The Development and Evaluation of a Statistical–Dynamical Tropical Cyclone Genesis Guidance Tool

DANIEL J. HALPERIN,[a] ROBERT E. HART, AND HENRY E. FUELBERG

Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, Florida

JOSHUA H. COSSUTH

Naval Research Laboratory, Monterey, California

ABSTRACT

The National Hurricane Center (NHC) has stated that guidance on tropical cyclone (TC) genesis is an operational forecast improvement need, particularly since numerical weather prediction models produce TC-like features and operationally required forecast lead times recently have increased. Using previously defined criteria for TC genesis in global models, this study bias corrects TC genesis forecasts from global models using multiple logistic regression. The derived regression equations provide 48- and 120-h probabilistic genesis forecasts for each TC genesis event that occurs in the Environment Canada Global Environmental Multiscale Model (CMC), the NCEP Global Forecast System (GFS), and the Met Office's global model (UKMET). Results show select global model output variables are good discriminators between successful and unsuccessful TC genesis forecasts. Independent verification of the regression-based probabilistic genesis forecasts during 2014 and 2015 are presented. Brier scores and reliability diagrams indicate that the forecasts generally are well calibrated and can be used as guidance for NHC's Tropical Weather Outlook product. The regression-based TC genesis forecasts are available in real time online.

1. Introduction

Although considerable progress has been made, many aspects of tropical cyclones (TCs), including their development and intensification, remain insufficiently understood and represent a continuing challenge to both the research and operational communities (Emanuel 1986; Rappaport et al. 2009). Researchers historically have focused on improving TC track and intensity forecast guidance (e.g., Rappaport et al. 2012; Gall et al. 2013). The resulting forecast improvements generally have been attributed to better operational model guidance and forecast tools available to forecasters (Rappaport et al. 2012). As

the forecast guidance has become more reliable, operational centers such as the National Hurricane Center (NHC) have increased their forecast lead times, which in turn, require more accurate genesis forecasts. For example, a tropical disturbance can develop near land, intensify, make landfall, and dissipate all within the current 5-day forecast window (e.g., Humberto in 2007). NHC's Tropical Weather Outlook (TWO), a product that provides categorical and probabilistic forecasts of TC genesis, also was extended from 2 to 5 days in August 2013 (Cangialosi and Franklin 2014). Thus, the ability to accurately predict TC genesis is an important operational need. The goal of the present study is to develop reliable probabilistic TC genesis forecasts based on global model output to serve as skillful guidance for NHC forecasters in preparing the TWO.

Several TC genesis guidance products have already been developed. DeMaria et al. (2001) used 5-day running averages of vertical wind shear, instability, and midlevel moisture over the tropical Atlantic to produce genesis probabilities relative to climatology. Their Tropical Cyclone Formation Probability (TCFP)

[a] Current affiliation: Department of Atmospheric and Environmental Science, University at Albany, State University of New York, Albany, New York.

Corresponding author address: Daniel J. Halperin, Dept. of Atmospheric and Environmental Sciences, University at Albany, SUNY, 1400 Washington Ave., Albany, NY 12222.
E-mail: danieljhalperin@gmail.com

product exhibited skill relative to climatology. Schumacher et al. (2009) described updates to the TCFP, employing screening and linear discriminant analysis of predictors averaged over $5° \times 5°$ areas to provide a 24-h probability of genesis. Both the TCFP and its revision use environmental conditions averaged over various spatial and temporal scales; they are not disturbance specific. They also do not use model forecast fields, although model analyses are included. Further updates to the TCFP (Schumacher et al. 2014) extended the product to 48 h and included information from model forecast fields. Cossuth et al. (2013) developed 48- and 120-h TC genesis probabilities using a climatology of pregenesis Dvorak classifications. The tropical cyclone genesis index (Dunion et al. 2013) is a statistical guidance tool that employs observations and GFS forecast fields to provide 48- and 120-h probabilistic genesis forecasts for NHC-designated invest areas over the North Atlantic (NATL) basin. Zhang et al. (2015) demonstrated the use of a decision tree to predict whether western North Pacific (WNP) tropical disturbances present in Navy Operational Global Atmospheric Prediction System (NOGAPS) analyses would develop into a TC within 24–48 h. Others have used global model ensembles to generate probabilistic genesis forecasts (e.g., Marchok 2002; Gall et al. 2013; Majumdar and Torn 2014), where the uncalibrated percentage of ensemble members that exceed specified genesis criteria defines the genesis probability. The current study is unique in that it presents the development of calibrated, disturbance-specific TC genesis probabilities for the 48- and 120-h forecast periods over the NATL and eastern North Pacific (EPAC) basins that rely solely on deterministic global model output. For some disturbances, both the proposed and previously developed products will provide TC genesis guidance. However, there are instances when the proposed products will be one of the only sources of TC genesis guidance (e.g., when the models forecast genesis at 120 h for a disturbance that does not yet exist at the initial time).

Global model forecasts provide important guidance for the TWO (R. Pasch 2012, personal communication). This begs the question, "How well do global models predict TC genesis?" Early studies (e.g., Beven 1999) showed that the models predicted too many spurious vortices to skillfully forecast TC genesis. Schumacher et al. (2009) suggested that global model forecasts have limited application in TC genesis forecasting because of uncertainties associated with their forecast skill and biases. However, multiyear, multimodel investigations of model-indicated TC genesis forecasts by Halperin et al. (2013, 2016, hereafter H13 and H16, respectively) revealed that the models' ability to predict TC genesis

has improved in recent years. Additionally, Komaromi and Majumdar (2015) used European Centre for Medium-Range Weather Forecasts (ECMWF) ensembles to demonstrate that TC genesis events exhibit some predictability out to 1 week. Furthermore, Elsberry et al. (2014) showed that the ECMWF ensembles were able to capture the genesis of some TCs during 2012 at 1–4-week lead times. The present study explores whether there is untapped predictability in the global model TC genesis forecasts that can be exploited by bias correction. For example, H13 and Cossuth et al. (2013) showed that genesis predictability varies regionally. Elsberry et al. (2014) found differences in model performance between TCs that formed from African easterly waves versus those that formed from baroclinic origins. H13 also showed that model performance varies by forecast hour (lead time) and month. The present study expands on H13 by using multiple logistic regression to bias correct global model–indicated TC genesis forecasts. The regression-based probabilistic TC genesis forecasts are produced in real time to provide objective genesis guidance to the Hurricane Specialist Unit at NHC. Please refer to appendix A for a description of the available products.

2. Methodology

The statistically derived TC genesis guidance products developed here are based on output from three numerical models: Environment Canada's Global Environmental Multiscale Model (CMC; Côté et al. 1998a,b), the National Centers for Environmental Prediction's (NCEP) Global Forecast System (GFS; Kanamitsu 1989), and the Met Office's global model (UKMET, also referred to as UKM; Cullen 1993). TC genesis guidance products were developed from the ECMWF (ECMWF 2016) model output. However, ECMWF-based statistical models are not discussed here because the real-time model output needed to test the guidance products was not available. Since NOGAPS (Rosmond 1992) was decommissioned in 2013, it was not included. The Navy Global Environmental Model (NAVGEM; Hogan et al. 2014), the replacement for NOGAPS, is not included since the sample size of archived forecasts currently is too small.

Operational global model data were available from a local archive during 2004–13, providing a sufficient sample of model genesis forecasts to construct a developmental/training dataset for the statistical analysis. The appendixes of H13 and H16 list select model upgrades that occurred during the period of study (e.g., resolution increases, changes in data assimilation, changes to convective parameterizations, etc.). Output for each global model was available for the 0000 and 1200 UTC initialization cycles. The present study used

the definition of TC genesis given in H13 and H16 to identify TC genesis events in the forecast fields out to 120 h. Each TC genesis forecast is verified against the best-track dataset (Jarvinen et al. 1984; McAdie et al. 2009; Torn and Snyder 2012; Landsea and Franklin 2013) and is classified as a "hit" or "false alarm" for the 48- and 120-h forecast windows according to the verification criteria in H16.

H13 tested numerous sets of TC genesis criteria and selected the set of criteria that optimized the probability of detection and the false alarm ratio. H13 and H16 found notable differences in model performance. On average, CMC had the smallest success ratio, but the greatest probability of detection compared to the other models. In contrast, ECMWF had the greatest mean success ratio, but the smallest mean probability of detection. All four models exhibited critical success index values near 0.2 over the NATL, indicating a trade-off between the success ratio and the probability of detection. Mean values of the critical success index over the EPAC generally were greater than over the NATL because of comparable success ratios and larger probabilities of detection (H16). The larger false alarm ratio exhibited by CMC does not hinder the performance of its regression equation. Smaller forecast probabilities occur more frequently, but the probabilities are fairly well calibrated (shown in section 5). However, smaller probabilities of detection may negatively impact the forecast products since they will result in a smaller developmental dataset and will reduce the number of cases for which the product will provide guidance. For example, GFS misses many best-track TCs poleward of 25°N (according to the H13 and H16 criteria), usually as a result of not exceeding the thickness threshold. Thus, the GFS-based regression model often is unable to provide guidance for Invests in that region. Future versions of this guidance product will examine the impact of altering the genesis criteria threshold values.

Theoretical and observational TC genesis studies have provided physically relevant variables and processes that are associated with TC genesis (e.g., Charney and Eliassen 1964; Gray 1968, 1979; Emanuel 1986; Ritchie and Holland 1997; Simpson et al. 1997; Bister and Emanuel 1997; Emanuel and Nolan 2004; Nolan 2007; Dunkerton et al. 2009). We use logistic regression to test whether these variables, when obtained from global models at the time of forecast genesis, are good discriminators between successful (i.e., hit) and unsuccessful (i.e., false alarm) forecasts. Some predictors that have not been discussed in the aforementioned literature also were considered. An averaged predictor refers to values averaged over the box area extending $\pm 5°$ from the model-indicated TC center and is denoted

by an overbar (e.g., $\overline{RH}$). A perturbation refers to the maximum value of a predictor within 5° of the model-indicated TC center minus the average value in that area and is denoted by the prime (′) symbol [e.g., $RH' = \max(RH) - \overline{RH}$]. Appendix B provides a list of symbols used in this paper. The multiple logistic regression (MLR) model is given by

$$E(y|x) = \frac{e^{g(x)}}{1 + e^{g(x)}}, \qquad (1)$$

where $E(y|x)$ is the expected probability of the outcome variable y (i.e., TC genesis), given a value of x, and

$$g(x) = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n. \qquad (2)$$

Equation (2) is the logit transformation, x_n is the value of the nth predictor, and β_n is the regression coefficient of the nth predictor (Hosmer et al. 2013). Peng et al. (2012) and Fu et al. (2012) found differences in the relative importance of various parameters in distinguishing between developing and nondeveloping tropical disturbances over the NATL and WNP. Given their results and the model-to-model and basin-to-basin differences revealed in H13 and H16, separate MLR equations were developed for each global model, each basin, and each forecast window (0–48 and 0–120 h). Development of the regression equations was limited by the data archive. For example, nearly all GFS output variables were available to test as predictors, but far fewer UKMET variables were available locally.

3. Univariable logistic regression equations and TC genesis theories

The first objective is to determine whether statistically significant relationships exist between individual environmental and storm-centered variables and the probability of genesis. The findings then are compared to theoretically proposed physical relationships based on the prior literature. This comparison will determine whether conditions found to be important for TC genesis in established theories also are good discriminators for genesis in the global models. Single-variable logistic regression equations are used to facilitate this comparison. This use of univariable equations provides a clear interpretation of the coefficients without having to consider potential interactions between predictors.

Tables 1 and 2 provide the list of predictors that are statistically significant (p value < 0.05) for at least one univariable regression equation (i.e., at least one global model, basin, and forecast window). Some physically relevant variables (e.g., wind shear) generally are not found to be significant predictors. These variables may be important in the genesis process, but simply are not good discriminators between the hit and false alarm outcomes.

TABLE 1. List of significant predictors ($p < 0.05$) for each NATL univariable regression equation using 2004–13 genesis events as the developmental set. The plus (+) and minus (−) symbols indicate the sign of a significant predictor coefficient. Exclamation points (!) indicate that the predictor was not in the data archive and significance testing was not possible. No symbol indicates that the predictor was tested, but not significant. Unless denoted with a prime (′) or a double asterisk (**), all variables are averaged over the box area extending $\pm 5°$ from the model-indicated TC center. A prime denotes a perturbation, which refers to the maximum value of the variable within $5°$ of the model-indicated TC center minus the average value in that area. A double asterisk denotes a value used for defining TC genesis in the models, as in H13 and H16.

Predictor	CMC48	CMC120	GFS48	GFS120	UKM48	UKM120
Forecast hour	−	−	−	−	−	−
Year			+	+		+
Lat (°)	−	−	+	+		
Lon (°)			−	−	−	
925-hPa $\mathbf{V}_{max}$ (m s^{-1})**	−	−		−		
250–850-hPa ΔZ (m)**	+	+	+	+	+	
850-hPa $\overline{\zeta}$ ($\times 10^{-5}$ s^{-1})	−	−	−	−		
850-hPa ζ' ($\times 10^{-5}$ s^{-1})	−	−	−	−		
850-hPa OW (s^{-2})	−	−	−	−		
700-hPa $\overline{RH}$ (%)		+	−	−	!	!
700-hPa RH′ (%)	−	−	+	+	!	!
600-hPa $\overline{RH}$ (%)	+	+	−	−	!	!
600-hPa RH′ (%)	−	−	+	+	!	!
200-hPa $\nabla_h \cdot \mathbf{V}$ ($\times 10^{-5}$ s^{-1})		+		−	!	!
$\mathbf{V}_{200} - \mathbf{V}_{850}$ (m s^{-1})		−			!	!
PWAT (kg m^{-2})	!	!		+	!	!
CAPE (J kg^{-1})	!	!	+	+	!	!
CIN (J kg^{-1})	!	!	+	+	!	!
Surface LH flux (W m^{-2})	!	!	+	+	!	!
Skin temp (K)	!	!	+	+	!	!

There are several notable similarities among the significant predictors for the univariable regression equations (Tables 1 and 2). Forecast hour is significant in each case and has a negative coefficient. The 250–850-hPa layer thickness ΔZ is significant for all regression equations except for the UKM 120-h NATL and exhibits the theoretically consistent positive coefficient. Latitude is significant with a positive coefficient for all EPAC forecasts. The NATL GFS and CMC regression equations all have negative coefficients for the following predictors: 850-hPa ζ, Okubo–Weiss (OW) parameter, and 925-hPa maximum wind speed. The negative sign is counterintuitive, suggesting that false alarms in these two models may be exaggerated because of erroneous positive feedback. For the EPAC forecasts, these same predictors only are significant at 120 h.

There are also several interesting differences in the significant predictors among the global models (Tables 1 and 2). For NATL forecasts, latitude has a positive coefficient for GFS, but is negative for CMC. This highlights some of the model biases: GFS produces a large number of false alarms equatorward of 10°N, while CMC produces numerous false alarms at higher latitudes. Both models agree that the 200-hPa divergence is statistically significant only at 120 h, but the signs of the coefficients are opposite. CMC's positive coefficient is consistent with theory: greater

upper-tropospheric divergence is related to greater outflow and a developing disturbance, which leads to a greater probability of genesis. However, GFS exhibits a counterintuitive negative coefficient.

There are also several notable differences in the coefficients between the NATL and EPAC basins (Tables 1 and 2). Fewer predictors are significant over the EPAC compared to the NATL, especially for 48-h forecasts. If a GFS 48-h predictor is significant over the NATL, it is also significant (with the same coefficient sign) at 120 h. Over the EPAC, nearly all meteorological predictors (i.e., not latitude, longitude, etc.) are significant at either 48 or 120 h, but not at both times.

Perhaps the most intriguing relationships are between RH and genesis probability. CMC over both basins exhibits a positive coefficient for the 600- and 700-hPa $\overline{RH}$, but a negative coefficient for 600- and 700-hPa RH′. This implies that greater environmental RH is associated with a greater probability of genesis [consistent with Gray (1968, 1979), Nolan (2007), and Helms and Hart (2015)]. However, GFS exhibits the opposite relationship over the NATL, with a positive coefficient for RH′ and a negative coefficient for $\overline{RH}$. One would expect that greater $\overline{RH}$ would lead to greater genesis probability. However, in a study of the NASA GEOS-5 climate model, Lim et al. (2015) suggests that enhanced entrainment of dry environmental

TABLE 2. As in Table 1, but for the EPAC basin.

Predictor	CMC48	CMC120	GFS48	GFS120	UKM48	UKM120
Forecast hour	−	−	−	−	−	−
Year	−		+	+		+
Month		−				
Lat (°)	+	+	+	+	+	+
Lon (°)	−		−	−		
925-hPa $\mathbf{V}_{max}$ (m s^{-1})**		+		−		
250–850-hPa ΔZ (m)**	+	+	+	+	+	+
850-hPa $\overline{\zeta}$ ($\times 10^{-5}$ s^{-1})		+		−		
850-hPa ζ' ($\times 10^{-5}$ s^{-1})				−		
850-hPa OW (s^{-2})		+		−		
700-hPa $\overline{RH}$ (%)	+	+	−		!	!
700-hPa RH′ (%)	−	−	+		!	!
600-hPa $\overline{RH}$ (%)	+	+		+	!	!
600-hPa RH′ (%)	−	−		−	!	!
850-hPa $\nabla_h \cdot \mathbf{V}$ ($\times 10^{-5}$ s^{-1})		−				
200-hPa $\nabla_h \cdot \mathbf{V}$ ($\times 10^{-5}$ s^{-1})		+			!	!
1000–700-hPa Γ (K km^{-1})		+	−		!	!
PWAT (kg m^{-2})	!	!		+	!	!
CAPE (J kg^{-1})	!	!		+	!	!
CIN (J kg^{-1})	!	!	−		!	!
Surface LH flux (W m^{-2})	!	!		+	!	!
Skin temp (K)	!	!		+	!	!

air can lead to drying of the midlevels and a positive latent heat (LH) flux in the lower levels. They found this scenario to be more conducive to TC genesis. This scenario also may be occurring in the GFS forecasts and might explain the observed statistical relationships between midlevel RH, surface LH flux, and genesis probability. It also may explain GFS's frequent TC genesis forecasts off the coast of Africa, where dry air may be entrained into the TC. Another possibility is that when the RH values are recorded (at the forecast genesis time), the model's TC secondary circulation is sufficiently established to produce subsidence surrounding the TC that results in smaller $\overline{RH}$. Regardless, this is an interesting disagreement that highlights the differences in how successful genesis forecasts are simulated in each model. Over the EPAC, GFS is not temporally consistent regarding the RH relationships. For 48-h forecasts, only 700-hPa RH is statistically significant (same coefficients as over the NATL). However, for the 120-h forecasts, only 600-hPa RH is significant, and the coefficient signs are opposite, such that they now agree with CMC.

Vertical wind shear between 200 and 850 hPa is significant only for CMC 120-h NATL forecasts. This does not mean that shear is unimportant for model-derived genesis. Rather, shear simply is not a good discriminator between the hit and false alarm outcomes for most of the models investigated.

To summarize, the models exhibit some of the expected statistical relationships between predictors and

genesis probability. However, there are also some surprising and counterintuitive relationships and notable differences among the models.

4. Multivariable logistic regression equation development

a. Regression equations for each individual global model

The univariable regression equations offered insight as to whether the statistical relationships between the various parameters and genesis probability are consistent with theory. Conversely, they showed which theoretically relevant variables are not useful discriminators between the hit and false alarm outcomes. This section develops multivariable regression equations to produce probabilistic genesis forecasts. The proper combination of multiple variables is expected to yield better-calibrated forecasts than any univariable regression equation. Separate MLR equations are developed for each global model, basin, and forecast window (0–48 and 0–120 h).

Details of developing the regression equations are presented using the GFS 120-h NATL forecast dataset as an example. Hosmer et al. (2013) recommend identifying the MLR equation predictors using the method of purposeful selection. This approach typically is used when it is well known which predictors have physically meaningful relationships to the outcome variable. While the literature points to several such predictors, section 3

showed that not all of these physically relevant variables are statistically significant predictors in the global models. Therefore, predictors were selected using the method of backward elimination combined with a multiple fractional polynomial analysis (Sauerbrei et al. 2006; Hosmer et al. 2013). The multiple fractional polynomial analysis assesses if the relationship between a predictor and the outcome variable is linear or if an exponential transformation of the predictor provides a better fit. This method makes no a priori assumptions about which predictors have physical relevance to the outcome variable. However, it does require that one specify the statistical significance level at which a predictor is to be removed from the regression model during the backward elimination step ($p > 0.15$) as well as the significance level for selecting a nonlinear transform ($p < 0.05$). All variables in Tables 1 and 2 are tested for significance while creating the MLR equations.

To ensure that the selected predictors are robust, cross validation (Wilks 2011) was performed. Specifically, the historical cases were split into a developmental set, which consisted of a randomly selected 95% of the events ($N = 679$), and a verification set, which comprised the remaining 5% ($N = 36$). A logistic regression equation is fit using the developmental set, and the significant predictors are recorded. This process is repeated for 20 iterations. Each time, a different, randomly selected set of events is used as the verification set. Thus, each case is used once in the verification set. This cross validation reveals in how many iterations each predictor is statistically significant. The predictors that are significant in at least 15 of the 20 iterations are denoted as the initial predictor set.

Once the initial predictor set is developed (Table 3), it is refined based on how well the predictors fit an independent/verification dataset. It is desirable to remove any predictors that do not impact the goodness of fit. Because of the potential interactions between the predictors, it is possible that a predictor may be statistically significant on its own, but is no longer significant when included with other covariates.

In an operational setting, the independent/verification set would be the current season's genesis forecasts (not a random 5% sample as used above). Thus, to refine the initial predictor set, the historical cases were split using 2004–10 cases ($N = 498$) as the developmental set and 2011–13 cases ($N = 217$) as the verification set. Three years' worth of data were chosen for the verification set to ensure a sufficient sample size. A regression equation using the initial predictors (Table 3) is fit based on the developmental dataset. This equation is tested on the events in the verification set. A reliability diagram

TABLE 3. Initial predictor set for the GFS 120-h NATL regression equation listed in order of increasing p value (except the intercept term).

Predictor	β_n	p value
(Intercept)	−581.9	5.31×10^{-6}
Forecast hour	−0.028	1.68×10^{-11}
Year	+0.283	8.55×10^{-6}
Surface LH flux	+0.148	1.43×10^{-5}
600-hPa RH′	+0.089	2.58×10^{-4}
PWAT	+0.274	2.68×10^{-4}
Lon (°)	−0.015	0.011
Relative Julian day	+0.001	0.697
CAPE	$+1.72 \times 10^{-4}$	0.706

reveals how well the regression equation fits the data. The regression-based probabilities ideally will lie along the line $y = x$ (i.e., where the forecast probability equals the verification probability). If the forecast probabilities are above (below) the line $y = x$, the regression equation underpredicts (overpredicts) the probability of TC genesis. Figure 1a shows that the initial predictor set fits the verification dataset reasonably well. Predictors with p values > 0.05 are removed from the regression model, one at a time, and a new model is fit and evaluated. This occurs until all remaining predictors have a p value < 0.05, and the goodness of fit suffers from removing any additional predictors. For example, Fig. 1b shows the reliability diagram after CAPE and the relative Julian day[1] were removed from the model. Since their removal has little negative impact on the goodness of fit, they were deleted from the final predictor set. Thus, Table 4 and Fig. 1b describe the final set of predictors for the GFS-based 120-h NATL regression equation. This process was conducted for each model (CMC, GFS, and UKMET), for each basin, and for each forecast time period (0–48 and 0–120 h).

The final set of predictors and their coefficients for each regression equation used in 2014's operational testing are presented in Table 5. While the predictors were selected using 2004–10 as the developmental set as described above, the predictor coefficients are recalibrated using 2004–13 as the developmental set for 2014 testing. The impact of adding 2011–13 to the developmental set is evident by comparing Table 4 with the GFS NATL 120-h column in Table 5. The coefficient values change slightly, but the equations are quite similar overall.

There are a number of interesting similarities and differences among the regression equations (Table 5).

---

[1] Relative Julian day is defined here as the difference between the current Julian day and the Julian day of the climatological peak of the TC season.
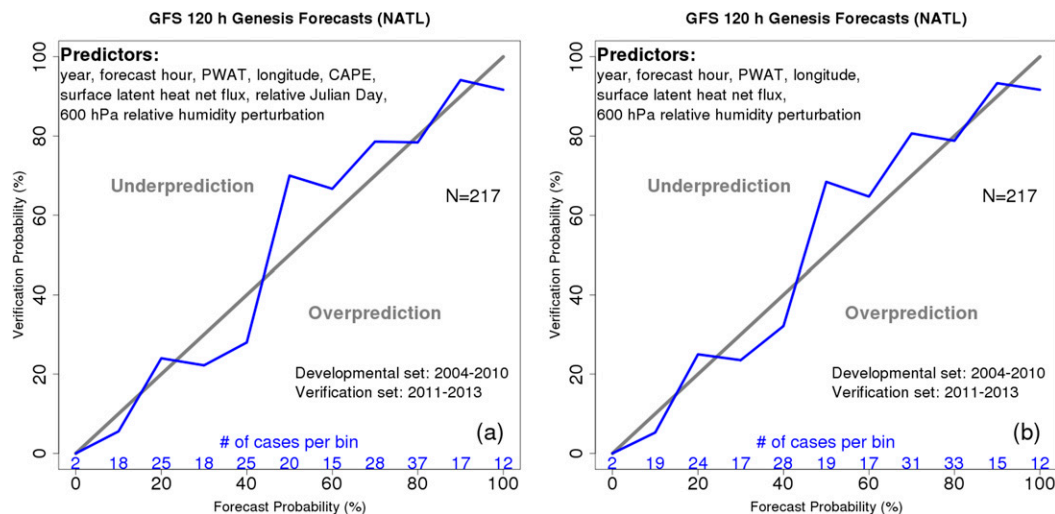
FIG. 1. Reliability diagram of GFS-based regression equation probabilities for 0–120-h forecasts over the NATL for the (a) initial and (b) final predictor sets. Results are from the independent data (2011–13 cases). Perfect reliability is given by the gray, diagonal line; above (below) this line indicates underprediction (overprediction).

Forecast hour is a significant predictor in each of the regression equations. While not a physical covariate, it does have predictive power in determining whether a given model genesis forecast is more likely to result in a hit or a false alarm. It also may suggest when physical biases become more pronounced. The negative coefficient indicates that as genesis is predicted later in the forecast cycle, the probability of genesis actually occurring decreases. This trend is confirmed by results shown in H13 and H16. Latitude and/or longitude are also included in most of the regression equations. Again, while these location covariates are not physically based, they do fit the data and may act as proxies for other variables. For example, longitude in the GFS-based NATL regression equations captures the large number of GFS false alarms over the main development region (MDR). Although some regression equations contain no physically based covariates, most regression equations contain at least one. For example, surface LH flux is significant in most of the GFS-based regressions. For the CMC-based EPAC 120-h regression, an increased 1000–700-hPa lapse rate and 850-hPa convergence yield a greater genesis probability. This is consistent with the upward vertical motion that is needed for TC development. As noted earlier, the limited archive of UKM model forecast fields reduces the number of predictors that are available for testing.

b. Regression equation for consensus of global models

H13 indicated that when multiple global models predict the same TC genesis event, the probability of actual genesis increases. Thus, in addition to the individual global model–based regression equations described in the previous section, a consensus-based regression equation (denoted CON) was developed. For the individual global model–based regression equations, the $g(x)$ term in (2) consisted of predictor values from that global model's forecast fields. However, the $g(x)$ term for the CON regression [(3)] is based on the appropriately weighted genesis probabilities from each of the three individual regression equations. That is,

$$g(x) = \beta_1[P(\mathrm{CMC})] + \beta_2[P(\mathrm{GFS})] + \beta_3[P(\mathrm{UKM})]. \quad (3)$$

If a global model does not predict genesis, the regression-based probability for that individual model term is zero. The TC genesis probabilities from all of the aforementioned multiple logistic regression equations, including CON, were generated in real time during 2014 and 2015. These probabilities were made available to NHC forecasters for evaluation. The verification of the regression equations is presented next.

TABLE 4. As in Table 3, except for the final predictor set.

Predictor	β_n	p value
(Intercept)	−588.5	1.16×10^{-6}
Forecast hour	−0.028	5.29×10^{-12}
Year	+0.286	2.14×10^{-6}
Surface LH flux	+0.015	1.44×10^{-5}
600-hPa RH′	+0.093	5.34×10^{-5}
PWAT	+0.284	5.75×10^{-5}
Lon (°)	−0.014	0.012

Table 5. Final predictor list with coefficients (β_n) for 2014 regression equations. Predictors are listed in order of decreasing significance based on their p values (except the intercept term).

NATL 48 h	NATL 120 h	EPAC 48 h	EPAC 120 h
CMC			
−174.6 (intercept)	+7.57 (intercept)	−2.81 (intercept)	−5.33 (intercept)
−0.04 × forecast hour	−0.03 × forecast hour	−0.03 × forecast hour	−0.01 × forecast hour
+0.02 × 250–850-hPa ΔZ	−0.07 × lat	+0.24 × lat	+0.22 × lat
−0.04 × lat	−0.02 × lon		+1.12 × 1000–700-hPa Γ
−0.01 × lon			−1.08 × 850-hPa $\nabla_h \cdot \mathbf{V}$
GFS			
−110.8 (intercept)	−623.3 (intercept)	−3.98 (intercept)	−3.21 (intercept)
−0.04 × forecast hour	−0.03 × forecast hour	−0.04 × forecast hour	−0.02 × forecast hour
+0.01 × 250–850-hPa ΔZ	+0.30 × yr	+0.29 × lat	+0.24 × lat
−0.03 × lon	−0.02 × lon	+0.01 × surface LH flux	+0.01 × surface LH flux
	+0.40 × PWAT		−0.02 × 850-hPa ζ'
	+0.01 × surface LH flux		
	+0.12 × 600-hPa RH′		
UKM			
+8.13 (intercept)	+3.71 (intercept)	−0.90 (intercept)	−1.13 (intercept)
−0.03 × forecast hour	−0.02 × forecast hour	−0.04 × forecast hour	−0.02 × forecast hour
−0.02 × lon	−0.02 × lon	+0.21 × lat	+0.22 × lat
	−4224 × lat^{-2}		
	+1898 × lat^{-2} × ln (lat)		

5. Verification

a. 2014

Verification of each regression equation first is presented for the 2014 hurricane season. For the remainder of the paper, when referencing a global model (e.g., GFS), the text discusses the regression equation developed from that model's output (unless specified otherwise). NHC's verification is obtained from Cangialosi and Franklin (2015). The nonhomogeneous[2] results are presented first using the Brier score. The Brier score is the mean of the square of the difference between the regression-based forecast genesis probability and the outcome, which equals 1 for a hit and 0 for a false alarm (Wilks 2011). It is expressed as

$$\text{Brier Score} = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2, \qquad (4)$$

where n is the number of forecast pairs, y_k is the forecast genesis probability for the kth forecast, and o_k is the outcome for the kth forecast (Wilks 2011, his Eq. 8.36). Given that the forecast probabilities range from 0 to 1

[2] 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.

and the outcome values are either 0 or 1, the Brier score values here will range from 0 to 1. A Brier score of 0 indicates a perfect forecast.

Wilks (2011) also shows that the Brier score can be expressed as the sum of three components: reliability, resolution, and uncertainty. The reliability term indicates how well calibrated the forecast probabilities are. As with the full Brier score, a reliability value of 0 is desirable and indicates a perfectly calibrated regression equation. The resolution term indicates how well the forecasts can distinguish between the hits and false alarms. For example, if the verification probability is 50% for all forecast probability bins, the regression equation has poor resolution. Finally, the amount of uncertainty in the forecast is given by the uncertainty term. This term depends on the climatological probability of the hit outcome. For models with a smaller success ratio (e.g., CMC), the uncertainty term is smaller. The uncertainty term ranges from 0 to 0.25; it is minimized when the success ratio is either 0 or 1 and maximized when the success ratio is 0.5. The Brier score can be calculated from these three terms as

$$\text{Brier Score} = \text{reliability} - \text{resolution} + \text{uncertainty}. \qquad (5)$$

Wilks (2011) provides a more detailed description of each of the Brier score components, but it is evident here

TABLE 6. Brier score and its components for each set of 2014 forecasts.

Model	Brier score	Reliability	Resolution	Uncertainty
		NATL 48 h		
CMC	0.156	0.032	0.026	0.150
GFS	0.295	0.176	0.032	0.151
UKM	0.281	0.205	0.098	0.174
CON	0.096	0.017	0.039	0.118
NHC	0.089	0.008	0.041	0.121
		NATL 120 h		
CMC	0.181	0.010	0.057	0.228
GFS	0.249	0.089	0.080	0.240
UKM	0.217	0.045	0.078	0.250
CON	0.210	0.011	0.042	0.241
NHC	0.126	0.014	0.083	0.195
		EPAC 48 h		
CMC	0.135	0.028	0.058	0.165
GFS	0.187	0.026	0.079	0.240
UKM	0.233	0.112	0.057	0.178
CON	0.104	0.011	0.062	0.155
NHC	0.117	0.008	0.074	0.184
		EPAC 120 h		
CMC	0.179	0.005	0.056	0.230
GFS	0.215	0.053	0.059	0.221
UKM	0.221	0.008	0.037	0.250
CON	0.200	0.019	0.069	0.250
NHC	0.200	0.026	0.075	0.250

that small values of reliability and uncertainty and a large value of resolution are needed for the desirable small Brier score. Table 6 gives the Brier score values with each of its components for all 2014 forecasts. For the NATL 48-h forecasts, NHC and CON have the smallest Brier scores, due in part to their smaller reliability and uncertainty values. Among the individual regression equations, CMC has the smallest Brier score because of its superior reliability, compared to GFS and UKM. At 120 h, NHC again exhibits the smallest Brier score. NHC's reliability is comparable to CON and CMC, and its resolution is comparable to those of GFS and UKM, but its uncertainty value is smaller than those of all the regression models. The uncertainty terms for the regression models are near the maximum value of 0.25, indicating that the success ratio is near 0.5. This contributes to the relatively large Brier scores for the regression equations.

CON exhibits the smallest Brier score for the EPAC 48-h forecasts. NHC is a close second, but suffers a bit from larger uncertainty values. CMC again has the smallest Brier score among the individual model regression equations. At 120 h, CMC has the smallest Brier score overall because of its good reliability and relatively smaller uncertainty. UKM, CON, and NHC all exhibit the maximum uncertainty value of 0.25, which largely causes the total Brier scores near 0.2.

The reliability and resolution components of the Brier score are also provided graphically using reliability diagrams (Fig. 2) that show the verification for each single-model regression equation (color coded), the CON regression equation (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 CMC, GFS, and NHC TWO. However, at probabilities ≥ 30%, CMC and NHC TWO underpredict genesis, while UKM and CON overpredict genesis. For the 120-h NATL (Fig. 2b), CMC is well calibrated, while CON exhibits some overprediction bias. GFS and UKM 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 CMC, CON, and NHC perform fairly well in the 0%–40% range, they stray from the perfect reliability line at the higher forecast probability bins, with CMC, GFS, and NHC (CON and UKM) underpredicting (overpredicting) genesis. For the 120-h EPAC forecasts (Fig. 2d), GFS and NHC generally underpredict genesis. CMC and UKM are fairly well calibrated.

Overprediction by UKM (Figs. 2a,b) may be due in part to a new global model configuration that was implemented 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 produce more intense disturbances or early-stage TCs, thus causing the UKM-based regression equation to overpredict genesis during 2014. Upgrades to all global models in the guidance suite undoubtedly impact the reliability of the regression equations. The UKM global model upgrade is the most obvious example during 2014.

The 120-h NATL GFS regression equation exhibits especially poor reliability during 2014 (Fig. 2b). The use of "year" as a predictor was a contributing factor. While the developmental dataset 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 2010–13. Thus, it became apparent in the postseason verification that the GFS-based
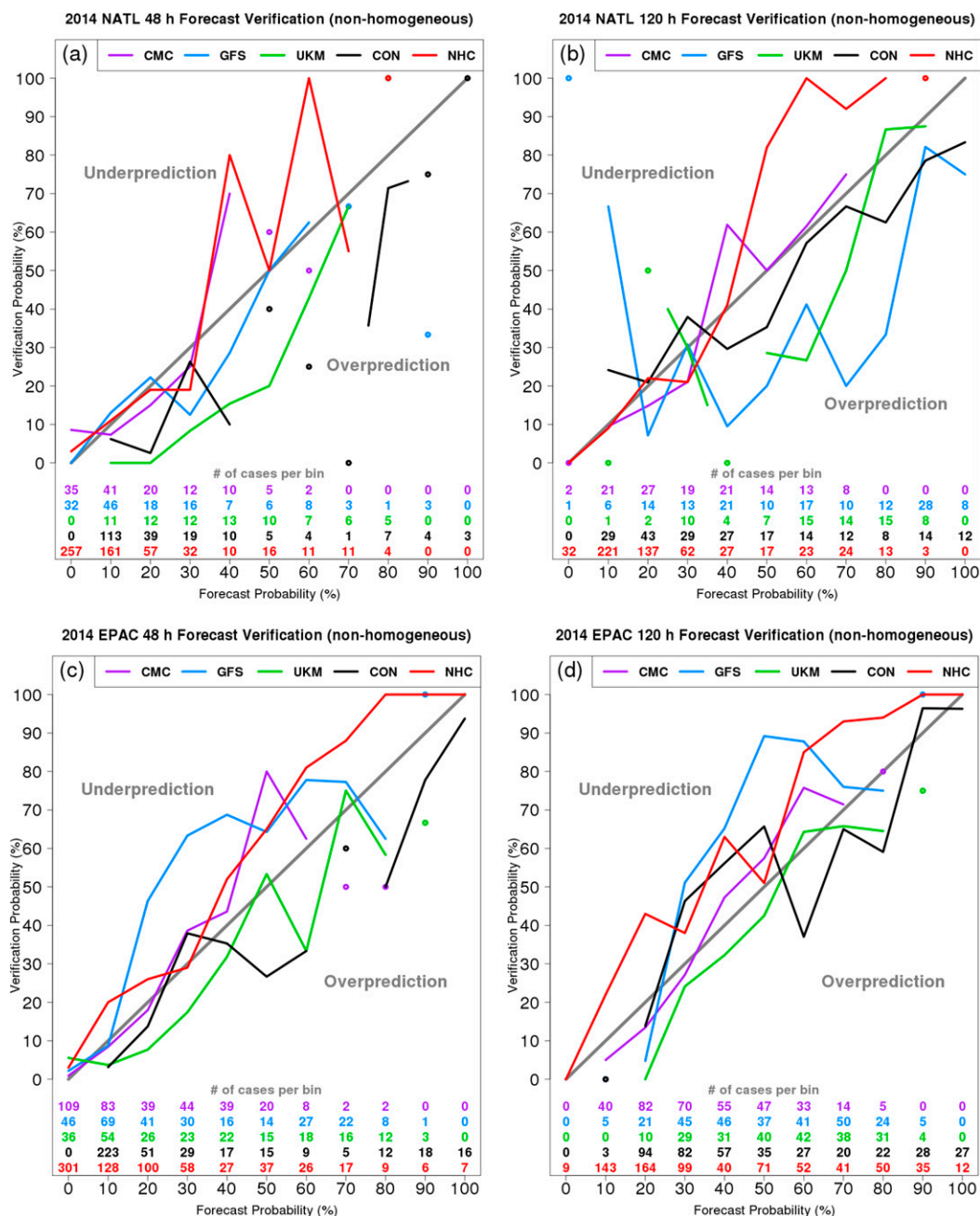
FIG. 2. Reliability diagram comparing the nonhomogeneous set of regression and NHC TWO probabilities for the (a) NATL 48-h, (b) NATL 120-h, (c) EPAC 48-h, and (d) EPAC 120-h forecasts during 2014. Verification is based on the best track. Perfect reliability is given by the gray, diagonal line; above (below) this line indicates underprediction (overprediction). Breaks in the lines indicate forecast probability bins with five or fewer cases.

regression probabilities were inflated by including year as a predictor. The operational GFS-based regression equation for the NATL at 120 h was compared to a regression equation that excluded year as a predictor. While still far from perfect, removing year as a predictor would have prevented the notable overprediction (not shown).

It is encouraging that the CMC-based regression equations performed well for both basins and forecast windows (Fig. 2). While historical verification indicates that the false alarm ratio for the CMC global model is greater than for the other global models (H16), it appears that the regression equations are able to correct
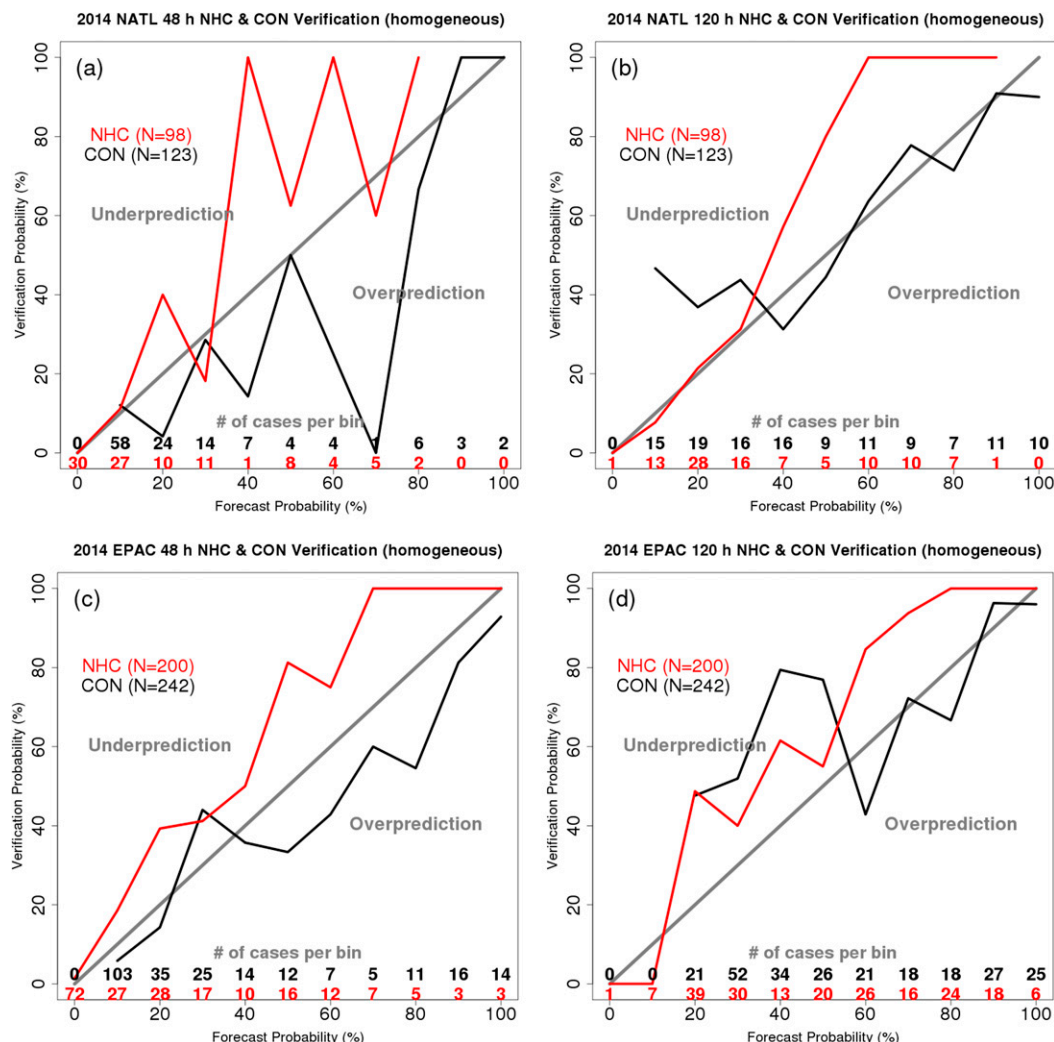
FIG. 3. Reliability diagram comparing the homogeneous set of 2014 NHC TWO (red line) and CON regression (black line) probabilities for the (a) NATL 48-h, (b) NATL 120-h, (c) EPAC 48-h, and (d) EPAC 120-h forecasts. The number of cases in each forecast probability bin is given in the corresponding red and black text. Verification is based on the best track.

for the global model's biases and provide well-calibrated probabilistic forecasts.

To provide a more direct comparison of the verification results, a set of homogeneous[3] NHC TWO and CON regression forecasts was constructed, with the associated reliability diagrams presented in Fig. 3. The verification of the NATL 48-h forecasts in the 10% and 30% forecast probability bins is comparable between CON and NHC (Fig. 3a). For probabilities exceeding 30%, the sample size—given by the black (CON) and red (NHC TWO) text—is too small. Using probability bins with a 20% interval reveals underprediction in the NHC forecasts and overprediction from the CON forecasts (not shown). However, the sample size still is fairly small even when using the 20% probability interval. Sample size is not an issue at 120 h (Fig. 3b). NHC TWO outperforms CON in the 0%–30% forecast probability range. However, at the higher probability bins, CON is better calibrated. Over the EPAC at 48 h (Fig. 3c), NHC TWO (CON) underpredicts (overpredicts) genesis. At 120 h, CON struggles in the 20%–50% forecast probability range, but is fairly well calibrated in the 70%–100% range (Fig. 3d). NHC TWO generally underpredicts genesis.

The sample sizes of the NHC TWO and CON probabilities are not equal. There are a few instances where

---

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

TABLE 7. As in Table 5, but for 2015 regression equations.

NATL 48 h	NATL 120 h	EPAC 48 h	EPAC 120 h
CMC			
-174.58 (intercept)	$+8.70$ (intercept)	-3.25 (intercept)	-6.61 (intercept)
$-0.04 \times$ forecast hour	$-0.03 \times$ forecast hour	$-0.03 \times$ forecast hour	$-0.016 \times$ forecast hour
$+0.02 \times 250\text{–}850\text{-hPa } \Delta Z$	$-0.07 \times$ lat	$+0.28 \times$ lat	$+0.24 \times$ lat
$-0.04 \times$ lat	$-0.02 \times$ lon		$+1.35 \times 1000\text{–}700\text{-hPa } \Gamma$
$-0.01 \times$ lon			$-1.08 \times 850\text{-hPa } \nabla_h \cdot \mathbf{V}$
GFS			
-245.01 (intercept)	-285.7 (intercept)	-4.02 (intercept)	-4.70 (intercept)
$-0.04 \times$ forecast hour	$-0.03 \times$ forecast hour	$-0.04 \times$ forecast hour	$-0.02 \times$ forecast hour
$+0.03 \times 250\text{–}850\text{-hPa } \Delta Z$	$-0.20 \times 200\text{–}850\text{-hPa shear}$	$+0.32 \times$ lat	$+0.27 \times$ lat
$-0.02 \times$ lon	$+0.03 \times 250\text{–}850\text{-hPa } \Delta Z$	$+0.01 \times$ surface LH flux	$+0.01 \times$ surface LH flux
$+0.01 \times$ surface LH flux	$-0.02 \times 850\text{-hPa } \zeta'$		$-0.02 \times 850\text{-hPa } \zeta'$
$-0.04 \times 850\text{-hPa } \zeta'$	$+0.02 \times$ surface LH flux		$+0.001 \times$ CAPE
UKM			
$+8.28$ (intercept)	$+3.71$ (intercept)	-3.30 (intercept)	-1.13 (intercept)
$-0.03 \times$ forecast hour	$-0.02 \times$ forecast hour	$-0.0003 \times$ forecast hour^2	$-0.02 \times$ forecast hour
$-0.02 \times$ lon	$-0.02 \times$ lon	$-63.9 \times \text{lat}^{-1}$	$+0.22 \times$ lat
	$-4224 \times \text{lat}^{-2}$	$+0.04 \times$ lon	
	$+1897 \times \text{lat}^{-2} \times \ln(\text{lat})$		

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, because each model genesis forecast occurs within the TWO-shaded potential genesis region, all three forecasts are included in the homogeneous verification. This issue 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. Since it does not occur frequently, however, this issue likely is not significant.

b. 2015

All regression equations were recalibrated prior to operational testing during 2015 to determine whether any predictors should be added or removed when the 2014 forecasts were added to the developmental dataset that originally consisted of the years 2004–13. The predictors used in 2015 are given in Table 7. The predictors for all CMC-based equations are unchanged (Tables 5 and 7), although the coefficients are slightly different in 2015 compared with 2014. The 2015 UKM-based predictors are similar to the 2014 predictors (Tables 5 and 7), except that longitude has been added to or removed from some equations. The greatest differences between the 2014 and 2015 versions of the regression equations occur with the GFS-based equations for the NATL (Tables 5 and 7). Regardless of potential significance, year was removed from the GFS-based regression

model since there was no guarantee that improved GFS genesis forecasts would occur during 2015. In fact, retrospective TC genesis forecasts for 2012–14 using the 2015 version of GFS reveals that the 2015 version of GFS exhibits a greater false alarm ratio and smaller probability of detection over the NATL compared with the 2012–14 operational versions of GFS (not shown).

Nonhomogeneous Brier scores with their three components for 2015 forecasts are given in Table 8. NHC exhibits the smallest Brier scores with CON generally a close second. NHC and CON have comparable reliability, but NHC produces better resolution. CMC has smaller Brier scores than GFS and UKM. The smaller success ratio for CMC provides it less uncertainty than GFS and UKM, especially at 120 h.

Nonhomogeneous verification of the 2015 regression equations was conducted using the best-track files[4] (Fig. 4). NHC's verification is obtained from Cangialosi and Franklin (2016). The 48-h NATL NHC verification (Fig. 4a) shows some well calibrated forecasts; however, CMC, UKM, and CON overpredict genesis. Meanwhile GFS underpredicts genesis in the 0%–20% forecast probability range. There are small sample sizes in the higher forecast probability bins of all regression equations. NHC's forecasts for the NATL at 120 h (Fig. 4b) are well calibrated in the 0%–40% range but underpredict genesis at the higher forecast probabilities. CON is reliable in the 10%–40% and 90% bins, but

---

[4] Best-track file for EP042015 is preliminary as of 9 Aug 2016.

TABLE 8. As in Table 6, but for 2015.

Model	Brier score	Reliability	Resolution	Uncertainty
		NATL 48 h		
CMC	0.115	0.047	0.028	0.096
GFS	0.142	0.033	0.030	0.139
UKM	0.268	0.132	0.041	0.177
CON	0.108	0.007	0.026	0.127
NHC	0.087	0.002	0.066	0.150
		NATL 120 h		
CMC	0.151	0.009	0.044	0.186
GFS	0.265	0.069	0.054	0.250
UKM	0.247	0.040	0.041	0.248
CON	0.190	0.010	0.044	0.224
NHC	0.145	0.007	0.077	0.215
		EPAC 48 h		
CMC	0.138	0.042	0.018	0.114
GFS	0.222	0.011	0.034	0.245
UKM	0.223	0.082	0.041	0.182
CON	0.111	0.008	0.036	0.139
NHC	0.110	0.004	0.085	0.191
		EPAC 120 h		
CMC	0.181	0.008	0.023	0.196
GFS	0.223	0.073	0.024	0.174
UKM	0.220	0.010	0.040	0.250
CON	0.169	0.008	0.073	0.234
NHC	0.166	0.012	0.095	0.249

overpredicts genesis in the 50%–60% and 100% forecast probability bins. UKM (GFS) generally overpredicts (underpredicts) genesis, similar to that at 48 h.

The regression equations generally are better calibrated over the EPAC than the NATL. All guidance except GFS performs well in the 0%–30% forecast range for the 48-h EPAC forecasts (Fig. 4c). NHC and GFS (CMC and CON) generally underpredict (overpredict) genesis at the higher forecast probabilities. UKM is well calibrated, except for overprediction in the 40%–50% range. At 120 h, UKM and CON are well calibrated, especially in the 30%–60% range (Fig. 4d). GFS (CMC) underpredicts (overpredicts) genesis. NHC's forecasts generally are reliable, with underprediction in some probability bins.

Verification of the homogeneous set of NHC TWO and CON forecasts also was conducted (Fig. 5). With NHC's increased use of the guidance products experimentally during 2015 compared with 2014 (E. Blake 2016, personal communication), this comparison becomes less independent, and it is increasingly difficult for the CON forecasts to outperform the NHC TWO forecasts. Small sample sizes preclude meaningful conclusions for the 48-h NATL forecasts (Fig. 5a), except in the 10%–30% forecast probability range, where CON exhibits slight underprediction. CON generally underpredicts genesis, with the best reliability in the 40%–60%

forecast probability range for the 120-h NATL forecasts (Fig. 5b). Unlike 2014, the NHC TWO forecasts are better calibrated than CON at the high forecast probabilities. The NHC TWO 48-h EPAC forecasts also are very reliable in the 80%–90% forecast probability range (Fig. 5c). Otherwise, NHC (CON) generally underpredicts (overpredicts) genesis for this time period. The verification for the 120-h EPAC forecasts (Fig. 5d) is fairly similar for both sets of forecasts, with a general underprediction bias.

c. Combined 2014–15

Given the relatively small sample size for some of the verification comparisons—possibly due in part to relatively quiet TC seasons over the NATL—results for the combined 2014–15 seasons are presented. Figure 6 shows the homogeneous NHC and CON verification for the combined 2014–15 period. The probabilities are comparable in the 10%–40% forecast probability range for the NATL 48-h forecasts (Fig. 6a). NHC (CON) tends to underpredict (overpredict) genesis at the larger forecast probabilities. At 120 h, CON is fairly well calibrated in the medium-to-high forecast probabilities, while NHC tends to underpredict genesis (Fig. 6b). NHC's forecasts in the 0%–40% probability range are quite reliable. NHC (CON) generally underpredicts (overpredicts) genesis for the EPAC 48-h forecasts (Fig. 6c). NHC underpredicts genesis at 120 h, with CON exhibiting good reliability in the 60%–90% forecast probability range.

The spatial distribution of the genesis forecasts is also of interest. Figure 7 shows the forecast genesis location, 120-h regression-based probability (color coded), and whether the forecast verified as a hit (filled circle) or false alarm (open circle) for each model. It also shows the best-track genesis locations during 2014–15 [black plus signs (+)]. For all of the models, the only clear spatial bias over the EPAC is that more false alarms occur at the more equatorward latitudes. Since latitude is a predictor in the EPAC regression equations, the probabilities for low-latitude genesis forecasts tend to be relatively small.

There are different spatial biases over the NATL. For example, the CMC forecast probabilities over the NATL are greatest over the MDR between 25° and 45°W (Fig. 7a). Meanwhile, there are many small forecast probability events over the northwest and northcentral Atlantic. Appropriately, most of those forecasts are false alarms. The majority of GFS forecast genesis events occur over the MDR. There appears to be a shift in predictability over the MDR near 30°W. East (west) of that longitude there are numerous (fewer) false alarms. UKM also forecasts many genesis events over
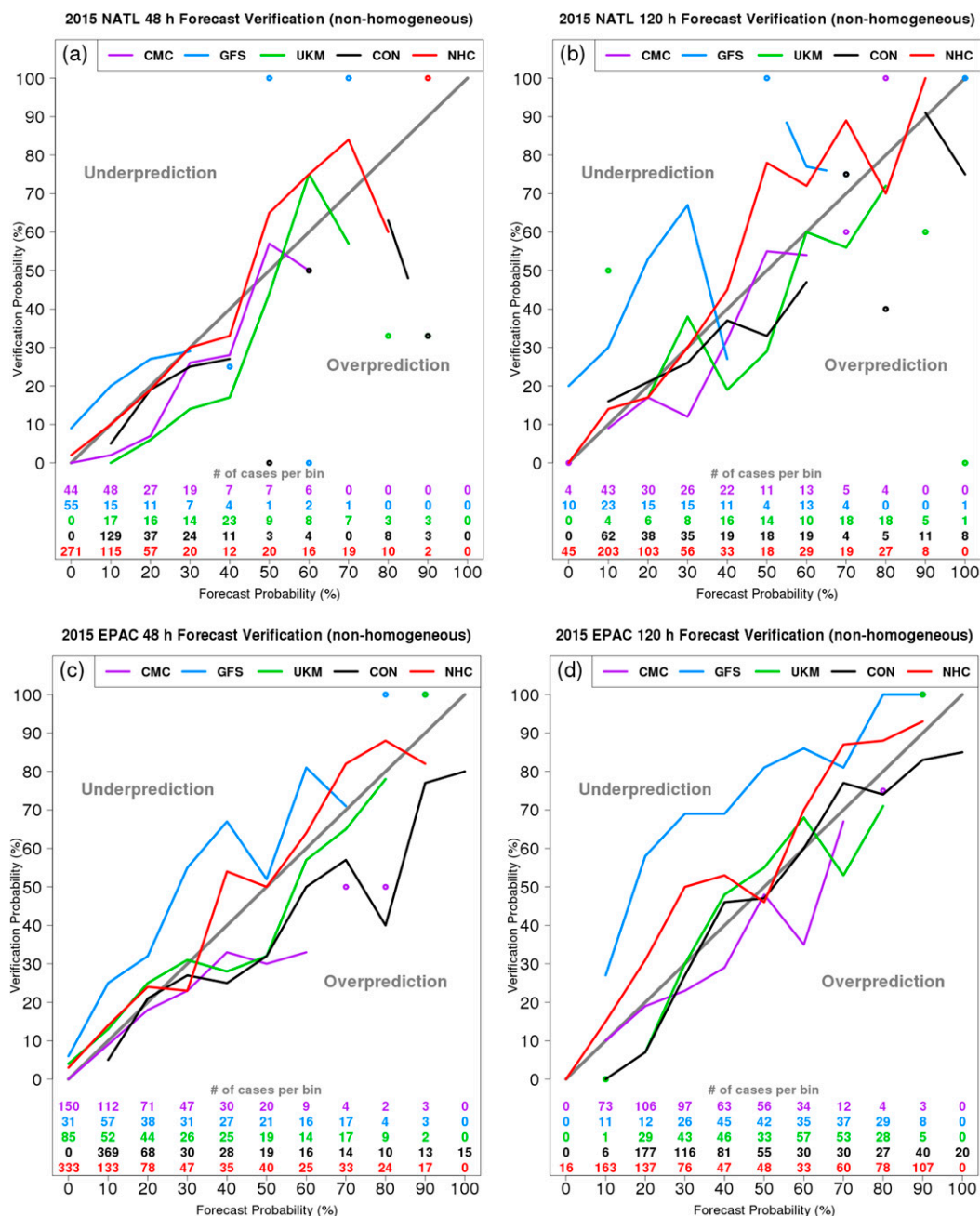
FIG. 4. As in Fig. 2, but for 2015 forecasts.

6. Summary and conclusions

The goal of this study was to develop a set of reliable operational TC genesis probability guidance products based on global model output. A decade's worth of historical model–indicated TC genesis forecasts was used as the developmental dataset for deriving multiple logistic regression equations. These equations attempt to bias correct the model genesis forecasts and provide probabilistic TC genesis forecasts within 48 and 120 h over the North Atlantic and eastern North Pacific basins.

Univariable logistic regression equations were developed to determine the significance of various predictors in discriminating between the hit and false alarm outcomes for global model TC genesis forecasts. The

the MDR. Most of its genesis forecasts poleward of 25°N verify as false alarms.

FIG. 5. As in Fig. 3, but for 2015 forecasts.

signs of the regression coefficients were used to determine whether the statistical relationships between a predictor and the probability of genesis were consistent with proposed genesis theories. Results showed that the models exhibited some of the expected statistical relationships. For example, a greater 250–850-hPa thickness (i.e., stronger warm core) resulted in greater genesis probabilities. However, several counterintuitive relationships were noted. For example, larger values of 850-hPa relative vorticity were associated with smaller genesis probabilities, contrary to the expected enhanced cyclonic relative vorticity being needed for genesis (Gray 1968, 1979). The statistical relationships between midlevel relative humidity and genesis probability were intriguing. CMC exhibited a positive coefficient for the 600- and 700-hPa environmental relative humidity and a negative

coefficient for 600- and 700-hPa relative humidity perturbation. This implies that greater environmental relative humidity was associated with a greater probability of genesis [consistent with Gray (1968, 1979), Nolan (2007), and Helms and Hart (2015)]. However, GFS exhibited the opposite relationship over the NATL, indicating that smaller average relative humidity was associated with greater genesis probability. This could be due to enhanced entrainment [similar to findings from Lim et al. (2015)] or the GFS's TC secondary circulation being sufficiently established at the genesis time to cause subsidence and reduced relative humidity in the TC environment. Regardless, it is an interesting disagreement among the models.

Multivariable logistic regression equations then were developed to provide probabilistic TC genesis guidance. Separate equations were developed for each global model, basin, and forecast window. Predictors were selected using
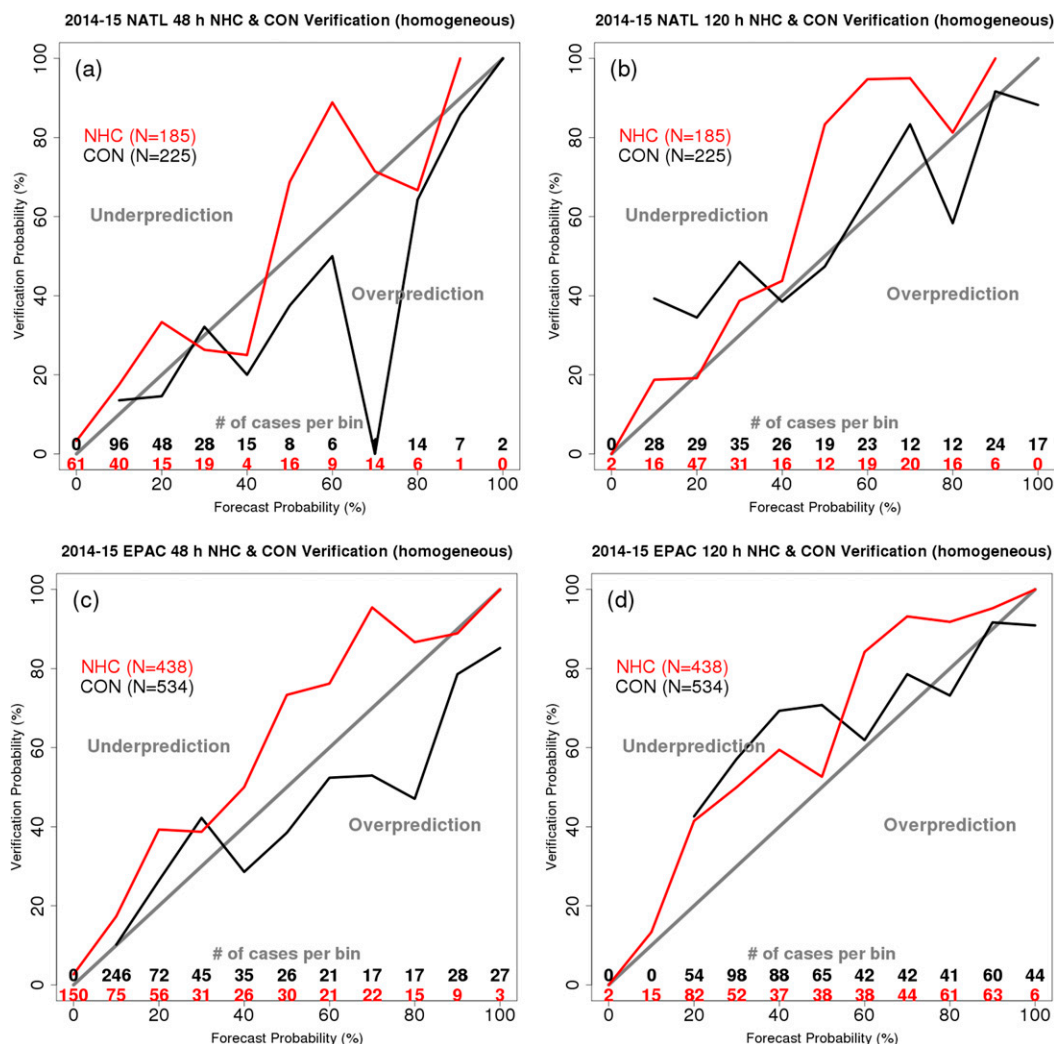
2014-15 NATL 48 h NHC & CON Verification (homogeneous)

(a)

NHC (N=185)
CON (N=225)

Underprediction

Overprediction

# of cases per bin

| 0 | 96 | 48 | 28 | 15 | 8 | 6 | | 14 | 7 | 2 |
| 61 | 40 | 15 | 19 | 4 | 16 | 9 | 14 | 6 | 1 | 0 |

2014-15 NATL 120 h NHC & CON Verification (homogeneous)

(b)

NHC (N=185)
CON (N=225)

Underprediction

Overprediction

# of cases per bin

| 0 | 28 | 29 | 35 | 26 | 19 | 23 | 12 | 12 | 24 | 17 |
| 2 | 16 | 47 | 31 | 16 | 12 | 19 | 20 | 16 | 6 | 0 |

2014-15 EPAC 48 h NHC & CON Verification (homogeneous)

(c)

NHC (N=438)
CON (N=534)

Underprediction

Overprediction

# of cases per bin

| 0 | 246 | 72 | 45 | 35 | 26 | 21 | 17 | 17 | 28 | 27 |
| 150 | 75 | 56 | 31 | 26 | 30 | 21 | 22 | 15 | 9 | 3 |

2014-15 EPAC 120 h NHC & CON Verification (homogeneous)

(d)

NHC (N=438)
CON (N=534)

Underprediction

Overprediction

# of cases per bin

| 0 | 0 | 54 | 98 | 88 | 65 | 42 | 42 | 41 | 60 | 44 |
| 2 | 15 | 82 | 52 | 37 | 38 | 38 | 44 | 61 | 63 | 6 |

FIG. 6. As in Fig. 3, but for 2014–15 forecasts.

backward elimination combined with a multiple fractional polynomial analysis. Cross validation was conducted to ensure that the predictor pool was robust.

Verification of the regression-based forecasts during the 2014 season revealed that some were well calibrated (Fig. 2). However, it appears that an upgrade to the UKM global model configuration caused the regression equations to overpredict genesis over the NATL. The GFS-based 120-h NATL regression equation also suffered from overprediction. A retrospectively developed regression equation with year removed as a predictor exhibited somewhat better reliability (not shown). The CMC-based regression equations were quite reliable. Although the CMC global model exhibited more false alarms compared to GFS and UKM, it appears that CMC had more consistent biases than the other models. Thus, the regression technique was able to correct for the biases and provide well-calibrated probabilistic genesis forecasts.

However, one should not solely rely on the CMC, since it does not often produce probabilities > 70%. The consensus forecast benefits from information from all three individual models and is able to capture the high genesis probability events. Homogeneous comparisons between the consensus regression equation and NHC TWO probabilities (Fig. 3) showed that NHC performed best in the 0%–30% forecast probability range at 120 h. However, the 120-h consensus regression equations were more reliable in the higher forecast probability ranges.

Verification during the 2015 season revealed some results similar to those from 2014. The GFS- (UKM-) based regression equations underpredicted (overpredicted) genesis over the EPAC (NATL). In contrast to 2014, the 120-h NATL GFS-based regression equation underpredicted genesis during 2015. The CON regression equation had a slight overprediction bias in both basins. NHC's forecasts generally were well calibrated,
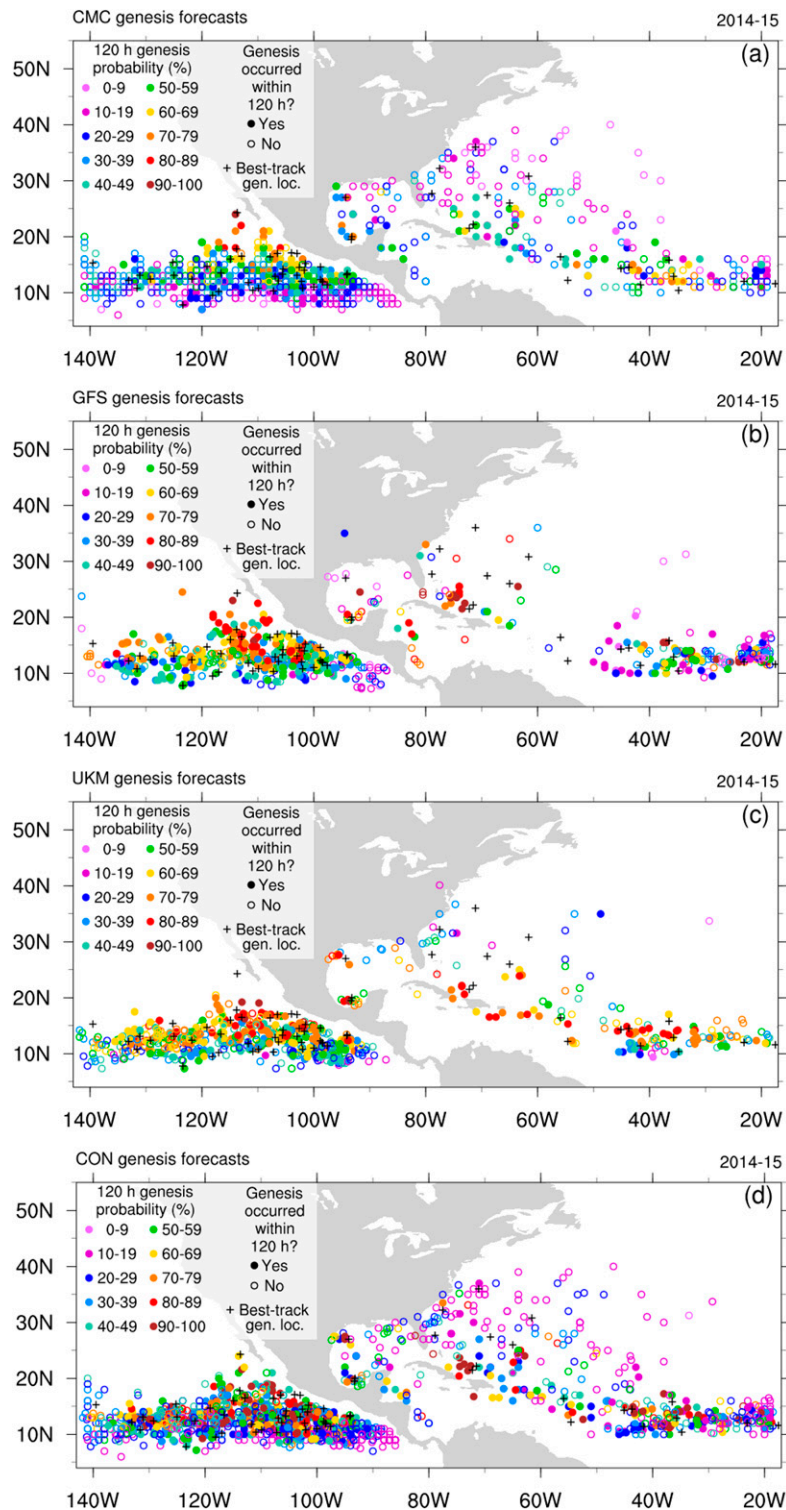
FIG. 7. Plot showing the locations of model-indicated genesis events, the regression-based 120-h genesis probability (color coded), and whether the forecast verified as a hit (filled circle) or false alarm (open circle) for (a) CMC, (b) GFS, (c) UKM, and (d) CON.
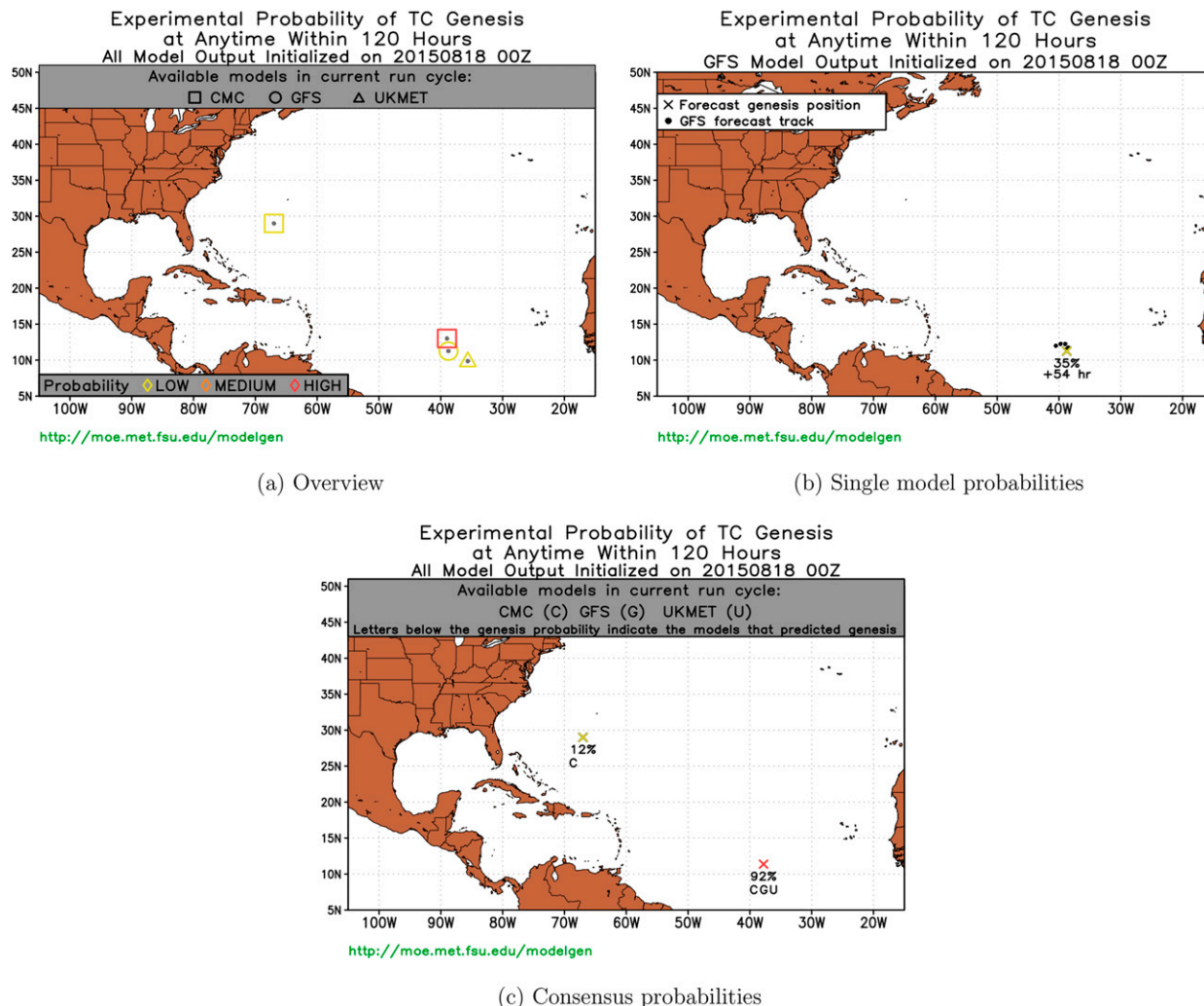
(a) Overview

(b) Single model probabilities

(c) Consensus probabilities

FIG. A1. Example graphics from the Tropical Cyclone Genesis Portal website. (a) Overview page of TC genesis forecasts for the NATL. (b) Graphic of GFS-indicated TC genesis events. The ex (×) indicates the forecast genesis location (colored by categorical genesis probability) and the filled circles (●) give the model-indicated forecast track. The "+54 hr" indicates the time with respect to model initialization at which the genesis criteria were met. This time is provided for situational awareness, but the probability of genesis is valid for the 0–120-h forecast window. The genesis probability is calculated from the multiple logistic regression equation for the GFS 120-h NATL forecasts. (c) Graphic of the consensus-based TC genesis events. The × gives the average genesis location (colored by categorical genesis probability). The genesis probability is calculated from the consensus-based logistic regression equation. The letters under the probability indicate the models that are predicting genesis to occur.

with some underprediction bias. The homogeneous comparison between the NHC and CON forecasts revealed that NHC's underprediction bias at the higher forecast probability bins was reduced in 2015 compared with 2014. NHC tested and evaluated these guidance products during 2014 and 2015 to determine whether they showed promise as an operational guidance tool.

APPENDIX A

Real-Time Guidance Products

Since this research was sponsored by the Joint Hurricane Testbed, the major goal was to create real-time

TC genesis guidance products that could be used by the Hurricane Specialist Unit at the National Hurricane Center. The multiple logistic regression equations developed in section 4 were used to determine 48- and 120-h genesis probabilities for each model-indicated TC genesis event. The process was fully automated, and output was displayed on a locally hosted website (http://moe.met.fsu.edu/modelgen) that was accessible by NHC and the general public. A suite of guidance products was provided.

An overview graphic of the current model initialization cycle (e.g., Fig. A1a) is shown on the website's home page. This graphic indicates which models are predicting genesis and the categorical genesis probability for each model-generated TC. The user can gain more detailed information by selecting an individual model on the left toolbar. For example, Fig. A1b shows the forecast genesis location, track, and genesis probability for each GFS-indicated genesis event. Multiple models sometimes predict the same genesis event. In those cases, the consensus product (Fig. A1c) provides a single genesis probability for each forecast TC.

Several text products also are available. All graphics have a corresponding text product that gives the genesis forecast time, location, and probability in a tabular format. A more detailed table provides the values of the genesis criteria and the predictors used in the regression equations at each 6-h model output time. Finally, a history file of each forecast TC displays how the forecast genesis time, location, and probabilities have changed with each model initialization cycle. This gives the forecasters trend information and allows them to see how consistently a given TC has been forecast.

APPENDIX B

Definition of Symbols

Table B1 provides the definitions of the symbols used in this paper.

TABLE B1. List of symbols.

Symbol	Definition
β_n	Estimated regression coefficient
Γ	Lapse rate
$\overline{RH}$	Relative humidity averaged within 5° from TC
RH'	Max − avg relative humidity
ΔZ	Layer thickness
$\nabla_h \cdot \mathbf{V}$	Divergence of the horizontal wind
$\mathbf{V}$	Total horizontal wind
ζ	Relative vorticity
$\overline{\zeta}$	Relative vorticity averaged within 5° from TC
ζ'	Max − avg relative vorticity

REFERENCES

Beven, J., 1999: The boguscane: A serious problem with the NCEP medium range forecast model in the Tropics. Preprints, *23rd Conf. on Hurricanes and Tropical Meteorology*, Dallas, TX, Amer. Meteor. Soc., 845–848.

Bister, M., and K. A. Emanuel, 1997: The genesis of Hurricane Guillermo: TEXMEX analyses and a modeling study. *Mon. Wea. Rev.*, **125**, 2662–2682, doi:10.1175/1520-0493(1997)125<2662:TGOHGT>2.0.CO;2.

Cangialosi, J. P., and J. L. Franklin, 2014: 2013 National Hurricane Center forecast verification report. National Hurricane Center Tech. Rep., 84 pp. [Available online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2013.pdf.]

——, and ——, 2015: 2014 National Hurricane Center forecast verification report. National Hurricane Center Tech. Rep., 82 pp. [Available online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2014.pdf.]

——, and ——, 2016: 2015 National Hurricane Center forecast verification report. National Hurricane Center Tech. Rep., 69 pp. [Available online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2015.pdf.]

Charney, J. G., and A. Eliassen, 1964: On the growth of the hurricane depression. *J. Atmos. Sci.*, **21**, 68–75, doi:10.1175/1520-0469(1964)021<0068:OTGOTH>2.0.CO;2.

Cossuth, J. H., R. D. Knabb, D. P. Brown, and R. E. Hart, 2013: Tropical cyclone formation guidance using pregenesis Dvorak climatology. Part I: Operational forecasting and predictive potential. *Wea. Forecasting*, **28**, 100–118, doi:10.1175/WAF-D-12-00073.1.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998a: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.*, **126**, 1373–1395, doi:10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2.

——, J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998b: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part II: Results. *Mon. Wea. Rev.*, **126**, 1397–1418, doi:10.1175/1520-0493(1998)126<1397:TOCMGE>2.0.CO;2.

Cullen, M., 1993: The unified forecast/climate model. *Meteor. Mag.*, **122** (1449), 81–94.

DeMaria, M., J. A. Knaff, and B. H. Connell, 2001: A tropical cyclone genesis parameter for the tropical Atlantic. *Wea. Forecasting*, **16**, 219–233, doi:10.1175/1520-0434(2001)016<0219:ATCGPF>2.0.CO;2.

Dunion, J., J. Kaplan, A. Schumacher, and J. Cossuth, 2013: NOAA Joint Hurricane Testbed (JHT) end of year progress report, year 2. NOAA/NHC Tech. Rep., 4 pp. [Available online at http://www.nhc.noaa.gov/jht/11-13reports/Final_Dunion_JHT13.pdf.]

Dunkerton, T. J., M. Montgomery, and Z. Wang, 2009: Tropical cyclogenesis in a tropical wave critical layer: Easterly waves. *Atmos. Chem. Phys.*, **9**, 5587–5646, doi:10.5194/acp-9-5587-2009.

ECMWF, 2016: ECMWF IFS documentation. [Available online at http://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation.]

Elsberry, R. L., H. C. Tsai, and M. S. Jordan, 2014: Extended-range forecasts of Atlantic tropical cyclone events during 2012 using the ECMWF 32-day ensemble predictions. *Wea. Forecasting*, **29**, 271–288, doi:10.1175/WAF-D-13-00104.1.

Emanuel, K. A., 1986: An air–sea interaction theory for tropical cyclones. Part I: Steady-state maintenance. *J. Atmos.*

Sci., **43**, 585–605, doi:10.1175/1520-0469(1986)043<0585:AASITF>2.0.CO;2.

——, and D. S. Nolan, 2004: Tropical cyclone activity and the global climate system. Preprints, *26th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 10A.2. [Available online at https://ams.confex.com/ams/26HURR/techprogram/paper_75463.htm.]

Fu, B., M. S. Peng, T. Li, and D. E. Stevens, 2012: Developing versus nondeveloping disturbances for tropical cyclone formation. Part II: Western North Pacific. *Mon. Wea. Rev.*, **140**, 1067–1080, doi:10.1175/2011MWR3618.1.

Gall, R., and Coauthors, 2013: 2012 HFIP R&D activities summary: Recent results and operational implementation. NOAA/HFIP Tech. Rep., 40 pp. [Available online at http://www.hfip.org/documents/HFIP_2012_Annual_Report_Final.pdf.]

Gray, W. M., 1968: Global view of the origin of tropical disturbances and storms. *Mon. Wea. Rev.*, **96**, 669–700, doi:10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2.

——, 1979: Hurricanes: Their formation, structure and likely role in the tropical circulation. *Meteorology over the Tropical Oceans*, D. B. Shaw, Ed., Royal Meteorological Society, 155–218.

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, doi:10.1175/WAF-D-13-00008.1.

——, ——, ——, and ——, 2016: Verification of tropical cyclone genesis forecasts from global numerical models: Comparisons between the North Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **31**, 947–955, doi:10.1175/WAF-D-15-0157.1.

Helms, C. N., and R. E. Hart, 2015: The evolution of dropsonde-derived kinematic and thermodynamic structures in developing and nondeveloping Atlantic tropical convective systems. *Mon. Wea. Rev.*, **143**, 3109–3135, doi:10.1175/MWR-D-14-00242.1.

Heming, J. T., 2014: The impact on tropical cyclone predictions of a major upgrade to the Met Office global model. *Proc. 31st Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., 11A.3. [Available online at https://ams.confex.com/ams/31Hurr/webprogram/Paper243428.html.]

Hogan, T. F., and Coauthors, 2014: The Navy Global Environmental Model. *Oceanography*, **27** (3), 116–125, doi:10.5670/oceanog.2014.73.

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant, 2013: *Applied Logistic Regression*. 3rd ed. Wiley Series in Probability and Statistics, John Wiley and Sons, 528 pp.

Jarvinen, B., C. Neumann, and M. Davis, 1984: A tropical cyclone data tape for the North Atlantic basin, 1886–1983: Contents, limitations, and uses. NOAA Tech. Memo. NWS NHC 22, 21 pp. [Available online at http://www.nhc.noaa.gov/pdf/NWS-NHC-1988-22.pdf.]

Kanamitsu, M., 1989: Description of the NMC Global Data Assimilation and Forecast System. *Wea. Forecasting*, **4**, 335–342, doi:10.1175/1520-0434(1989)004<0335:DOTNGD>2.0.CO;2.

Komaromi, W. A., and S. J. Majumdar, 2015: Ensemble-based error and predictability metrics associated with tropical cyclogenesis. Part II: Wave-relative framework. *Mon. Wea. Rev.*, **143**, 1665–1686, doi:10.1175/MWR-D-14-00286.1.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, doi:10.1175/MWR-D-12-00254.1.

Lim, Y.-K., S. D. Schubert, O. Reale, M.-I. Lee, A. M. Molod, and M. J. Suarez, 2015: Sensitivity of tropical cyclones to param-

eterized convection in the NASA GEOS-5 model. *J. Climate*, **28**, 551–573, doi:10.1175/JCLI-D-14-00104.1.

Majumdar, S. J., and R. D. Torn, 2014: Probabilistic verification of global and mesoscale ensemble forecasts of tropical cyclogenesis. *Wea. Forecasting*, **29**, 1181–1198, doi:10.1175/WAF-D-14-00028.1.

Marchok, T. P., 2002: How the NCEP tropical cyclone tracker works. Preprints, *25th Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., P1.13. [Available online at https://ams.confex.com/ams/pdfpapers/37628.pdf.]

McAdie, C., C. Landsea, C. Neumann, J. David, E. Blake, and G. Hammer, 2009: *Tropical Cyclones of the North Atlantic Ocean, 1851–2006*. NOAA Historical Climatology Series, No. 6-2, NOAA/NCDC, 238 pp.

Nolan, D. S., 2007: What is the trigger for tropical cyclogenesis? *Aust. Meteor. Mag.*, **56**, 241–266.

Peng, M. S., B. Fu, T. Li, and D. E. Stevens, 2012: Developing versus nondeveloping disturbances for tropical cyclone formation. Part I: North Atlantic. *Mon. Wea. Rev.*, **140**, 1047–1066, doi:10.1175/2011MWR3617.1.

Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395–419, doi:10.1175/2008WAF2222128.1.

——, J.-G. Jiing, C. W. Landsea, S. T. Murillo, and J. L. Franklin, 2012: The Joint Hurricane Test Bed: Its first decade of tropical cyclone research-to-operations activities reviewed. *Bull. Amer. Meteor. Soc.*, **93**, 371–380, doi:10.1175/BAMS-D-11-00037.1.

Ritchie, E. A., and G. J. Holland, 1997: Scale interactions during the formation of Typhoon Irving. *Mon. Wea. Rev.*, **125**, 1377–1396, doi:10.1175/1520-0493(1997)125<1377:SIDTFO>2.0.CO;2.

Rosmond, T. E., 1992: The design and testing of the Navy Operational Global Atmospheric Prediction System. *Wea. Forecasting*, **7**, 262–272, doi:10.1175/1520-0434(1992)007<0262:TDATOT>2.0.CO;2.

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. *Comput. Stat. Data Anal.*, **50**, 3464–3485, doi:10.1016/j.csda.2005.07.015.

Schumacher, A. B., M. DeMaria, and J. A. Knaff, 2009: Objective estimation of the 24-h probability of tropical cyclone formation. *Wea. Forecasting*, **24**, 456–471, doi:10.1175/2008WAF2007109.1.

——, ——, ——, L. Ma, and H. Syed, 2014: Updates to the NESDIS Tropical Cyclone Formation Probability product. *Proc. 31st Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., 1C.6. [Available online at https://ams.confex.com/ams/31Hurr/webprogram/Paper244172.html.]

Simpson, J., E. Ritchie, G. Holland, J. Halverson, and S. Stewart, 1997: Mesoscale interactions in tropical cyclone genesis. *Mon. Wea. Rev.*, **125**, 2643–2661, doi:10.1175/1520-0493(1997)125<2643:MIITCG>2.0.CO;2.

Torn, R. D., and C. Snyder, 2012: Uncertainty of tropical cyclone best-track information. *Wea. Forecasting*, **27**, 715–729, doi:10.1175/WAF-D-11-00085.1.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.

Zhang, W., B. Fu, M. S. Peng, and T. Li, 2015: Discriminating developing versus nondeveloping tropical disturbances in the western North Pacific through decision tree analysis. *Wea. Forecasting*, **30**, 446–454, doi:10.1175/WAF-D-14-00023.1.