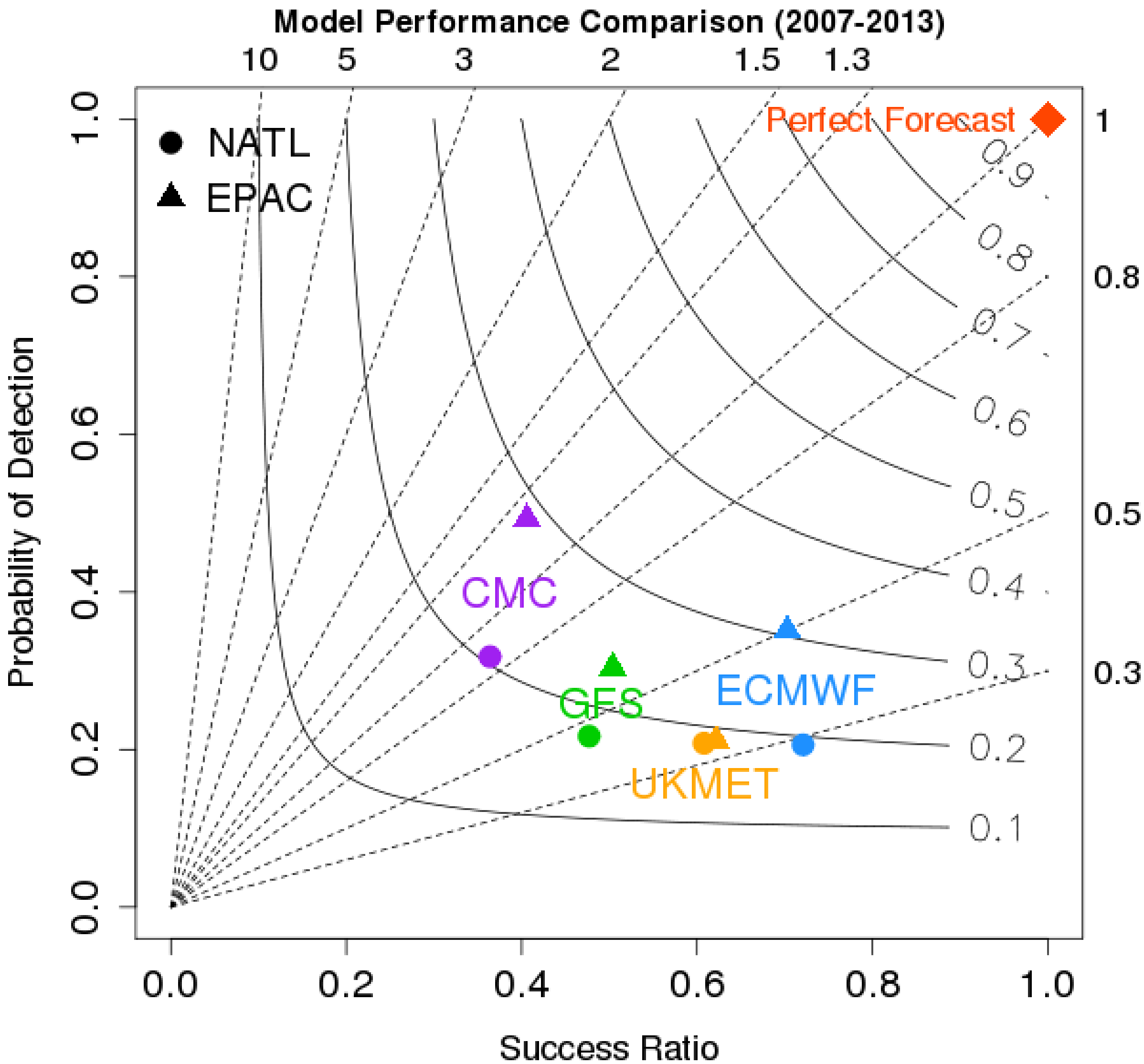

Figure 1. 2007-2013 average performance diagram for all models. Success Ratio (i.e., 1-False Alarm Ratio) is given on the x-axis; POD on the y-axis. Bias values are indicated by the dashed lines, and CSI values are indicated by the curved, solid lines. A "perfectly" performing model would be in the upper-right corner of the plot. Circles indicate the NATL basin; triangles indicate the EPAC basin.

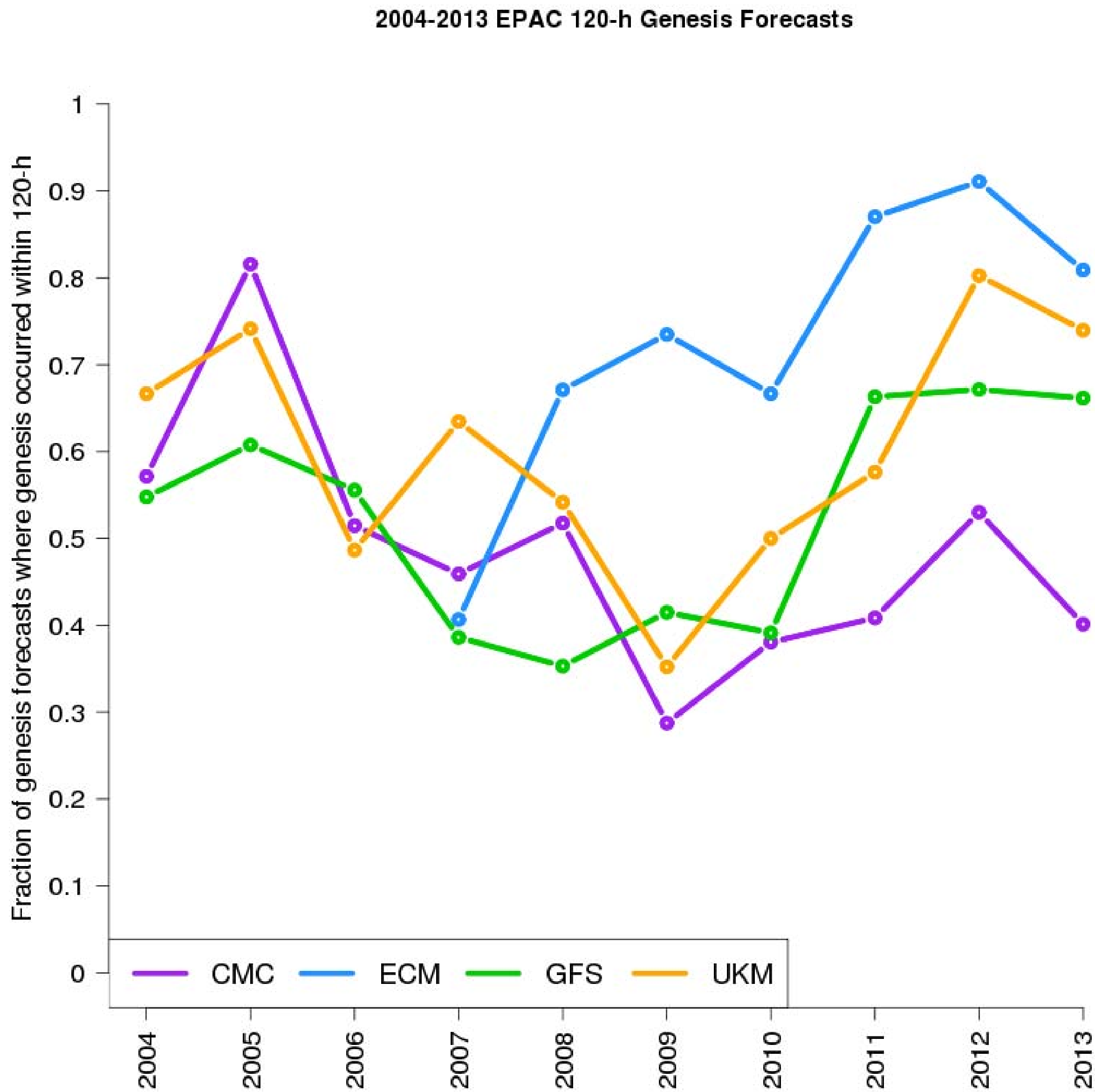

Figure 2. The fraction of EPAC model TC genesis forecasts that result in BT genesis within 120 h of the initialization time.

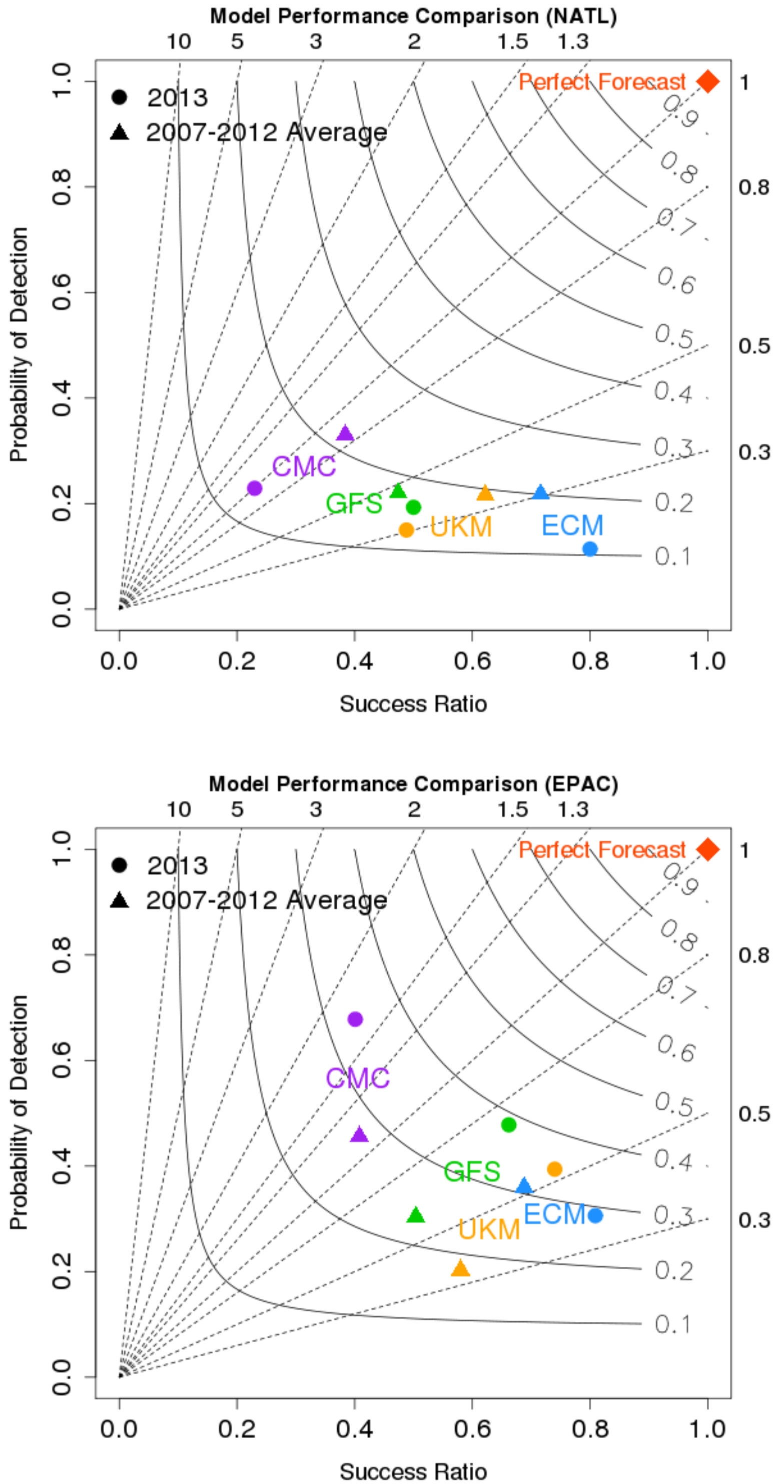

Figure 3. Performance diagram showing verification statistics for 2013 (circles) versus the 2007-2012 average (triangles) for the NATL (top) and EPAC (bottom). Lines and axes are as in Fig. 1.

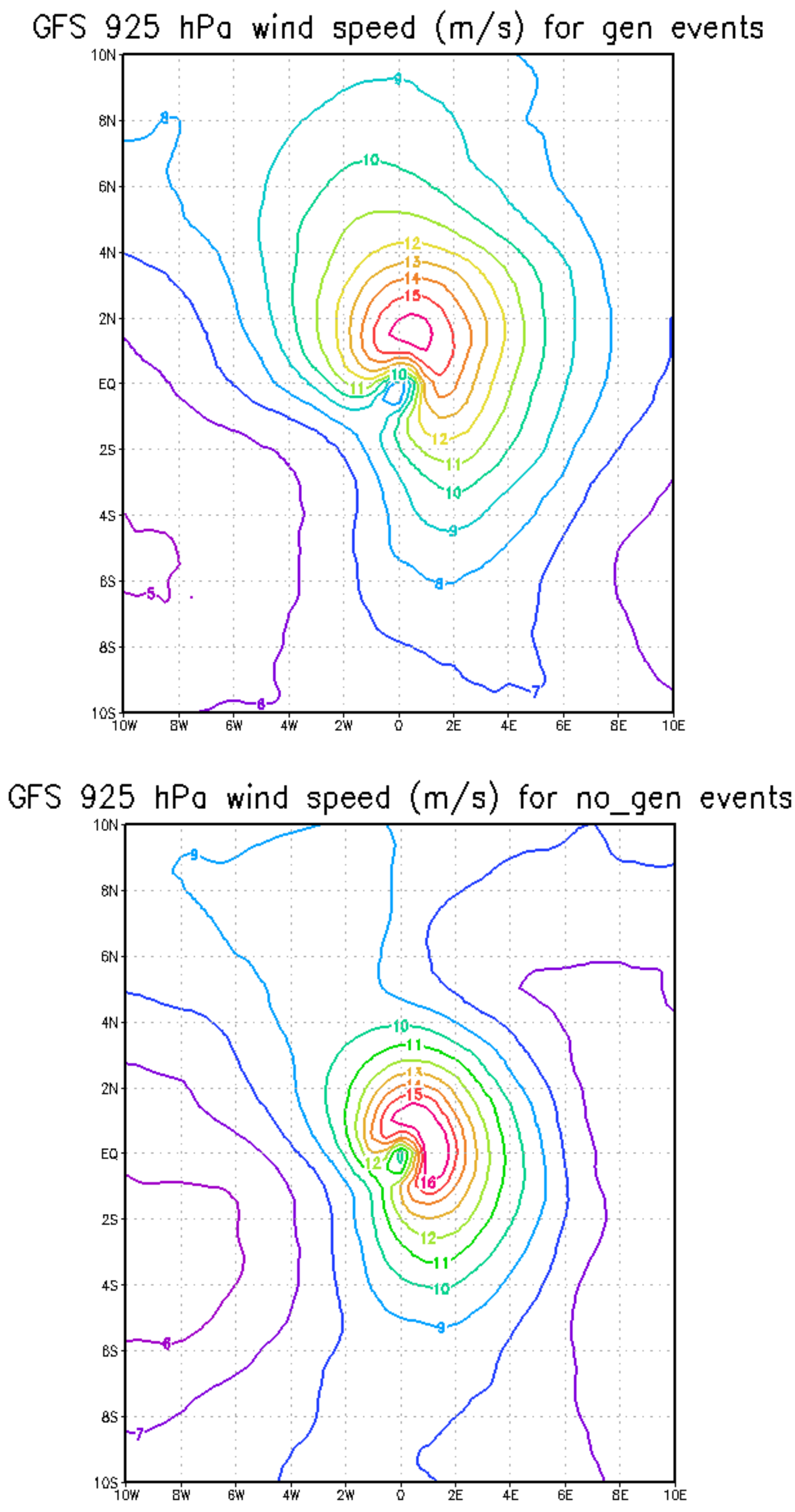

Figure 4. 925 hPa wind speed composites for GFS genesis events (top) and no genesis events (bottom).

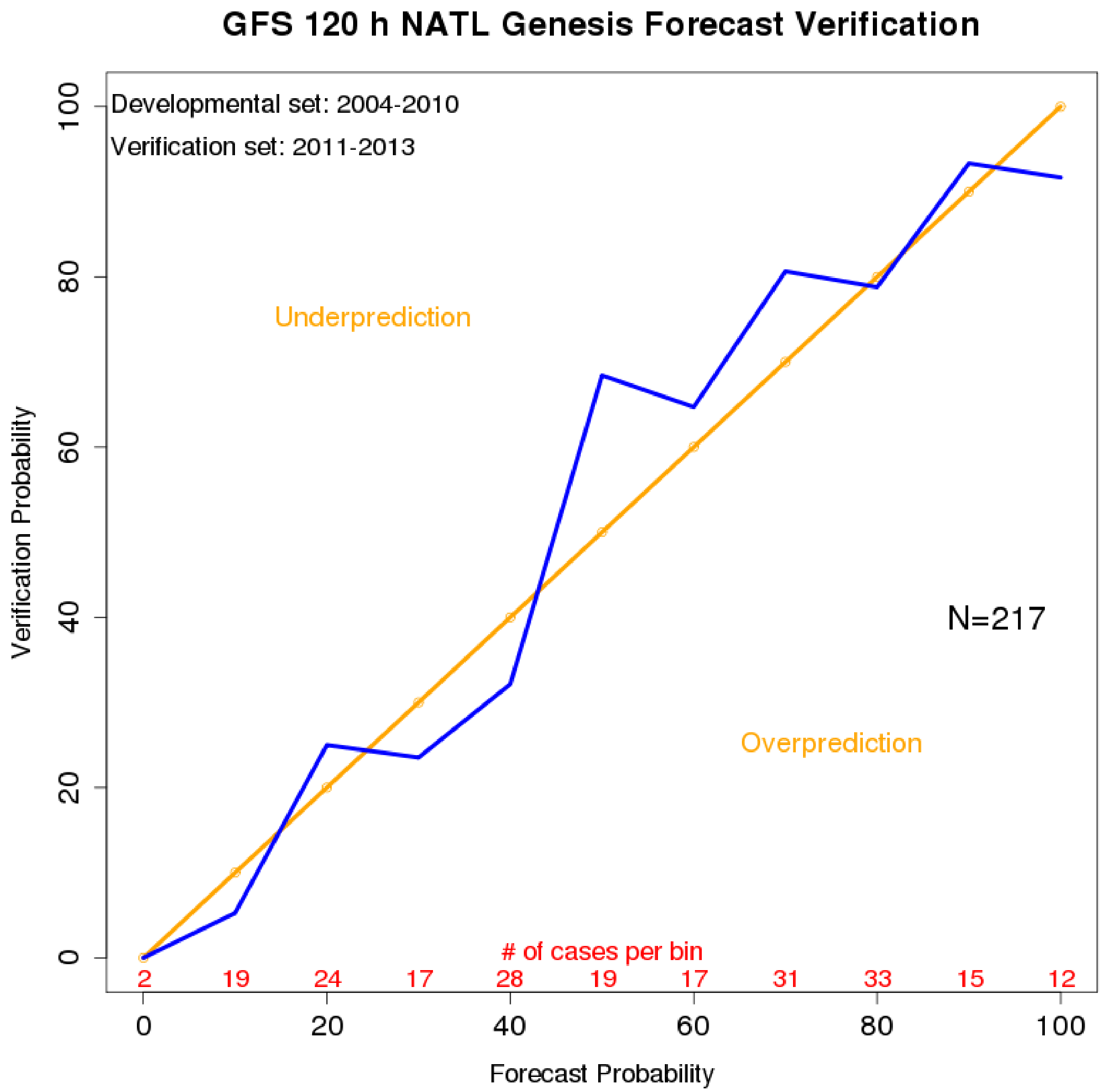

Figure 5. Reliability diagram for the 0-120 h GFS-based regression model for the NATL basin. The regression model was developed with 2004-2010 events and verified on 2011-2013 events. "Perfect reliability" is denoted by the orange diagonal line. Results are given by the blue line in 10% bins, consistent with NHC's presentation of their seasonal verification.