Joint Hurricane Testbed Final Report
1 September 2013 - 31 August 2015
A probabilistic TC genesis forecast tool utilizing an ensemble of global models

Principal investigator: Robert Hart, Florida State University (FSU)
Co-principal investigator: Henry Fuelberg, Florida State University (FSU)
Co-investigator: Daniel Halperin, University at Albany, SUNY
Co-investigator: Joshua Cossuth, Naval Research Laboratory

Project Overview:

A disturbance-specific statistical-dynamical tropical cyclone (TC) genesis forecast tool based on global model output is developed and tested. The tool generates 0-48 h and 0-120 h probabilistic TC genesis forecasts for the North Atlantic (NATL) and eastern North Pacific (EPAC) basins. The goal is for this tool to provide useful guidance to the Hurricane Specialist Unit (HSU) in preparing the Tropical Weather Outlook (TWO). Output from the NCEP Global Forecast System (GFS), U.K. Met Office global model (UKM), and Environment Canada's Global Environmental Multiscale model (CMC) is used.

Accomplishments:

All of the proposed tasks were completed in accordance with the project timeline. The investigators also accommodated additional requests from the HSU during the project cycle to address evolving model developments and NHC-desired online product modifications. Details regarding each accomplishment are provided below.

1. Determined 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 (e.g., relative vorticity, thermal structure, etc.) and to verify the model genesis forecasts against the Best-Track. Accordingly, Fig. 1 shows a performance diagram of 2007-2014 mean model performance over the NATL and EPAC basins. Results show that on average over this eight year period the models perform better over the EPAC than the NATL, due mainly to an increased probability of detection (POD). It is possible that the models are better able to predict TC genesis originating from pathways more commonly found over the EPAC than the NATL (McTaggart-Cowan et al. 2008, 2013), which may explain the difference in POD between the basins. As in the NATL, the best ranking model varies from season to season. It is interesting that although

the POD and False Alarm Ratio (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 insight into the limits of predictability of genesis forecasting (especially over the NATL) or insight into the physics of TC genesis itself.

2. 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. 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), the differing quality of TCs in gridded datasets (Schenkel and Hart 2012), and the differing genesis pathways between the basins (McTaggart-Cowan et al. 2008, 2013), predictors were tested separately for each global model, each basin, and each forecast window (i.e., 48 and 120 h). 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 contained 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 most 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. Again, this process was completed separately for each global model, basin, and forecast window.

A consensus (CON) logistic regression model also was developed. The predictors were the genesis probabilities from each of the single model (i.e., CMC, GFS, UKM) regression equations. If a single model did not predict genesis for a given disturbance, zero was assigned as the corresponding predictor value.

3. Verification from 2014 quasi-operational testing

The regression-based TC genesis probabilities were generated quasi-operationally during 2014. Each regression equation was verified at the conclusion of the 2014 hurricane season. Unless specified otherwise, when referring to a global model, the statistical guidance products based on the referenced global model are being discussed (e.g., when referencing the GFS, the text is discussing the GFS based regression model). The non-homogeneous[1]

[1]This refers to all available results from each technique. For example, cases where NHC was issuing probabilities on a given disturbance that the models did not detect and vice versa are included.

results are presented first using reliability diagrams (Fig. 2). The reliability diagrams show the verification for each single model based regression (color coded), the CON based regression (black line), and NHC TWO forecasts (red line). Breaks in the lines indicate that five or fewer cases are available in a given forecast bin, yielding a sample size that is too small to draw meaningful conclusions. The 48 h NATL verification (Fig. 2a) shows well calibrated forecasts in the 0-20% probability bins for the CMC and GFS based regression models, as well as the NHC TWO forecasts. However, at probabilities $\geq$ 30%, the CMC regression and NHC TWO forecasts underpredict genesis, while the UKM and CON regression models overpredict genesis. For the 120 h NATL forecasts (Fig. 2b), the CMC based regression model is well calibrated, and the CON regression model only slightly overpredicts 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 the 48 h EPAC forecasts (Fig. 2c). While the CMC and CON regression models and NHC forecasts perform well in the 0-40% range, they stray from the perfect reliability line at the higher probability bins. The NHC forecasts, along with the CMC and GFS regression models underpredict genesis, while the CON and UKM regression models overpredict genesis. For the 120 h EPAC forecasts (Fig. 2d), the regression models and NHC TWO forecasts generally underpredict genesis, but the UKM regression model is well calibrated.

Overprediction by the UKM (Figs. 2a,b) may be due in part to a new global model configuration that was implemented operationally during July 2014. Heming (2014) noted that reforecasts of TCs using the new UKM global model configuration generally yielded stronger forecast intensities of mature TCs compared to the prior configuration. While the impact to genesis forecasts was not explicitly discussed, it is possible that the new configuration of the UKM global model 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 the guidance suite (Table 1) undoubtedly impact the reliability of the regression equations. The relative stability of the CON probability should help mitigate the impact of global model upgrades. Also, the season-to-date verification product can show how the reliability of a particular regression equation changes following a global model upgrade.

The 120 h GFS based regression for the NATL stands out due to its especially poor reliability (Fig. 2b). The use of year as a predictor was a contributing factor. While the developmental set for the logistic regression model did indicate an improvement in GFS global model TC genesis forecasts over time, there was no guarantee that these improvements would continue during 2014. Indeed, the GFS global model success ratio during 2014 was less than during 2010-2013. Thus, the GFS based regression probabilities were unnecessarily inflated by including year as a predictor. Figure 3 compares the verification of the quasi-operational GFS based regression for the NATL at 120 h (blue line) with a regression model configuration that excluded year as a predictor (gray line). While still far from 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. Given its lack of physical relevance and predictive power, year will not be a predictor in future versions of the regression equations.

Historical verification indicates that the false alarm ratio for the CMC global model is greater than for the other global models (Fig. 1). However, the CMC based regression equations are among the best calibrated (Fig. 2), implying that the global model contains common/repeatable biases that the regression equations are able to correct.

To provide a more direct comparison of the verification results, a set of homogeneous[2] NHC TWO and CON regression forecasts was constructed. The associated reliability diagrams are presented in Fig. 4. The verification is quite comparable at the 10% and 30% forecast probability bins for the NATL 48 h forecasts (Fig. 4a). For probabilities exceeding 30%, the sample size – given by the blue (CON) and red (NHC TWO) text – is too small. Sample size is not an issue at 120 h (Fig. 4b). The NHC TWO forecasts outperform the CON regression model in the 0-30% forecast probability range. However, at the higher probability bins, the CON regression model is better calibrated. Over the EPAC at 48 h (Fig. 4c), NHC TWO forecasts underpredict genesis, while the CON regression model generally overpredicts genesis. At 120 h, the CON regression model struggles in the 20-50% forecast probability range, but is fairly well calibrated in the 70-100% range (Fig. 4d). The NHC TWO forecasts underpredict genesis in the 20-90% range.

The reader may be wondering why the sample sizes of the TWO and CON 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 5 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 three CON genesis forecasts and one TWO forecast contributing to the homogeneous verification. This issue likely is not significant, but generally causes the lower forecast probability bins to contain more cases and to underpredict genesis, and causes the higher forecast probability bins to have fewer cases and potentially overpredict genesis.

4. 2015 regression equation development

Prior to quasi-operational testing during 2015, all regression models were re-calibrated to determine whether any predictors should be added or removed after the 2014 forecasts were added to the developmental dataset. Regardless of potential significance, year was

[2]This indicates that only cases where both NHC and the CON regression were issuing probabilities for the same genesis event are included.

removed from the GFS based regression model since there was no guarantee that improved GFS genesis forecasts would occur during 2015. Tables 2-4 provide a list of the predictors included in the regression equations used for 2015 quasi-operational testing.

5. Preliminary verification from 2015 quasi-operational testing

Preliminary non-homogeneous verification of the 2015 regression equations was conducted using the working (i.e., preliminary) Best-Track files (Fig. 6). The 48 h NATL verification (Fig. 6a) shows well calibrated forecasts from NHC as well as the CMC and CON regression equations. The UKM regression equation overpredicts genesis, while the GFS regression equation generally underpredicts genesis. The sample size is small at the higher forecast probability bins for the regression equations. NHC and the CMC regression equation again are well calibrated for the 120 h NATL forecasts (Fig. 6b). The CON regression equation is reliable at the 10-40% and 90% forecast probability bins, but overpredicts genesis at the 50-60% and 100% probability bins. The UKM (GFS) generally overpredicts (underpredicts) genesis, similar to at 48 h.

The regression equations generally are better calibrated over the EPAC compared to the NATL. The CMC, GFS, and UKM all are well calibrated for the 48 h EPAC forecasts, except for overprediction at the 50% forecast probability bin (Fig. 6c). The CON regression equation overpredicts genesis at the mid and high forecast probability bins. NHC's forecasts are quite reliable. At 120 h, the CMC, UKM, and CON regression equations are well calibrated (Fig. 6d). The GFS regression equation and NHC's TWO forecasts generally underpredict genesis.

Preliminary verification for the homogeneous NHC TWO and CON regression equation set also was conducted (Fig. 7). With NHC's increased use of the guidance products in 2015 (compared to 2014), this comparison is not independent and it is increasingly difficult for the CON regression equation to outperform NHC's TWO forecasts. Small sample size precludes meaningful conclusions for the 48 h NATL forecasts (Fig. 7a). The CON regression equation is reliable in the 20-60% forecast probability range for the 120 h NATL forecasts (Fig. 7b). Unlike 2014, NHC's TWO forecasts are better calibrated than CON at the high forecast probabilities. NHC's forecasts also are very reliable for the 48 h EPAC forecasts in the 80-100% probability range (Fig. 7c). Otherwise, NHC (CON) generally underpredicts (overpredicts) genesis for this time period. The verification for the 120 h EPAC forecasts (Fig. 7d) is quite similar for both sets of forecasts, with a general underprediction bias.

6. Quasi-operational products tested during 2014 and 2015

The simplest way to output the quasi-operational products for viewing by the Hurricane Specialist Unit is via a website hosted by the PI at FSU (http://moe.met.fsu.edu/modelgen). The following products currently are available on the site:

- Overview of each basin that shows the location and categorical 0-48 h and 0-120 h

genesis probability of each model-indicated TC as well as the models available in the current initialization cycle (displayed as a graphic).

- 0-48 h and 0-120 h genesis probabilities for each model-indicated TC (displayed both graphically and via text). This is available for each individual model and for the multi-model consensus.

- Model-indicated tracks (according to the developed tracking algorithm) for each model-indicated TC out to 144 h (graphic and text; Fig. 8).

- The values of the criteria for defining a TC—including whether the values exceed the threshold required—for each 6 h forecast interval (text; Fig. 8).

- The values of each predictor used in the regression equations (text; Fig. 8).

- A history of the forecast genesis time, location, and probabilities for each model-indicated TC (text; Fig. 9).

- Real-time season-to-date verification of each regression model (reliability diagrams and geographical plots).

- Historical verification of each regression model using 2011-2013 and 2013 only as the verification set (reliability diagrams).

- An archive of all images.

- A brief description of each product.

7. ECMWF regression equations developed and tested

Real-time ECMWF forecast fields were not available during the 2014 and 2015 quasi-operational testing that was conducted at FSU. In preparation for the possibility that this project will be selected for operational implementation, regression models were developed and tested based on the historical ECMWF genesis forecasts[3]. 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, which is somewhat surprising given how intense mature TCs can be depicted in the analysis and forecast fields. 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. 10. Note that the sample size over the three year independent period still is fairly small. Given the ECMWF's low false alarm ratio, most of the regression based genesis probabilities are at or above 70%. In the 70-100% forecast probability range, the regression model is

[3]Historical ECMWF data were obtained from the TIGGE archive.

fairly well calibrated. However in the 0-60% forecast probability range, the small sample size precludes meaningful conclusions.

8. Case study of Sean (2011)

During a meeting at NHC, members of the HSU recommended interesting case studies. Several of these case studies were TCs of baroclinic origins that the developed tracking algorithm did not detect. Since the genesis criteria were specifically calibrated to detect purely tropical cyclones, it is not surprising that the guidance products struggled to detect tropically transitioning cyclones. Sean (2011) proved to be an interesting case study and the analysis guided potential future modifications to the genesis criteria to better detect tropical transition.

The GFS correctly captured many aspects of the forecast for Sean, but the genesis criteria were never met because the 250-850 hPa thickness threshold was never satisfied. The large pre-Sean disturbance was easily trackable several days prior to tropical transition (1800 UTC 8 November 2011) in the GFS fields. A well defined shortwave moved off the southeast coast of the U.S. and became cut off from the longwave trough. The occluded cyclone meandered over SSTs $> 26°$C for several days over the northwest Atlantic, consistent with the behavior of many tropically transitioning cyclones (Davis and Bosart 2003, 2004). The 200-850 hPa wind shear decreased notably very near the time of tropical transition, which also is consistent with Davis and Bosart (2003, 2004).

Thickness forecasts from the early GFS model runs (Figs. 11a, 11b) did not show the cold to warm core transition. However, mid range runs (Fig. 11c) did exhibit some larger thickness values aligning with the surface low. A few runs just before the tropical transition time showed the development of a weak warm core (Fig. 11d). Forecast 200-850 hPa shear at the tropical transition time decreased with successive model cycles (Fig. 12), also indicating that the GFS struggled to simulate tropical transition early, but began to show signs of the transition with ~ 2 days lead time.

The 600 hPa relative humidity (RH) field also was analyzed. The early model runs showed large RH values away from the low level center, but small RH close to the center (Fig. 13a). This decreased moisture near the center may imply reduced convection, a lack of resulting latent heat release, and thus no warm core development. In short, the early runs appear to show a structure that is more typical of subtropical cyclones (Hart 2003; Evans and Guishard 2009; Guishard et al. 2009). In the later runs (where a weak warm core was observed), large RH values wrapped around the entire low level center (Fig. 13d). This short lead time, or lag in the GFS predicting tropical transition, may be similar to a bias found in reanalyses, where prepeak TC intensification occurred more slowly than in the Best-Track (Schenkel and Hart 2012).

Despite these signs of tropical transition in the later GFS runs, the genesis criteria never were met because the thickness threshold value was not exceeded. One possible way to increase detection of tropical transition events may be to include SST (or skin tempera-

ture) and wind shear criteria (consistent with the forecast rules of Davis and Bosart (2004)). Another possibility would be to use the parameters of the cyclone phase space (Hart 2003) to determine whether a cyclone has become tropical. Such modifications to the genesis criteria only would be applied to cyclones that meet all requirements except the thickness threshold.

9. Comparison of operational and reforecast GFS configurations

A new version of the GFS was implemented operationally during early 2015. Reforecasts (i.e., retrospective runs) of the 2011-2014 TC seasons were available for model evaluation. Some members of the HSU expressed interest in how the 2015 GFS handled TC genesis relative to previous operational configurations. Thus, a comparison of 2011-2014 operational GFS genesis forecasts and 2015 GFS reforecasts of the 2011-2014 seasons was conducted[4]. Several measures were taken to ensure a homogeneous comparison:

- Both datasets were on 0.5° output grids.
- Both datasets used the same genesis criteria, including identical threshold values. The threshold values were calibrated from 2008-2010 Best-Track tropical depressions, so the 2011-2014 seasons were independent of this calibration.
- The same code was used to identify and verify the forecast genesis events from both datasets.
- Only model cycles where data were available from both datasets were included.

Figures 14 and 15 show the results from this comparison. The reforecasts for 2011 yielded a greater success ratio/smaller false alarm ratio over the North Atlantic. *Otherwise, the operational GFS forecasts outperformed the reforecasts in probability of detection, success ratio, and critical success index* (Fig. 14). Fewer false alarms were forecast over the Northwest Atlantic and – to a lesser extent – over the western Caribbean Sea in the reforecast dataset. The 2015 reforecasts have a large number of false alarms over the eastern main development region, consistent with the operational GFS. Slightly more false alarms were forecast over the Gulf of Mexico in the reforecast dataset compared to the operational GFS (Fig. 15).

The reforecast genesis events over the eastern North Pacific generally yielded a smaller probability of detection, but greater success ratio. The critical success index is lower for the reforecast dataset compared to the operational GFS (Fig. 14). The smaller probability of detection is evident in the geographic plots (Fig. 15).

These results were presented at a NOAA EMC Model Evaluation Group discussion during September 2015. Feedback from EMC was positive and the investigators have been asked to conduct a similar analysis on the GFS configuration that is scheduled to be implemented operationally in 2016.

[4]Reforecast model data were provided by NHC using both PI and NHC owned external hard drives.

Conducting the 2015 reforecast verification also provided the opportunity to update the developmental set of our GFS based regression equations to include the reforecasts. Theoretically, a regression equation with these reforecasts in the developmental set should produce well calibrated probabilistic genesis forecasts since the regression equation will be applied to the same global model configuration in real-time (i.e., the independent data). However, attempts to develop regression equations with only the reforecasts as the developmental set failed to produce a well calibrated set of forecasts. This likely is due to an insufficient sample size.

10. Code provided to JHT staff

The scripts for the GFS and UKM guidance products have been provided to JHT staff to test for compatibility on NHC's IT platform. The investigators will continue to assist with this testing and provide additional components of the program at NHC's request. The investigators will be available to assist with the transition to operations should the project be accepted for implementation.

Comments regarding quasi-operational testing

The relatively quiet TC seasons over the NATL during 2014-2015 made it somewhat difficult to evaluate the reliability of the real-time products. The active EPAC, however, provided an excellent opportunity to test the real-time guidance. The verification statistics for 2014-2015 quasi-operational testing show promise regarding the usefulness of the guidance probabilities.

Products usually are available 30-75 min prior to the synoptic/TWO issuance time, though some operational challenges were encountered. For example, the increased resolution and output file size of the upgraded UKM global model during 2014 initially caused the UKM based guidance products to be generated after the synoptic/TWO time. After consulting directly with Julian Heming of UKMO, some modifications at both his end (converting files from GRIB1 to GRIB2) and at FSU (cron timing) 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, no changes were made to the UKM based regression models. The mid-season global model upgrade was an opportunity to test the robustness of the regression models.

The 2015 upgrade of the GFS presented similar challenges as the output grids changed from 0.5° to 0.25°. However, a few enhancements to the TC identification script resulted in the GFS based guidance products being available at least 30 min prior to the synoptic/TWO time. The biggest operational challenge during 2015 was beyond our control. During the latter part of the season, the download speed was quite slow for acquiring GFS fields from

NCEP and was noted by several emails released to the NCEP email listserv. The later-than-usual arrival of the GFS output fields caused the GFS and CON guidance products to complete after the synoptic/TWO issuance time. This should not be a problem if the product is selected for operational implementation. A summary of typical product generation times is given in Table 5.

Summary

All proposed tasks – as well as additional tasks requested by the HSU during the project cycle – have been completed. A real-time statistical-dynamical genesis guidance tool based on global model output was developed and tested quasi-operationally during 2014 and 2015. The products were formatted for easy interpretation by the HSU to accommodate their time constraints. The genesis probabilities were available at least 30 minutes prior to the TWO issuance time for $\sim 95\%$ of the model cycles. The primary challenge to operational reliability (slow download times for GFS fields from NCEP) should not be an issue if the project is chosen for operational implementation. Verification of the quasi-operational products revealed generally well calibrated probabilistic genesis forecasts.

The investigators very much appreciate the productive discussions with the NHC points-of-contact. The project results would not be possible without those exchanges. Their feedback has contributed to a better, more reliable guidance tool.

Publications

Building upon Halperin et al. (2013):

Halperin, D.J., H.E. Fuelberg, R.E. Hart, J.H. Cossuth, 2015: Verification of tropical cyclone genesis forecasts from global numerical models: Comparisons between the North Atlantic and eastern North Pacific basins. *Wea. Forecasting*, in review.

Halperin, D.J., R.E. Hart, H.E. Fuelberg, J.H. Cossuth, 2016: The development and evaluation of a statistical-dynamical tropical cyclone genesis guidance tool. *Wea. Forecasting*, in preparation.

Halperin, D.J., J.H. Cossuth, A. Penny, R.E. Hart, 2016: Quantifying the impact of model upgrades on tropical cyclone genesis forecasts: Comparisons between the GFS operational and reforecast configurations. *Mon. Wea. Rev.*, in preparation.

Presentations

Halperin, D.J., R.E. Hart, H.E. Fuelberg, J.H. Cossuth, 2015: Experimental tropical cyclone genesis forecast guidance products using global models. 2015 Hurricane Forecast Improvement Program Annual Review Meeting, Miami, FL, 17-19 November 2015. (invited)

Halperin, D.J., R.E. Hart, H.E. Fuelberg, J.H. Cossuth, 2015: A probabilistic tropical cyclone genesis guidance tool using global numerical models. NOAA Virtual Lab Forum, 21 October 2015. (invited)

Halperin, D.J., R.E. Hart, H.E. Fuelberg, J.H. Cossuth, 2015: GFS tropical cyclone genesis verification. NOAA EMC Model Evaluation Group discussion, 24 September 2015. (invited)

Halperin, D.J., R.E. Hart, H.E. Fuelberg, J.H. Cossuth, 2015: Developing a tropical cyclone genesis forecast tool: Preliminary results from 2014 quasi-operational testing. *69th Interdepartmental Hurricane Conference*, Miami, FL, 2-5 March 2015, S03a-04.

Halperin, D.J., R.E. Hart, H.E. Fuelberg, J.H. Cossuth, 2015: Developing a tropical cyclone genesis forecast tool using global model output. *5th Conference on Transition of Research to Operations*, Amer. Meteo. Soc., Phoenix, AZ, 4-8 January 2015, J2.2.

Halperin, D.J., R.E. Hart, H.E. Fuelberg, J.H. Cossuth, 2014: Determining predictors for a statistical tropical cyclone genesis tool based on GFS output. *31st Conference on Hurricanes and Tropical Meteorology*, Amer. Meteo. Soc, San Diego, CA, 31 March-4 April 2014, 1C.7.

Halperin, D.J., H.E. Fuelberg, R.E. Hart, J.H. Cossuth, 2014: Verification of TC genesis forecasts from global models: Updates and real-time applications. *31st Conference on Hurricanes and Tropical Meteorology*, Amer. Meteo. Soc, San Diego, CA, 31 March-4 April 2014, P146.

Halperin, D.J., R.E. Hart, H.E. Fuelberg, J.H. Cossuth, 2014: Tropical cyclone genesis: Verifying historical forecasts and developing a statistical model from NWP output. *68th Interdepartmental Hurricane Conference*, Miami, FL, 3-6 March 2014, S02-08.

Halperin, D.J., H.E. Fuelberg, R.E. Hart, J.H. Cossuth, 2013: Determining relevant parameters for a statistical tropical cyclone genesis tool based upon global model output. *2013 AGU Fall Meeting*, San Francisco, CA, 9-13 December 2013, A23E-0324.

Meetings with NHC points-of-contact

- November 2013: Present research completed to date; discuss methodology for future

research; discuss desired output products and operational constraints; identify potentially interesting case studies.

- March 2014: Provide a progress report during the Interdepartmental Hurricane Conference (IHC); receive feedback on sample output products for 2014 quasi-operational testing.

- March 2015: Provide a progress report during the IHC; discuss the operational utility/reliability of the guidance products during 2014; receive feedback on proposed changes for 2015 quasi-operational testing; receive request to conduct the GFS operational vs reforecast verification comparison; discuss IT compatibility issues and transitioning code for testing on NHC's JHT desktop.

Data Sources and Computation Resources

- All disk space and computing resources to obtain the global model output and run the forecast tool were volunteered by the PI.

- CMC, GFS operational: PI's local archive as well as TIGGE and NOMADS archives.

- UKM: Julian Heming of the UK Met Office.

- ECM: TIGGE archive (ECMWF portal).

- GFS reforecasts: Andy Penny, Mike Brennan, Richard Pasch (all NHC).

References

Davis, C. A., and L. F. Bosart, 2003: Baroclinically induced tropical cyclogenesis. *Mon. Wea. Rev.*, **131 (11)**, 2730–2747.

Davis, C. A., and L. F. Bosart, 2004: The TT problem: Forecasting the tropical transition of cyclones. *Bulletin of the American Meteorological Society*, **85 (11)**, 1657–1662.

Evans, J. L., and M. P. Guishard, 2009: Atlantic subtropical storms. Part I: Diagnostic criteria and composite analysis. *Mon. Wea. Rev.*, **137 (7)**, 2065–2080.

Guishard, M. P., J. L. Evans, and R. E. Hart, 2009: Atlantic subtropical storms. Part II: Climatology. *J. Climate*, **22 (13)**, 3574–3594.

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 (6)**, 1423–1445.

Hart, R. E., 2003: A cyclone phase space derived from thermal wind and thermal asymmetry. *Mon. Wea. Rev.*, **131 (4)**, 585–616.

Heming, J. T., 2014: The impact on tropical cyclone predictions of a major upgrade to the Met Office global model. *Preprints, 31st Conf. on Hurricanes and Tropical Meteorology, San Diego, CA, Amer. Meteor. Soc.*

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). *Mon. Wea. Rev.*, **136 (4)**, 1284–1304.

McTaggart-Cowan, R., T. J. Galarneau Jr, L. F. Bosart, R. W. Moore, and O. Martius, 2013: A global climatology of baroclinically influenced tropical cyclogenesis. *Mon. Wea. Rev.*, **141 (6)**, 1963–1989.

Sauerbrei, W., C. Meier-Hirmer, A. Benner, and P. Royston, 2006: Multivariable regression model building by using fractional polynomials: description of SAS, STATA and R programs. *Computational Statistics & Data Analysis*, **50 (12)**, 3464–3485.

Schenkel, B. A., and R. E. Hart, 2012: An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets. *Journal of Climate*, **25 (10)**, 3453–3475.

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

Table 1: Horizontal grid spacing of the global models since 2005. Multiple model physics and data assimilation changes also occurred during this time period.

Year	CMC	GFS	UKM
2005	100 km	35 km	40 km
2006	33 km	35 km	40 km
2010	33 km	27 km	25 km
2014	33 km	27 km	17 km
2015	33 km	13 km	17 km

Table 2: 2015 predictors with coefficient signs for CMC based regression models.

NATL 48 h	NATL 120 h	EPAC 48 h	EPAC 120 h
- forecast hour	- forecast hour	- forecast hour	- forecast hour
+ 250-850 hPa thickness	- latitude	+ latitude	+ latitude
- latitude	- longitude		+ 1000-700 hPa lapse rate
- longitude			- 850 hPa divergence

Table 3: As in Table 2, except for GFS based regression models.

NATL 48 h	NATL 120 h	EPAC 48 h	EPAC 120 h
- forecast hour	- forecast hour	- forecast hour	- forecast hour
+ 250-850 hPa ΔZ	- 200-850 hPa shear	+ latitude	+ latitude
- longitude	+ 250-850 hPa ΔZ	+ sfc latent heat flux	+ sfc latent heat flux
+ sfc latent heat flux	- 850 hPa ζ'		- 850 hPa ζ'
- 850 hPa ζ'	+ sfc latent heat flux		+ CAPE

Table 4: As in Table 2, except for UKM based regression models.

NATL 48 h	NATL 120 h	EPAC 48 h	EPAC 120 h
- forecast hour	- forecast hour	- forecast hour^2	- forecast hour
- longitude	- longitude	- latitude^{-1}	+ latitude
	- latitude^{-2}	+ longitude	
	+ latitude^{-2} $\times$ ln(latitude)		

Table 5: Typical guidance product generation times (cycle) in UTC.

Model	Guidance products available (cycle)	Guidance products available (cycle)
	NATL	**EPAC**
CMC	0501 (00); 1701 (12)	0458 (00); 1658 (12)
GFS	0516 (00); 1116 (06)	0453 (00); 1053 (06)
	1716 (12); 2316 (18)	1653 (12); 2253 (18)
UKM	0515 (00); 1715 (12)	0457 (00); 1657 (12)
CON	0520 (00); 1720 (12)	0500 (00); 1700 (12)

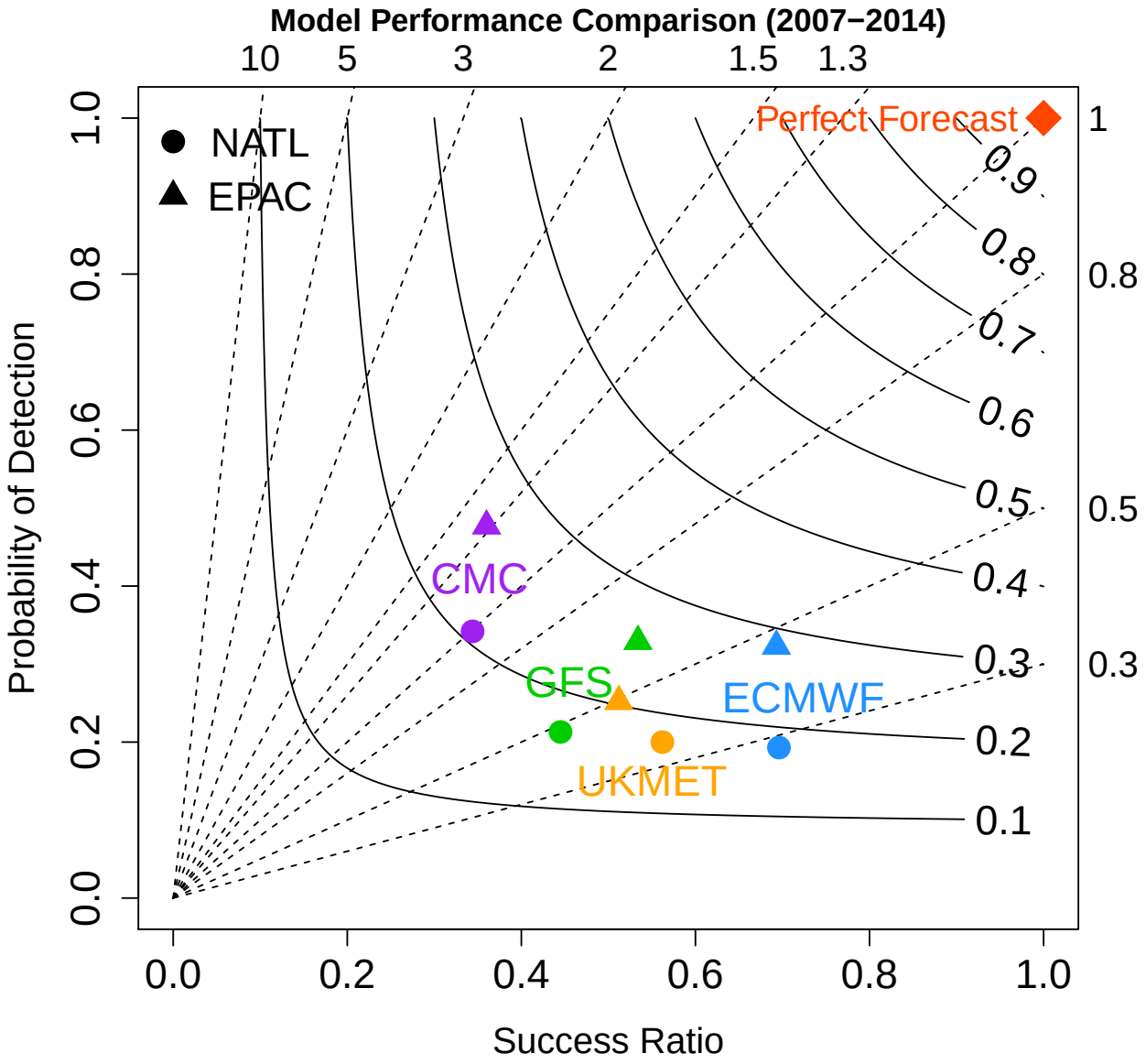
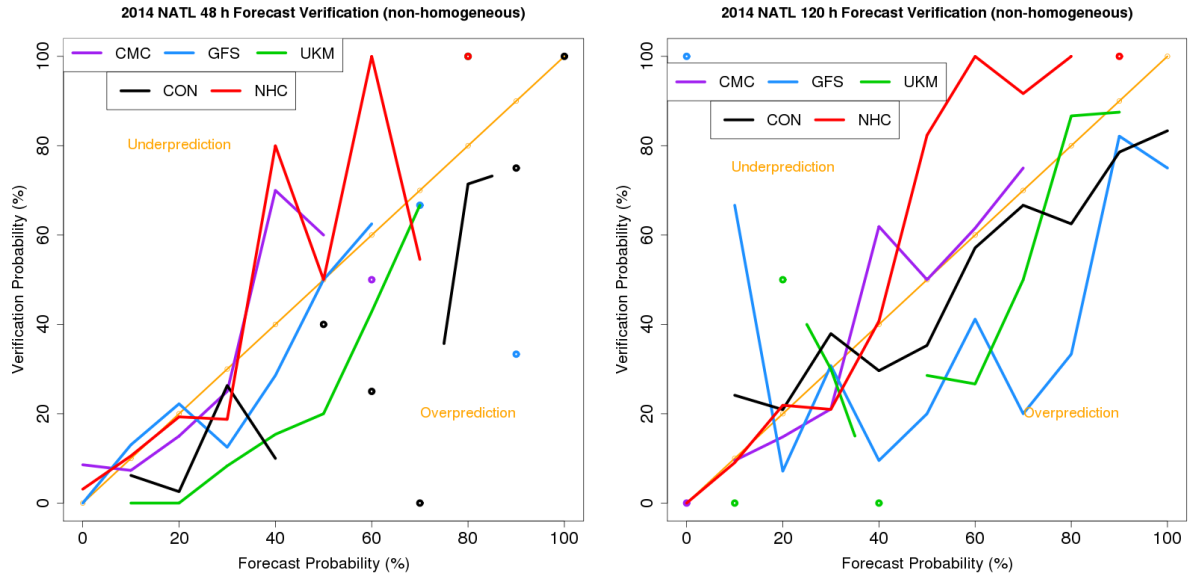
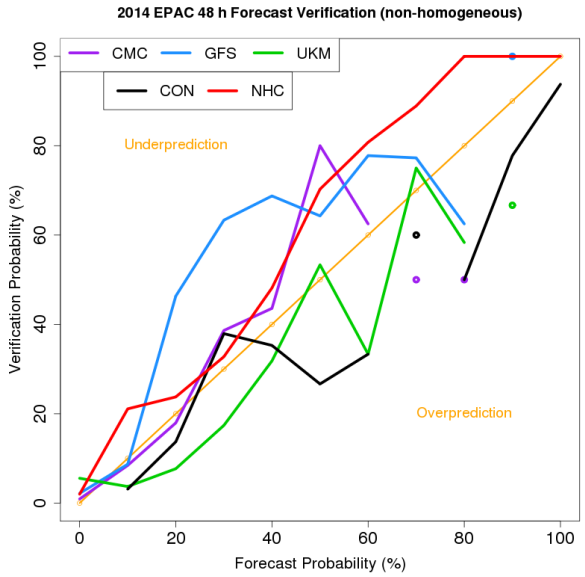


Figure 1: Performance diagram of 2007-2014 mean forecast performance for each model over the NATL (circle) and EPAC (triangle) basins. Success ratio is given on the x axis; probability of detection on the y axis. Bias values are indicated by the dashed lines, and critical success index values are indicated by the curved, solid lines. A "perfectly" performing model would be in the top-right corner of the plot. Genesis events from all forecast hours (6-120) are included.

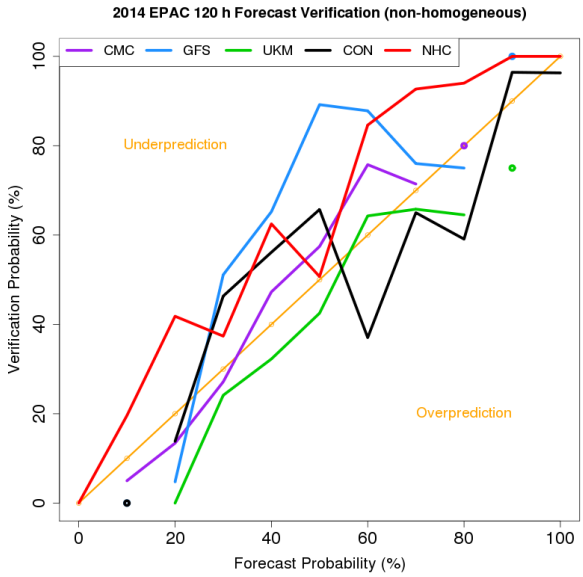


(a) NATL 48 h

(b) NATL 120 h



(c) EPAC 48 h



(d) EPAC 120 h

Figure 2: Reliability diagram comparing the non-homogeneous set of regression and NHC TWO probabilities for the (a) NATL 48 h window, (b) NATL 120 h window, (c) EPAC 48 h window, and (d) EPAC 120 h window. Verification is based on the 2014 Best-Track. Perfect reliability is given by the orange, diagonal line; above (below) this line indicates underprediction (overprediction). Breaks in the lines indicate forecast probability bins with five or fewer cases.

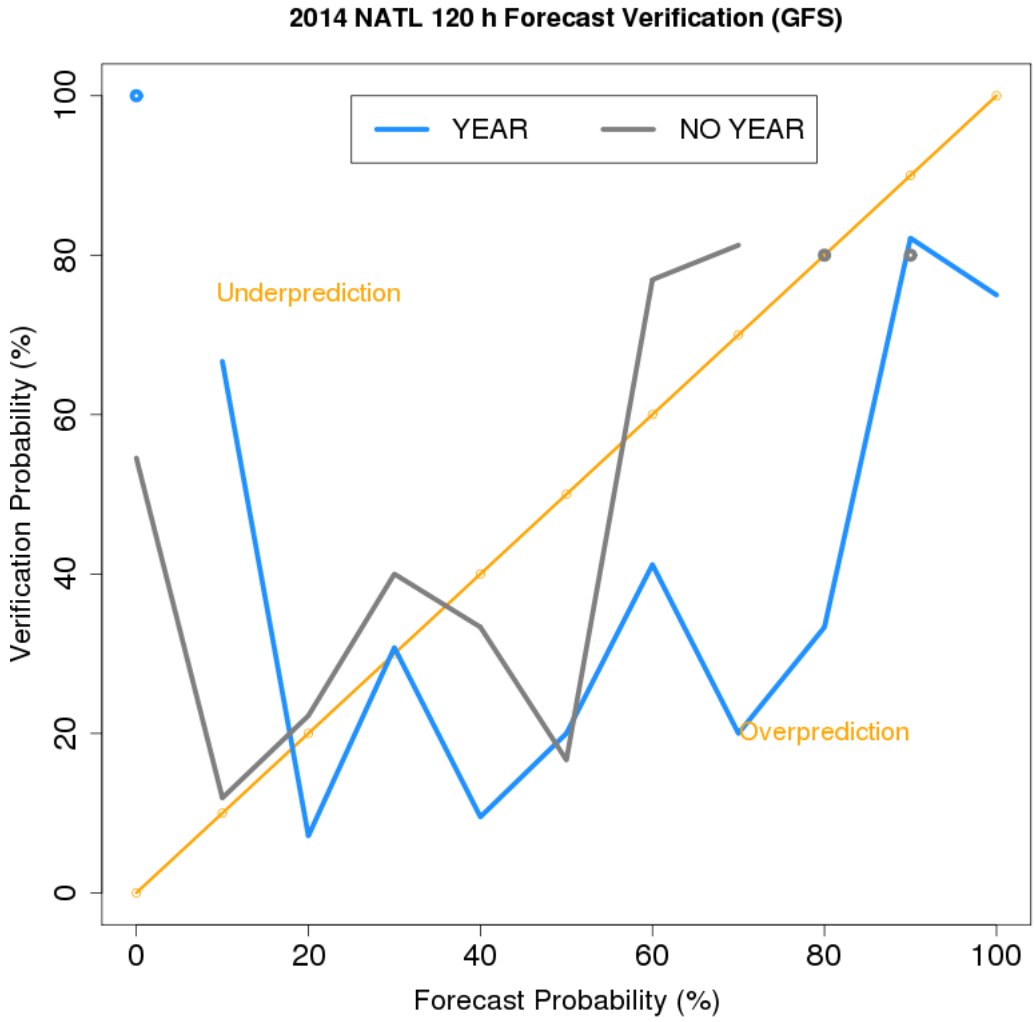
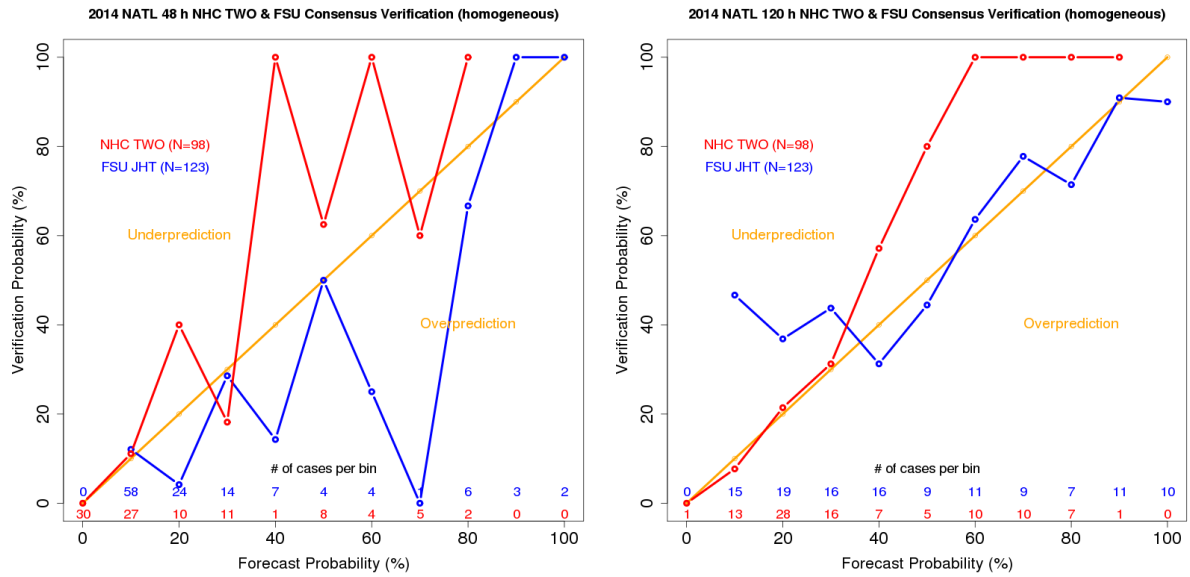
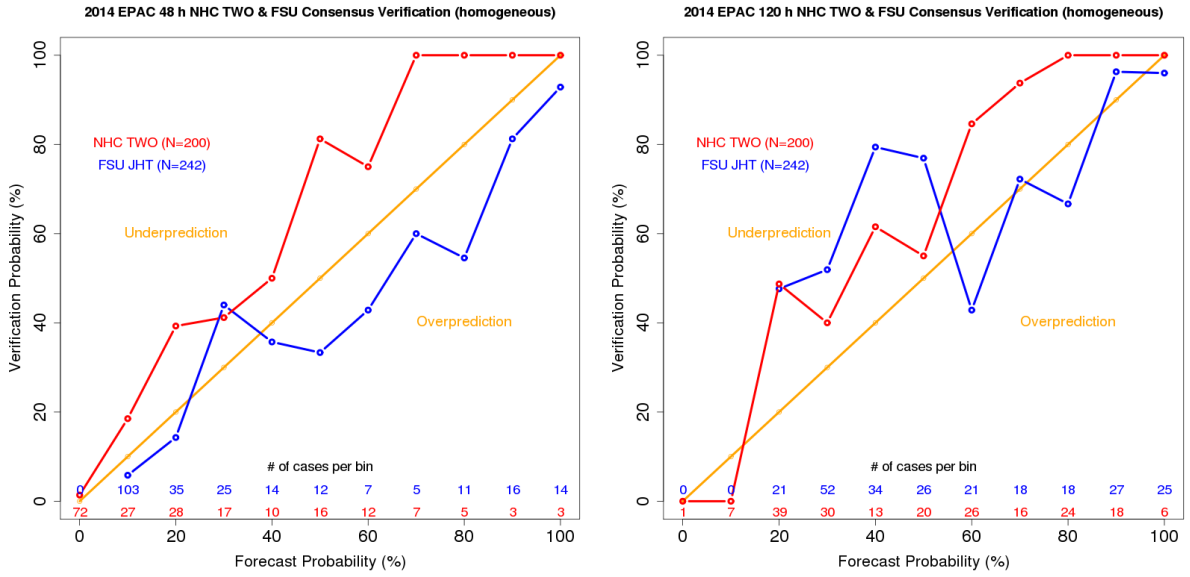


Figure 3: Reliability diagram of 2014 GFS based NATL 120 h regression model forecasts. The quasi-operational configuration (blue line) and a reforecast configuration of the regression model with "year" removed as a predictor (gray line) are compared.



(a) NATL 48 h

(b) NATL 120 h



(c) EPAC 48 h

(d) EPAC 120 h

Figure 4: Reliability diagram comparing the homogeneous set of 2014 NHC TWO (red line) and FSU JHT CON regression (blue line) probabilities for the (a) NATL 48 h window, (b) NATL 120 h window, (c) EPAC 48 h window, and (d) EPAC 120 h window. The number of cases in each forecast probability bin are shown in the corresponding red and blue text. Verification is based on the 2014 Best-Track.

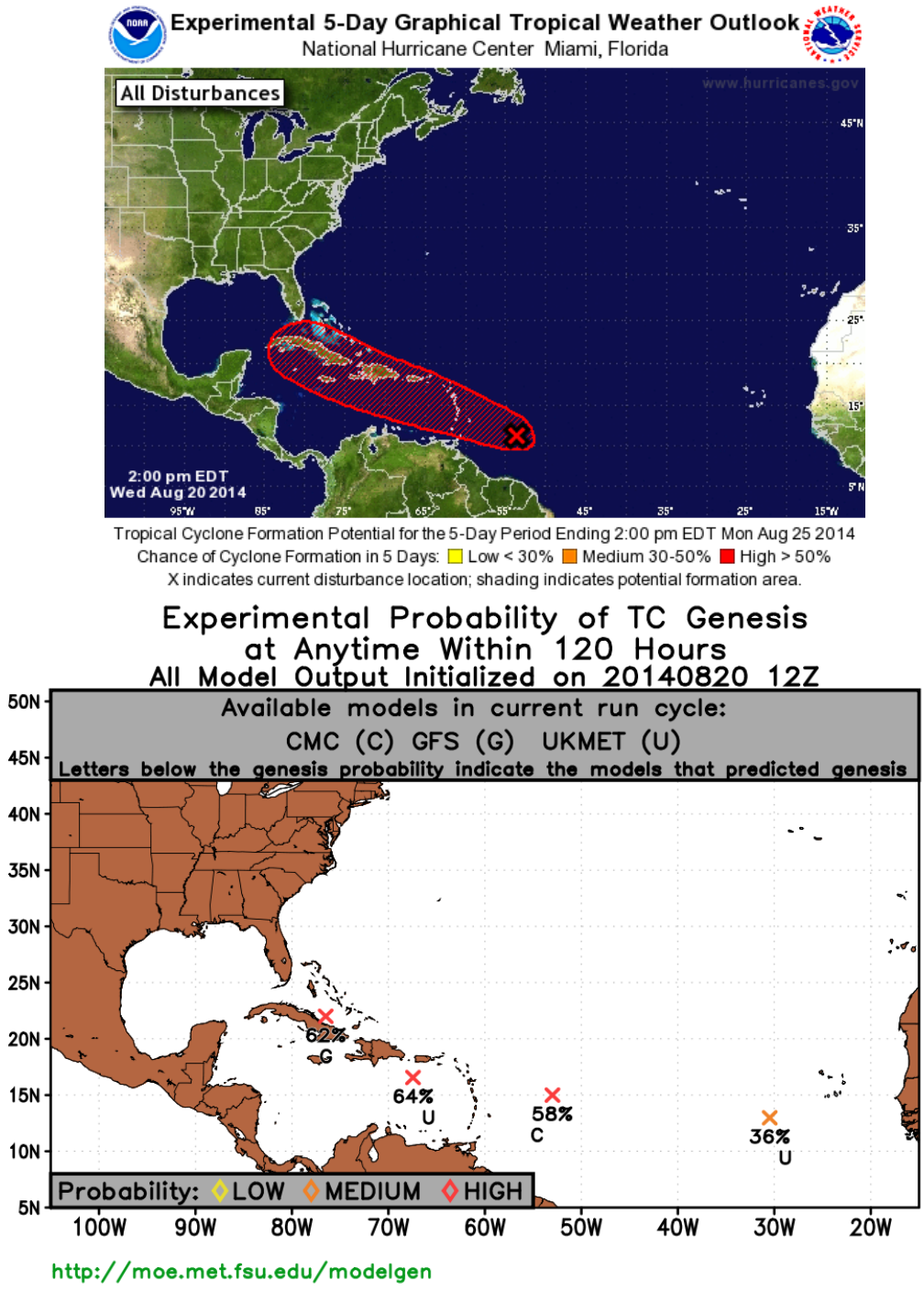
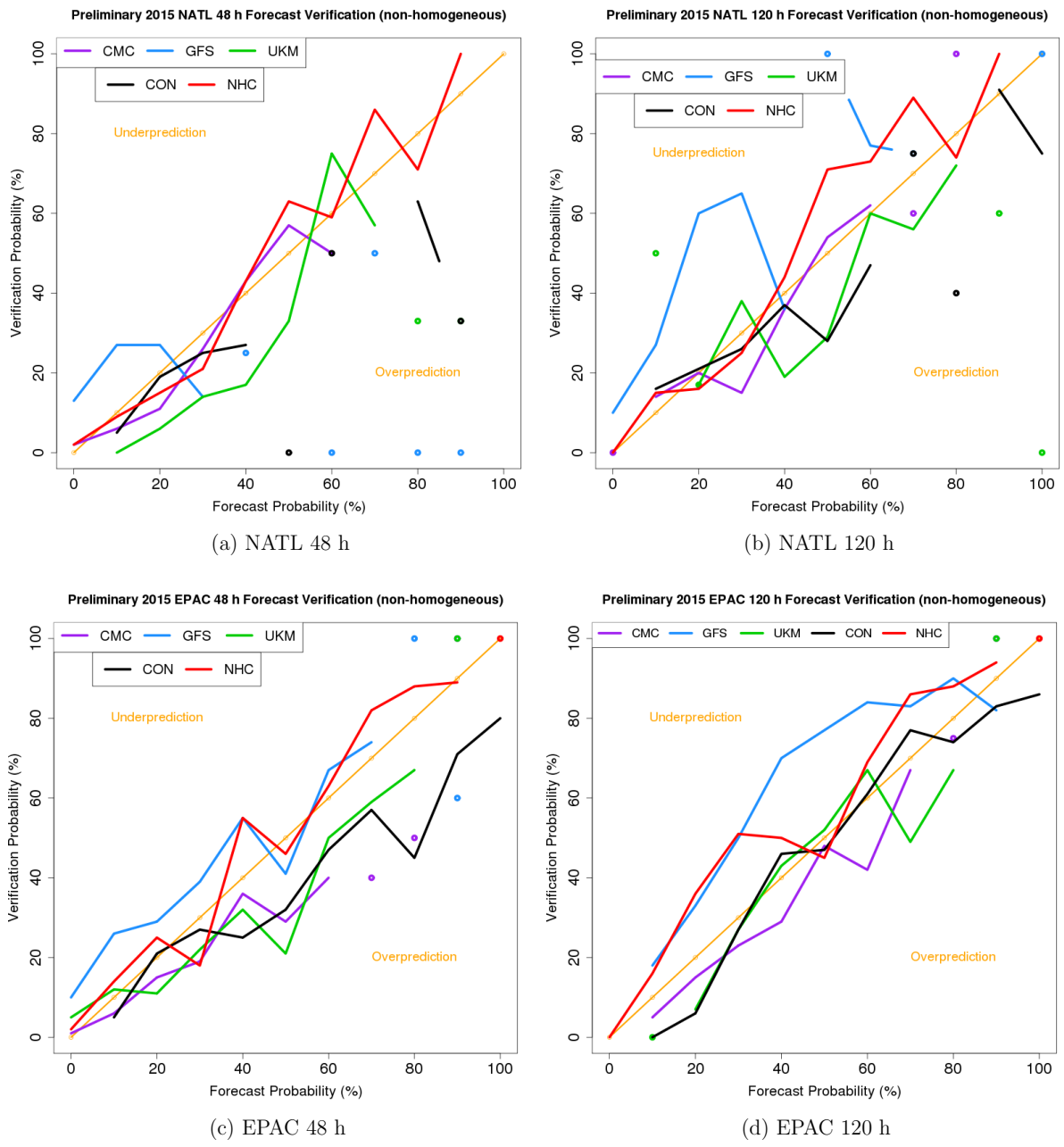


Figure 5: NHC TWO forecast (top) and CON regression probabilities (bottom). All three global models disagree on the timing and location of genesis for the disturbance highlighted in the TWO. As a result, the CON tracking algorithm treats each forecast as a separate TC. However, since all three forecast genesis locations are within the TWO shaded region, they all are included in the homogeneous verification comparison.



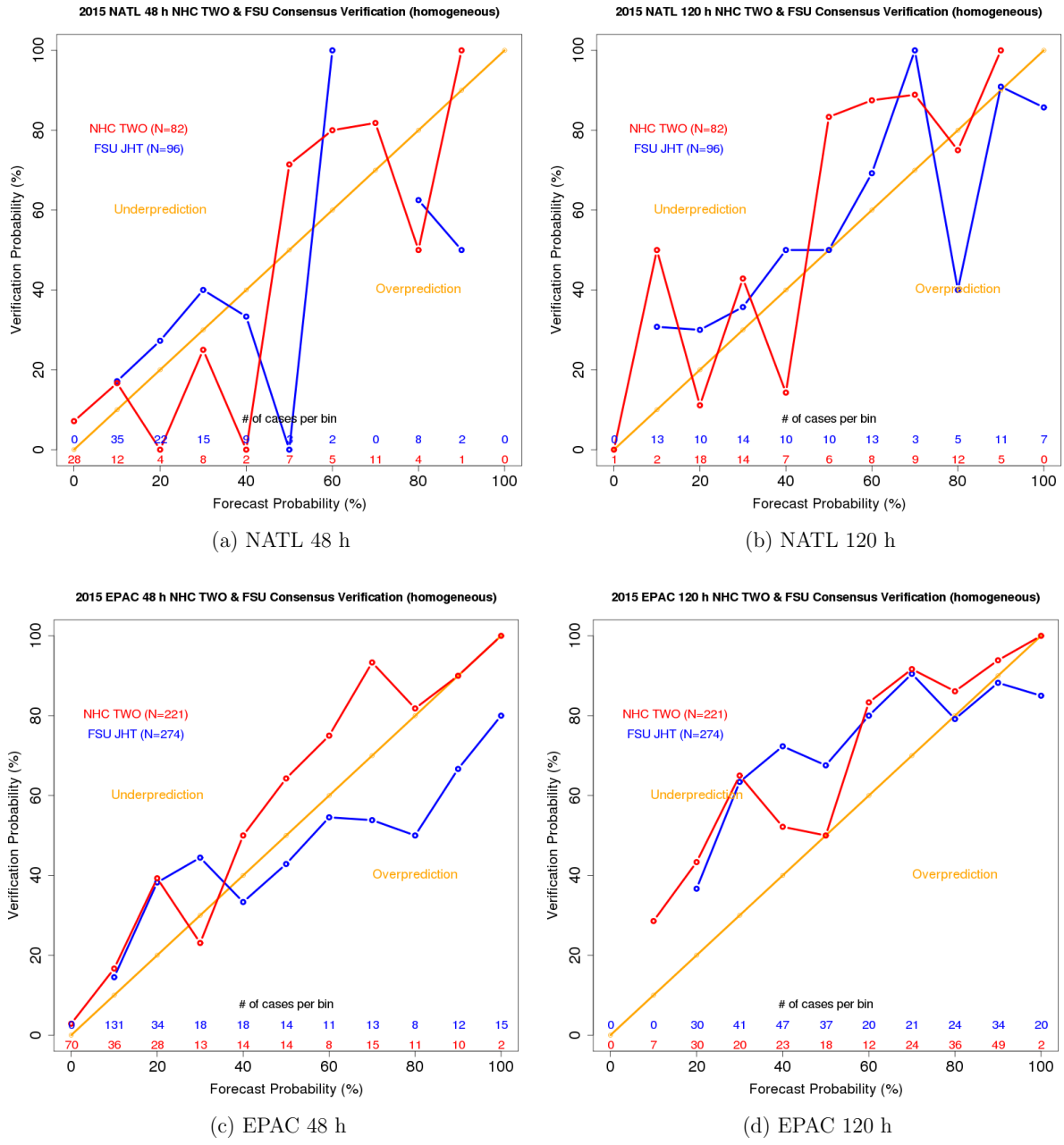
(a) NATL 48 h

(b) NATL 120 h

(c) EPAC 48 h

(d) EPAC 120 h

Figure 6: Reliability diagram comparing the non-homogeneous set of regression and NHC TWO probabilities for the (a) NATL 48 h window, (b) NATL 120 h window, (c) EPAC 48 h window, and (d) EPAC 120 h window. Verification is preliminary and based on the 2015 working Best-Track files. Perfect reliability is given by the orange, diagonal line; above (below) this line indicates underprediction (overprediction). Breaks in the lines indicate forecast probability bins with five or fewer cases. NHC TWO data were available through 1 November 2015.



(a) NATL 48 h

(b) NATL 120 h

(c) EPAC 48 h

(d) EPAC 120 h

Figure 7: Reliability diagram comparing the homogeneous set of 2015 NHC TWO (red line) and FSU JHT CON regression (blue line) probabilities for the (a) NATL 48 h window, (b) NATL 120 h window, (c) EPAC 48 h window, and (d) EPAC 120 h window. The number of cases in each forecast probability bin are shown in the corresponding red and blue text. Verification is preliminary and based on the 2015 working Best-Track files. NHC TWO data were available through 1 November 2015.

** Tropical Disturbance #2 Information **
** GFS output initialized 2015091500 **

GENESIS PREDICTED? YES, at forecast hour 60

TIME (hr)	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
GEN PROB (%)			20												51
CRIT MET?	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
LAT (N)	N/A	11.25	11.50	11.75	12.25	12.50	13.00	13.50	14.00	14.25	14.75	15.00	15.25	15.50	16.00
LON (W)	N/A	29.00	30.00	30.80	31.80	32.50	33.30	34.00	34.80	35.50	36.30	37.30	38.50	39.30	40.30
925VMAX (m/s)	N/A	(14.61)	(15.78)	(15.70)	17.65	16.58	16.97	17.69	21.87	24.38	24.62	24.32	27.58	26.05	26.02
850RV (*10^-5 1/s)	N/A	26.04	24.55	22.29	20.33	19.67	17.89	19.06	22.88	34.78	32.53	26.51	38.59	33.23	26.25
H25-85 THCK (m)	N/A	9487	9484	9476	9481	9490	9486	9479	9479	9491	9487	9479	9478	9489	9479
SFC LH FLUX	N/A	152.78	160.26	174.45	172.48	161.52	150.34	147.25	141.78	136.15	131.53	137.98	141.31	133.79	131.60
850RVPERT	N/A	24.66	22.98	20.50	18.37	17.81	17.39	17.30	21.32	33.44	31.27	25.36	37.46	32.30	25.14
H2-85 SHEAR	N/A	10.12	11.11	12.32	12.77	12.51	12.74	13.72	14.22	14.18	14.58	15.47	15.12	13.09	12.66

( ) indicates that the threshold value was not met

Figure 8: Example of the tropical disturbance information text product. This product shows for each 6 h model output time (1) the disturbance position; (2) the values of the genesis criteria and whether each threshold value was met; (3) the values of the predictors used in the regression equations; and (4) the 48 h and 120 h genesis probabilities (assuming the genesis criteria were met). In this example, a disturbance is identified at forecast hour 42, but the 925 hPa wind speed criterion is not met until forecast hour 60. Starting at forecast hour 60, all genesis criteria are met for at least 24 consecutive hours in the model cycle. This product includes all forecast hours from 0 to 144, but was truncated here to fit on the page.

** TC Genesis Summary for UKMAL9018**

INIT TIME	FHR	LAT(N)	LON(W)	PROB48(%)	PROB120(%)
15091812	0	12.66	37.10	N/A	N/A
15091800	12	13.29	36.60	65	86
15091712	24	13.60	37.50	56	84
15091700	36	13.44	36.80	47	79
15091612	18	11.73	31.90	58	73
15091600	30	12.51	32.20	49	75
15091512	42	12.98	31.70	40	73
15091500	36	12.66	29.40	43	73
15091412	42	12.51	28.90	38	69
15091400	36	11.42	27.20	42	58
15091312	42	11.26	26.30	37	51
15091300	54	11.10	28.00	30	42
15091212	90	11.88	32.90	14	38

Figure 9: Example of the TC history product. For each TC, this product shows for each initialization time (1) the forecast genesis time relative to the initialization time; (2) the forecast genesis location; and (3) the 48 and 120 h genesis probabilities. This product gives the forecasters trend information for a particular model-indicated TC. In this example, the forecast genesis time trends later, farther west, and the genesis probabilities increase. It also shows that the UKM forecasts genesis for this system in consecutive model cycles for 6 days.

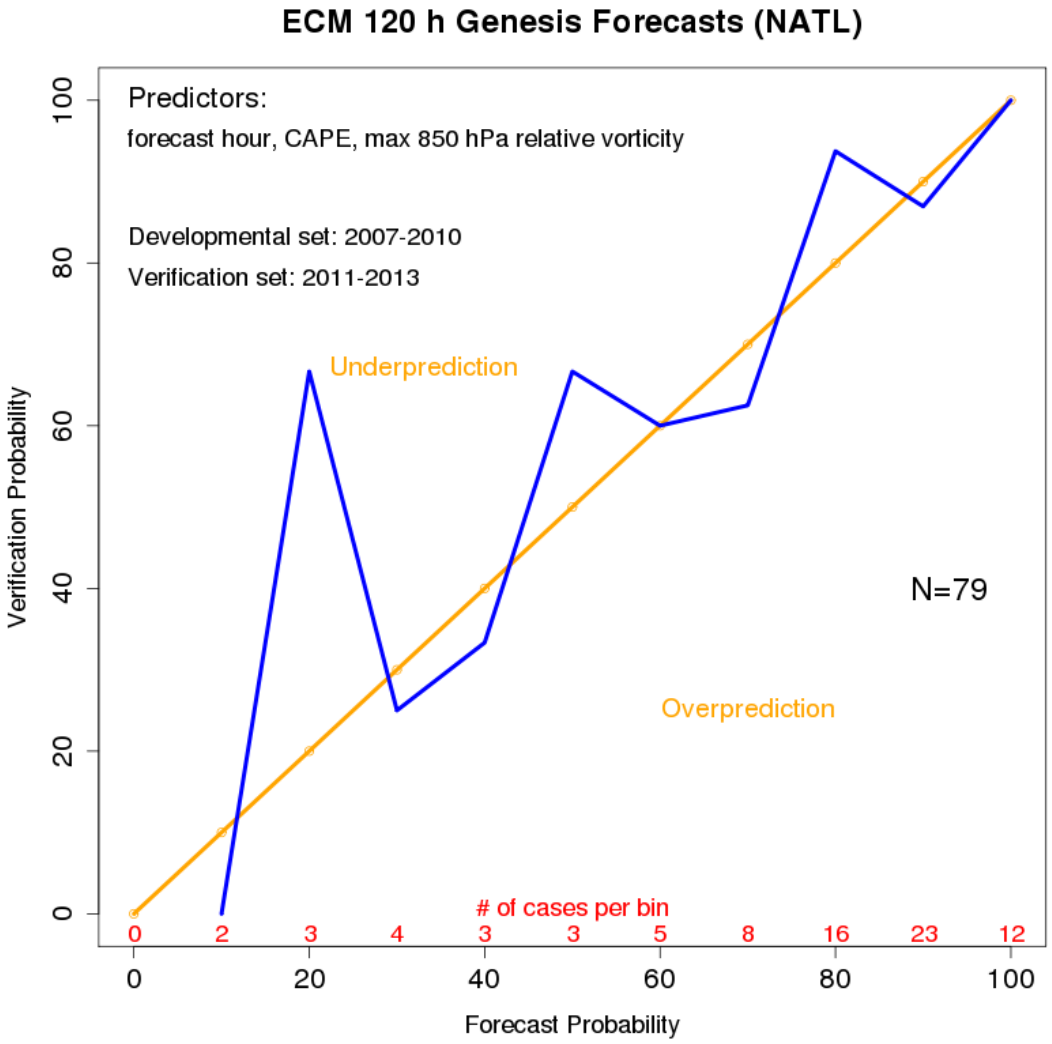
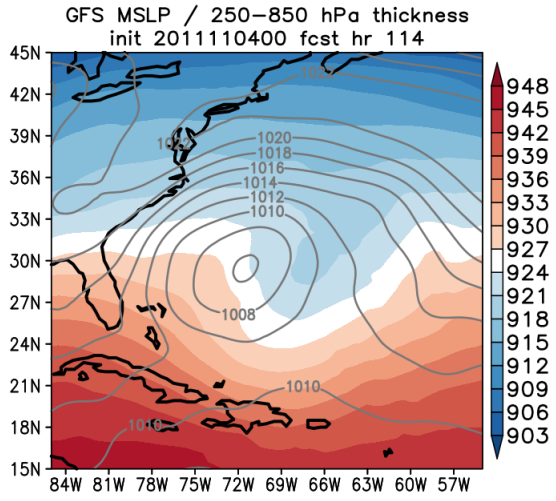
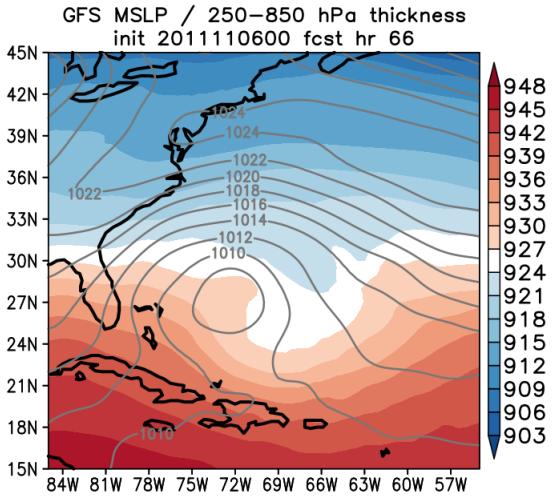


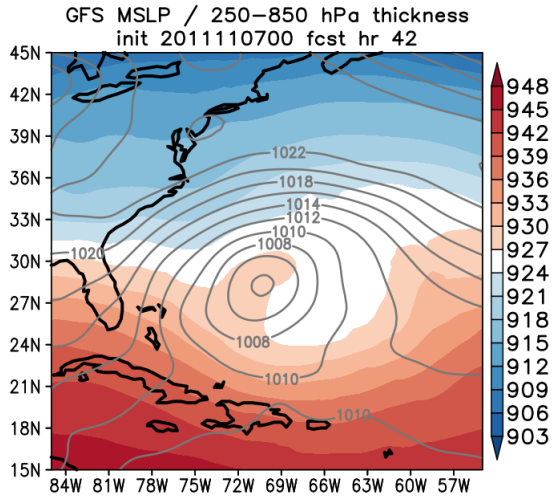
Figure 10: Reliability diagram evaluating the fit of ECMWF based regression model forecasts. The regression equation was based on 2007-2010 cases and tested on 2011-2013 cases. The larger range of years for the independent set is necessary given the small sample size of cases during 2013 alone.



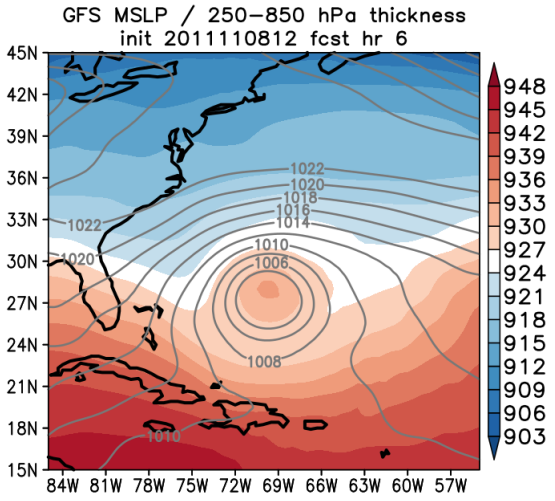
(a) 114 h forecast.



(b) 66 h forecast.

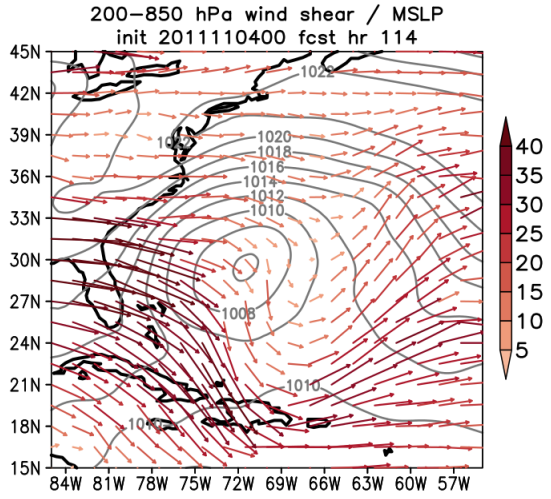


(c) 42 h forecast.

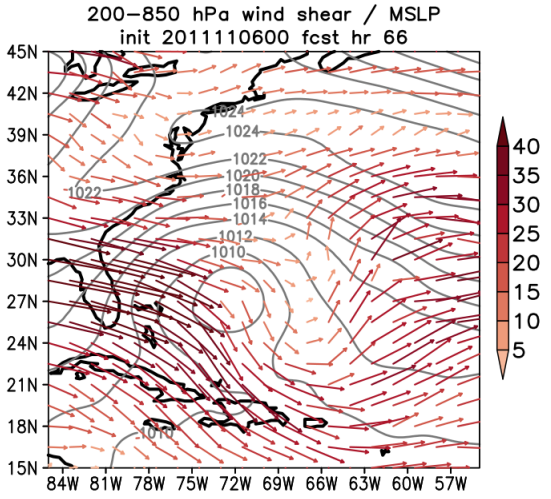


(d) 6 h forecast.

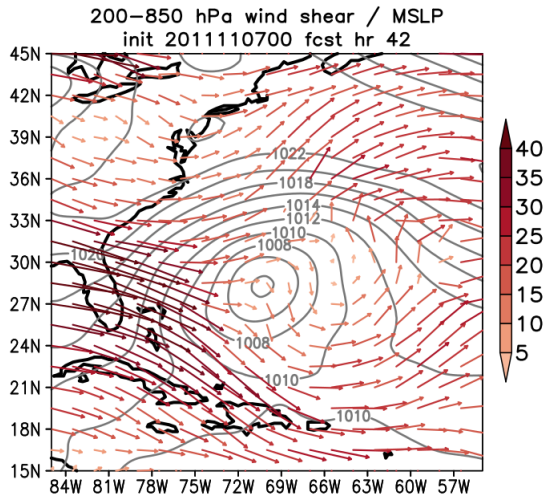
Figure 11: GFS forecasts for Sean showing 250-850 hPa thickness (dam; shaded) and MSLP (hPa; contoured). All forecasts are valid at 1800 UTC 8 November 2011 (first entry as a tropical storm in the Best-Track). Note the evolution of the developing warm core by model cycle.



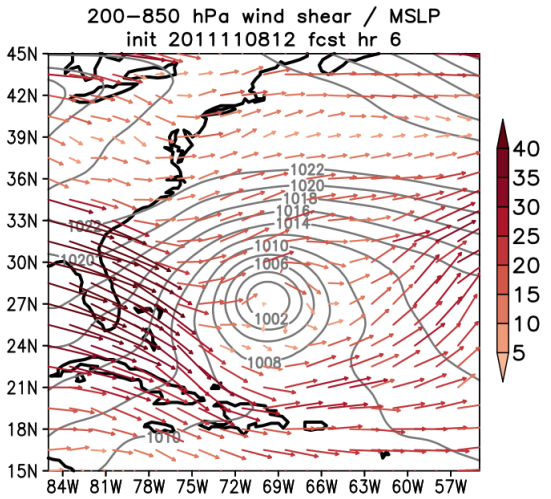
(a) 114 h forecast.



(b) 66 h forecast.

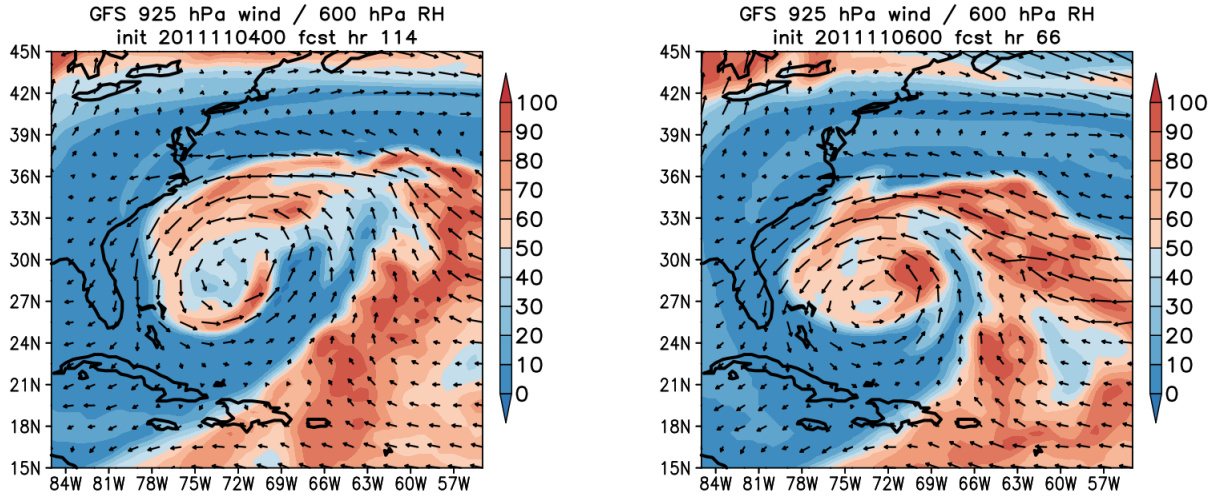


(c) 42 h forecast.



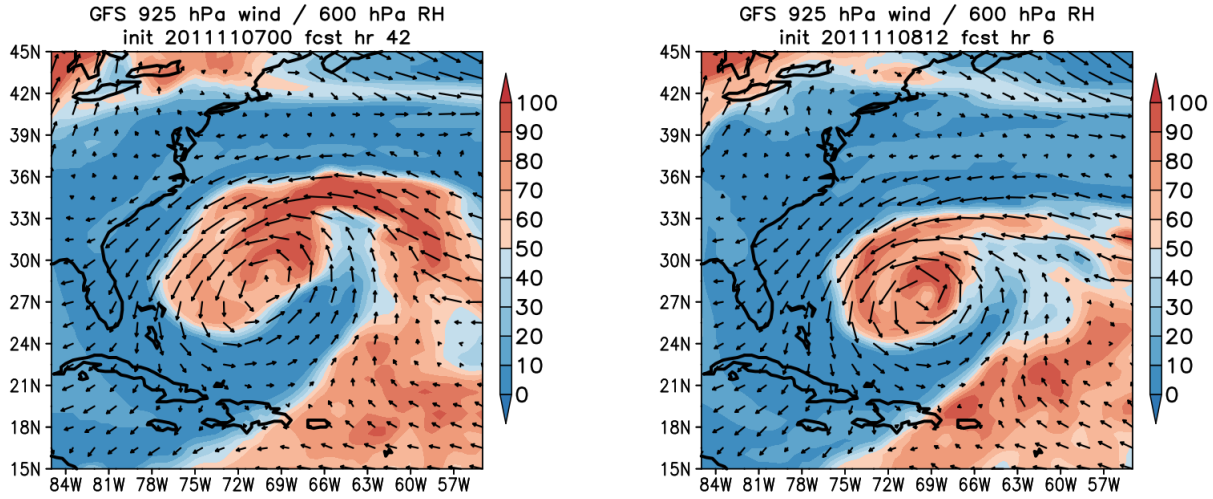
(d) 6 h forecast.

Figure 12: GFS forecasts for Sean showing 200-850 hPa wind shear (m s^{-1}; vectors) and MSLP (hPa; contoured). All forecasts are valid at 1800 UTC 8 November 2011 (first entry as a tropical storm in the Best-Track). Note the decrease in forecast wind shear over the cyclone between panels (a-b) and (c-d).



(a) 114 h forecast.

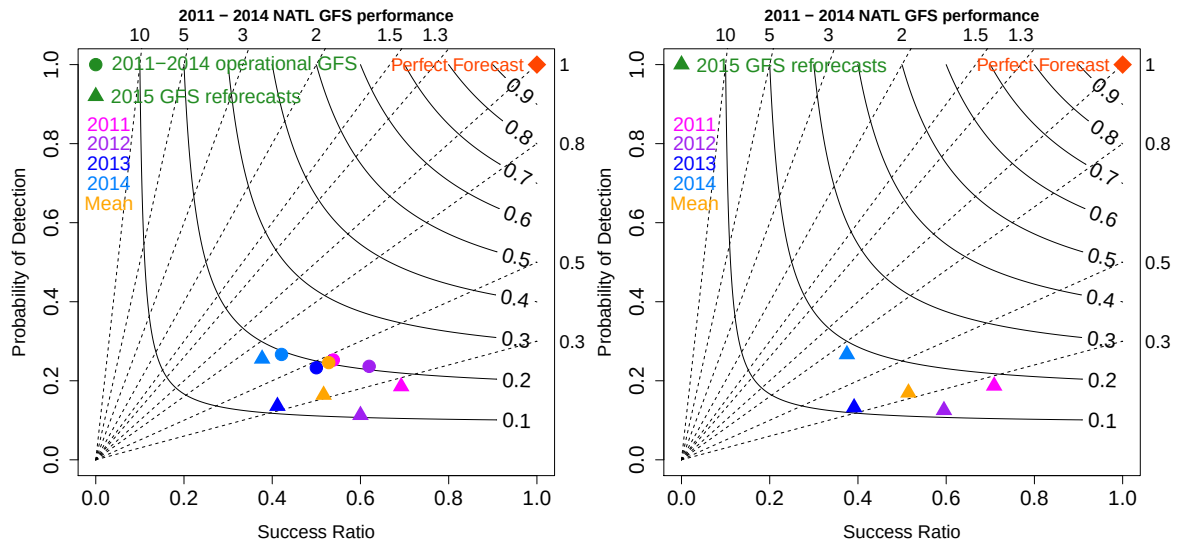
(b) 66 h forecast.



(c) 42 h forecast.

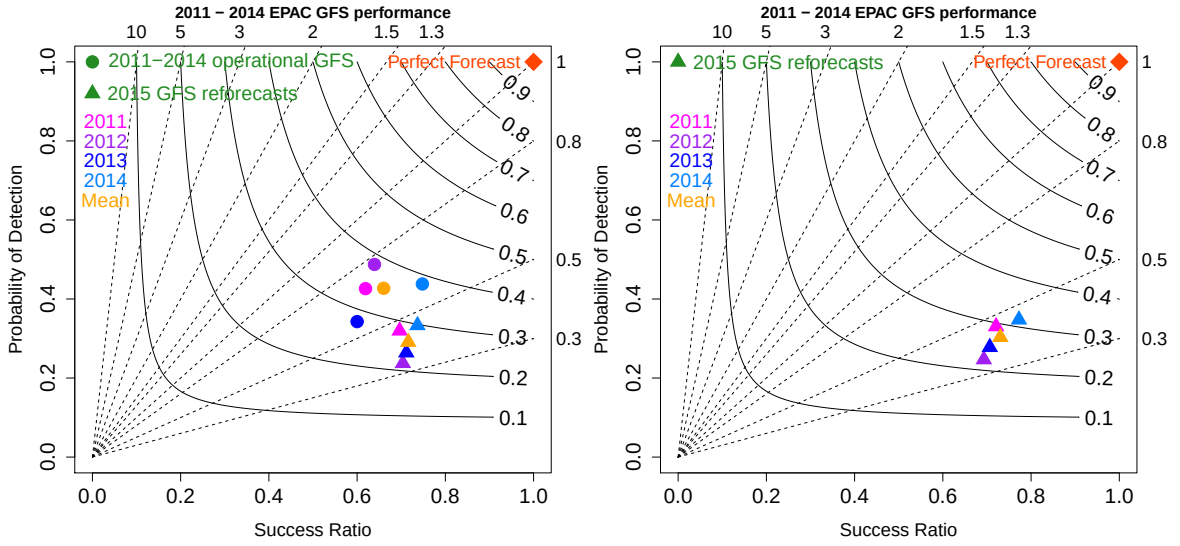
(d) 6 h forecast.

Figure 13: GFS forecasts for Sean showing 600 hPa RH (%; shaded) and 925 hPa wind (m s^{-1}; vectors). All forecasts are valid at 1800 UTC 8 November 2011 (first entry as a tropical storm in the Best-Track). Note the smaller (larger) RH values near the low level center in the earlier (later) model runs.



(a) NATL homogeneous set

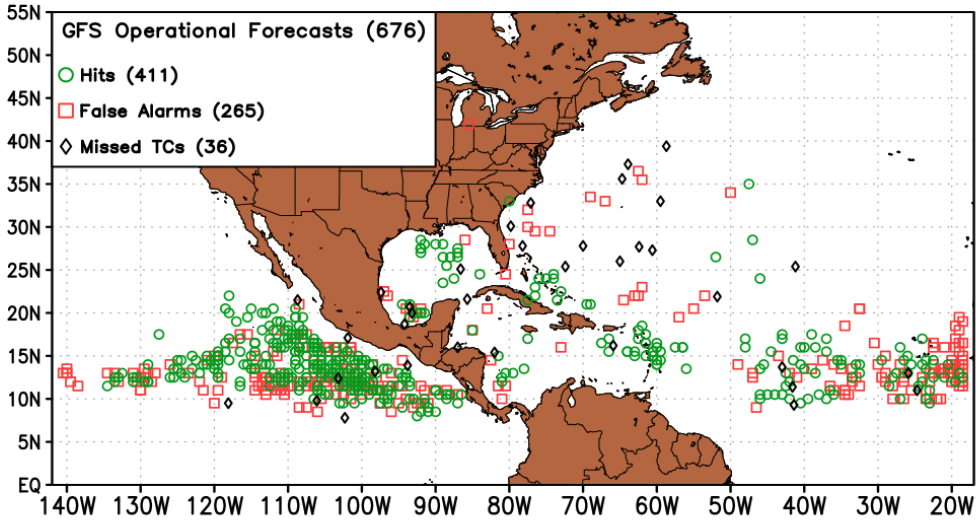
(b) NATL all reforecast set



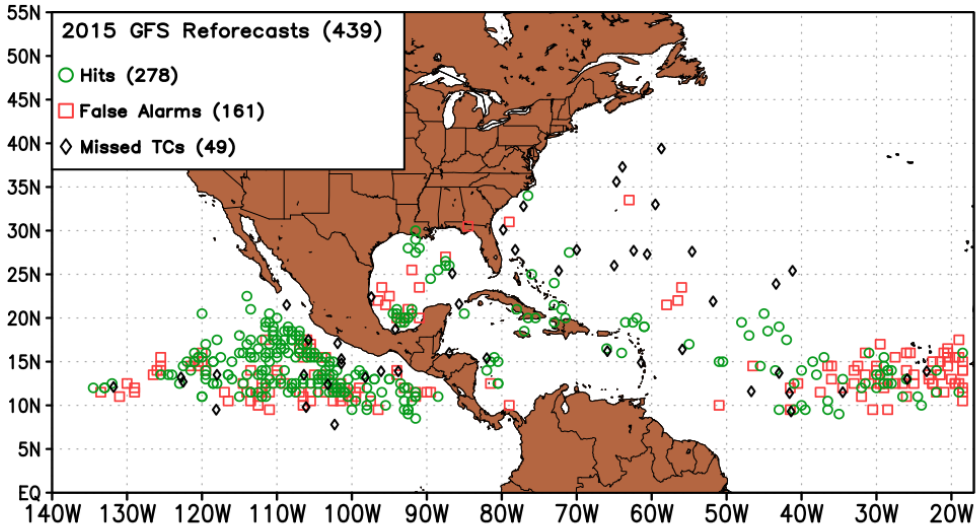
(c) EPAC homogeneous set

(d) EPAC all reforecast set

Figure 14: Performance diagrams of the operational and reforecast GFS configurations for the (a) NATL homogeneous set, (b) NATL all reforecast set, (c) EPAC homogeneous set, and (d) EPAC all reforecast set. Success ratio is given on the x axis; probability of detection on the y axis. Bias values are indicated by the dashed lines, and the critical success index values are indicated by the curved, solid lines. A "perfect" performing model would be in the top-right corner of the plot. Genesis events from all forecast hours (6-120) are included. Average statistics for each year are plotted in different colors. The operational (reforecast) configuration is indicated by the circles (triangles).



(a) Operational configuration.



(b) Reforecast configuration.

Figure 15: A geographic plot of GFS-indicated TC genesis events for (a) the operational configuration (homogeneous set), and (b) the reforecast configuration (homogeneous set).