JHT Mid-Year 2 Report
1 September 2014 – 28 February 2015

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, Florida State University
Joshua Cossuth, Naval Research Laboratory

Project Overview:

The goal of this JHT project is to create a disturbance-specific statistical tropical cyclone (TC) genesis guidance product to aid the Hurricane Specialists Unit (HSU) in preparing the Tropical Weather Outlook (TWO). The guidance product consists of separate genesis probabilities for the 0-48 h and 0-120 h time periods. It covers NHC's entire area of responsibility.

Accomplishments:

The guidance products were tested quasi-operationally during the beginning of year 2. Regression equations for each global model (CMC, GFS, and UKM), each basin (NATL and EPAC), and each forecast window (48 and 120 h) were developed and applied to real-time global model indicated TC genesis forecasts. Several text and graphical products were available, some of which were created at the request of the HSU. The overall operational reliability of the guidance was good, with all products usually available at least 35 minutes prior to the synoptic/TWO time. Preliminary verification of the regression probabilities indicate that—with some exceptions—the guidance generally is well calibrated. Homogeneous comparisons of the verification of consensus regression probabilities and TWO probabilities suggest that these guidance products may be most useful to the HSU at the higher forecast probability bins (at least for the 120 h forecast window). Finally, ECMWF based regression equations were developed and will be ready for inclusion in the guidance suite for 2015 quasi-operational testing (assuming real-time ECMWF data are available).

1. Quasi-operational products tested during 2014

The simplest way for the HSU to view the quasi-operational products is via a locally-hosted website. The URL is http://moe.met.fsu.edu/modelgen and the following products currently are available:

a. Overviews of each basin that show the location and categorical 0-48 and 0-120 h genesis probability of each model-indicated TC and list 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 model and for the multi-model consensus.

c. Model-indicated tracks for each model-indicated TC, out to 144 h (graphic and text).
d. Values of the criteria for defining a TC—including whether the values exceed the required thresholds—for each 6 h forecast interval (text).
e. 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.

2. Preliminary verification from 2014 quasi-operational testing

Using the working Best-Track files, we completed a preliminary verification of each regression model. Unless specified otherwise, for the remainder of this section, when referring to a global model, the authors are discussing their statistical guidance products based on the referenced global model. The non-homogeneous results (i.e., all available results for each technique) are presented first using reliability diagrams (Figs. 1-4). The verification for each individual model based regression, the consensus based regression (black line), and NHC TWO forecasts (red line) are plotted. Points/breaks in the lines indicate that five or fewer cases were available in a given forecast bin – a sample size too small to draw meaningful conclusions. The North Atlantic (NATL) 48 h verification (Fig. 1) shows well calibrated forecasts in the 0-30% probability bins for the CMC and GFS based regression models, as well as the NHC TWO forecasts. At probabilities exceeding 30%, the GFS based regression model remains well calibrated. However, the UKM and consensus regression models overpredict genesis, and the CMC regression and NHC TWO forecasts underpredict genesis. For the NATL at 120 h (Fig. 2), the CMC based regression model is well calibrated, with the consensus regression model only slightly overpredicting genesis. The GFS and UKM regression models generally overpredict genesis. NHC's TWO forecasts are very reliable in the 0-40% forecast probability bins, but genesis is underpredicted in the higher forecast probability bins.

Verification of the guidance was mixed for 48 h forecasts in the eastern North Pacific (EPAC) (Fig. 3). While the CMC, consensus regression model, and NHC TWO forecasts perform well in the 0-40% range, they stray from the "perfect reliability" line at the higher probability bins. The NHC TWO forecasts, along with the CMC and GFS regression models underpredict genesis, while the consensus and UKM regression models overpredict it. For the EPAC 120 h forecasts (Fig. 4), the regression models and NHC TWO forecasts generally underpredict genesis, but the UKM regression model is well calibrated.

Overprediction by the UKM regression may be due in part to a new global model configuration that was implemented operationally during July 2014. Heming (2014) noted that re-forecasts of TCs using the new UKM global model configuration generally yielded stronger forecast intensities of mature TCs than the prior configuration. While they did not explicitly discuss the impact on genesis forecasts, it is possible that the new UKM global model configuration also may be producing more intense disturbances or early-stage TCs, thus causing the UKM based regression model to overpredict genesis during 2014. Upgrades to all global models in our guidance suite undoubtedly impact the reliability of the regression equations. The UKM global model upgrade is the most apparent example for 2014. It will be interesting to see how the recent GFS global model upgrade affects the GFS based regression equations during 2015 quasi-operational testing.

The GFS based regression for the NATL at 120 h stands out due to its especially poor reliability. The use of "year" as a predictor was a contributing factor. While the developmental set for the logistic regression model did indicate a linear improvement in GFS global model genesis forecasts over time, there was no guarantee that these improvements would continue during 2014. Indeed, the GFS global model hit rate during 2014 was lower than during 2010-2013. Thus, the GFS based regression probabilities were unnecessarily inflated by the choice to include "year" as a predictor. Figure 5 compares the quasi-operational verification of the GFS based regression for the NATL at 120 h (blue line) with the re-forecast verification if "year" was not included as a predictor (gray line). While still not perfect, removing "year" as a predictor would have prevented the notable overprediction. The underprediction observed for the "year removed" regression in the 0% forecast probability bin is attributable to low genesis probabilities for what became Bertha.

It is encouraging that the CMC based regression models performed quite well for both basins and forecast windows. While historical verification indicates that the CMC global model false alarm rate is higher than the other global models, it appears that the regression equations are able to correct for the CMC global model's biases and provide well calibrated probabilistic forecasts. That is the goal of each of our regression models.

Prior to the upcoming 2015 quasi-operational testing, the predictor selection process for all regression models will be repeated to determine whether any predictors should be added or removed when the 2014 forecasts are added to the developmental dataset. The authors will discuss with the NHC points of contact whether there is value in developing two regression equations – one with and one without "year" as a predictor – for the GFS NATL 120 h forecasts. If the 2015 GFS global model upgrade results in improved TC genesis forecasts, then "year" will again be a useful predictor for calibrating the GFS based NATL 120 h regression equation.

For a more direct comparison of the verification results, a set of homogeneous NHC TWO and consensus regression forecasts was constructed. These are instances when NHC and the consensus product issue probabilities for the same disturbance. These forecasts were verified against the working Best-Tracks and the associated reliability diagrams are presented in Fig. 6-9. For the NATL 48 h product, meaningful comparisons can only be made in the 10-30% forecast probability bins where the forecast performance is fairly comparable (Fig. 6). At probabilities exceeding 30%, the sample size – given explicitly by the blue (consensus) and red (NHC TWO) text – is too small to make conclusions. At 120 h (Fig. 7), sample size is not an issue. Here it is apparent that the NHC TWO forecasts outperform the consensus regression model in the 0-30% forecast probability range. However, at the higher probability bins, the consensus regression model is better calibrated. Over the EPAC at 48 h (Fig 8), NHC TWO forecasts underpredict genesis, while the consensus regression model generally overpredicts it. However, neither is far from the "perfect reliability" line. At 120 h (Fig. 9), the consensus regression model struggles in the 20-50% forecast probability range, but is fairly well calibrated in the 70-100% range. The NHC TWO forecasts generally underpredict genesis.

The reader may be wondering why the sample size of the TWO and consensus regression probabilities are not equal. There are a few instances where the global models disagree on the timing and location of genesis for a particular disturbance. This causes the automated tracking algorithm to assume that these are forecasts of two or three different TC genesis events. However, each model genesis forecast occurs within the TWO shaded potential genesis region. Figure 10 illustrates one such occurrence. Since the forecast genesis locations are far apart, the consensus tracker treats each genesis forecast as a separate TC. But, because all three forecasts occur within the TWO shaded genesis region, all three forecasts are included in the homogeneous verification. In this particular example, there are 3 consensus genesis forecasts and 1 TWO forecast contributing to the homogeneous verification.

3. ECMWF regression equations developed and tested

Real-time ECMWF forecast fields were not available during the 2014 quasi-operational testing that was conducted at FSU. There is a possibility that quasi-operational testing during 2015 will occur on the JHT workstation at NHC, where real-time ECMWF data would be available. In preparation for this possibility, regression models were developed and tested based on the historical ECMWF genesis forecasts. One of the challenges in developing the regression models is the relatively small sample size of historical forecasts. The ECMWF has a low probability of detection for TC genesis events. In addition, the ECMWF historical archive of genesis forecasts only dates back to 2007, whereas the archive for the other models extends back to 2004. An example of how well the 2011-2013 data fit a regression model developed from the 2007-2010 forecasts is presented in Fig. 11. Note that the sample size over the three year independent period is still fairly small. Given the ECMWF's low false alarm rate, most of the regression based genesis probabilities are at or above 70%. In the 70-100% forecast probability range, the regression model is fairly well calibrated. However in the 0-60% forecast probability range, the small sample size precludes meaningful conclusions. A consensus regression model that includes ECMWF forecasts is currently being developed. Should real-time ECMWF data be available during the upcoming hurricane season, ECMWF based products will be included in the guidance suite.

Comments on quasi-operational testing:

The below normal TC activity over the NATL during 2014 made it somewhat difficult to evaluate the real-time products. The active EPAC, however, provided an excellent opportunity to test the real-time guidance. The preliminary verification statistics show promise regarding the usefulness of the guidance probabilities.

We generally are pleased with the timely generation of the guidance products. Products usually are available 30-75 min prior to the synoptic/TWO issuance time. Our biggest operational challenge thus far was accounting for the 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 identification 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 did not make any changes to the UKM based regression models. We viewed the mid-season global model upgrade as an opportunity to test the robustness of the regression models.

With the recent upgrade of the GFS, we anticipated similar challenges for 2015 quasi-operational testing. However, the model output now arrives at FSU earlier than it did during 2014. This earlier availability, combined with a few modifications to the TC identification script, resulted in the GFS based guidance products being available 90-120 min prior to the synoptic/TWO time. Thus, we do not anticipate any problems with the timeliness and operational reliability of the guidance products during 2015 quasi-operational testing. An updated summary of product generation times is given in Table 1.

Summary:

We are slightly ahead of our originally proposed schedule and are pleased with how the work is proceeding. Despite the inactivity in the NATL last season, the active EPAC provided ample opportunity to test the effectiveness of the regression models and the timely generation of the guidance products. We look forward to another round of quasi-operational testing this season with the potential inclusion of ECMWF based products.

References:

Heming, J., 2014: The impact on tropical cyclone predictions of a major upgrade to the Met Office global model. *31st Conference on Hurricanes and Tropical Meteorology*, Amer. Meteo. Soc., San Diego, CA, 31 March-4 April 2014, 11A.3.

Table 1. Timeline showing when the guidance products are available (all times given in UTC). These times occasionally change due to delays in data transfer. The global model initialization time is given in parentheses. NHC TWO issuance times are 0000, 0600, 1200, and 1800 UTC.

Model	Guidance products available (cycle)
NATL	
CMC	0516 (00); 1716 (12)
GFS	0412 (00); 1012 (06); 1612 (12); 2212 (18)
UKM	0501 (00); 1721 (12)
CONSENSUS	0520 (00); 1725 (12)
EPAC	
CMC	0514 (00); 1714 (12)
GFS	0402 (00); 1002 (06); 1602 (12); 2202 (18)
UKM	0439 (00); 1659 (12)
CONSENSUS	0515 (00); 1715 (12)

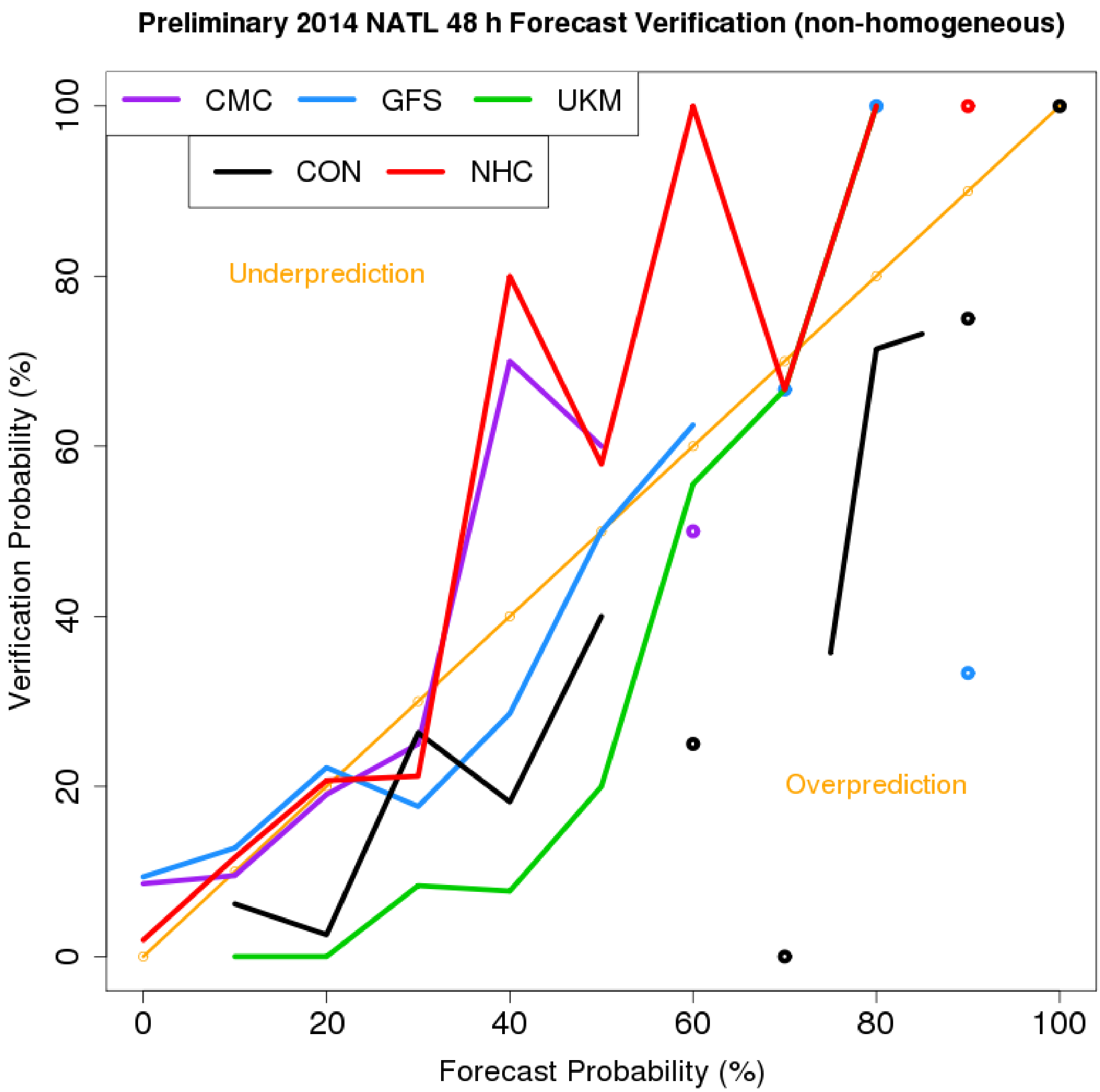


Figure 1. Reliability diagram of the NATL 48 h regression models and NHC TWO forecasts (red line). Verification is preliminary and based on the 2014 working Best-Tracks. The results are based on a non-homogeneous set of forecasts. “Perfect reliability” is given by the orange, diagonal line; above (below) this line indicates underprediction (overprediction).

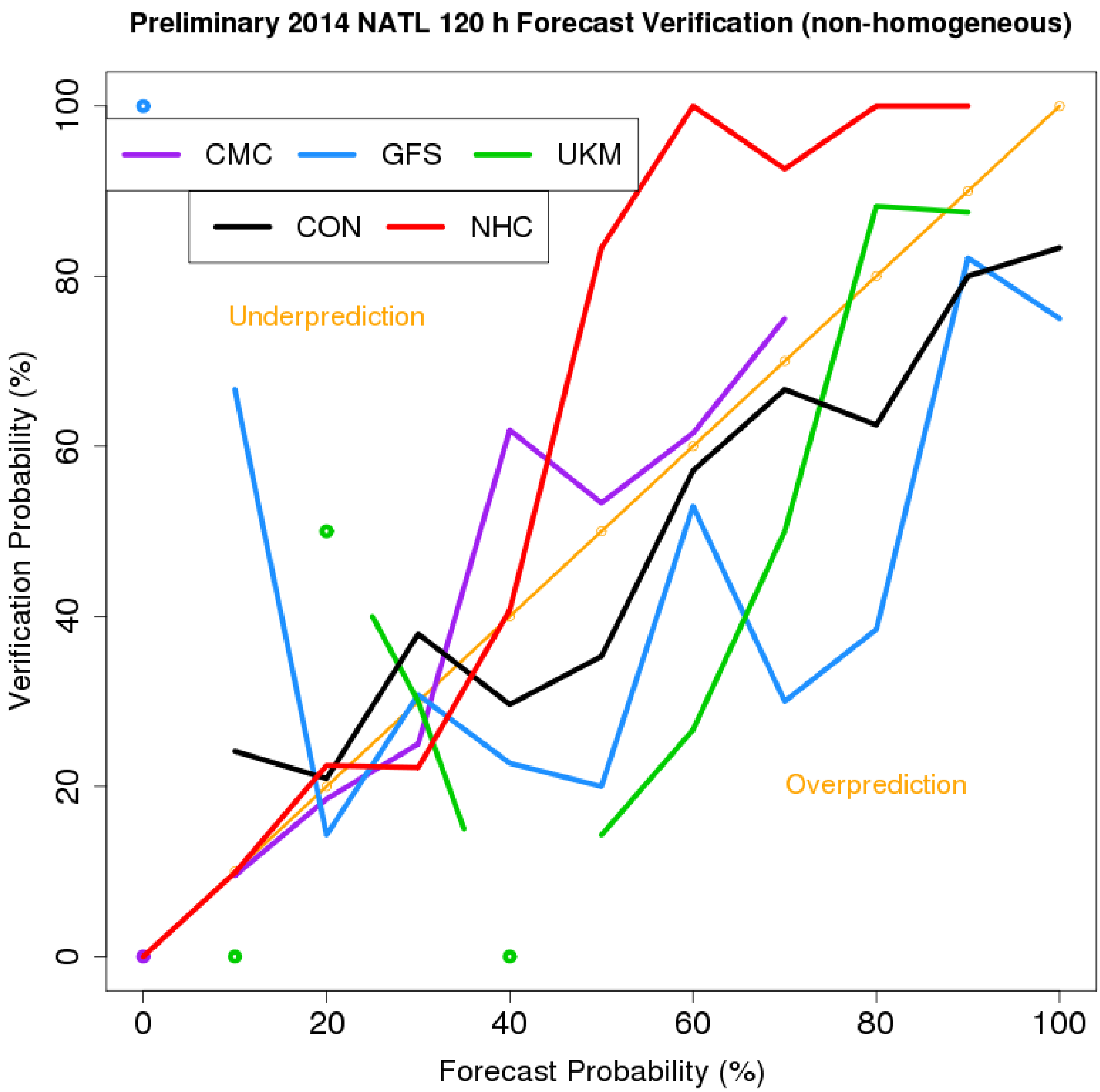


Figure 2. As in Fig. 1, except for NATL 120 h forecasts.

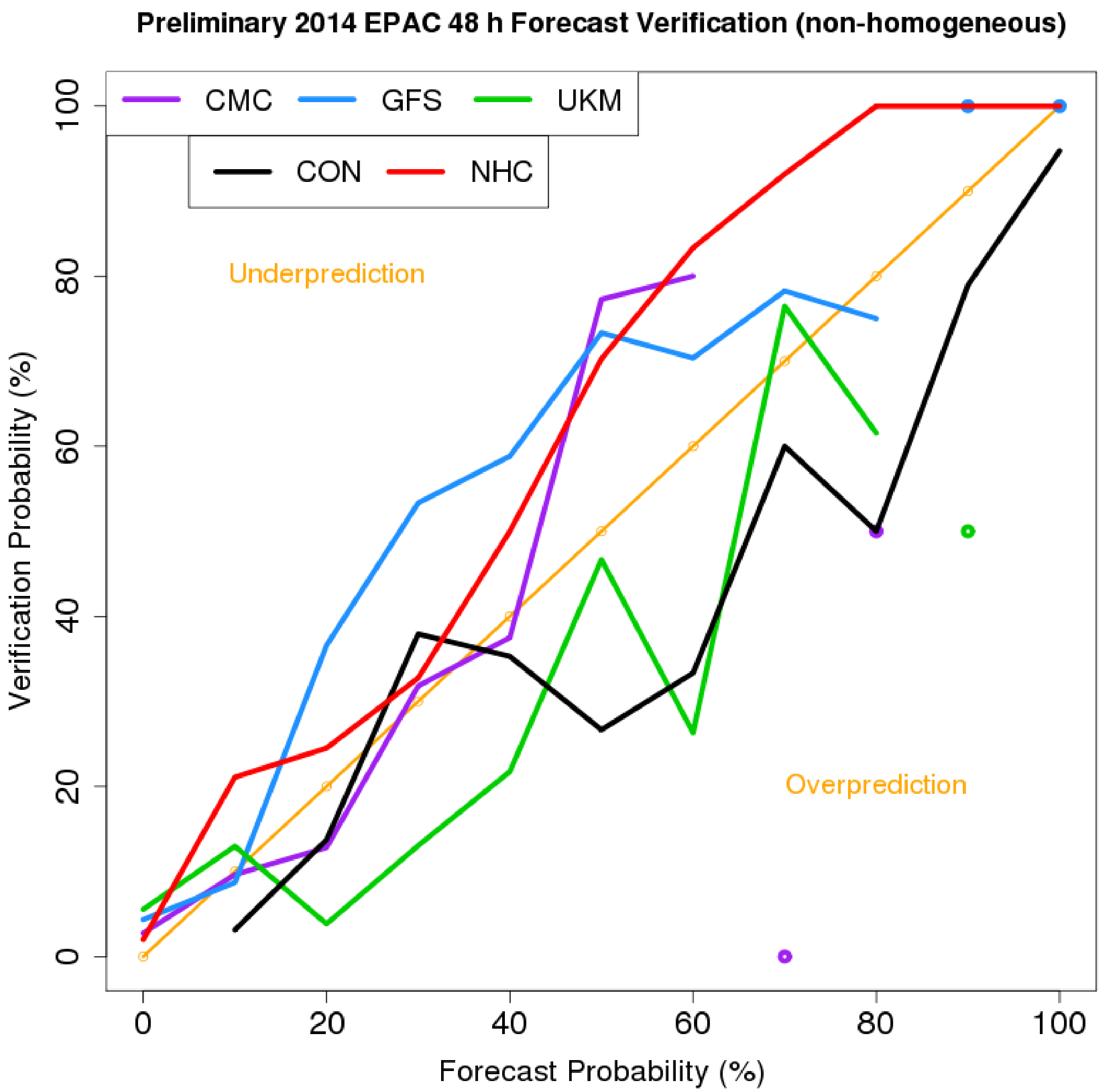


Figure 3. As in Fig. 1, except for EPAC 48 h forecasts.

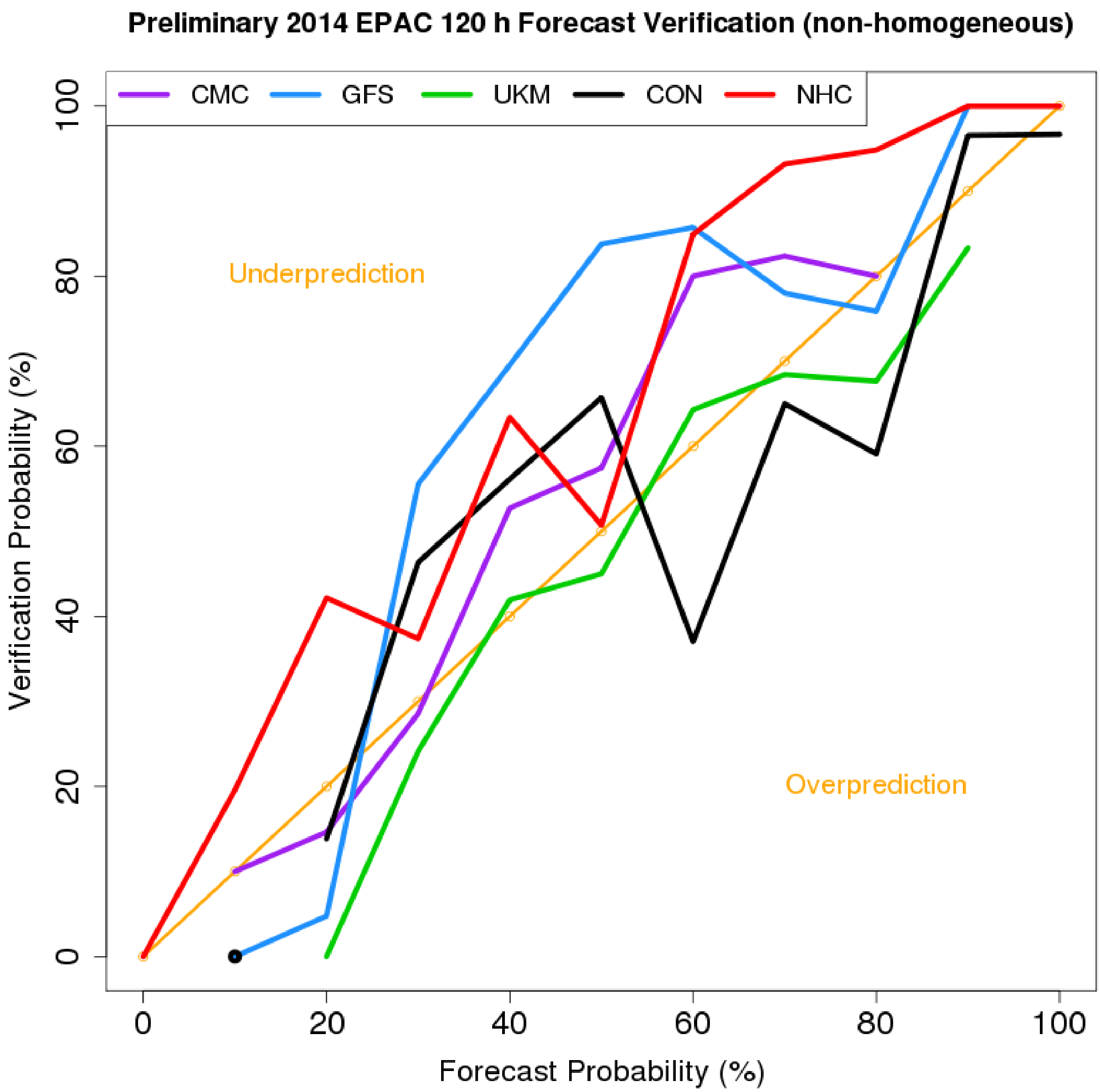


Figure 4. As in Fig. 1, except for EPAC 120 h forecasts.

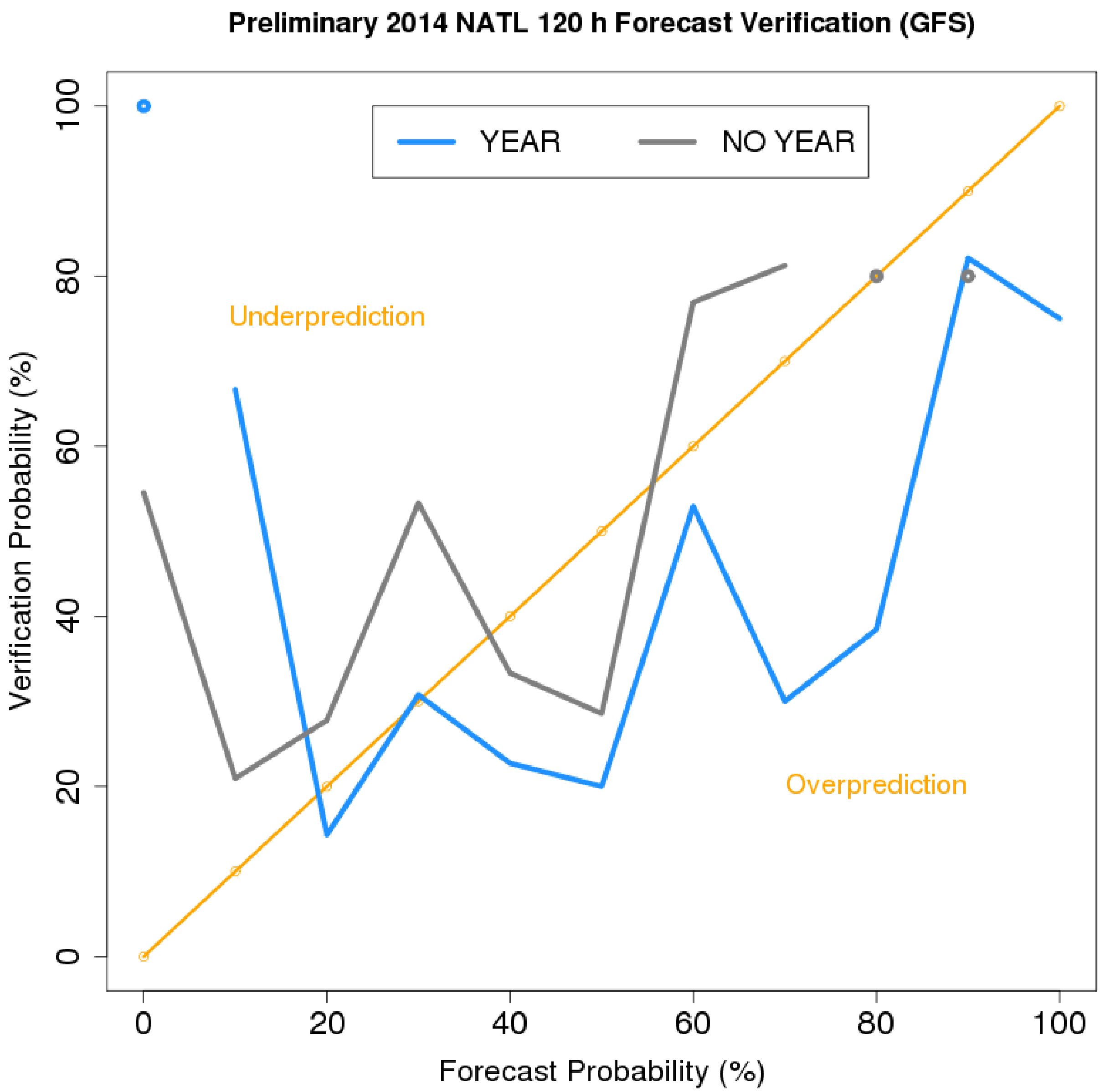


Figure 5. Reliability diagram of the GFS based regression probabilities for the NATL 120 h forecasts. The quasi-operational regression model configuration (blue line) and a re-forecast regression model configuration with “year” removed as a predictor (gray line) are compared.

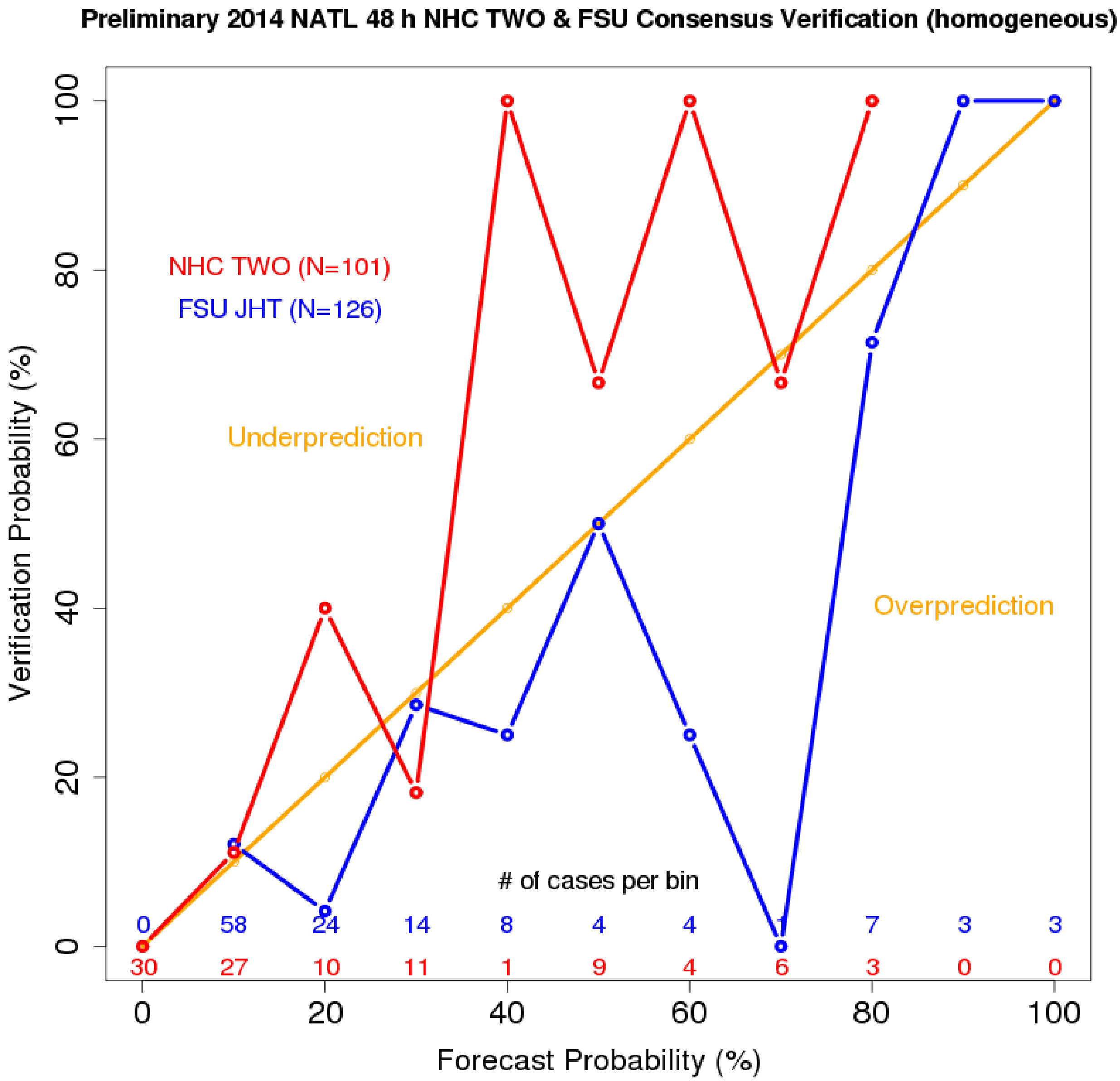


Figure 6. Reliability diagram of 2014 NHC TWO forecasts (red line) and FSU JHT consensus regression model forecasts (blue line). The number of cases in each forecast probability bin are provided at the bottom in the corresponding red and blue text. Verification is preliminary and based on the 2014 working Best-Tracks. The results are based on a homogeneous set of forecasts. "Perfect reliability" is given by the orange, diagonal line; above (below) this line indicates underprediction (overprediction).

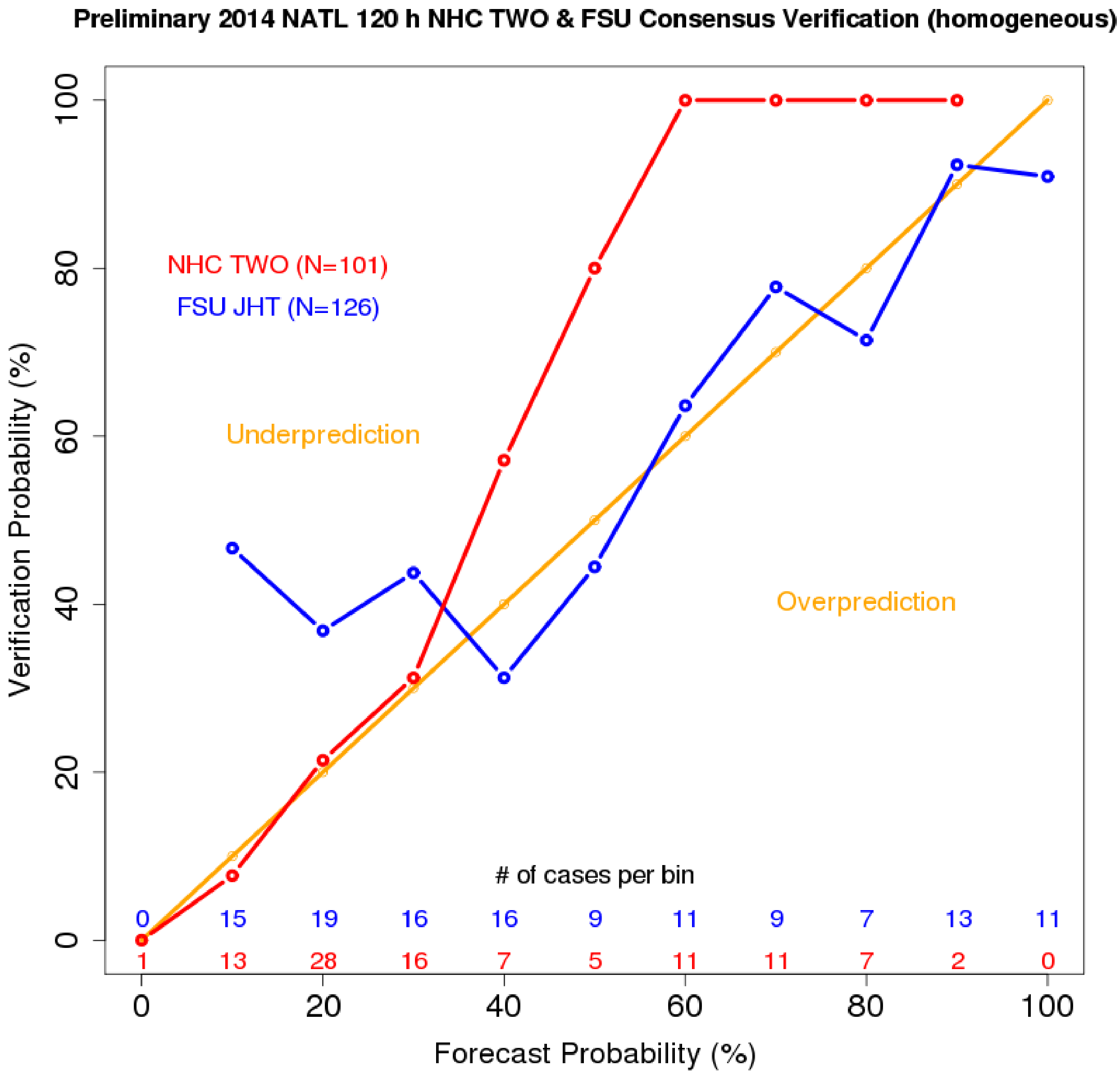


Figure 7. As in Fig. 6, except for the NATL 120 h forecasts.

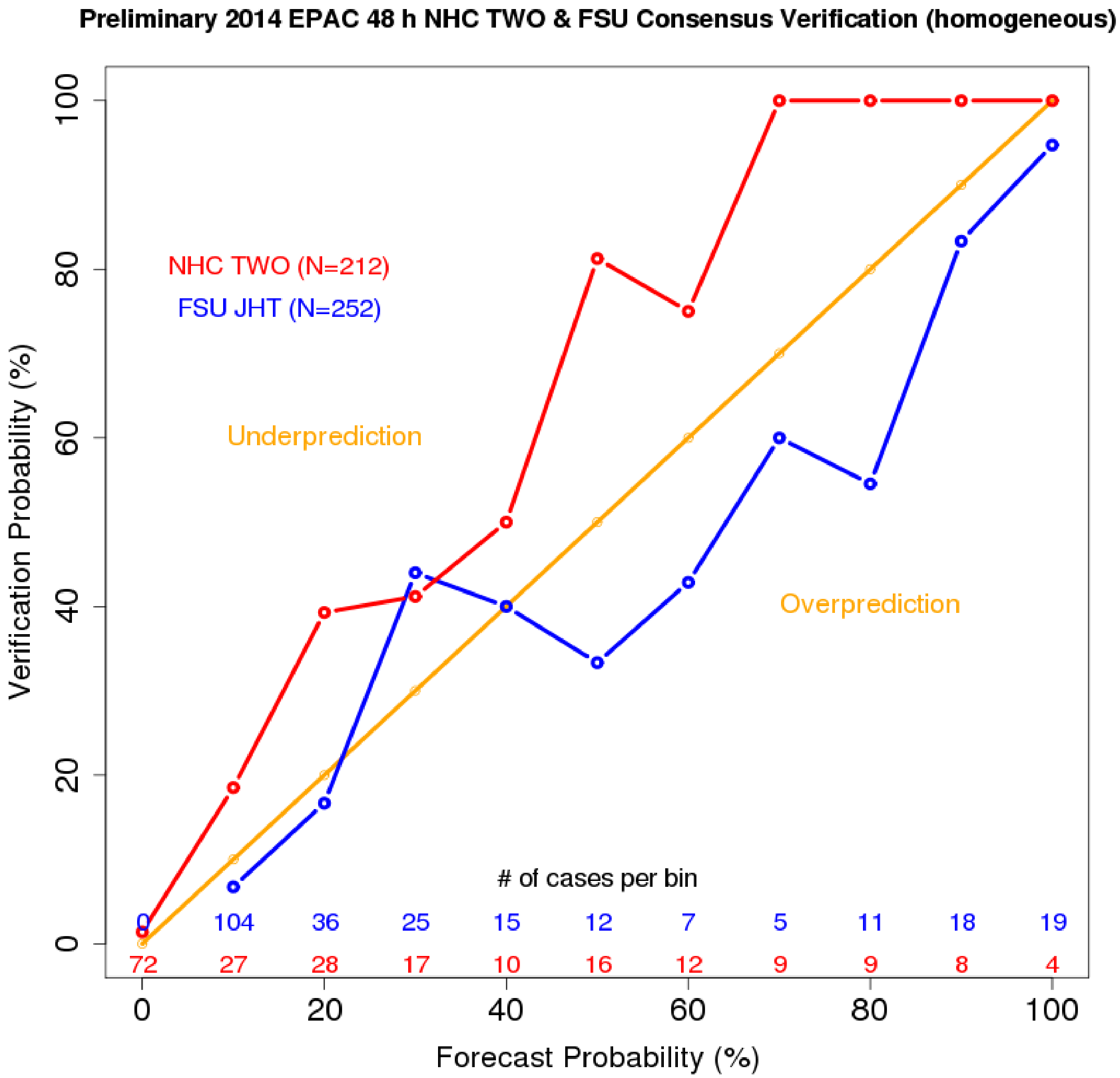


Figure 8. As in Fig. 6, except for the EPAC 48 h forecasts.

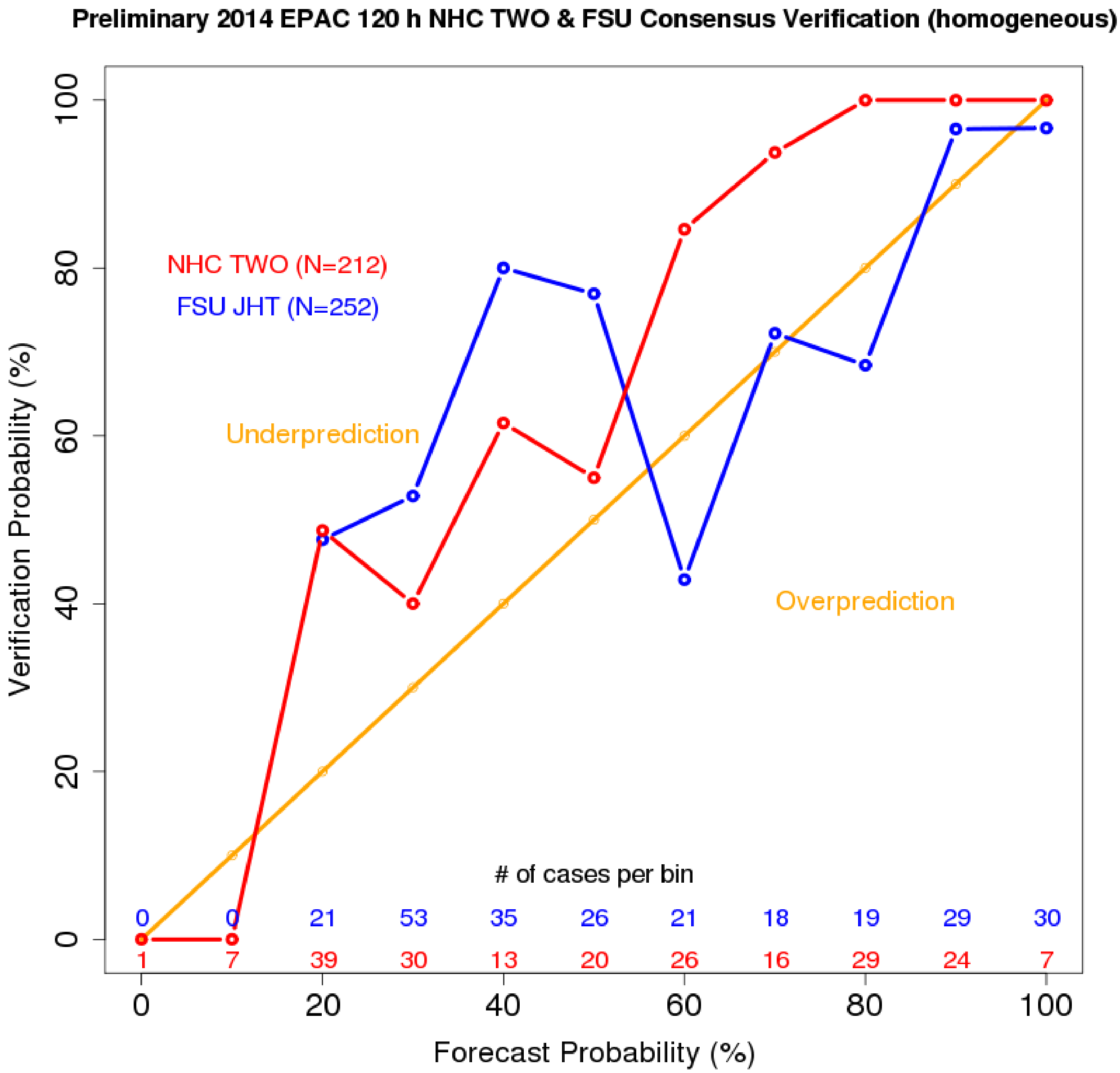
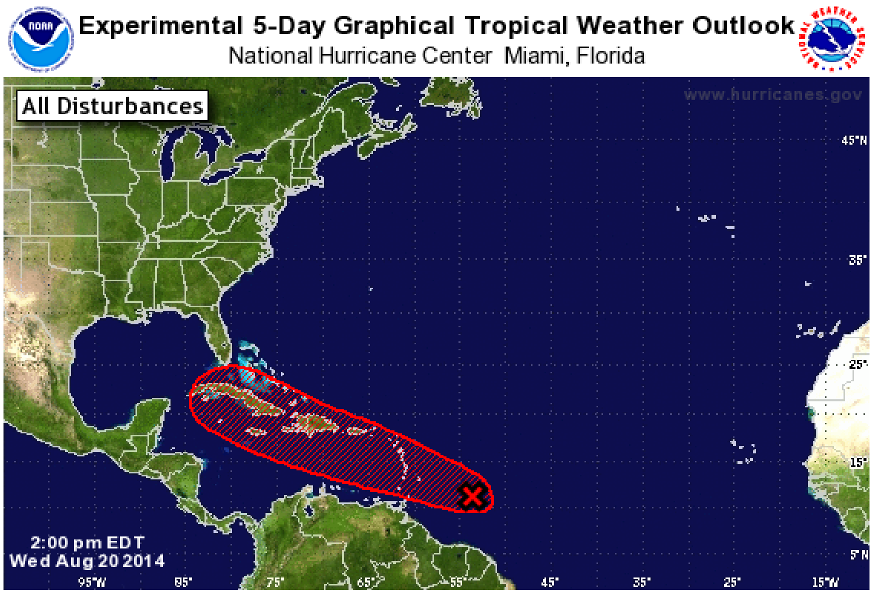


Figure 9. As in Fig. 6, except for the EPAC 120 h forecasts.



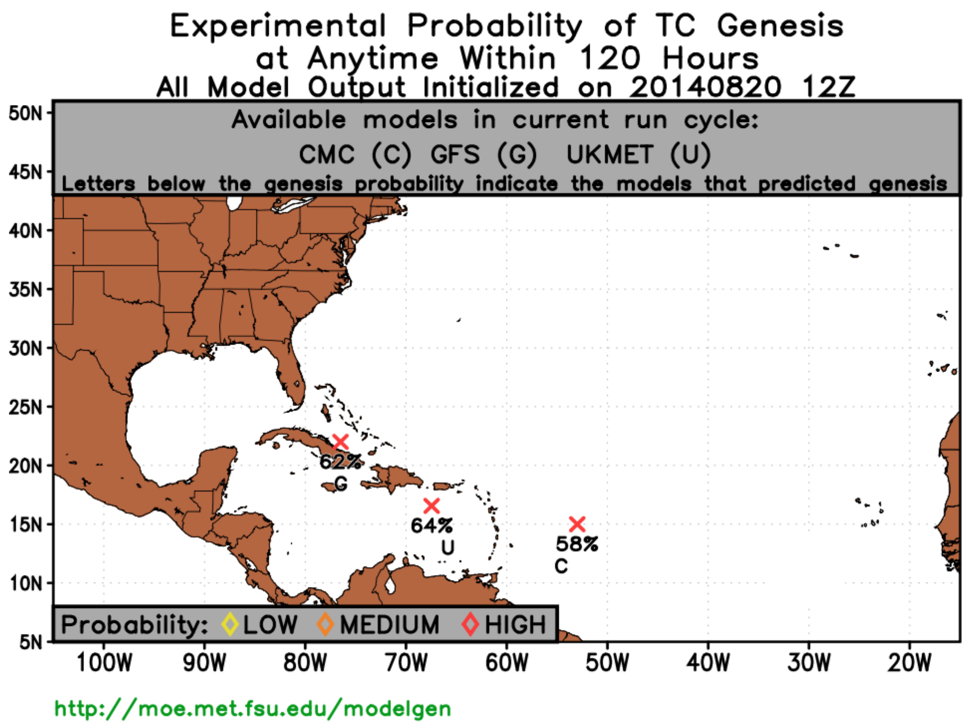


Figure 10. NHC TWO forecast (top) and consensus regression probabilities (bottom). All three global models disagree on the timing and location of genesis for the disturbance highlighted in the TWO. As a result, the consensus tracking algorithm treats each forecast as a separate TC. But, since all three forecast genesis locations are within the TWO shaded region, they are all included in the homogeneous verification comparison.

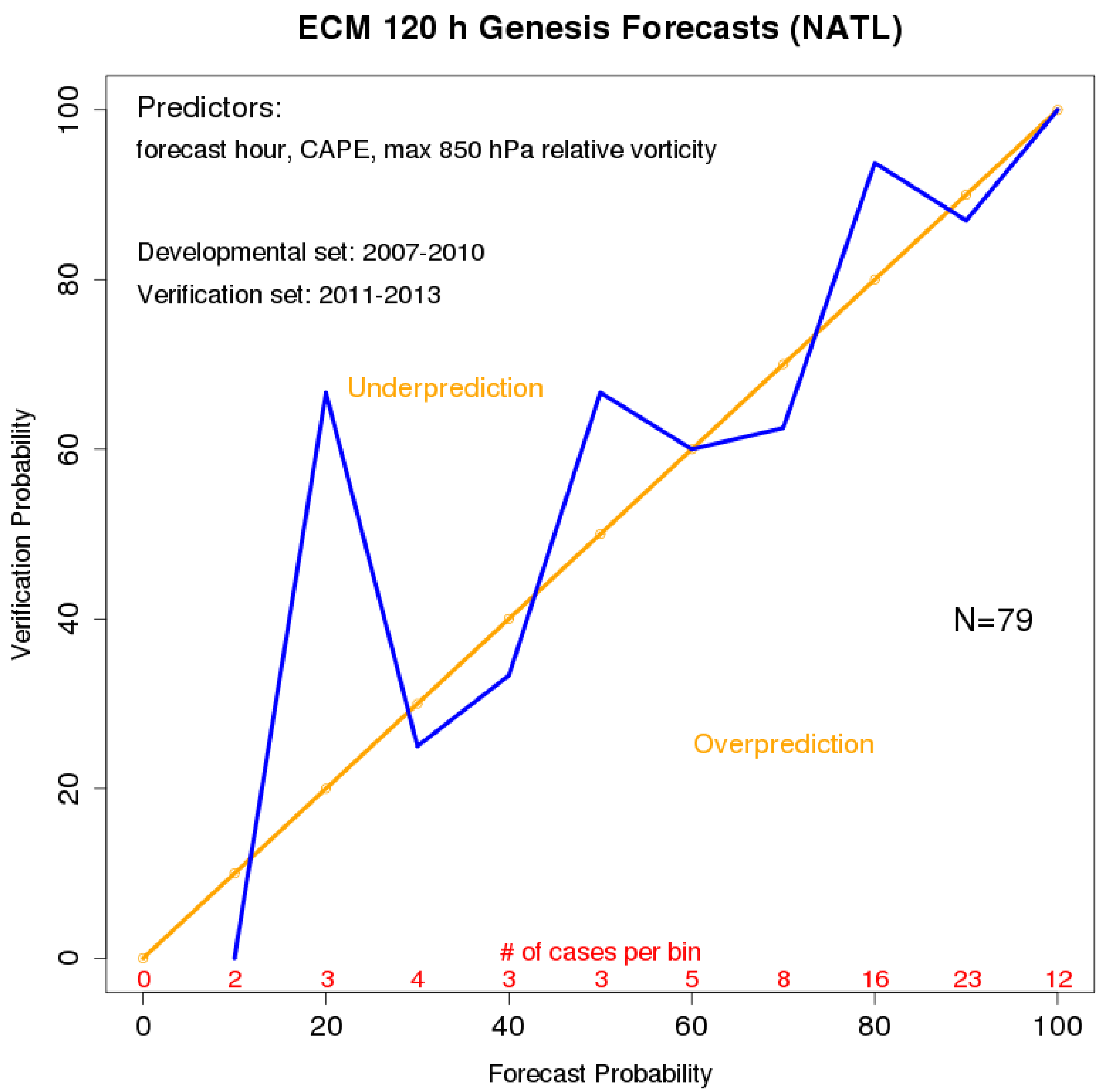


Figure 11. Reliability diagram evaluating the fit of ECMWF based regression model forecasts over the NATL at 120 h. The regression equation was developed with 2007-2010 cases and tested on 2011-2013 cases.