

A Comparison of Tropical Cyclone Genesis Forecast Verification from Three Global Forecast System (GFS) Operational Configurations

DANIEL J. HALPERIN

*Department of Applied Aviation Sciences, Embry-Riddle Aeronautical University,
Daytona Beach, Florida*

ANDREW B. PENNY

*National Hurricane Center, Miami, Florida, and CPAESS, University Corporation for
Atmospheric Research, Boulder, Colorado*

ROBERT E. HART

Florida State University, Tallahassee, Florida

(Manuscript received 13 March 2020, in final form 18 June 2020)

ABSTRACT

Operational forecasting of tropical cyclone (TC) genesis has improved in recent years but still can be a challenge. Output from global numerical models continues to serve as a primary source of forecast guidance. Bulk verification statistics (e.g., critical success index) of TC genesis forecasts indicate that, overall, global models are increasingly able to predict TC genesis. However, as global model configurations are updated, TC genesis verification statistics will change. This study compares operational and retrospective forecasts from three configurations of NCEP's Global Forecast System (GFS) to quantify the impact of model upgrades on TC genesis forecasts. First, bulk verification statistics from a homogeneous sample of model initialization cycles during the period 2013–14 are compared. Then, composites of select output fields are analyzed in an attempt to identify any key differences between hit and false alarm events. Bulk statistics indicate that TC genesis forecast performance decreased with the implementation of the 2015 version of the GFS, but then modestly recovered with the 2016 version of the model. In addition, the composite analysis suggests that false alarm forecasts in the 2015 version of the GFS may have been the result of inaccurately forecasting the location and/or strength of upper-level troughs poleward of the TC. There is also evidence of convective feedbacks occurring, such as ridging above the low-level circulation and upper-level convective outflow that were too strong, in this same set of false alarm forecasts. Overall, analyzing retrospective forecasts can assist forecasters in determining the strengths and weaknesses associated with a new configuration of a global model with respect to TC genesis.

1. Introduction

Accurately forecasting tropical cyclone (TC) genesis can be a challenge (e.g., [Blake 2019](#); [Cangialosi and Ramos 2019](#); [Avila et al. 2020](#)), especially when global models provide insufficient guidance that a TC will soon develop. Several probabilistic TC genesis forecast products have been developed recently to provide guidance to forecasters (e.g., [Schumacher et al. 2009](#); [Cossuth et al. 2013](#); [Dunion 2017](#); [Halperin et al. 2017](#); [Yamaguchi and Koide 2017](#); [Tsai and](#)

[Elsberry 2019](#)), many of which rely at least in part on output from global numerical models. Indeed, the global model output itself is a primary source of forecast guidance. Studies have documented to what extent global models are able to predict TC genesis (e.g., [Briegel and Frank 1997](#); [Beven 1999](#); [Chan and Kwok 1999](#); [Cheung and Elsberry 2002](#); [Pratt and Evans 2009](#); [Tsai et al. 2011](#); [Halperin et al. 2013, 2016](#)), with the general consensus that the guidance overall has improved in recent years. For example, while [Beven \(1999\)](#) documented the deficiencies with TC genesis forecasts from deterministic global model output on a time scale of a few days, more recent studies (e.g., [Elsberry et al. 2014](#); [Komaromi and Majumdar 2015](#);

Corresponding author: Daniel J. Halperin, daniel.halperin@erau.edu

[Lee et al. 2018](#); [Wang et al. 2018](#)) have shown that TC genesis may be predictable on a scale of a week or more using output from global model ensembles.

[Halperin et al. \(2013, 2016](#), hereafter H13 and H16, respectively) showed that the bulk verification statistics of TC genesis forecasts from global models exhibit some interannual variability. Some of this change in performance also coincides with upgrades to the global model itself. However, these studies do not quantify how much of the improvement or degradation is due to the change in model configuration or simply due to a given year containing TC genesis events that occur from more or less predictable genesis pathways ([McTaggart-Cowan et al. 2013](#)). For example, [Wang et al. \(2018\)](#) suggest that North Atlantic TCs that develop via tropical transition pathways exhibit less predictability with respect to genesis than those that develop from the nonbaroclinic pathway. Therefore, it is possible that a year with more TCs that develop via tropical transition may yield worse TC genesis forecast verification statistics, even if updates to the model configuration had a positive overall impact on other TC and non-TC related metrics. Recent studies have shown the sensitivity of TC formation forecasts to changes in environmental and storm-scale parameters (e.g., [Fritz and Wang 2013](#); [Penny et al. 2016a,b](#)). It is difficult to attribute a change in model performance to a specific change in the global model configuration because numerous changes occur at once. However, it is desirable to look beyond the bulk statistics and analyze composites of the forecast TCs to try to determine how model configuration changes may be impacting TC structure.

When a new version of NCEP's Global Forecast System (GFS) model is being considered for operational implementation, retrospective forecasts using the proposed version of the GFS were created for comparison with the current operational configuration of the model. These retrospective forecast datasets typically span at least three TC seasons and provide a unique opportunity to quantify how changes in model configuration impact forecast performance.

Using retrospective forecasts of three different configurations of the GFS initialized over the period 2013–14, the goals of this study are 1) to quantify the impact of model upgrades on bulk TC genesis forecast verification statistics and 2) to try to identify potential causes of improved or degraded forecast performance using composite analysis. This type of analysis can be considered by the National Hurricane Center (NHC) as a part of their overall assessment of a proposed model upgrade.

2. Methodology

Three configurations of the GFS were compared: 1) the 2013–14 operational configuration (hereafter “GFS 1314”), 2) retrospective forecasts using the 2015 GFS configuration (i.e., v12.0.0; hereafter “GFS 2015”), and 3) retrospective forecasts using the 2016 GFS configuration (i.e., v13.0.2; hereafter “GFS 2016”). Select model changes are provided in the [appendix](#). For the bulk verification statistics calculations, only forecasts from the years 2013–14 were considered because these forecasts were available for all three GFS configurations. Furthermore, the subset of 1279 initialization cycles where data were available for all three configurations was used to ensure a homogeneous comparison. All TC genesis forecasts were identified using the TC tracking algorithm described in H13 and H16. In addition, all model data were output to a 0.5° latitude–longitude grid.

a. TC genesis verification

The verification criteria employed here were identical to those defined in H16: “a successful genesis forecast (i.e., “hit”) is defined when best track ([Jarvinen et al. 1984](#); [McAdie et al. 2009](#); [Landsea and Franklin 2013](#)) TC genesis occurred within 120 h of the model initialization time [(i.e., the start time of the model)] and when the model forecast genesis location was within 5° latitude and longitude of the best track location at the corresponding time. [...] For model genesis forecasts with valid times prior to the best track genesis time, combined automated response to query (CARQ) entries in the Automated Tropical Cyclone Forecasting (ATCF) system a-deck files ([Sampson and Schrader 2000](#)) were used to verify the forecast TC location. [...] A genesis forecast that did not result in best track genesis was classified as a false alarm (FA).”

A miss was defined as the case where output exists from a model cycle with an initialization time within 120 h of best track genesis, but TC genesis was not forecast. For best track TCs with no data gaps in the 120 h preceding genesis, there are 20 model cycles where TC genesis may be forecast (i.e., four model runs per day for five days). However, since some data gaps exist in the homogeneous model initialization cycle subset, some best track TCs may have a maximum number of miss events less than 20.

b. Composites

To facilitate a comparison of the average storm environment during genesis forecasts, composites were constructed for each version of the GFS (retrospective) forecasts (GFS 1314, GFS 2015, and GFS 2016).

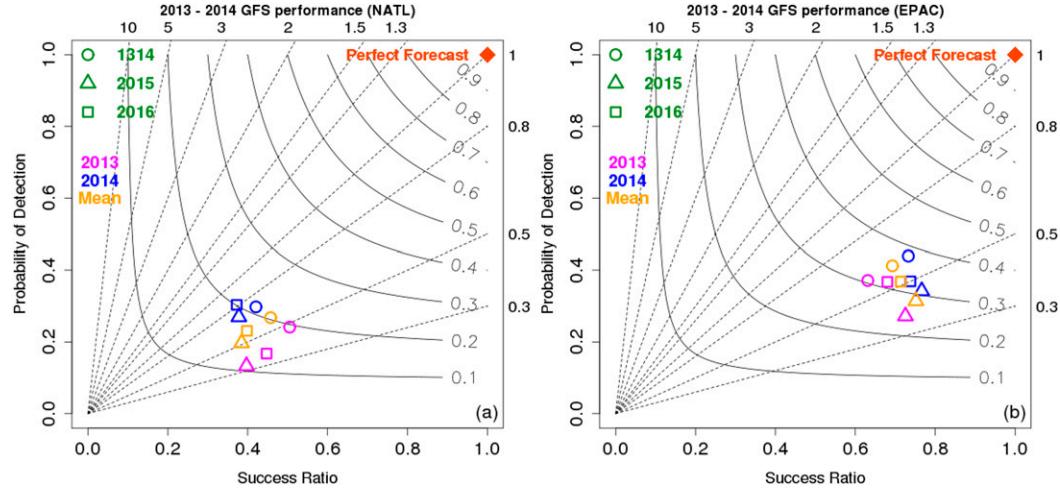


FIG. 1. Performance diagrams for the (a) NATL and (b) EPAC TC genesis forecasts during 2013–14. Success ratio is given on the x axis and probability of detection is given on the y axis. Frequency bias values are indicated by the dashed lines, and the critical success index values are indicated by the 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 GFS 1314, GFS 2015, and GFS 2016 configurations are indicated by the circles, triangles, and squares, respectively.

Separate sets of composites were created for hits and false alarms.

Before composites were created, the TC center locations of each genesis event (hit or false alarm) were visually inspected using 10-m wind speed and sea level pressure (SLP; GRIB variable name “Pressure Reduced to Mean Sea Level” or “PRMSL”). An attempt was then made to identify center locations for the 12 h period prior to the genesis event at 6 h intervals. Beginning at the time of the genesis event and

moving backward in time, center locations were identified based on the minimum SLP within 5° of the previously identified center location. Center locations were visually inspected once more; cases were excluded if the center locations appeared questionable based on the meteorological features (e.g., no closed circulation) or if the center locations were not consistent in time leading up to the genesis event. Not surprisingly the number of “good” center locations decreased when moving backward in time and farther away from the

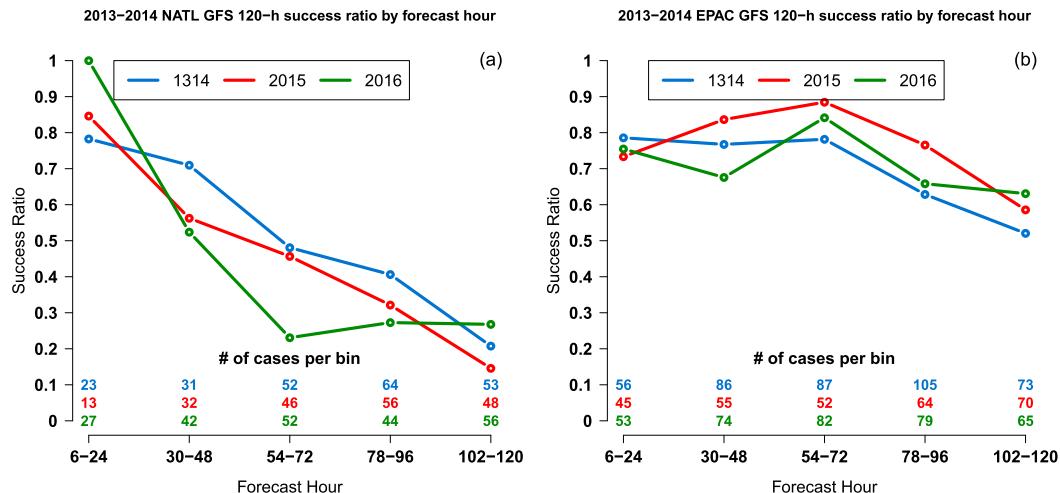


FIG. 2. A comparison of SR as a function of forecast hour among the model configurations for (a) NATL and (b) EPAC forecasts. The GFS 1314, GFS 2015, and GFS 2016 model configuration is in blue, red, and green, respectively.

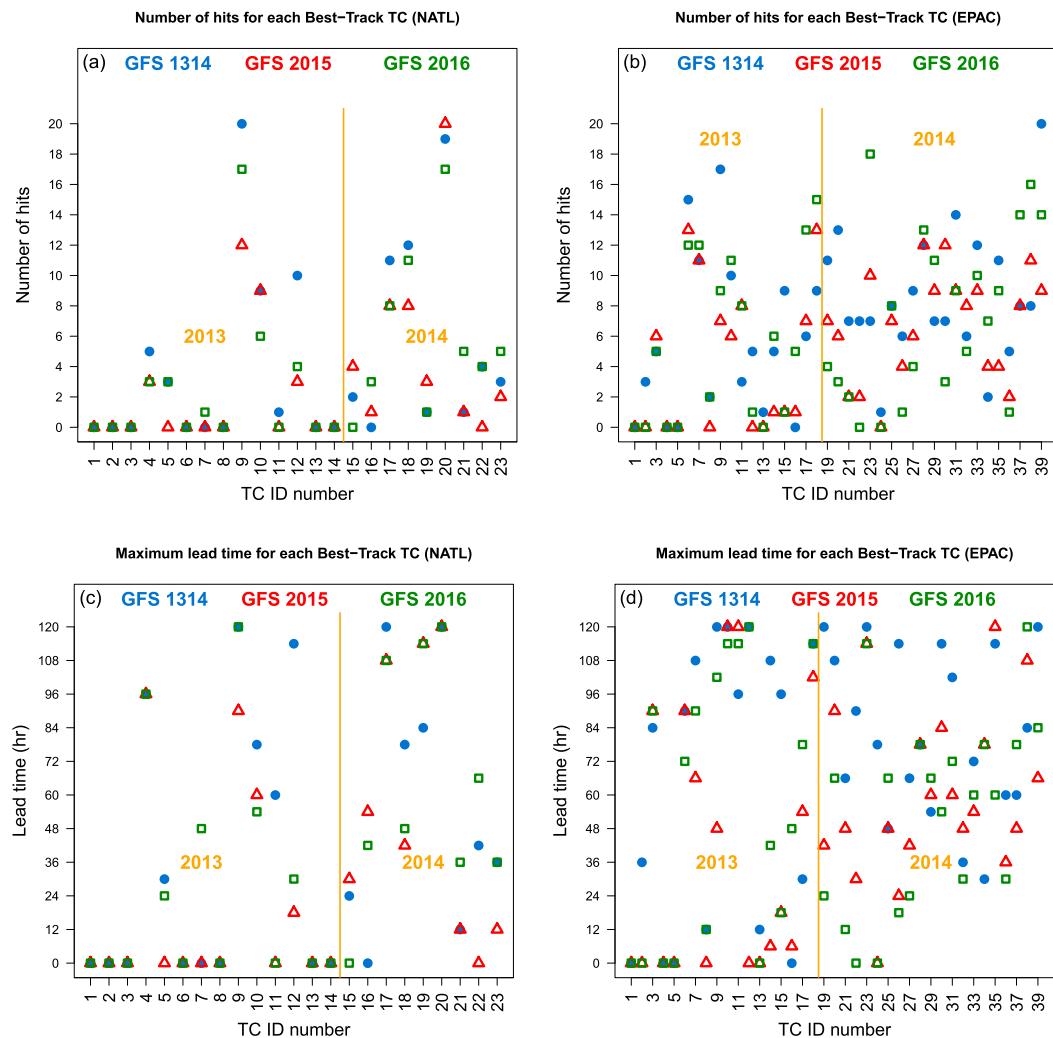


FIG. 3. A comparison of the number of hits for each best track TC among the model configurations for (a) NATL and (b) EPAC forecasts, and the maximum lead time for each best track TC among the model configurations for (c) NATL and (d) EPAC forecasts. The GFS 1314, GFS 2015, and GFS 2016 model configuration is in blue, red, and green, respectively.

genesis event. After constructing a list of genesis event center locations for each version of the GFS (retrospective) forecasts, a similar method was used to compile a list of center locations from the verifying analyses of each version of the GFS.¹

Based on the “good” center positions of the (retrospective) forecasts and verifying analyses, meteorological data were then interpolated to storm-centered

0.25° horizontal resolution grids with data extending $\pm 15^\circ$ from the center position. Average storm-centered composites were then computed for each of the GFS versions at 6-h intervals leading up to the genesis event time. Since some center locations could not be tracked at 6 and 12 h before the forecast genesis time and were excluded from the sample, the sample size for each composite event type often decreased at 6 and 12 h prior to the forecast genesis time.

In contrast to the homogeneous comparison used to compute the bulk verification statistics shown in section 3a, composites were constructed using all of the available cases for each version of the GFS. This was done to maximize the sample size for the composites at times prior to the genesis event. Due to the

¹For example, the center-finding algorithm would occasionally identify the center to be between two low pressure areas if the system of interest in the model forecasts was weakening while another low pressure area just outside of the search radius was deepening. These center locations were excluded from the sample.

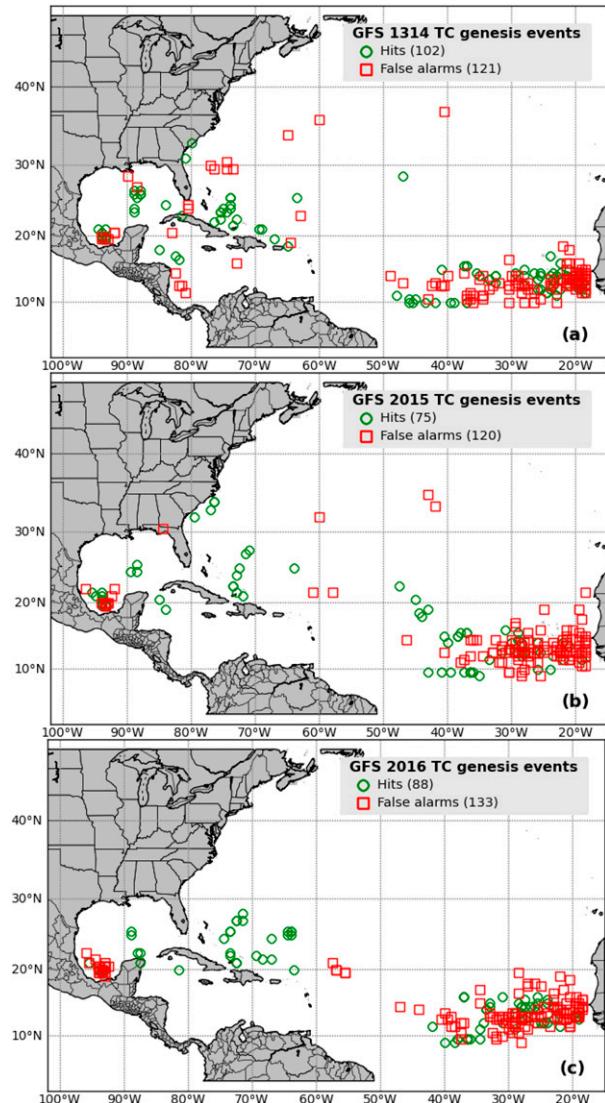


FIG. 4. A geographic plot of GFS-indicated TC genesis events over the NATL for (a) 2013–14, (b) 2015, and (c) 2016 model configurations.

limited sample size, separate composites were not constructed based on forecast hour or location within a basin.

3. Results

a. TC genesis verification

Performance diagrams (Roebber 2009) are used to compare the success ratio (SR), probability of detection (POD), frequency bias, and critical success index (CSI) among the model configurations (Fig. 1). The CSI is greatest over the North Atlantic basin (NATL) for GFS 1314 (Fig. 1a). GFS 2015 exhibits the smallest

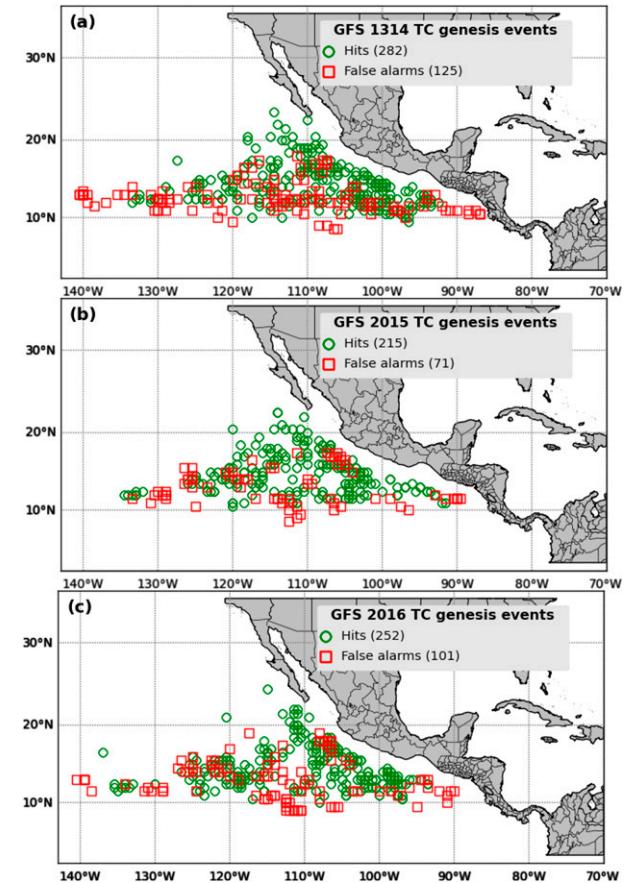


FIG. 5. As in Fig. 4, but for the EPAC basin.

SR, POD, and CSI relative to GFS 1314 and GFS 2016. In general, the results suggest that TC genesis forecasts degraded from GFS 1314 to GFS 2015, then improved from GFS 2015 to GFS 2016. However, the GFS 2016 performance values were still worse than the GFS 1314 values. Results over the eastern North Pacific basin (EPAC) were less consistent than over the NATL. Over the EPAC, GFS 2015 was less cyclogenetic compared to GFS 1314, which resulted in GFS 2015 exhibiting a greater SR, but a smaller POD compared to GFS 1314. The GFS 2016 statistics generally fall in between the GFS 1314 and GFS 2015 values. GFS 1314 exhibits the largest CSI for 2014 and 2013–14 mean forecasts. Meanwhile, GFS 2016 exhibits the largest CSI during 2013.

The SR decreases with increasing forecast hour over the NATL (Fig. 2a). GFS 2015 has smaller SR values than GFS 1314 after 24 h, which is consistent with the degraded performance noted in Fig. 1a. GFS 2016 exhibits the largest relative SR for 6–24- and 102–120-h forecasts, but the smallest relative SR for 30–96-h forecasts. The SR degradation is far less

2015

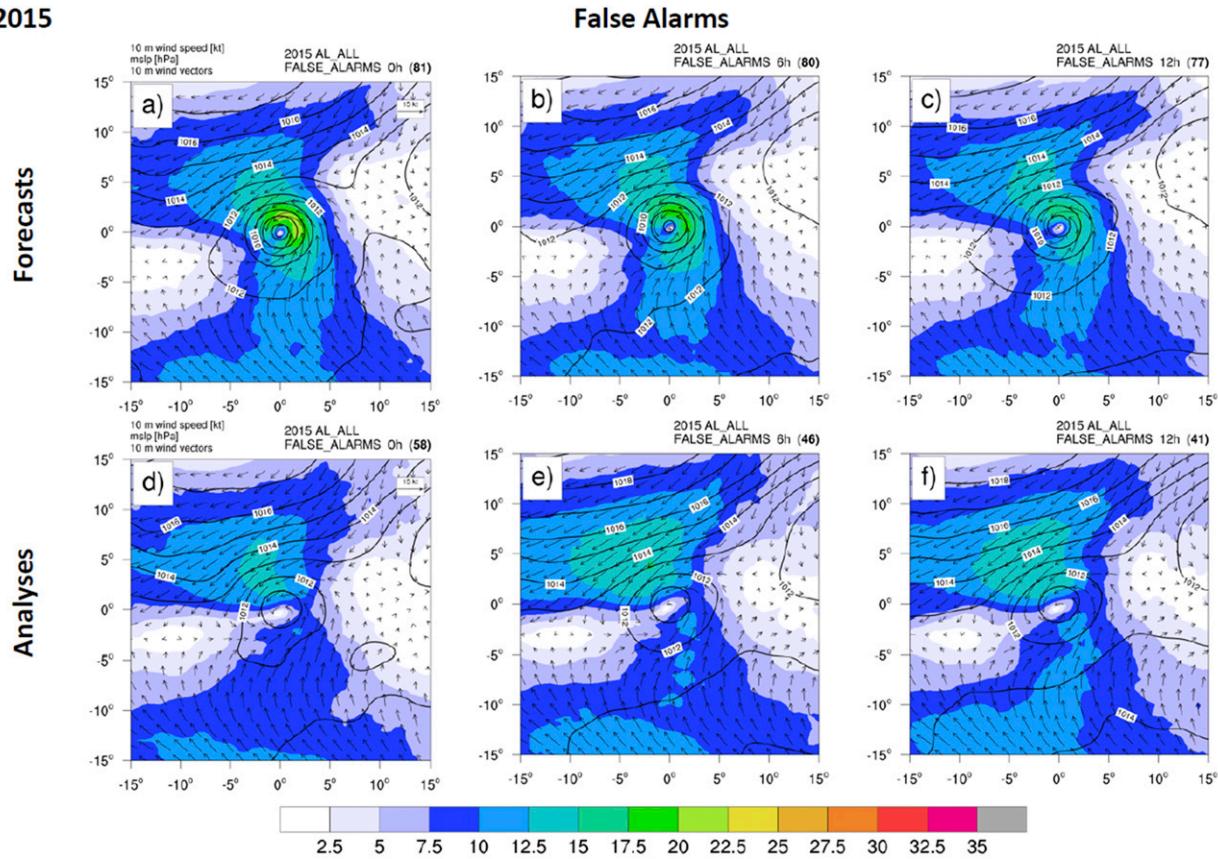


FIG. 6. “False alarm” NATL composites of 10-m wind speed (kt; $1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$; shading and vectors) and mean SLP (hPa; contours) from the GFS 2015 configuration for (top) forecasts and (bottom) analyses. Composites are shown at the (a),(d) genesis event time as well as (b),(e) 6 h prior and (c),(f) 12 h prior. The number of samples included in each composite is indicated in boldface at the top right of each panel.

pronounced over the EPAC (Fig. 2b). The less-cyclogenetic GFS 2015 exhibits the largest relative SR for 30–96-h forecasts.

The relatively small POD values are the result of best track TCs that the models completely fail to forecast and the fact that the models are typically unable to capture genesis in every initialization cycle five days before genesis occurs (H16). Plotting the number of hits per best track TC provides more insight into the low POD issue (Figs. 3a,b). The differences in the mean number of hits per best track TC among the configurations is not statistically significant for either basin. However, there is notable variability among the model configurations in the number of hits for a given TC. This also leads to a variability in the maximum lead time for each best track TC (Figs. 3c,d) (e.g., Chen et al. 2019). The difference in median of the maximum lead time among the model configurations is not significantly different over the NATL (GFS 1314: 30 h; GFS 2015: 12 h; GFS 2016: 36 h). However, the median of the maximum lead time over the EPAC for GFS

1314 (84 h) is significantly longer than for GFS 2015 (48 h) and GFS 2016 (60 h) with 95% confidence, according to a Wilcoxon rank-sum test (e.g., Wilks 2011).

In terms of geographical differences between the three model configurations (Figs. 4 and 5), while the majority of hits and false alarms over the NATL (Fig. 4) occur in the main development region (MDR; 5° – 20° N, 50° W– 0°), it is noteworthy that GFS 2015 and GFS 2016 appear to have fewer false alarms in the western Atlantic and Caribbean compared to the GFS 1314. When comparing the TC genesis forecasts of the GFS model with other global models for the period of 2004–11, H13 found a large number of GFS-generated false alarms in the western Atlantic and Caribbean basin.

A comparison of hits and false alarms in the EPAC (Fig. 5) basin reveals that apart from GFS 1314 having a larger number of hits and false alarms, there does not appear to be noticeable differences in the geographic distributions of hits and false alarms among the different versions of the GFS.

2015

Forecasts

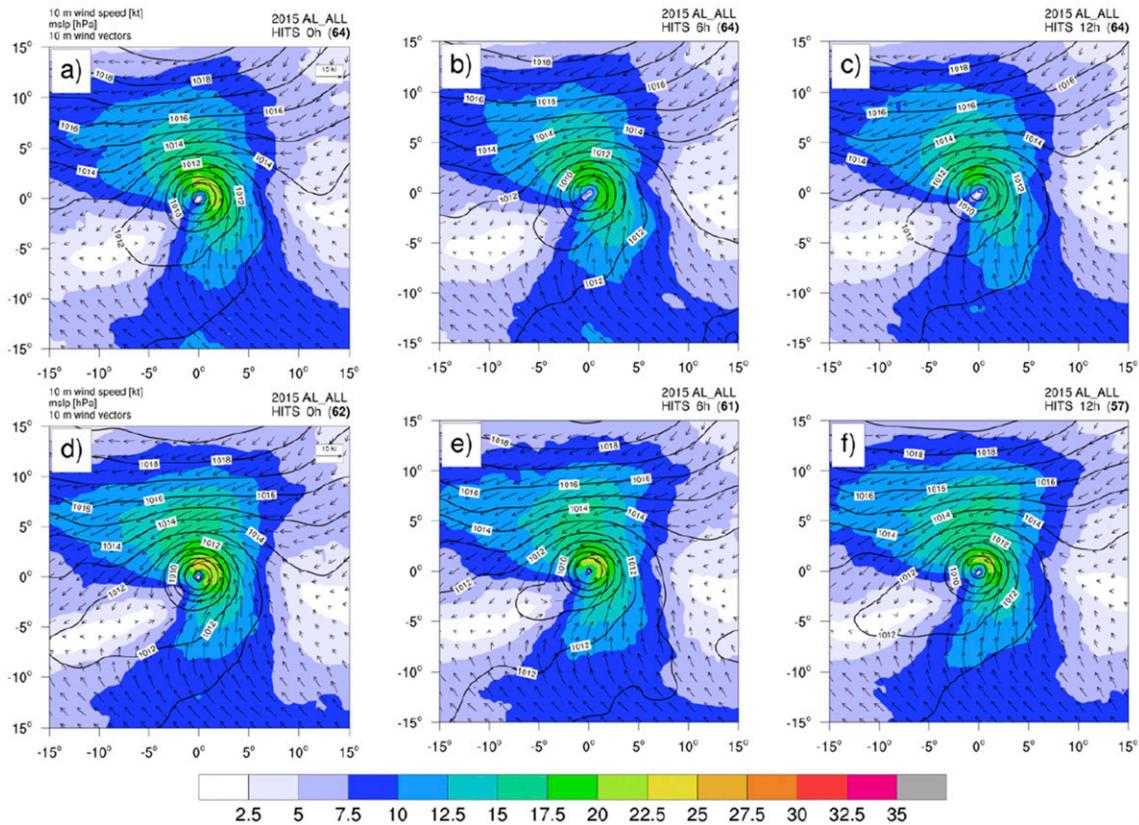


FIG. 7. As in Fig. 6, but for NATL cases identified as hits.

b. Composites

False alarm composites of 10-m wind speed for the GFS 2015 (Fig. 6), which had the lowest bulk verification scores, reveal considerable differences between the forecasts and analyses at each 6-h interval leading up to the genesis event time. Not surprisingly, the 10-m wind speeds are much larger near the center of the storm in the forecasts (Figs. 6a–c) compared to the analyses (Figs. 6d–f) and the minimum SLP is lower, indicating overdevelopment in the false alarm cases. The strongest winds are along the eastern semicircle of the circulation. While forecast composites look progressively more organized leading up to the time of the genesis event, the wind field in the analyses tends to look less organized at the genesis event time compared to 12 h prior. Outside of the inner core, one of the largest differences between the analyses and forecasts is that the southerly winds to the south of the circulation are much stronger in the forecasts, especially near the time of the genesis event. To the north of the circulation, the opposite is generally true: northeasterly winds are stronger in the analyses compared to the

forecasts, and extend over a broader area. The enhanced northeasterlies in the analyses might be due to a stronger ridge of high pressure to the north of the disturbance compared to the forecasts.

For the genesis events identified as “hits” in the GFS 2015, the strongest winds in both the forecasts (Figs. 7a–c) and analyses (Figs. 7d–f) exist along the eastern and northern semicircle of the circulation, while the weakest winds are southwest of the center. Although the forecasts and analyses for the hits look remarkably similar in terms of the spatial structure of the low-level wind field, a careful inspection reveals that the circulation in the analyses is stronger and more compact than in the forecasts at each of the composite times. This is indicative of an underforecast bias for hits.

Although only composites from the 2015 GFS are shown, a similar pattern is also present when examining the GFS 1314 and GFS 2016 cases, but to a slightly lesser extent, especially for the GFS 1314 cases (i.e., the underforecast bias for hits and overforecast bias for forecasts was not as pronounced for the GFS 1314 cases).

2015

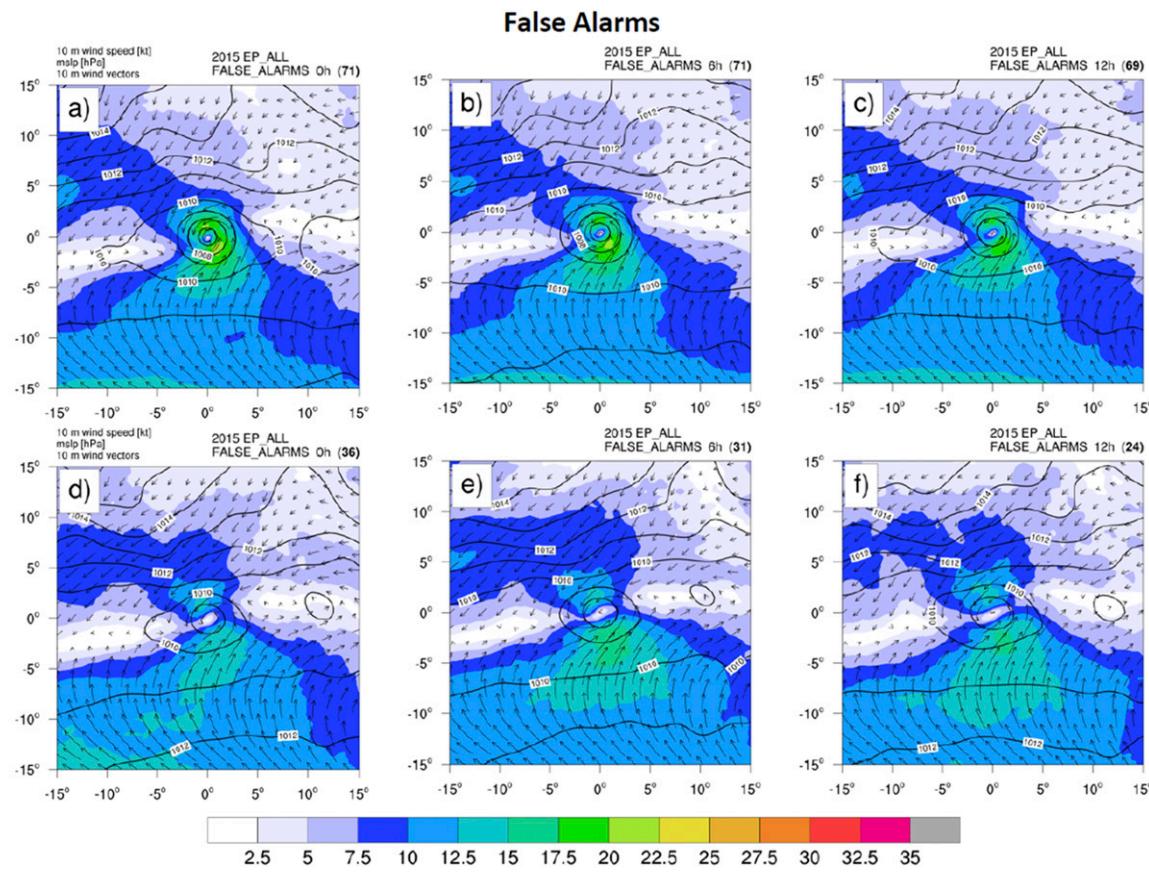
Forecasts
Analyses

FIG. 8. As in Fig. 6, but for EPAC cases identified as false alarms.

While the general patterns of forecast overdevelopment for false alarms and underdevelopment for hits are evident in the eastern North Pacific composites (Figs. 8 and 9) there were several notable differences compared to the Atlantic composites. The southwesterly flow south of the circulation center is stronger in the analyses for both false alarms and hits in the EPAC cases compared to NATL cases. The enhanced southwesterly flow is consistent with the monsoon trough environment, which is an important pathway for genesis in the EPAC (e.g., McTaggart-Cowan et al. 2013).

Although the NATL composites are predominantly made up of genesis events from within the MDR where easterly waves are the primary genesis mechanism, there are a number of cases included from higher latitudes. A separate composite of only NATL MDR cases reveals stronger southerly flow to the south of the circulation center compared to the composite that includes all NATL cases (Figs. 10 and 11). In fact, the strength of this southerly flow in the NATL MDR cases is more similar to the EPAC composites. However, the northeasterly flow to the north of the circulation is still much stronger in the NATL MDR composites compared to the EPAC

composites, possibly due to a stronger ridge of high pressure to the north of the MDR cases and/or the inverted trough structure of the easterly waves. The size of the disturbance is smaller in the EPAC composites, which is consistent with observations that indicate that TC size is smallest in the EPAC (Chavas and Emanuel 2010; Chan and Chan 2015).

To assess the upper-level structural differences between the forecasts and analyses for false alarm cases, 200-hPa wind speed and geopotential height were also composited (Fig. 12). An inspection of false alarm composites from the GFS 1314 indicates that the area encompassed by the 12 440-m geopotential height contour is much larger in the forecasts (Figs. 12a–c) compared to the analyses (Figs. 12d–f), which indicates that a stronger upper-level ridge was present in the forecasts for false alarm cases. In fact, the upper-level ridge in the analyses appears to weaken somewhat leading up to the genesis event time. However, despite the stronger ridge in the forecast composites, wind speeds to the north of the ridge are noticeably weaker in the forecasts. This likely indicates that there were lower geopotential heights to the north of the domain in the

2015

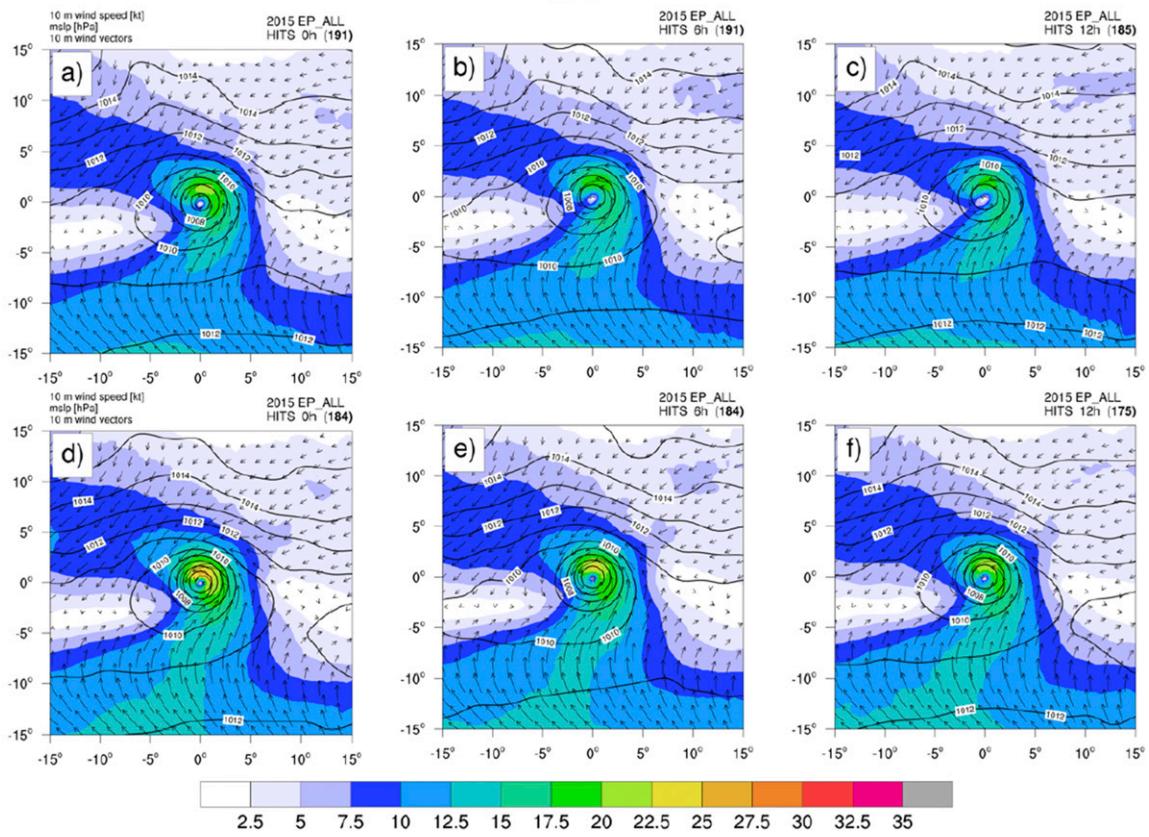
Forecasts
Analyses

FIG. 9. As in Fig. 6, but for EPAC cases identified as hits.

analyses, which would explain the tighter pressure gradient and stronger westerly winds in this area.

The impact of the stronger upper-level winds north of the circulation in the analyses of the GFS 1314 false alarm cases (Figs. 12d–f) can be seen when comparing composites of deep-layer (200–850 hPa) vertical wind shear (Fig. 13), as the vertical wind shear to the north of the circulation center is much stronger in the analyses (Figs. 13d–f) compared to the forecasts (Figs. 13a–c). Interestingly, the vertical wind shear is noticeably weaker in the analyses north of the circulation center at the genesis event time compared to 6 and 12 h prior. Based on a comparison of the low-level winds (similar to that shown for GFS 2015 in Figs. 6d–f) and the decreasing amplitude of the upper-level ridge (Figs. 9d–f), the decrease in shear may be due in part to the weakening of the TC circulation at the genesis event time in the analyses. In contrast, the vertical wind shear in the forecast composites (Figs. 13a–c) increases slightly to the north and south of the circulation center leading up to the genesis event time. The increase in vertical wind shear to the south and southeast of the circulation center (Figs. 12a–c) is due

to enhanced upper-level northeasterly winds that appear to be the result of increasing convective organization and upper-level outflow to the south of the circulation center in the forecast composites. Although not shown, these features are also evident in the GFS 2015 and GFS 2016 composites.

The vertical wind shear underforecast bias for false alarm cases is likely due to the poor representation of large-scale environmental features that impart vertical shear over the incipient disturbances (e.g., the location and intensity of tropical upper-tropospheric troughs is poorly forecast). However, based on the enhanced upper-level outflow noted in the forecasts (Figs. 12a–c), it also appears likely that convective feedbacks are too strong and are acting to partially mitigate the negative impacts of vertical wind shear (Corbosiero and Molinari 2002; Rappin et al. 2011; Penny et al. 2016b; Ryglicki et al. 2018).

4. Summary and conclusions

Assessing the quality of the TC genesis forecasts has become part of the scientific evaluation process for the

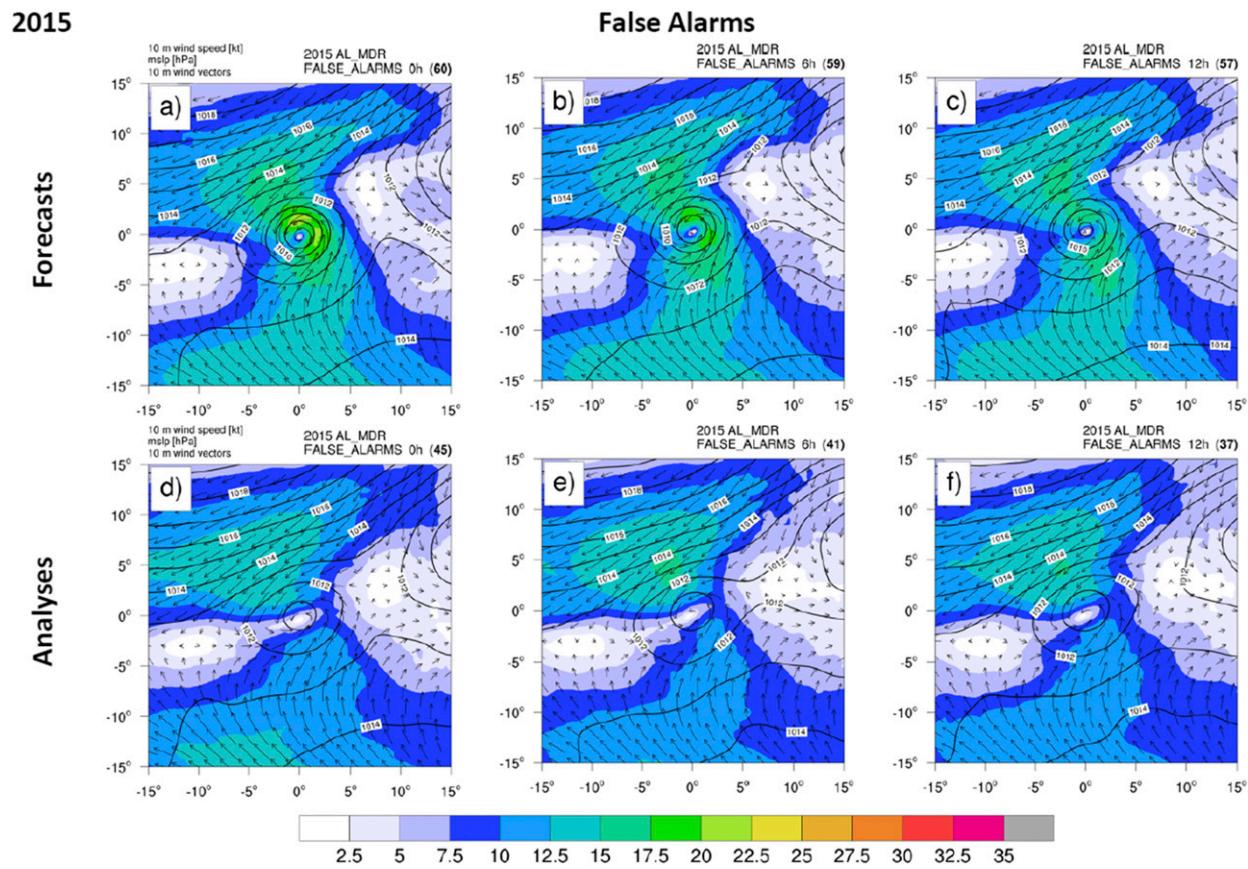


FIG. 10. As in Fig. 6, but for NATL MDR cases identified as false alarms.

last several implementations of the GFS global model. To evaluate whether the quality of these forecasts has changed over the last several GFS model upgrades, objective and subjective metrics are used to compare the quality of the genesis forecasts between three different versions of the GFS global model: 1) the 2013–14 operational configuration (GFS 1314), 2) the 2015 configuration (GFS 2015), and 3) the 2016 configuration (GFS 2016).

Bulk statistics from a homogeneous comparison of NATL forecasts indicate that GFS 1314 (GFS 2015) performed the best (worst) in terms of CSI. Results from the EPAC were more varied. While GFS 2015 generally had the highest success ratio for EPAC cases at most forecast lead times, it was less cyclogenetic compared to the other two versions (reduction in both the FAR and POD). However, in terms of the median of the maximum lead time of TC genesis, which is perhaps one of the most important metrics for operational forecasting, GFS 2015 was the worst performing.

NATL composites of 10-m wind speed and minimum SLP from GFS 2015 indicate an overdevelopment bias for false alarms (which is expected since, by definition,

no TC formed in the best track), and an underdevelopment bias for hits. A similar pattern was observed for GFS 1314 and GFS 2016, but to a lesser extent. In addition to the over and underdevelopment biases, differences in the surrounding low-level environment were apparent. Composites of the verifying analyses for false alarm cases exhibit stronger winds to the northwest of the circulation compared to forecast composites, which indicates that the subtropical ridge to the north and northwest of the circulation is also being under forecast.

Although the upper-level ridge centered above the low-level circulation was better-developed in forecast composites of false alarm cases, the geopotential height was much lower in the analyses north of the circulation, which explains the much stronger upper-level winds in this region. This pattern suggests that the strength and/or location of upper-level troughs north of the low-level circulation were not being well forecast. The inability to accurately forecast these upper-level features resulted in an underestimate of the vertical wind shear. A more realistic forecast of vertical wind shear would have resulted in an environment much less favorable for development, especially for systems that were not well

2015

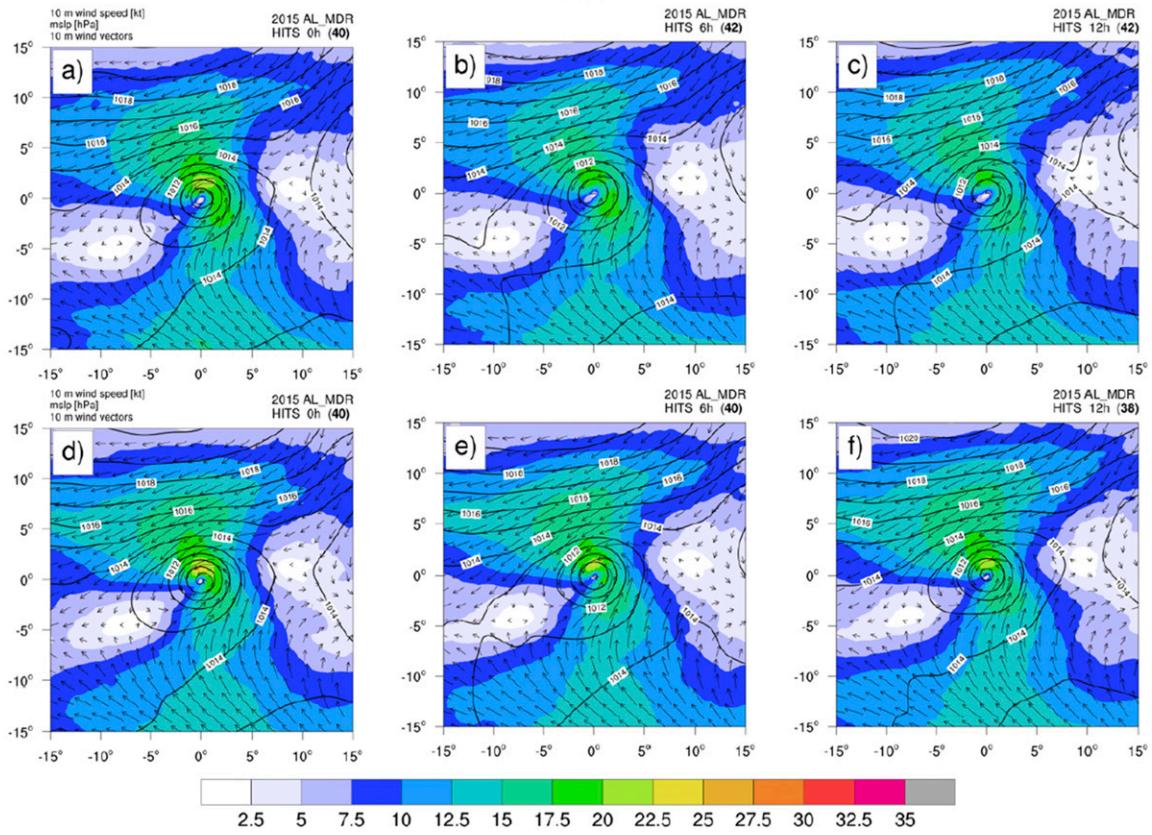
Forecasts
Analyses

FIG. 11. As in Fig. 6, but for NATL MDR cases identified as hits.

organized and may have been near a threshold for development.

Forecast composites of false alarms also exhibited signs that the convective feedbacks were too strong, especially for GFS 2015 and GFS 2016. In addition to the better developed upper-level ridge above the low-level circulation, the upper-level outflow channels north and south of the circulation were much stronger in the false alarm forecast composites.

The evidence from composites suggests that convection was stronger for the GFS 2015 and GFS 2016 configurations compared to GFS 1314, although this hypothesis is speculative. It appears that strong convective feedbacks led to over development of the disturbances, which was most pronounced for GFS 2015. In addition, the unrealistically strong convection may have indirectly led to development by protecting disturbances from the negative effects of vertical wind shear (Corbosiero and Molinari 2002; Rappin et al. 2011; Penny et al. 2016b; Ryglicki et al. 2018). This led to an increase in the number of false alarms relative to the number of hits (reduced success ratio) for NATL forecasts. However, one would expect

that if the convective feedbacks in GFS 2015 were always too strong, both false alarms and hits would be over forecast. Being that this was not the case points to a more complicated picture, and indicates that other sources of model error also contributed to the differences between the forecasts and analyses.

The degradation in the quality of the GFS 2015 TC genesis forecasts is likely related in part to the changes made to this version of the GFS. The horizontal resolution was increased from T574 (~27 km) to T1534 (~13 km), and a semi-Lagrangian advection scheme was introduced. In addition, the ice and water cloud conversion rates were adjusted and changes were made to the drag coefficient at high wind speeds (McClung 2014). Previous studies have shown that TC forecasts can be affected by the horizontal resolution of the model (Fierro et al. 2009; Davis et al. 2010; Gopalakrishnan et al. 2011). Given that the changes made to GFS 2016 were primarily related to the data assimilation component of the forecast system (the 3D hybrid ensemble–variational technique was updated to 4D hybrid ensemble–variational; McClung 2016), it is not surprising that the bulk verification statistics and composites were more similar to those of GFS 2015 than to GFS 1314.

1214

Forecasts

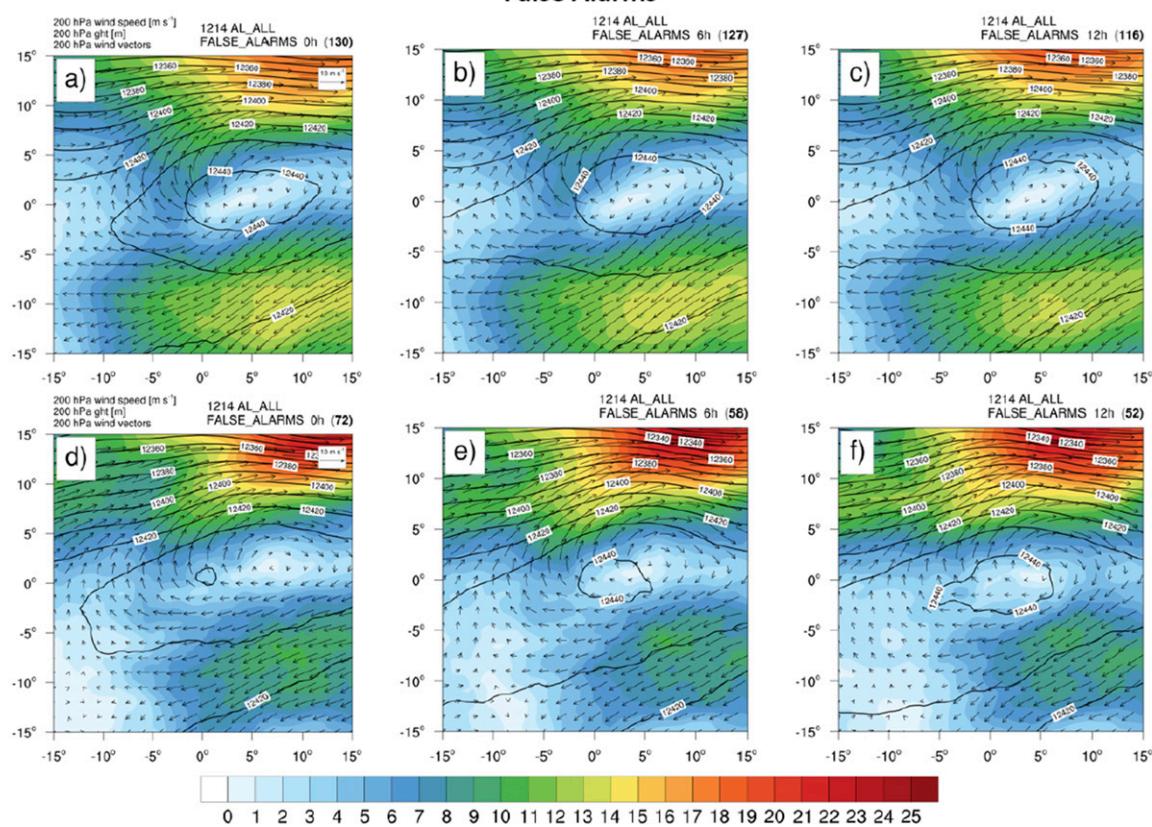


FIG. 12. As in Fig. 6, but for NATL false alarm composites from GFS 1314 of 200-hPa wind speed (m s^{-1} ; shading and vectors) and 200-hPa geopotential height (m; contours).

A notable aspect of the bulk verification results (Figs. 1–3) is the large difference between the characteristics of the NATL and EPAC forecasts. While the SR declines dramatically with forecast hour for NATL forecasts for all three GFS configurations, the SR remains fairly steady with forecast hour for EPAC cases. This indicates that, on average, there is greater predictability for TC genesis in the EPAC than in the NATL, at least for the GFS model, potentially due to the increased percentage of TC genesis events from tropical transition pathways in the NATL (Wang et al. 2018). The greater predictability translates into an almost doubling of the median of the maximum lead time for EPAC forecasts relative to NATL forecasts. Despite the increased predictability, it is interesting that, similar to the NATL, there was also evidence that the convective feedbacks were too strong for EPAC false alarm cases (not shown), which was most pronounced for GFS 2015. This, combined with evidence that GFS 2015 was less cyclogenetic than the other two versions of the GFS suggest that the EPAC TC genesis forecasts may not be as sensitive to the strength of convection, and that other factors may be more important in terms of affecting the predictability of the EPAC TC genesis forecasts.

While the degradation in the quality of the TC genesis forecasts from GFS 1314 to GFS 2015/GFS 2016 may seem small in terms of the SR or the CSI, it translates into a significant decrease in the maximum lead time for EPAC forecasts. Although the differences in maximum lead time for NATL forecasts were not considered statistically significant among the three versions of the GFS, GFS 2015 had the shortest maximum lead time. Ample lead time is especially important in situations where development occurs close to land, otherwise there may not be adequate time to effectively warn the public and to allow emergency management personnel to take necessary actions to keep the public safe.

NHC relies heavily on the GFS global model forecasts to provide guidance for TC genesis probabilities over the NATL and EPAC basins, especially for the 2–5-day forecasts. To assess the potential for TC development, forecasters often compare and contrast the genesis forecasts from several of the best performing global models. When the quality or characteristics of the TC genesis forecasts change considerably from one model configuration to the next, it undermines forecaster confidence and requires forecasters to recalibrate how

1214

Forecasts

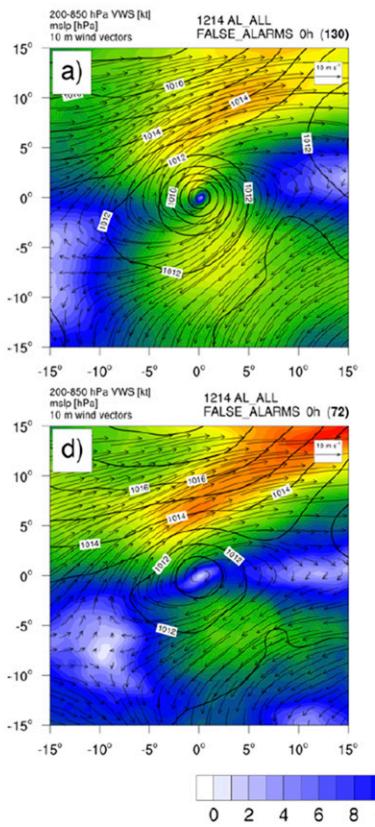
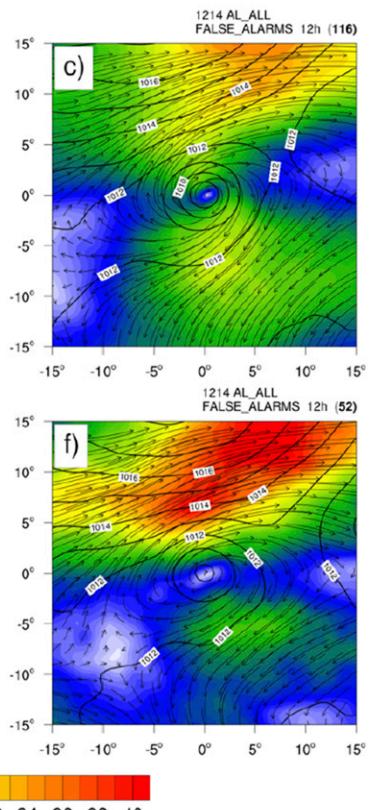
**False Alarms**

FIG. 13. As in Fig. 6, but for NATL false alarm composites from GFS 1314 of 200–850-hPa vertical wind shear (kt, shading and vectors) and mean SLP (hPa; contours).

they use the forecast guidance relative to the many other sources of information available to them. Therefore, it is important to understand how a model upgrade may affect the quality of the TC genesis forecasts. This information is not only useful for model developers that are working to improve the forecasts, it also allows forecasters to know what to expect during the upcoming season so they can make the best use of the guidance available to them.

Acknowledgments. The authors thank the anonymous AMS reviewers for their thoughtful feedback on this manuscript. The authors also thank Eric Blake, Mark DeMaria, Ed Rappaport, and Brian Zachry for conducting the NHC internal review of this manuscript. This research was supported by NOAA Grants NA17OAR4590141 and NA18NWS4680066. Funding for ABP was provided by NOAA's Hurricane Forecast Improvement Program.

Data availability statement. The data used in this study were provided by NOAA's Environmental Modeling Center.

APPENDIX**Select GFS Model Changes***a. GFS 1314 (McClung 2012)*

1. MODEL CONFIGURATION

- T574 Eulerian model resolution (~ 27 km).
- 1° Reynolds weekly sea surface temperature (SST) observations.

2. DATA ASSIMILATION SYSTEM

- 3D hybrid ensemble–variational data assimilation system.
- T254L64 ensemble Kalman filter resolution.

b. GFS 2015 (McClung 2014)

1. MODEL CONFIGURATION

- T1534 semi-Lagrangian model resolution (~ 13 km).
- 5' real-time global daily SST observations.
- Reduced drag coefficient was implemented at high wind speeds.

- “Modify initialization of forecast state variables to reduce a sharp decrease in cloud water in the first model time step.”
- “Use hybrid eddy-diffusivity mass-flux planetary boundary layer scheme and turbulent kinetic energy dissipative heating.”

2. DATA ASSIMILATION SYSTEM

- T574L64 ensemble Kalman filter resolution.
- Updated version of the Community Radiative Transfer Model that contains improved analysis of near-surface temperature over water.

c. GFS 2016 ([McClung 2016](#))

DATA ASSIMILATION SYSTEM

- 4D hybrid ensemble–variational data assimilation system.
- Advanced Very High Resolution Radiometer winds assimilated.

REFERENCES

- Avila, L. A., S. R. Stewart, R. Berg, and A. B. Hagen, 2020: National Hurricane Center tropical cyclone report: Hurricane Dorian (24 August–7 September 2019). NHC Tech. Rep. AL052019, 74 pp.
- Beven, J. L., 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.
- Blake, E. S., 2019: National Hurricane Center tropical cyclone report: Tropical Storm Kirk (22–28 September 2018). NHC Tech. Rep. AL122018, 15 pp.
- Briegel, L. M., and W. M. Frank, 1997: Large-scale influences on tropical cyclogenesis in the western North Pacific. *Mon. Wea. Rev.*, **125**, 1397–1413, [https://doi.org/10.1175/1520-0493\(1997\)125<1397:LSIOTC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1397:LSIOTC>2.0.CO;2).
- Cangialosi, J. P., and N. A. Ramos, 2019: National Hurricane Center tropical cyclone report: Tropical Storm Karen (22–27 September 2019). NHC Tech. Rep. AL122019, 20 pp.
- Chan, J. C., and R. H. Kwok, 1999: Tropical cyclone genesis in a global numerical weather prediction model. *Mon. Wea. Rev.*, **127**, 611–624, [https://doi.org/10.1175/1520-0493\(1999\)127<0611:TCGIAG>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<0611:TCGIAG>2.0.CO;2).
- Chan, K. T., and J. C. Chan, 2015: Global climatology of tropical cyclone size as inferred from QuikSCAT data. *Int. J. Climatol.*, **35**, 4843–4848, <https://doi.org/10.1002/joc.4307>.
- Chavas, D. R., and K. A. Emanuel, 2010: A QuikSCAT climatology of tropical cyclone size. *Geophys. Res. Lett.*, **37**, L18816, <https://doi.org/10.1029/2010GL044558>.
- Chen, J.-H., S.-J. Lin, L. Zhou, X. Chen, S. Rees, M. Bender, and M. Morin, 2019: Evaluation of tropical cyclone forecasts in the Next Generation Global Prediction System. *Mon. Wea. Rev.*, **147**, 3409–3428, <https://doi.org/10.1175/MWR-D-18-0227.1>.
- Cheung, K. K., and R. L. Elsberry, 2002: Tropical cyclone formations over the western North Pacific in the Navy Operational Global Atmospheric Prediction System forecasts. *Wea. Forecasting*, **17**, 800–820, [https://doi.org/10.1175/1520-0434\(2002\)017<0800:TCFOTW>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0800:TCFOTW>2.0.CO;2).
- Corbosiero, K. L., and J. Molinari, 2002: The effects of vertical wind shear on the distribution of convection in tropical cyclones. *Mon. Wea. Rev.*, **130**, 2110–2123, [https://doi.org/10.1175/1520-0493\(2002\)130<2110:TEOVWS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2110:TEOVWS>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, <https://doi.org/10.1175/WAF-D-12-00073.1>.
- Davis, C., W. Wang, J. Dudhia, and R. Torn, 2010: Does increased horizontal resolution improve hurricane wind forecasts? *Wea. Forecasting*, **25**, 1826–1841, <https://doi.org/10.1175/2010WAF222423.1>.
- Dunion, J., 2017: NOAA JHT final report: Improvement to the Tropical Cyclone Genesis Index (TCGI). NOAA/NHC Tech. Rep., 16 pp.
- 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, <https://doi.org/10.1175/WAF-D-13-00104.1>.
- Fierro, A. O., R. F. Rogers, F. D. Marks, and D. S. Nolan, 2009: The impact of horizontal grid spacing on the microphysical and kinematic structures of strong tropical cyclones simulated with the WRF-ARW model. *Mon. Wea. Rev.*, **137**, 3717–3743, <https://doi.org/10.1175/2009MWR2946.1>.
- Fritz, C., and Z. Wang, 2013: A numerical study of the impacts of dry air on tropical cyclone formation: A development case and a nondevelopment case. *J. Atmos. Sci.*, **70**, 91–111, <https://doi.org/10.1175/JAS-D-12-018.1>.
- Gopalakrishnan, S. G., F. Marks Jr., X. Zhang, J.-W. Bao, K.-S. Yeh, and R. Atlas, 2011: The experimental HWRF system: A study on the influence of horizontal resolution on the structure and intensity changes in tropical cyclones using an idealized framework. *Mon. Wea. Rev.*, **139**, 1762–1784, <https://doi.org/10.1175/2010MWR3535.1>.
- 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, <https://doi.org/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, <https://doi.org/10.1175/WAF-D-15-0157.1>.
- , R. E. Hart, H. E. Fuelberg, and J. H. Cossuth, 2017: The development and evaluation of a statistical–dynamical tropical cyclone genesis guidance tool. *Wea. Forecasting*, **32**, 27–46, <https://doi.org/10.1175/WAF-D-16-0072.1>.
- 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, 24 pp., <https://www.nhc.noaa.gov/pdf/NWS-NHC-1988-22.pdf>.
- 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, <https://doi.org/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, <https://doi.org/10.1175/MWR-D-12-00254.1>.
- Lee, C.-Y., S. J. Camargo, F. Vitart, A. H. Sobel, and M. K. Tippett, 2018: Subseasonal tropical cyclone genesis prediction and MJO in the S2S dataset. *Wea. Forecasting*, **33**, 967–988, <https://doi.org/10.1175/WAF-D-17-0165.1>.

- McAdie, C., C. Landsea, C. Neumann, J. David, E. Blake, and G. Hammer, 2009: Tropical cyclones of the North Atlantic Ocean, 1851–2006. Historical Climatology Series 6-2, NCDC, 238 pp., https://www.nhc.noaa.gov/pdf/TC_Book_Atl_1851-2006_lowres.pdf.
- McClung, T., 2012: Technical implementation notice 12-22, Amended. NOAA/NWS Tech. Rep., 4 pp., https://www.weather.gov/media/notification/tins/tin12-22gfs_hybridaab.pdf.
- , 2014: Technical implementation notice 14-46 Corrected. NOAA/NWS Tech. Rep., 8 pp., https://www.weather.gov/media/notification/tins/tin14-46gfs_ccaa.pdf.
- , 2016: Technical implementation notice 16-11, Amended. NOAA/NWS Tech. Rep., 6 pp., https://www.weather.gov/media/notification/tins/tin16-11gfs_gdasaaa.pdf.
- 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**, 1963–1989, <https://doi.org/10.1175/MWR-D-12-00186.1>.
- Penny, A. B., J. P. Hacker, and P. A. Harr, 2016a: Analysis of tropical storm formation based on ensemble data assimilation and high-resolution numerical simulations of a nondeveloping disturbance. *Mon. Wea. Rev.*, **144**, 3631–3649, <https://doi.org/10.1175/MWR-D-16-0100.1>.
- , P. A. Harr, and J. D. Doyle, 2016b: Sensitivity to the representation of microphysical processes in numerical simulations during tropical storm formation. *Mon. Wea. Rev.*, **144**, 3611–3630, <https://doi.org/10.1175/MWR-D-15-0259.1>.
- Pratt, A. S., and J. L. Evans, 2009: Potential impacts of the Saharan air layer on numerical model forecasts of North Atlantic tropical cyclogenesis. *Wea. Forecasting*, **24**, 420–435, <https://doi.org/10.1175/2008WAF2007090.1>.
- Rappin, E. D., M. C. Morgan, and G. J. Tripoli, 2011: The impact of outflow environment on tropical cyclone intensification and structure. *J. Atmos. Sci.*, **68**, 177–194, <https://doi.org/10.1175/2009JAS2970.1>.
- Roeber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, <https://doi.org/10.1175/2008WAF2222159.1>.
- Ryglicki, D. R., J. D. Doyle, Y. Jin, D. Hodyss, and J. H. Cossuth, 2018: The unexpected rapid intensification of tropical cyclones in moderate vertical wind shear. Part II: Vortex tilt. *Mon. Wea. Rev.*, **146**, 3801–3825, <https://doi.org/10.1175/MWR-D-18-0021.1>.
- Sampson, C. R., and A. J. Schrader, 2000: The automated tropical cyclone forecasting system (version 3.2). *Bull. Amer. Meteor. Soc.*, **81**, 1231–1240, [https://doi.org/10.1175/1520-0477\(2000\)081<1231:TATCFS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2).
- 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, <https://doi.org/10.1175/2008WAF2007109.1>.
- Tsai, H.-C., and R. L. Elsberry, 2019: Combined three-stage 7-day weighted analog intensity prediction technique for western North Pacific tropical cyclones: Demonstration of optimum performance. *Wea. Forecasting*, **34**, 1979–1998, <https://doi.org/10.1175/WAF-D-19-0130.1>.
- , K.-C. Lu, R. L. Elsberry, M.-M. Lu, and C.-H. Sui, 2011: Tropical cyclone-like vortices detection in the NCEP 16-day ensemble system over the western North Pacific in 2008: Application and forecast evaluation. *Wea. Forecasting*, **26**, 77–93, <https://doi.org/10.1175/2010WAF2222415.1>.
- Wang, Z., W. Li, M. S. Peng, X. Jiang, R. McTaggart-Cowan, and C. A. Davis, 2018: Predictive skill and predictability of North Atlantic tropical cyclogenesis in different synoptic flow regimes. *J. Atmos. Sci.*, **75**, 361–378, <https://doi.org/10.1175/JAS-D-17-0094.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- Yamaguchi, M., and N. Koide, 2017: Tropical cyclone genesis guidance using the early stage Dvorak analysis and global ensembles. *Wea. Forecasting*, **32**, 2133–2141, <https://doi.org/10.1175/WAF-D-17-0056.1>.