JHT Mid-Year Report (Year 1)
1 September 2013 – 28 March 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

Accomplishments:

1. *Determine eastern North Pacific (EPAC) criteria thresholds for genesis and verify 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 the 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, Figure 1 shows a performance diagram of 2007-2013 average model performance over the North Atlantic (NATL) and EPAC basins. Overall, the models perform better over the EPAC compared to 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 as it may suggest physical insight into the limits of predictability of genesis forecasting (especially over the NATL).

2. *Began composite studies*

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 and no genesis outcomes, composites of relevant variables were created for each outcome. The goal was to see whether there are structural differences between the genesis and no genesis events. Any differences found could be used as categorical predictors in the regression model. So far, we have found few notable structural differences between outcome types. One example is in the 925 hPa wind field (Fig. 3). For the genesis outcome, the maximum wind speed is north of the center, but for the no genesis outcome, it is to the north and east of the center. This does not appear to be a result of forward motion (speed or direction) alone, as the mean forward motion vector for each of the outcome variables is between 270 and 300°. We plan to examine other variables over the next several weeks.

3. Began 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 were also tested for some variables. Given the basin-to-basin differences in model performance (Fig. 1), 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 in the verification set once. Several predictors were significant during all iterations, and those were used as our initial predictor set (Table 2). So far, predictor testing and selection has only been completed for the GFS.

4. Began regression model development and testing

With the significant predictors identified, logistic regression models based on GFS output were created for the NATL and EPAC. Each model was fit with 2004-2010 events as the developmental set while 2011-2012 events were reserved as the out-of-sample verification set. Figure 4 shows the resulting reliability diagram for the NATL, where "perfect" reliability is indicated by the orange, diagonal line. Results using 10% bins (consistent with NHC; blue line) and 20% bins (to compensate for smaller sample sizes; red line) indicate that the model performed well, especially in the 0.4 to 0.8 forecast probability range – which is the most critical decision range where human forecasts have shown greatest genesis forecasting difficulty. Overall, this is an encouraging result and suggests that there is some predictive capability based on the chosen predictors. Figure 5 shows a similar diagram for the EPAC. Results show that this regression model underpredicts genesis at nearly all forecast probability bins.

These regression models were recently verified with the 2013 genesis forecasts. Over the NATL, results were not as reliable as during the prior years, but it is likely that a small sample size (N=54) was a contributing problem. Six of the forecast probability bins contained 5 or fewer cases – and those bins where there were more cases had greater reliability. Nevertheless, there is an apparently overprediction problem during 2013. Over the EPAC, reliability is better during 2013 than during the period when the regression model was developed (2011-2012).

5. Verified 2013 genesis forecasts

A verification of 2013 genesis forecasts was recently completed. Figure 8 shows how the models' performance in 2013 compares with the 2007-2012 average. The models were less reliable during 2013 compared to the 2007-2012 average over the NATL. The lower POD values may be a result of the 2013 TCs forming from a genesis pathway (McTaggart-Cowan 2013) that is inherently less predictable. There were also 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. We are also examining whether 2013's events had a statistically significant smaller storm size that might suggest greater difficulty in prediction. We will investigate if there are differences in peak intensity and TC duration between detected TCs and missed TCs in the models. Early indications are that when numerous consecutive model runs all indicate genesis, the resulting TC is very likely to become a hurricane (and often a major one). In light of this, we will be examining the value of adding as a predictor the number of prior consecutive model runs indicating genesis – and/or the percentage those prior regressions provided. In contrast, the models performed better during 2013 compared to the 2007-2012 average. The improved performance here may be attributed to the same reasons as the NATL—perhaps 2013 EPAC TCs resulted from genesis pathways that are well handled by the models. The EPAC also contained stronger, longer-lived TCs than the NATL.

Plans for the remainder of year 1:

1. Finish composite studies.
2. Finish predictor testing and selection for all models.
3. Develop and evaluate regression models for CMC and UKM.
4. Test and evaluate the regression models in a quasi-operational setting during the 2014 hurricane season.
5. Refine output products based on feedback from the Hurricane Specialists Unit.
6. Develop a season-to-date verification to give forecasters real-time feedback regarding the performance of each model during 2014.

Summary:

Overall, we are ahead of schedule slightly on our original proposed plan and are pleased with how the work is proceeding. We look forward to seeing the experimental products tested for the upcoming season – assuming that a strong El Nino does not prevent storms from being tested.

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. Initial predictor set for NATL and EPAC GFS-based regression models.

Predictor (NATL)		Predictor (EPAC)	
Forecast hour	20	Forecast hour	20
Year	20	850 mb ζ perturbation	18
250-850 mb ΔZ	20	Sfc latent heat net flux	20
Tropopause temp	20	Latitude	20
850 mb ζ perturbation	20	PWAT	16
CAPE	20	CIN	16
Sfc latent heat net flux	20		
Longitude	20		

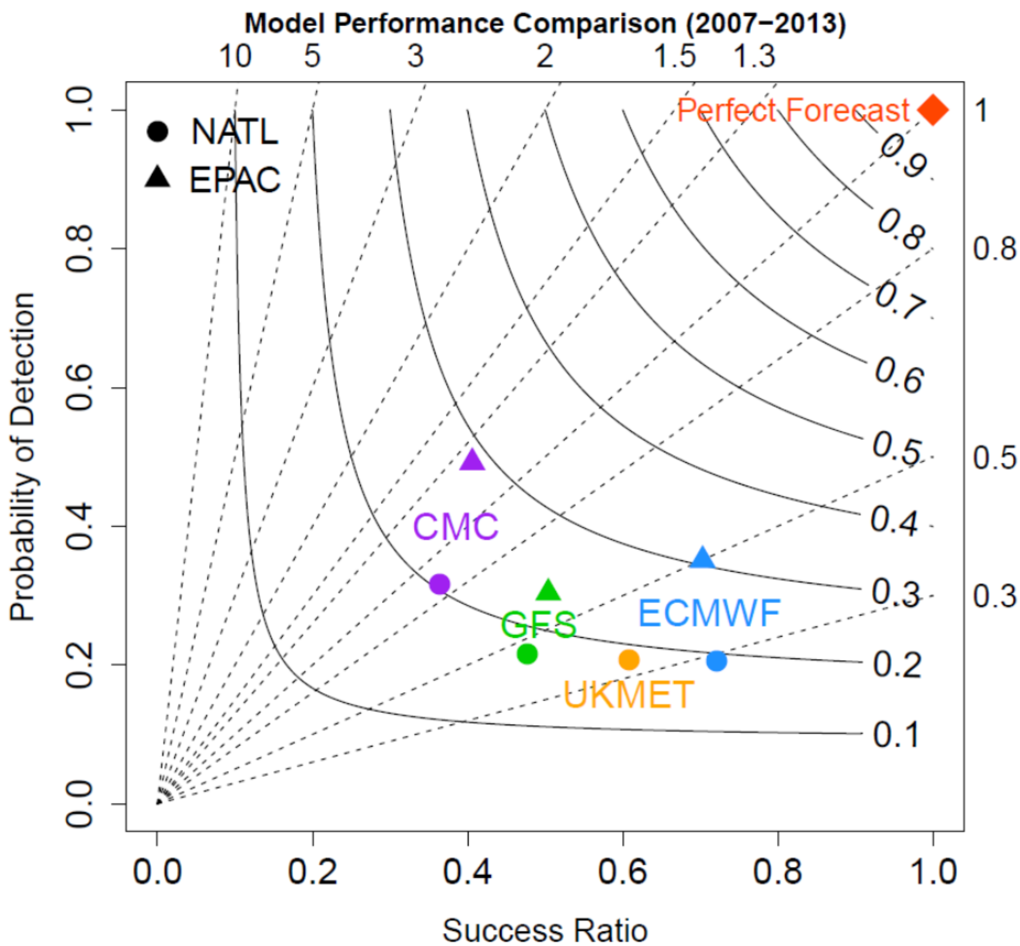

Figure 1. 2004-2013 average performance diagram for all models. SR is given on the x-axis; POD on the y-axis. Bias values are indicated by the dashed lines, and the CSI values are indicated by the curved, solid lines. A "perfect" performing model would be in the upper-right corner of the plot. Circles indicate the NATL basin; triangles indicate the EPAC basin.

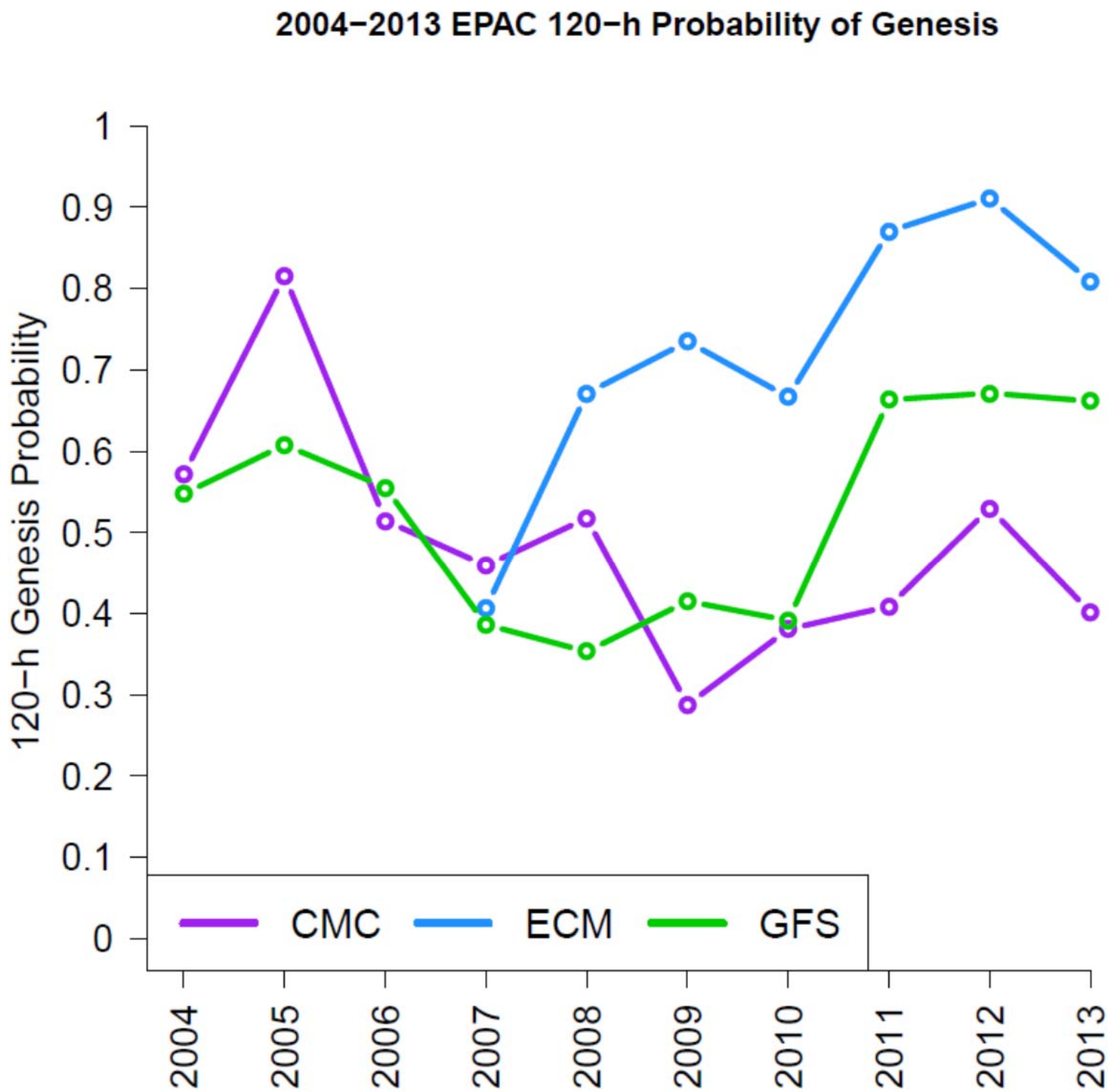

Figure 2. The average probability of genesis within 120-h for each season.

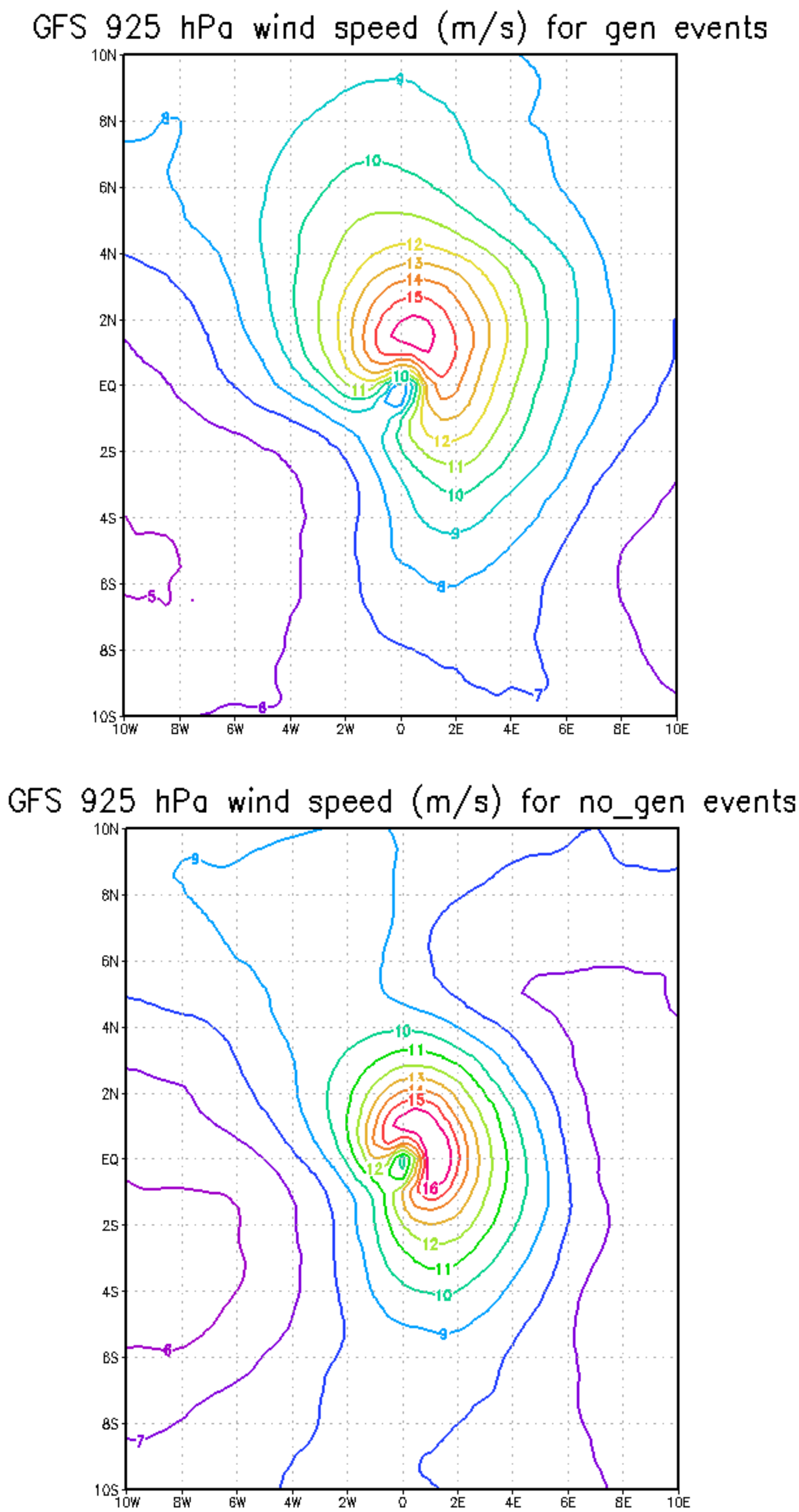

Figure 3. 925 hPa wind speed composites for GFS genesis events (top) and no genesis events (bottom).

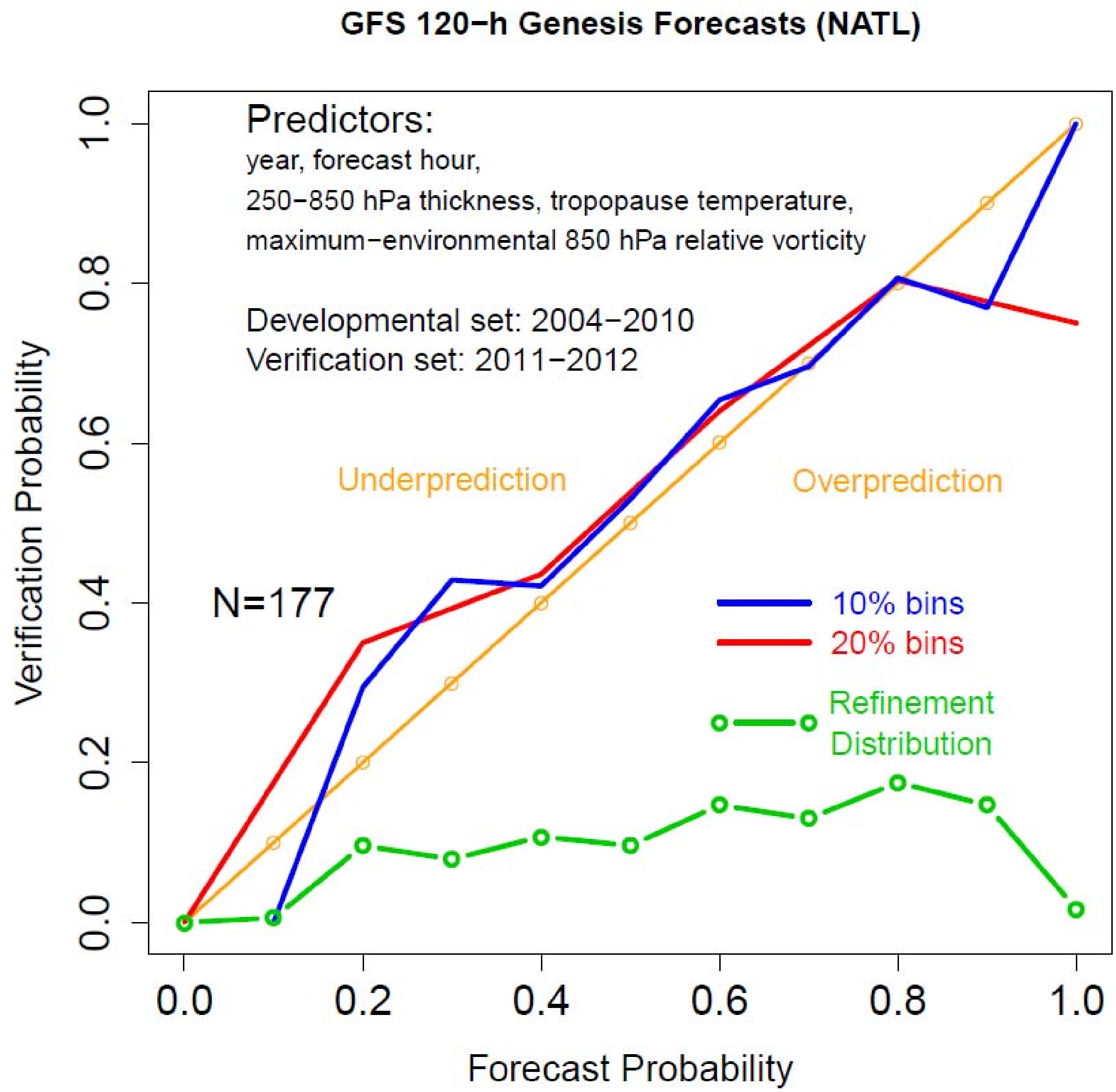

Figure 4. Reliability diagram for GFS-based regression model for the NATL basin. The regression model was developed with 2004-2010 events and verified on 2011-2012 events. "Perfect reliability" is given by the orange diagonal line.

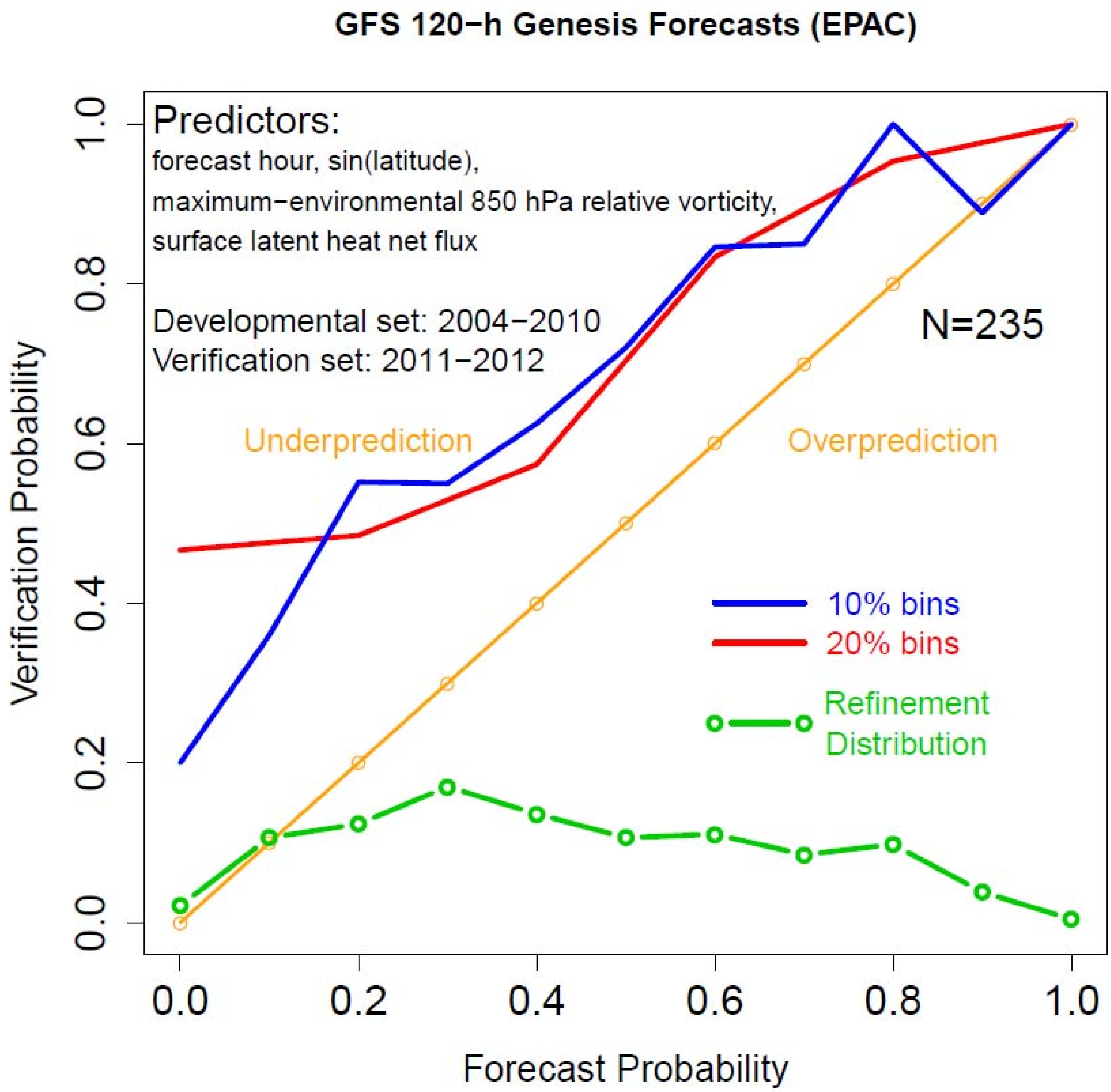

Figure 5. As in Fig. 4, but for the EPAC basin.

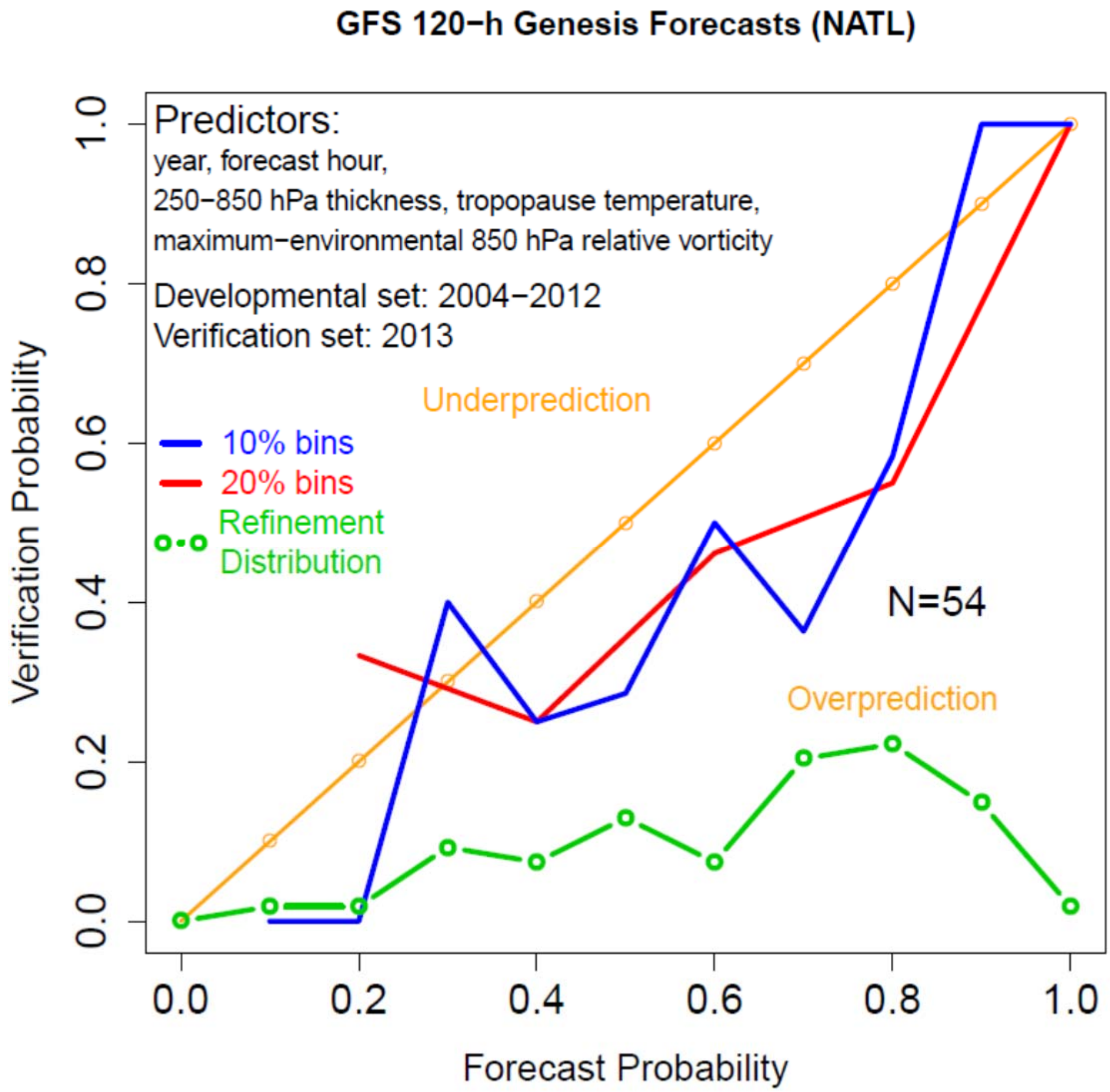

Figure 6. As in Fig. 4, except the developmental set is 2004-2012 and the verification set is 2013.

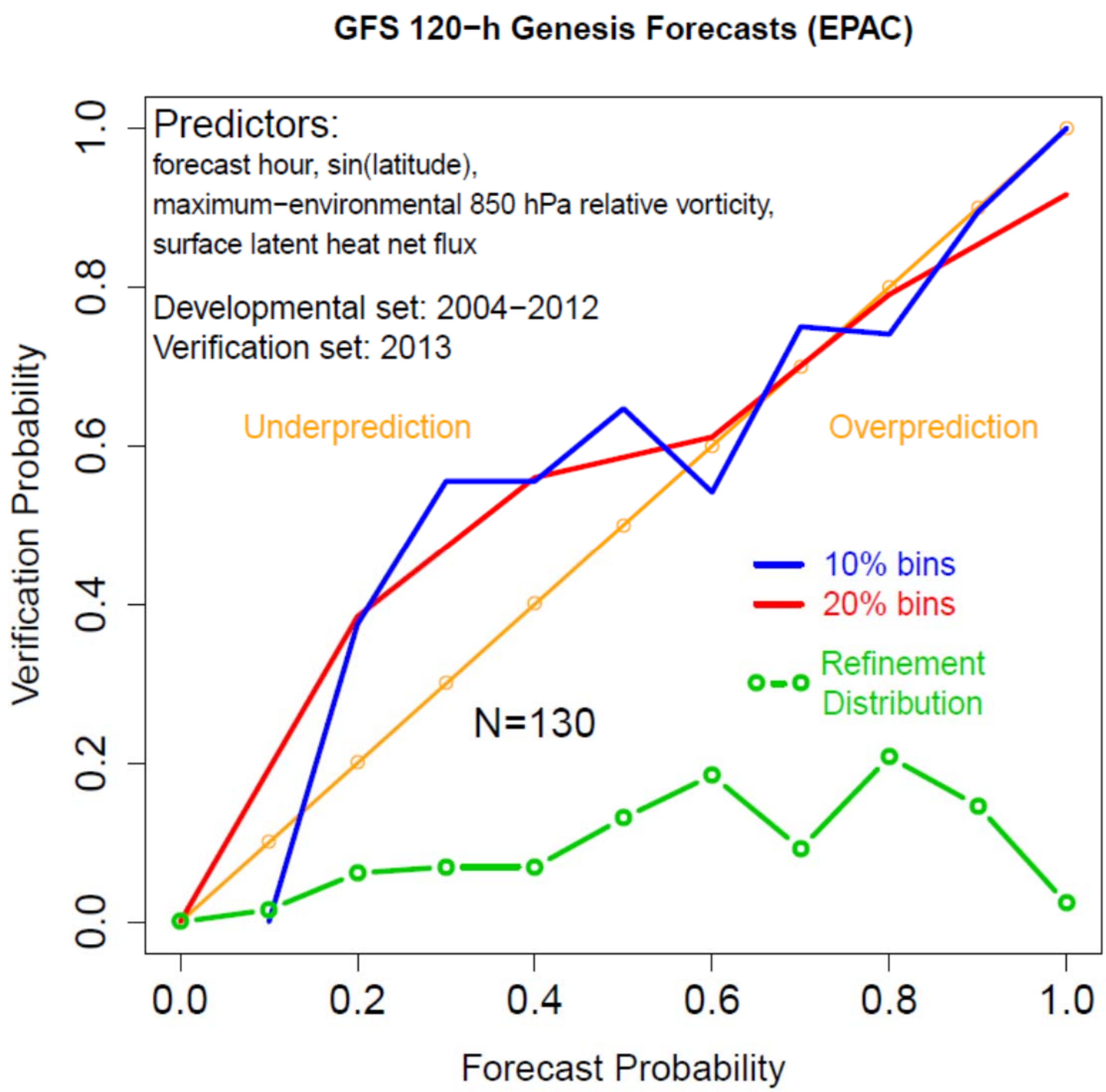

Figure 7. As in Fig. 6, except for the EPAC basin.

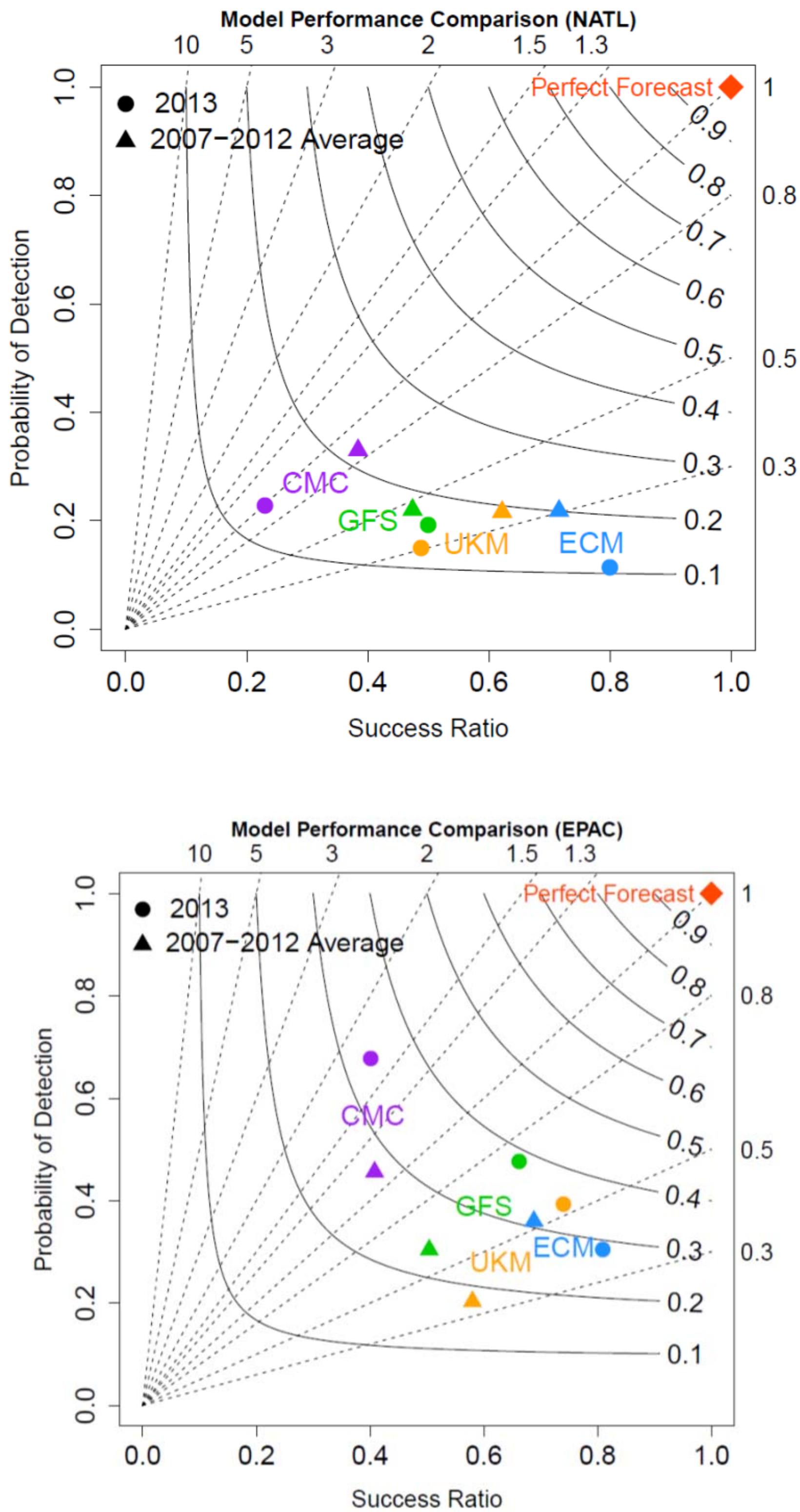

Figure 8. Performance diagram showing how each model performed in 2013 (circles) versus the 2007-2012 average (triangles) for the NATL (top) and EPAC (bottom).