Performance of Basin-Scale HWRF Tropical Cyclone Track Forecasts

GHASSAN J. ALAKA JR. AND XUEJIN ZHANG

Cooperative Institute for Marine and Atmospheric Studies, University of Miami, and NOAA/Atlantic Oceanographic and Meteorological Laboratory/Hurricane Research Division, Miami, Florida

SUNDARARAMAN G. GOPALAKRISHNAN, STANLEY B. GOLDENBERG, AND FRANK D. MARKS

NOAA/Atlantic Oceanographic and Meteorological Laboratory/Hurricane Research Division, Miami, Florida

(Manuscript received 18 August 2016, in final form 4 April 2017)

ABSTRACT

The Hurricane Weather Research and Forecasting (HWRF) Model is a dynamical model that has shown annual improvements in its tropical cyclone (TC) track forecasts as a result of various modifications. This study focuses on an experimental version of HWRF, called the basin-scale HWRF (HWRF-B), configured with 1) a large, static outer domain to cover multiple TC basins and 2) multiple sets of high-resolution movable nests to produce forecasts for several TCs simultaneously. Although HWRF-B and the operational HWRF produced comparable average track errors for the 2011–14 Atlantic hurricane seasons, strengths of HWRF-B are identified and linked to its configuration differences. HWRF-B track forecasts were generally more accurate compared with the operational HWRF when at least one additional TC was simultaneously active in the Atlantic or east Pacific basins and, in particular, when additional TCs were greater than 3500 km away. In addition, at long lead times, HWRF-B average track errors were lower than for the operational HWRF for TCs initialized north of 25°N or west of 60°W, highlighting the sensitivity of TC track forecasts to the location of the operational HWRF's outermost domain. A case study, performed on Hurricane Michael, corroborated these HWRF-B strengths. HWRF-B shows the potential to serve as an effective bridge between regional modeling systems and next-generational global efforts.

1. Introduction

Tropical cyclone (TC) model forecasts are critical tools that help forecasters, emergency managers, and citizens prepare for potentially devastating landfall events. Currently, the National Hurricane Center (NHC) of the National Weather Service (NWS) relies on both coarse-scale global models and finer-scale regional models to provide numerical guidance for TC forecasters. Two of the primary National Oceanic and Atmospheric Administration (NOAA) dynamical models used by NHC for numerical forecast guidance are the Global Forecast System (GFS) for track forecasts and the Hurricane Weather Research and Forecasting (HWRF) Model for intensity forecasts and, to some extent, track forecasts. With support from NOAA's Hurricane Forecast Improvement Project (HFIP; Gall et al. 2013), HWRF intensity forecasts have improved and met HFIP's 5-yr intensity goal (Atlas et al. 2015; Gopalakrishnan et al. 2016). HWRF track forecasts have also improved under HFIP. Recently, Goldenberg et al. (2015) demonstrated track forecast improvements by configuring a convection-permitting third nest (3-km horizontal grid spacing) in the 2012 version and improved intensity forecasts by adding more frequent physics calls in the 2013 version.

Yet, despite recent advancements to the HWRF system (Tallapragada et al. 2014b), operational HWRF track guidance has generally remained inferior to global model track guidance. The operational HWRF has two limitations: it is storm centric and has a small outermost domain. Also, the operational HWRF is only able to use

Supplemental information related to this paper is available at the Journals Online website: http://dx.doi.org/10.1175/WAF-D-16-0150.s1.

Corresponding author: Ghassan J. Alaka, Jr., ghassan.alaka@noaa.gov

DOI: 10.1175/WAF-D-16-0150.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

one set of high-resolution movable multilevel nests (MMLNs[1]; i.e., for one TC or invest[2]) per forecast cycle (Zhang et al. 2016). This necessitates individual runs for each TC/invest (hereafter, tropical systems). In addition to the one-MMLN limitation, the operational HWRF features an outermost domain that lacks the horizontal expanse to capture critical synoptic-scale features and multiscale interactions that can influence TC tracks. TC tracks are primarily determined by the large-scale environment, especially the large-scale wind fields (e.g., Franklin et al. 1996). Therefore, it is crucial to accurately initialize and predict the large-scale environment in order to reduce track errors at longer lead times (Gopalakrishnan et al. 2016).

In an effort to build a more robust HWRF system that can accommodate future enhancements, an experimental and parallel version of HWRF, called the basin-scale HWRF (HWRF-B; Zhang et al. 2016), has been developed by NOAA's Hurricane Research Division (HRD) at the Atlantic Oceanographic and Meteorological Laboratory (AOML), with its partners at the National Centers for Environmental Prediction (NCEP) Environmental Modeling Center (EMC) and the Developmental Testbed Center (DTC). HWRF-B addresses some of the shortcomings of the current operational HWRF configuration with two enhancements. First, by utilizing a much larger outermost domain that covers most of the North Atlantic,[3] east Pacific, and central Pacific hurricane basins (Fig. 1), the HWRF-B domain includes most of the North American continent, which may reduce errors related to topography, land–atmosphere interactions, and the timing/amplitude of midlatitude weather systems. Second, HWRF-B may be configured with multiple sets of MMLNs, giving it the ability to produce high-resolution forecasts for several tropical systems simultaneously. Consequently, the importance of multistorm (i.e., TC–TC) and multiscale interactions may be more rigorously tested in HWRF-B. The importance of multiple sets of MMLNs is supported by Zhang et al. (2016), who found that 81.0% (44.3%) of verifiable forecast cycles in the Atlantic basin from 2011 to 2014 featured at least two (three) tropical systems in the Atlantic/east Pacific basins. Forecast cycles that were limited to a single TC in the Atlantic basin were uncommon (19.0%).

TC–TC interactions can be split into three categories: direct, indirect, and remote. Fujiwhara (1921) first documented direct interactions between TCs, called the Fujiwhara effect. These binary TC–TC interactions typically occur when two TCs are less than 1400 km (~12°) apart (Brand 1970; Lander and Holland 1993; Carr et al. 1999). In these direct TC–TC interactions, TCs can influence the steering upon and even rotate around one another, which can significantly alter tracks. Indirect TC–TC interactions typically occur when two TCs are 15°–30° apart (Carr et al. 1999), so that the environmental outflow of one TC can induce vertical wind shear over the other, thus affecting the intensity of the second TC (Susca-Lopata et al. 2015). Remote TC–TC interactions are less intuitive, as information is quickly shared over thousands of kilometers via synoptic-scale teleconnections. Remote TC–TC interactions are not mentioned in the extensive review by Carr et al. (1999). Previous studies have noted improved TC track forecasts when far-field TCs are present in a global model domain (Aberson and DeMaria 1994; Aberson 2010, 2011). The HWRF-B configuration can help to provide insights into all three categories of TC–TC interactions.

This study focuses on how TC forecast tracks are impacted by the core modifications to HWRF implemented for HWRF-B: 1) a large outermost domain and 2) multiple sets of MMLNs. Section 2 introduces the two HWRF model configurations used in this study and additional data used to supplement the analysis. Section 3 documents track errors for the overall sample and for individual TCs during the period 2011–14 for HWRF-B versus the operational HWRF. Section 4 analyzes the sensitivity of track errors to the number of MMLNs, the number of tropical systems, and the location of the outermost domain. Section 5 investigates track errors for a specific Hurricane Michael forecast and section 6 discusses our conclusions and future areas of study.

2. Model configurations and methodology

The HWRF system was developed at NOAA/NWS/NCEP in 2007 (Gopalakrishnan et al. 2013) to improve operational track and intensity forecasts available to operational forecasters and to increase the HFIP community's understanding of TCs. A version of this incrementally evolving system is available at DTC of the National Center for Atmospheric Research (NCAR), which maintains a version of the HWRF system and the corresponding scientific documentation (Gopalakrishnan et al. 2010). HWRF-B and the operational HWRF represent two unique configurations of the same modeling system (see Table 1 and Fig. 1). Specifically, HWRF-B

[1] Refer to Zhang et al. (2016) for technical details on MMLNs.

[2] An invest is a weak tropical system that lacks the necessary organization and structure to be considered a TC, but is tracked by NHC for potential development. Many invests are characterized by disorganized positive vorticity or intermittent deep convection (Gray 1998). More information is available online (http://www.nhc.noaa.gov/aboutgloss.shtml#invest).

[3] The North Atlantic hurricane basin includes the North Atlantic Ocean, the Caribbean Sea, and the Gulf of Mexico.

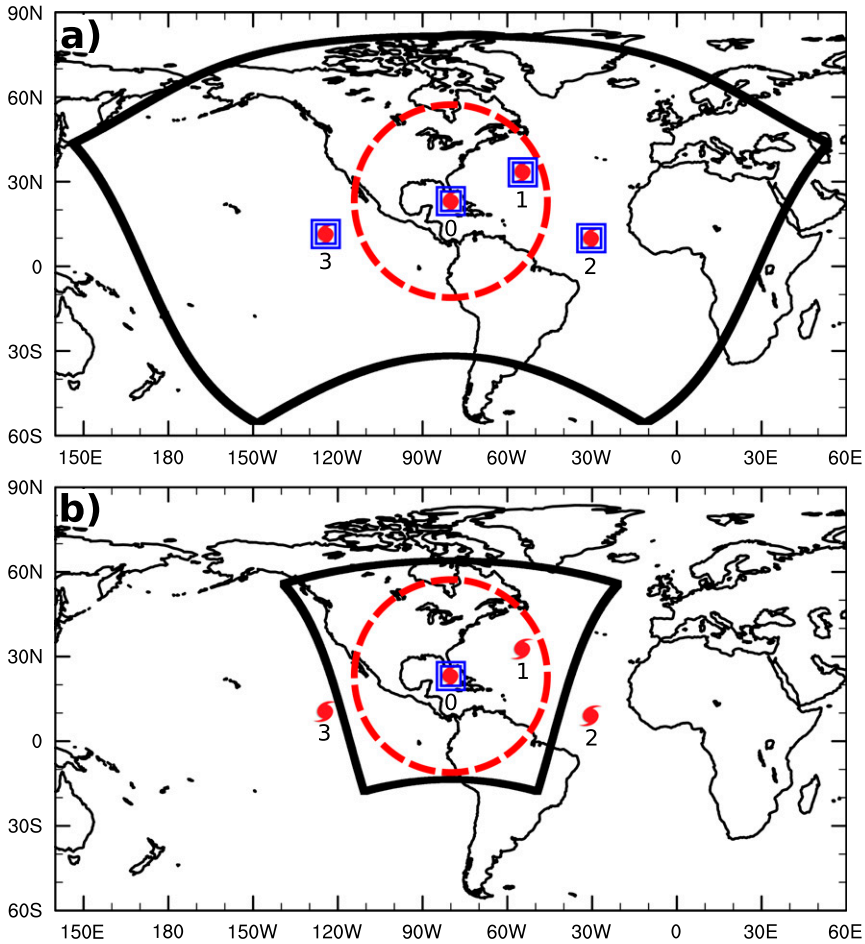

FIG. 1. Examples of outermost domains and MMLNs in (a) HWRF-B and (b) operational HWRF. The red circle denotes a radius of 3500 km around TC 0. In (a), MMLNs are configured for all four TCs. In (b), one MMLN is configured for TC 0.

features a large outermost domain and up to nine sets of MMLNs (computer resources permitting), while the operational HWRF has a much smaller outermost domain and only one set of MMLNs. Each MMLN is centered over, and follows, the tropical system of interest to better resolve its inner-core structure, while the outermost domain captures the large-scale environment that the tropical system interacts with throughout the forecast (Zhang et al. 2016).

The operational HWRF and HWRF-B versions are constantly evolving. In this study, the closest versions of these two systems available at the time were chosen to investigate track forecasts: the 2013 HWRF-B (HB13; Zhang et al. 2016) and the 2014 operational HWRF (H214; Tallapragada et al. 2014a). Similarities and differences between the HB13 and H214 configurations are provided in Table 1. HB13 and H214 are triply nested systems with a grid spacing of 27 km for the outermost domain and grid spacings of 9 and 3 km for the MMLN (Zhang et al. 2016). In HB13, the vertical resolution was increased from 42 vertical levels with a model top of 50 hPa to 61 vertical levels with a model top of 2 hPa. This increase in vertical resolution motivated the implementation of these enhancements in H214 (and subsequent operational versions of HWRF) and led to significant track improvements in H214 over previous versions of HWRF.

HB13 and H214 use the same physics packages (Table 1). The PBL scheme used in all model domains is a modified version from the GFS (Hong and Pan 1996; Gopalakrishnan et al. 2013). Convection is simulated by a simplified Arakawa–Schubert scheme (SAS; Pan and Wu 1995). The SAS scheme is turned off in the innermost domain where clouds are explicitly resolved. The microphysics package is the modified Ferrier scheme (Ferrier 2005). The shortwave and longwave radiation packages are Geophysical Fluid Dynamic Laboratory (GFDL) schemes (Lacis and Hansen 1974; Fels and Schwarzkopf 1975; Schwarzkopf and Fels

TABLE 1. Configuration options for HB13 and H214. Options that differ between HB13 and H214 are set in boldface.

Configuration options	HB13	H214
Domains	**27 km: 178.2° × 77.6°** **9 km: 10.6° × 10.2°** **3 km: 6.1° × 5.4°**	**27 km: 77.6° × 77.6°** **9 km: 12.0° × 12.0°** **3 km: 7.1° × 7.1°**
Data assimilation	**No GSI**	**Hybrid GSI**
Ocean coupling	**Static SST**	**27/9 km: yes** **3 km: downscaled**
Multistorm	**Yes (up to four)**	**No**
Model top	2 hPa	Same as HB13
Vertical levels	61	Same as HB13
Vortex initialization	At 3 km	Same as HB13
Physics schemes:		
Microphysics	Modified Ferrier (high resolution)	Same as HB13
Radiation	GFDL	Same as HB13
Convection	27/9 km: SAS, shallow 3 km: none	Same as HB13
Boundary layer	Modified GFS	Same as HB13
Land surface	GFDL slab	Same as HB13
Surface layer	GFDL	Same as HB13

1991). The surface-layer package is also a GFDL scheme (Kurihara and Tuleya 1974; Sirutis and Miyakoda 1990). Finally, the land surface is represented by a simple GFDL slab scheme (Tuleya 1994). A detailed discussion of the physics schemes used in the HWRF systems is provided in Gopalakrishnan et al. (2011), Yeh et al. (2012), Zhang et al. (2016), and the online HWRF scientific documentation.

Table 1 and Fig. 1 highlight the four important distinctions of HB13 versus H214. HB13 has 1) a large, static outermost domain; 2) the capacity to forecast multiple tropical systems simultaneously through the use of several sets of MMLNs; 3) no ocean coupling; and 4) no data assimilation. The first two differences are the primary focuses of this study. The H214 outermost domain (~80° × ~80°) is centered on the TC of interest at the initial time of each forecast cycle (Fig. 1b). The HB13 outermost domain spans almost 180° in longitude and 80° in latitude, which covers close to a quarter of the globe (Fig. 1a). Current computational resources allow HB13 to support up to four sets of MMLNs (Fig. 1a), which are assigned to active tropical systems at the initial forecast time.[4] H214, which is configured with just one MMLN, can only forecast one tropical system at high resolution per model run (e.g., TC 0 in Fig. 1b). For example, if four TCs are active for a particular forecast cycle (as in Fig. 1b), H214 would need to run four times to produce a high-resolution forecast for each tropical system. Tropical systems that developed after the forecast initial time are not assigned MMLNs in either model.

Operational versions of HWRF (e.g., H214) are configured with ocean coupling (i.e., two-way interaction between the ocean and atmosphere). In H214, the Princeton Ocean Model (POM) is coupled to the atmosphere in the 27- and 9-km domains, is downscaled in the 3-km domain, and is capable of forecasting the impact of the atmosphere, including TCs, on the ocean (Yablonsky et al. 2015). HB13 does not include ocean coupling and, therefore, is limited to one-way ocean-to-atmosphere forcing (i.e., the ocean is not modified during the forecast). However, with regard to TC forecasts, the issue is not how the TC modifies the ocean (e.g., cold wake), but, rather, how the modified ocean then impacts the TC forecast. As some studies have shown, the cold wake produced by a TC can have a significant impact on a TC when the TC is slow moving, large, and/or intense, particularly when the mixed layer is shallow (Bender and Ginis 2000; Yablonsky and Ginis 2009; Halliwell et al. 2015). For example, Zhang et al. (2016) found that intensity forecasts for Leslie, a large and slow-moving hurricane, dramatically improved when ocean coupling was employed. The data assimilation package featured in H214 is also absent in HB13. In H214, a gridpoint statistical interpolation (GSI) method is used to assimilate data in the 9- and 3-km domains (i.e., the MMLN). Zou et al. (2013) showed that the impact of data assimilation on track forecasts is minimal. Both models include the vortex-relocation-based methodology proposed in Kurihara et al. (1993).

It is important to note that the individual impacts of the configuration differences between HB13 and H214 (i.e., size of outermost domain, number of MMLNs,

[4] If more than four tropical systems are active, priority for MMLN is Atlantic TCs, eastern Pacific TCs, Atlantic invests, and eastern Pacific invests, in that order.

ocean coupling, and data assimilation) are difficult to resolve in this analysis because of the nature of the model runs. Limited computing resources prevented rigorous investigation into the sensitivity of track forecasts to these separate configuration options.

The NHC postprocessed best track ("BEST") is used as the true TC location and intensity to calculate track errors in model forecasts (Rappaport et al. 2009). Track or intensity improvement,[5] defined as

$$\text{Improvement} = 100 \times \left(1 - \frac{\overline{\text{HB13}}}{\overline{\text{H214}}}\right), \qquad (1)$$

is used to compare track and intensity forecasts from these two models, where $\overline{\text{HB13}}$ and $\overline{\text{H214}}$ represent the average track or intensity errors from each model, respectively, and $\overline{\text{H214}}$ is the baseline by which improvement is measured. Negative improvement is referred to as degradation. In this study, track/intensity improvement (degradation) indicates that HB13 produces lower (higher) track/intensity errors than H214. Statistical significance is calculated via a two-tailed Student's *t* test, for which sample sizes are reduced based on 24-h serial correlation (e.g., Aberson and DeMaria 1994; Goldenberg et al. 2015). The thresholds for statistical significance are set to 90% and 95% throughout the study. The frequency of superior performance (FSP; Velden and Goldenberg 1987; Goldenberg et al. 2015) indicates how often each model produces better forecasts.[6]

Finally, to illustrate the importance of the outermost domain size and multiple sets of MMLNs, a case study is performed on Hurricane Michael for the forecast cycle initialized at 0600 UTC 4 September 2012. GFS analysis (GFSA; http://www.emc.ncep.noaa.gov/GFS/doc.php) represents concurrent observations at all model valid times. GFSA fields have a horizontal grid spacing of 0.5°. TC vortices are removed using the methodology of Kurihara et al. (1993).

3. Track verification

Track forecast errors for most 2011–14 Atlantic basin TCs (58 of 63[7]) are examined in this study (see the online supplement to this article for more information about these 58 TCs). For these TCs, tracks are provided in Fig. 2. The 4-yr sample includes 1119 individual forecasts, of which 861 (76.9%) and 296 (26.5%) verify at 24 and 120 h, respectively. The TCs cover a large portion of the North Atlantic basin and include tropical depressions, tropical storms, hurricanes (categories 1–5), subtropical depressions, and subtropical storms. Although forecasts were made for all tropical systems, the results in this study follow NHC verification guidelines (http://www.nhc.noaa.gov/verification/), which state that location (i.e., track) and intensity are verified only for the cases where a system is a TC at the initial forecast time and the verification time (Cangialosi and Franklin 2016).

Absolute track errors are comparable for HB13 and H214 at all lead times (Fig. 3a). In Fig. 3b, track improvements versus H214 and FSP are presented for the full sample of TCs in the Atlantic basin during the period 2011–14 (figure adapted from Zhang et al. 2016). For 1119 forecasts, HB13 and H214 track errors are comparable at most lead times. HB13 shows track improvement compared with H214 by 8% at 12 h, which is significant at the 95% confidence level. Upon closer inspection, HB13 track improvements at 12 h could be related to a better representation of the TC intensity without data assimilation (see Fig. 7a in Zhang et al. 2016). Issues with TC intensities in the earliest H214 lead times might be related to the "spindown" problem documented in the HWRF system (Gopalakrishnan et al. 2012; Vukicevic et al. 2013). The FSP closely follows the shape of the HB13 track improvement, which points to the overall similarity of the performance of the two models, even at longer lead times.

HB13 and H214 track errors were also examined for every Atlantic basin TC individually for 2011–14 (not shown). It is clear that track errors and the relative performance of the two models vary considerably from one TC to the next. For some TCs, a given model may repeatedly issue forecasts with large or small track errors. From the perspective of a TC forecaster, one model might be preferred over another based on which performs better on a TC-by-TC (i.e., storm by storm) basis. Accordingly, a storm-by-storm FSP (SFSP) is introduced to compare track errors averaged for each TC at all lead times. This statistic is closely related to FSP, except that it compares the performance of the two models storm by storm rather than case by case. SFSP (Fig. 4) gives the percentage of TCs for which HB13 produces lower mean track errors compared with H214. To select TCs with a significant number of forecasts, TCs with small case numbers (the bottom third) were dropped at each lead time from the SFSP analysis. The minimum case number threshold is given for each forecast lead time in Fig. 4.

[5] The equation for "improvement" is identical to the equation for "skill score" (http://glossary.ametsoc.org/wiki/Skill).

[6] One count is given to a model for each superior forecast and 0.5 count is given to both models for a tie. For example, an FSP of 50% indicates that HB13 has the same number of superior and inferior forecasts as H214. An FSP of >50% (<50%) indicates HB13 has more (fewer) superior forecasts than H214.

[7] The five TCs not included in this study were minor, short-lived systems that had minimal impact on the overall average track errors, especially at medium- and long-range lead times.

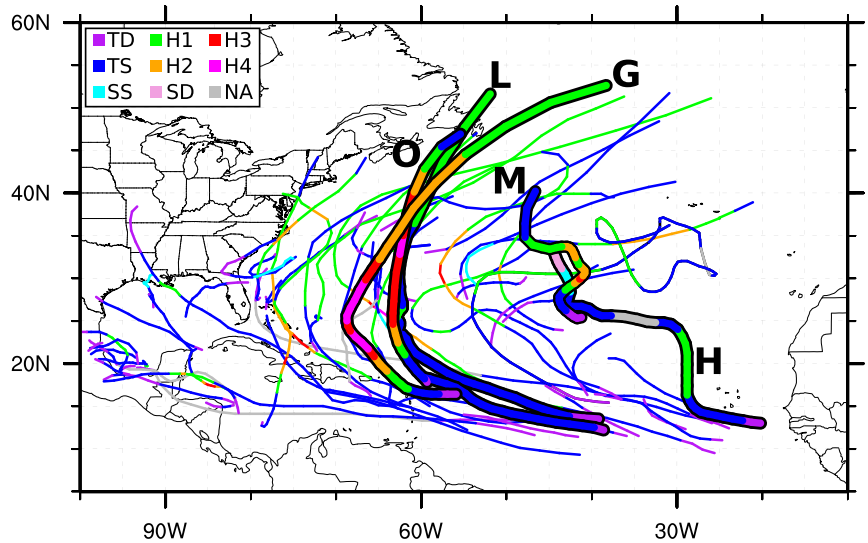

FIG. 2. Complete tracks and intensities for the 58 Atlantic basin TCs from 2011 to 2014 used in this study. BEST intensities are color coded based on classification. TC classifications include tropical depression (TD), tropical storm (TS), subtropical depression (SD), subtropical storm (SS), category 1 hurricane (H1), category 2 hurricane (H2), category 3 hurricane (H3), and category 4 hurricane (H4). The classification is considered not applicable (NA) if a system loses TC status during its lifetime. Thicker tracks and boldface letters represent Michael (M) (see section 5) and the four "problem" TCs (see section 3), Gonzalo (G), Humberto (H), Leslie (L), and Ophelia (O).

On a TC-by-TC basis, HB13 track errors are smaller for 70% of TCs at 12 h. From 12 to 96 h, SFSP exhibits a notable downward trend. By 120 h, SFSP increases up to 60%. This is an interesting result since HB13 track errors, averaged for all forecasts, showed degradation at 120 h (Fig. 3b). HB13 track degradation and higher SFSP at 120 h indicate that HB13 produces relatively large track errors for a few TCs (see Fig. 5). But, overall, HB13 track forecasts are more accurate than H214 at 120 h for the majority of TCs.

HB13 and H214 track errors at 120 h are compared for each individual TC (Fig. 5), which illustrates the

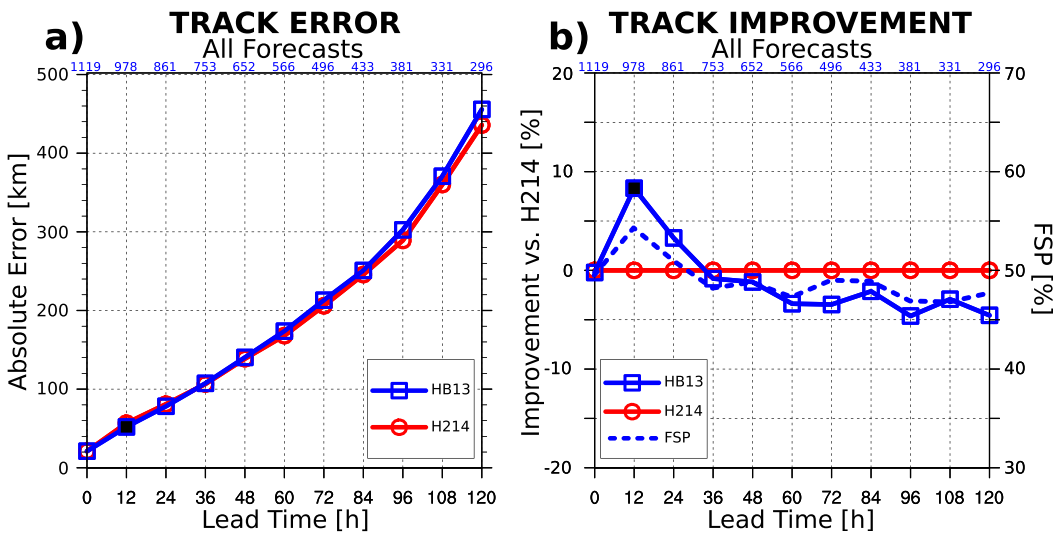

FIG. 3. (a) Track error and (b) track improvement (vs H214) are shown for Atlantic basin TC forecasts from HB13 (blue squares) and H214 (red circles) for 2011–14. FSP is shown for HB13 vs H214 (blue dashed). The number of cases at each lead time is given at the top of each column. HB13 track improvements that are significant at the 95% (90%) confidence level are marked with black (cyan).

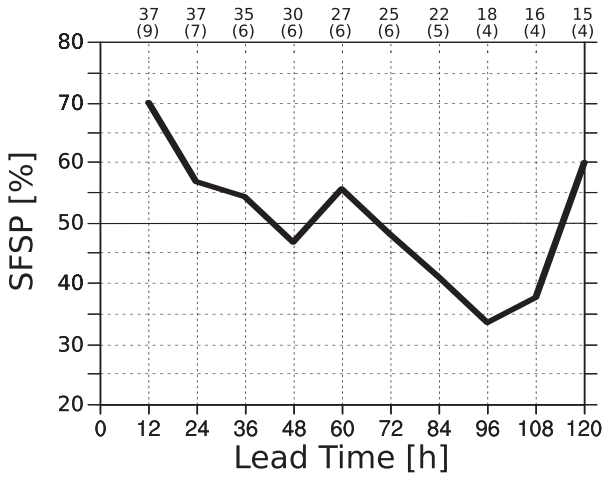

FIG. 4. The SFSP for HB13 vs H214. Values above 50% indicate forecasts where HB13 is superior for more TCs than H214. The number of TCs at each lead time is given at the top of each column with the minimum number of cases required for each TC to be included in the analysis at each lead time in parentheses.

variability of average and relative track errors from one TC to the next. For 2011–14, the average NHC official track error was ~400 km at 120 h. For some TCs (e.g., Edouard), both models produced track errors less than 400 km at 120 h, while, in the case of Philippe, both models produced track errors over 540 km. In a few cases, such as Michael and Irene, HB13 (H214) produced track errors less (greater) than 400 km at 120 h. The opposite is true for Katia.

Stratifications can help identify meaningful subsets of the full sample. For example, it is useful to identify cases with relatively poor ("outlier") track forecasts that might be associated with HB13 configuration deficiencies (i.e., no ocean coupling and no data assimilation). Although these deficiencies are important to address in future versions of HWRF-B, the goal of this study is to assess the impact of HWRF-B's core modifications on track forecasts. Intensity forecast degradation at 12- or 24-h lead times and overall track forecast degradation were used to identify four "problem" TCs that may have been impacted by a lack of ocean coupling and data assimilation: Humberto, Gonzalo, Leslie, and Ophelia (see Fig. 2 for tracks). These four problem TCs represent 13.5% of the total sample, and, yet, they have a disproportionately large negative impact on the overall HB13 track errors. While HB13 track forecasts degrade for each of these TCs for most lead times (Figs. 6a–d), HB13 intensity errors are much more variable (Figs. 6e–h). For example, the HB13 intensity improvement for Ophelia is >20% at 48–72 h. Conversely, the intensity degradation for Leslie is >120% at 60–96 h. Zhang et al. (2016) showed that extremely poor intensity forecasts for Leslie, a large, slow-moving TC, were associated with a lack of ocean coupling in HB13. Without the damping effects of the cold wake, Leslie forecasts with HB13 had a large positive intensity bias. Further research is necessary to understand the impact of ocean coupling and/or data assimilation on HWRF-B forecasts. These outlier TCs should be leveraged to advance future versions of HWRF-B.

4. HWRF-B track sensitivities

a. Sensitivity to multiple TCs

An important difference between HWRF-B and the operational HWRF is the number of tropical systems captured by the outermost domain. Since the operational HWRF runs independently for each tropical system at each forecast cycle, a storm-centric approach was

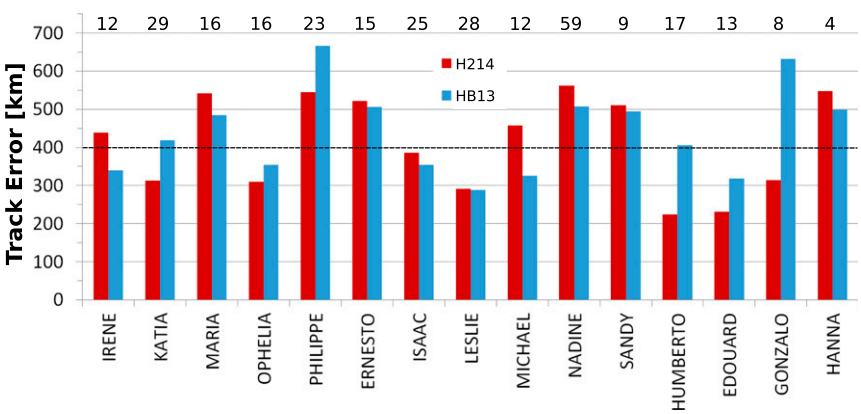

FIG. 5. HB13 (blue) and H214 (red) 120-h track errors for individual TCs. Those with fewer than four cases at 120 h are ignored. The number of cases for each TC is given at the top of each column. The NHC official average annual track error (400 km) at 120 h for the 2015 season is indicated by a dashed line.

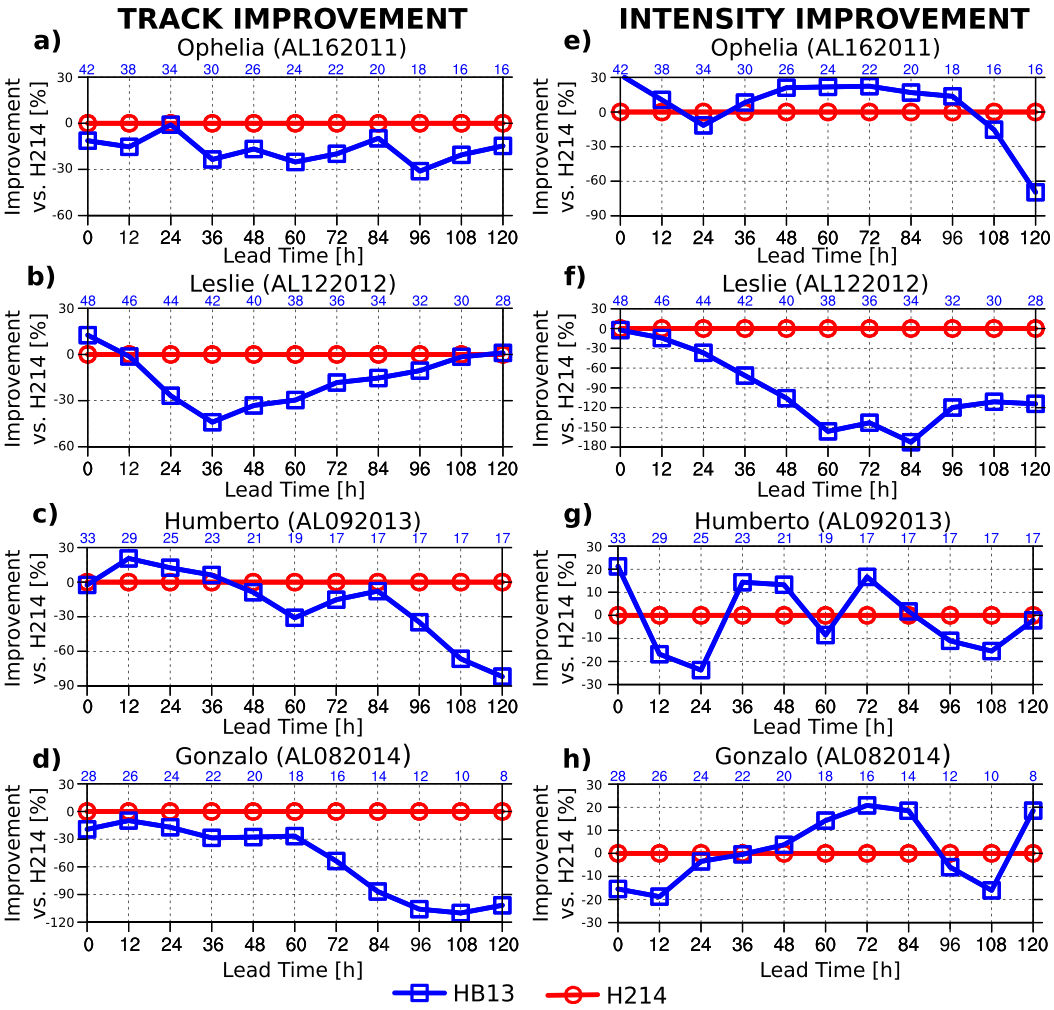

FIG. 6. (left) Track improvement and (right) intensity improvement from HB13 (blue squares) and H214 (red circles) for the four problem TCs: (a),(e) Ophelia, (b),(f) Leslie, (c),(g) Humberto, and (d),(h) Gonzalo. The number of cases at each lead time is given at the top of each column.

employed to study the impact of multiple tropical systems on track forecasts in the Atlantic basin. The number of additional tropical systems that were present in the Atlantic and east Pacific basins at any lead time (nTC) was counted for each verifiable TC forecast in the Atlantic basin. When stratifying by initial time, MMLNs were initialized for up to four tropical systems in HB13. When stratifying by verification time, the number of MMLNs still depends on the number of tropical systems present at the initial time and, therefore, was variable. Tropical systems that developed after the forecast initial time were not assigned MMLNs in either model. Simply stratifying by nTC at the initial time resulted in comparable HB13 and H214 track errors (not shown), which casts doubt on the importance of multiple MMLNs to track forecasts in HB13. However, stratifying by nTC at the verification time appears to have a positive impact on track forecasts and the remainder of this section examines these results.

When nTC ≥ 1, the vast majority of the HB13 forecasts were retained. Compared with the Fig. 3b results, the HB13 track forecasts showed improvement at medium-range (i.e., 48–84 h) lead times (Fig. 7a). Half of all cases had nTC ≥ 2, and HB13 track forecasts showed slightly larger improvements at medium-range lead times. In fact, HB13 track forecasts show improvement and FSP is >50% between 12 and 84 h for nTC ≥ 2. In both stratifications, HB13 shows significant improvement over H214 at 12 h. Results for nTC ≥ 3 are not shown because of the low number of cases.

An interesting stratification is to consider how the interaction between TCs that are thousands of kilometers apart (i.e., remote TC–TC interactions) impacts track improvement. The red-dashed circle in Fig. 1 is

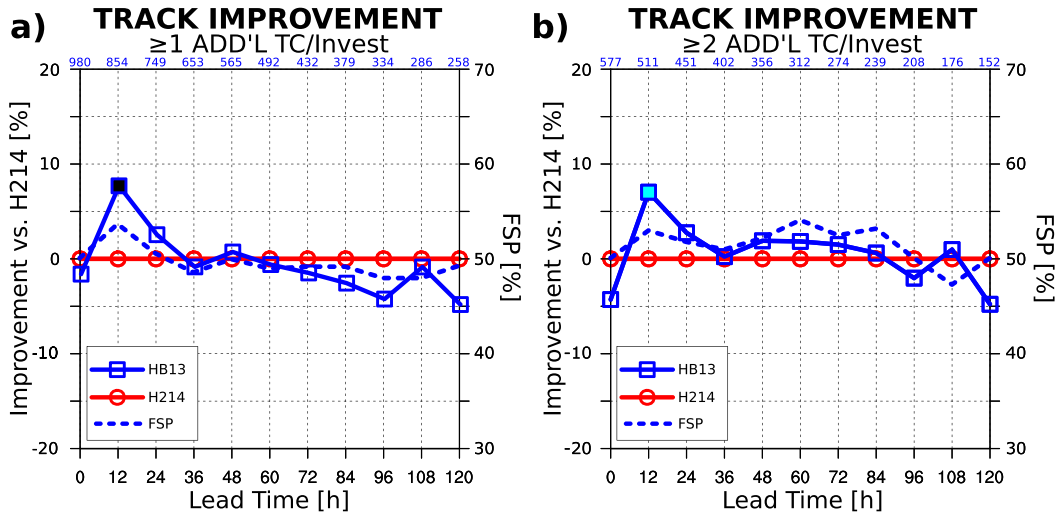

FIG. 7. As in Fig. 3b, but for (a) ≥1 additional TC/invest in the Atlantic or eastern Pacific basins and (b) ≥2 additional TCs/invests in the Atlantic or eastern Pacific basins.

centered on TC 0 (i.e., the TC being verified), has a radius of 3500 km, and is an approximate threshold for the region in which the operational HWRF can fully resolve a tropical system in its outermost domain, albeit at low resolution. Additional tropical systems outside of this 3500-km radius are considered ''far-field tropical systems'' (e.g., TC 2 and TC 3). The operational HWRF relies on the GFS boundary conditions to deliver information from far-field tropical systems to its outermost domain. A number of different radii were tested to define far-field tropical systems, with robust results between 3000 and 4000 km. Although TC 1 does not have its own MMLN in the operational HWRF, this system is still captured fully by the outermost domain and is not considered a far-field tropical system (Fig. 1b). The HWRF-B configuration fosters more natural information exchange between far-field tropical systems and the TC being verified by eliminating boundary-induced errors and by using the high-resolution MMLN for tropical systems present at the initial time.

The number of far-field tropical systems present at the verification time was used to stratify track improvements (Fig. 8). When one or more far-field tropical systems were present, ~⅔ of the forecasts were retained (Fig. 8a). HB13 track improvement at 12 h is significant at the 95% confidence level. Compared with forecast results for the full sample (Fig. 3b), where HB13 showed

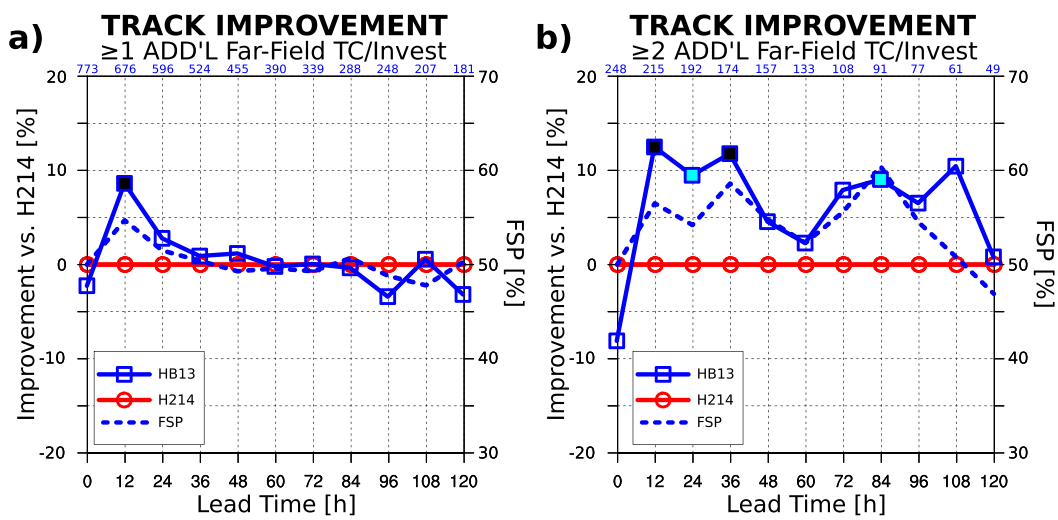

FIG. 8. As in Fig. 3b, but for (a) ≥1 additional far-field TC/invest in the Atlantic or eastern Pacific basins and (b) ≥2 additional far-field TCs/invests in the Atlantic or eastern Pacific basins.

degradation for almost all lead times after 24 h, HB13 showed less track degradation and even slight track improvement for the stratified sample at most lead times (Fig. 8a). Upon closer inspection, higher average H214 track errors (not lower average HB13 track errors) in the presence of far-field tropical systems are responsible for the increase in HB13 track improvement (not shown). When two or more far-field tropical systems were present (Fig. 8b), the sample was reduced by ~¾, and the HB13 track forecasts show improvement over H214 at all forecast times. In fact, HB13 track improvements in Fig. 8b are significant at the 90% confidence level at 12, 24, 36, and 84 h. HB13 track improvement at long lead times is considerably higher for this stratification than for the full sample (see Fig. 3b).

b. Sensitivity to outer domain location

In the operational HWRF, the outermost domain is roughly centered on each individual TC at the initial time for each forecast cycle. Therefore, the initial location of each TC can be used as a proxy for the center of the outermost domain. In Fig. 9, TC initial locations are binned into 5° × 5° tiles spanning the Atlantic basin, and the corresponding track improvements for various lead times are averaged for each tile. Differences between HB13 and H214 are then calculated for each 5° × 5° tile at 24, 72, and 120 h. Regardless of lead time, track improvement exhibits strong variability based on the longitudinal location of the H214 outermost domain. While clusters of track improvements and track degradations are highlighted below, the details reveal sharp gradients. In addition, track improvement in some tiles reflects only one TC. For example, the tiles centered on 37.5°N, 27.5°W and 32.5°N, 27.5°W include only Nadine forecasts.

At 24 h, HB13 track forecasts show improvement over H214 in the central North Atlantic, and, more generally, the east Atlantic (Fig. 9a). By 120 h, this relationship has reversed and H214 exhibits greater than 20% track improvement over HB13 for many tiles in this same region (Fig. 9c). In particular, most tiles in the central North Atlantic show track degradation at 120 h. In the Caribbean Sea, the opposite trend is observed, with HB13 track degradation (track improvement) at 24 h (120 h). Track improvements in the Gulf of Mexico exhibit sharp gradients at 24 h, with HB13 track improvement (track degradation) in the northeast and southwest (northwest and southeast) Gulf of Mexico. By 72 h, HB13 track forecasts show degradation for most of the Gulf of Mexico, but the case numbers are very low (Fig. 9b). A few tiles with a large number of cases consistently exhibit track improvement or track degradation (Fig. 9). For example, HB13 track forecasts consistently show improvement of at least 15% in the western Atlantic to the north of Hispanola (22.5°N, 72.5°W). On the other hand, HB13 track forecasts consistently show degradation of at least 10% to the east of the Lesser Antilles (17.5°N, 52.5°W).

For the most part, HB13 track forecasts demonstrate improvement in the western portion (west of 60°W) of the Atlantic basin (i.e., western Atlantic, Caribbean Sea, and Gulf of Mexico) at longer lead times (Fig. 9c). Track forecasts exhibit improvement in 10 of 15 tiles in this region and might reflect a benefit of the large HB13 outermost domain. When TCs are initialized in the western portion of the basin, surface fluxes from land (i.e., North America) may influence TC tracks. Since the H214 outermost domain may only include a portion of North America, it relies upon the GFS boundary conditions to deliver crucial information about the land surface. Conversely, HB13 includes most of North America in every forecast and may produce more realistic land–atmosphere interactions at long lead times. HB13 track forecasts show degradation in the eastern Atlantic at long lead times. In this region, the H214 outermost domain is typically situated over ocean, which limits errors related to land processes. Interestingly, when tiles are calculated for the eastern Pacific (not shown), a similar polarity in track improvements is observed, with HB13 track forecasts outperforming H214 near North America at longer lead times. Preliminary analysis of HB13 track forecasts in the eastern Pacific was presented in Zhang et al. (2016).

Although not immediately evident in Fig. 9, track forecasts are sensitive to the latitude of the outermost domain center. HB13 track forecasts show improvement of >5% after 72 h for TCs that were initialized at or north of 25°N (Fig. 11a). Track forecasts exhibit improvement at most lead times when the minimum latitude threshold is moved to 30°N (Fig. 10b), with HB13 track improvement near 20% at 96 h. Although the influences of the larger outermost domain and multiple MMLNs cannot be separated in this study, it is likely that the larger outermost domain in HB13 captures more of the critical evolution in the midlatitude synoptic environment.

In the northern portion of the Atlantic basin, ocean coupling in H214 does not appear to be as important to TC track forecasts as the synoptic-scale interactions captured by the HB13 outermost domain. TCs located in the northern portion of the Atlantic basin typically accelerate as they are steered by the midlatitude flow. A faster forward speed drastically decreases the impact of a TC on the ocean and also reduces the chance that any TC-induced changes in the ocean will impact the

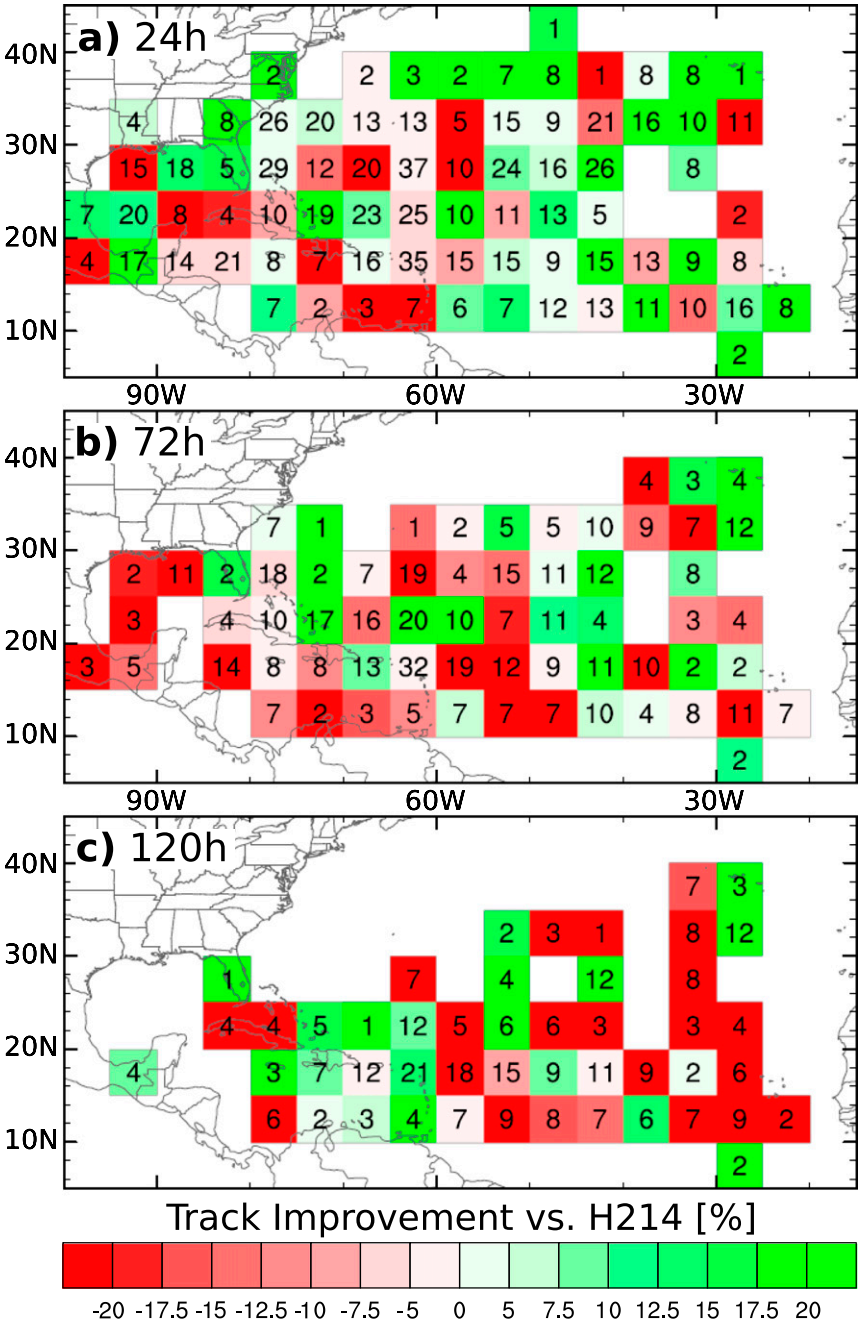

FIG. 9. Spatial variability of HB13 track improvement (%) based on TC initial location divided into 5° × 5° tiles. Track improvement is indicated by color and case numbers are shown in the center of each tile.

TC, thus reducing the importance of ocean coupling for track forecasts in those cases (Halliwell et al. 2015).

5. Hurricane Michael case study

The stratifications investigated in section 4 showed track improvement for HB13 relative to H214 in forecasts initialized in the northern part of the Atlantic basin and those with multiple tropical systems. These stratifications were used to identify interesting TC track forecasts. Several of these TC track forecasts from HB13 were investigated in detail. For example, HB13 showed significant track improvement for Hurricane Nadine, even producing more accurate track forecasts than GFS.

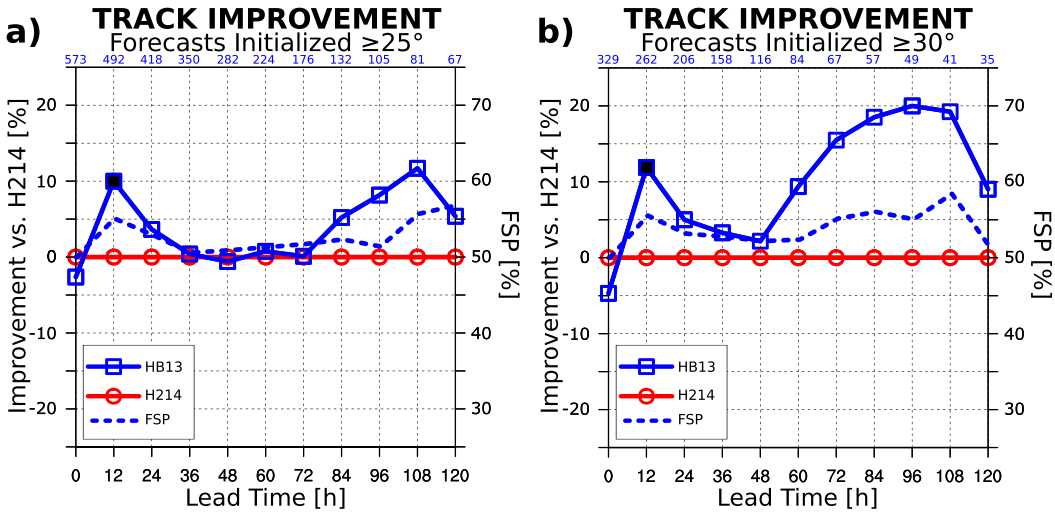

FIG. 10. As in Fig. 3b, but for TCs that are initialized at or north of (a) 25° and (b) 30°N. Note: the *y*-axis limits have changed from Fig. 3b.

Cases for Hurricane Michael clearly demonstrate the track improvement that is possible in the HWRF-B system and this improvement is discussed in greater detail here. Michael consistently coexisted with additional tropical systems and was initialized north of 25°N for every cycle in its lifetime (Kimberlain and Zelinsky 2012). Michael, only the fifth major hurricane in the satellite era with nontropical origins, moved generally northward in the North Atlantic before dissipating near 40°N (see Fig. 2) and interacted with several synoptic-scale features, including Hurricane Leslie to its west.

The forecast cycle initialized at 0600 UTC 4 September 2012 was a difficult track forecast that exemplified the two stratifications discussed in section 4. The initial latitude of Michael was 26.2°N and two additional TCs were present at the initial time: Hurricane Leslie in the Atlantic basin and Tropical Storm John in the eastern Pacific. Leslie was previously identified as a problem TC for HB13 forecasts (see section 3). John was classified as a far-field TC (>3500 km) relative to Hurricane Michael. In this forecast cycle, Michael was initialized as a minimal tropical storm and intensified by 65 kt (1 kt = 0.51 m s^{-1}) (to major hurricane status) in the next 54 h (Kimberlain and Zelinsky 2012). The track forecast was complex with Michael propagating to the northwest for the first 24 h, turning to the northeast from 24 to 72 h, and, finally, curving back to the northwest for the last 48 h of the forecast (Fig. 11a). This forecast represents HB13's lowest 120-h track error in the sample (12 km). Although the GFS 120-h track error is in excess of 200 km, this model captures the "S" shape of the track better than HB13. It should be noted that the 72-h HB13 track error is more than double that from GFS. By contrast, the H214 forecast was quite poor, with a track error of 284 km (347 km) at 72 h (120 h). H214 failed to predict the northeast propagation of Michael. Additionally, HB13 intensity errors were much lower than those from H214 (Fig. 11b). Although HB13 missed the rapid intensification of Michael between 36 and 60 h, the intensity at long lead times is very close to BEST. Conversely, H214 predicted a weak system through 90 h and never intensified Michael above 60 kt.

Geopotential height at 500 hPa (HGT500) may be employed as a first guess of the environmental steering flow to predict TC motion (Riehl et al. 1956; Pike 1985). GFSA HGT500, which represents the observations, is shown for 72- and 120-h lead times (Figs. 12a,b). A midlatitude trough was responsible for the northeast motion of Michael, which is observed downstream of this TC by 72 h. Prior to 72 h, this midlatitude trough provided a region of weak steering and low vertical wind shear, which allowed Michael to reach its maximum intensity. By 120 h, a short-wave ridge developed to the west of Michael, which had steered this TC back to the northwest.

The HGT500 fields for HB13 and H214, and differences of those fields from GFSA at the corresponding valid times, are compared to understand how these models predicted the synoptic-scale environment in the vicinity of Michael (Figs. 12c–f). North of ~40°N, HGT500 errors were similar in amplitude and location. In fact, these errors were of slightly higher amplitude in HB13, despite the better track forecast. South of ~40°N, H214 produced negative HGT500 errors (i.e., a deeper trough) northeast of Michael at 120 h (Fig. 12f).

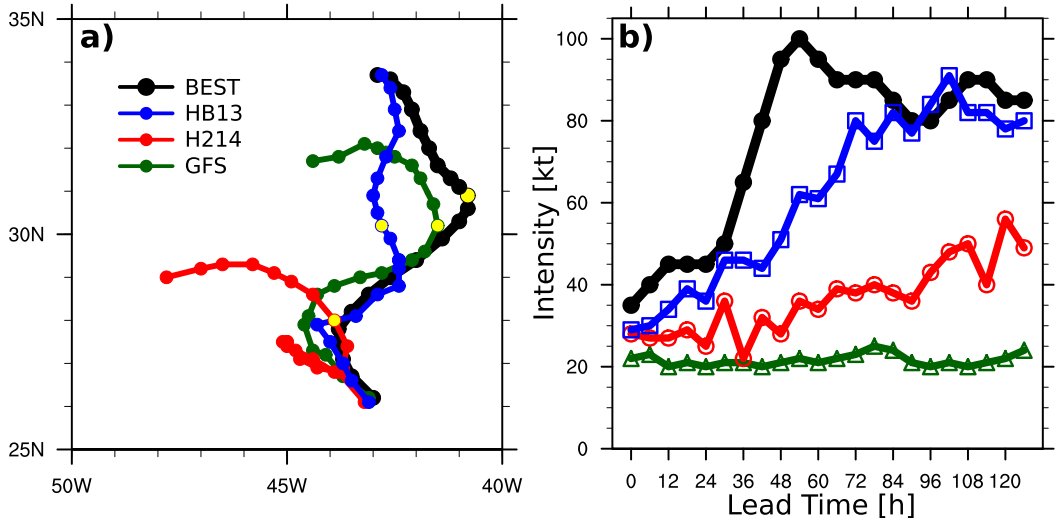

FIG. 11. (a) Track forecasts and (b) intensity forecasts for Hurricane Michael (AL132012) initialized at 0600 UTC 4 Sep 2012 for three different models: HB13 (blue), H214 (red), and GFS (green). BEST is shown in black. As a reference, 72-h locations are marked in yellow.

However, it is unclear if this feature was responsible for the large track errors associated with H214. Both HB13 and H214 exhibit negative HGT500 errors to the west of Michael in association with Leslie.

Given the difficulties involved with the interpretation of the HGT500 errors, additional tropospheric information provided by the 850–250-hPa layer-mean winds, or deep steering flow, was necessary to better understand model track forecasts. Deep-layer mean flows, used here as a proxy for TC steering (Miller 1958), at 96 h reveal consistently weak steering flow amplitudes $< 2.5\,\mathrm{m\,s^{-1}}$ in the vicinity of Michael (Fig. 13). In GFSA, Michael moved along the eastern flank of a deep-layer trough (Fig. 13a). While HB13 predicted this deep-layer trough accurately (Fig. 13b), the H214 forecast produced a cutoff trough (28°N, 47°W) to the southwest of Michael (Fig. 13c). The west-northwestward propagation of H214 is consistent with the deep steering flow. These results are robust for mean winds with a top layer $\geq$ 500 hPa.

Wu and Wang (2000) proposed a unique approach for diagnosing TC motion based on potential vorticity (PV). Following this method, TC motion is estimated using a least squares method and the wavenumber 1 PV tendency $(d\mathrm{PV}/dt)_1$. TC speed and direction predicted by 500 hPa $(d\mathrm{PV}/dt)_1$ were consistent with forecast tracks. At 96 h, the HB13 motion was $1.89\,\mathrm{m\,s^{-1}}$ at a heading of 32° and the H214 motion was $3.16\,\mathrm{m\,s^{-1}}$ at 288° (Fig. 14). The slow TC speeds were reflective of the weak steering region in which Michael was embedded, and TC directions were consistent with the deep steering flow (Fig. 13). The success of this method extends to most other lead times for H214 and HB13 (not shown).

6. Discussion and conclusions

HWRF-B, a parallel effort to the operational HWRF, shows promise as a TC-forecasting system and targets TC interactions with other TCs and with the large-scale environment. HWRF-B is configured with a large, static outermost domain and multiple sets of MMLNs, which allows for simultaneous high-resolution forecasts of several tropical systems, whereas the operational HWRF is restricted to a smaller storm-centric outermost domain and one set of MMLNs, allowing a high-resolution forecast of only one active tropical system per run. In this study, the 2013 HWRF-B (HB13) and the 2014 operational HWRF (H214) were directly compared to assess the impact of the HWRF-B enhanced configurations. Despite comparable overall mean track forecast errors to H214 in the Atlantic basin, HB13 is still preferable since a single run can simultaneously accommodate high-resolution forecasts for multiple tropical systems in the Atlantic and eastern Pacific basins. Meaningful stratifications were identified to test the impact of these configuration differences on HB13 track improvement (vs H214) and to find strengths and weaknesses of HB13 that will guide future HWRF-B developments.

Relative to H214, HB13 track forecasts are more accurate when TCs are 1) accompanied by far-field tropical systems (>3500 km from the TC being verified) and 2) initialized in the northern (north of 25°N) or western

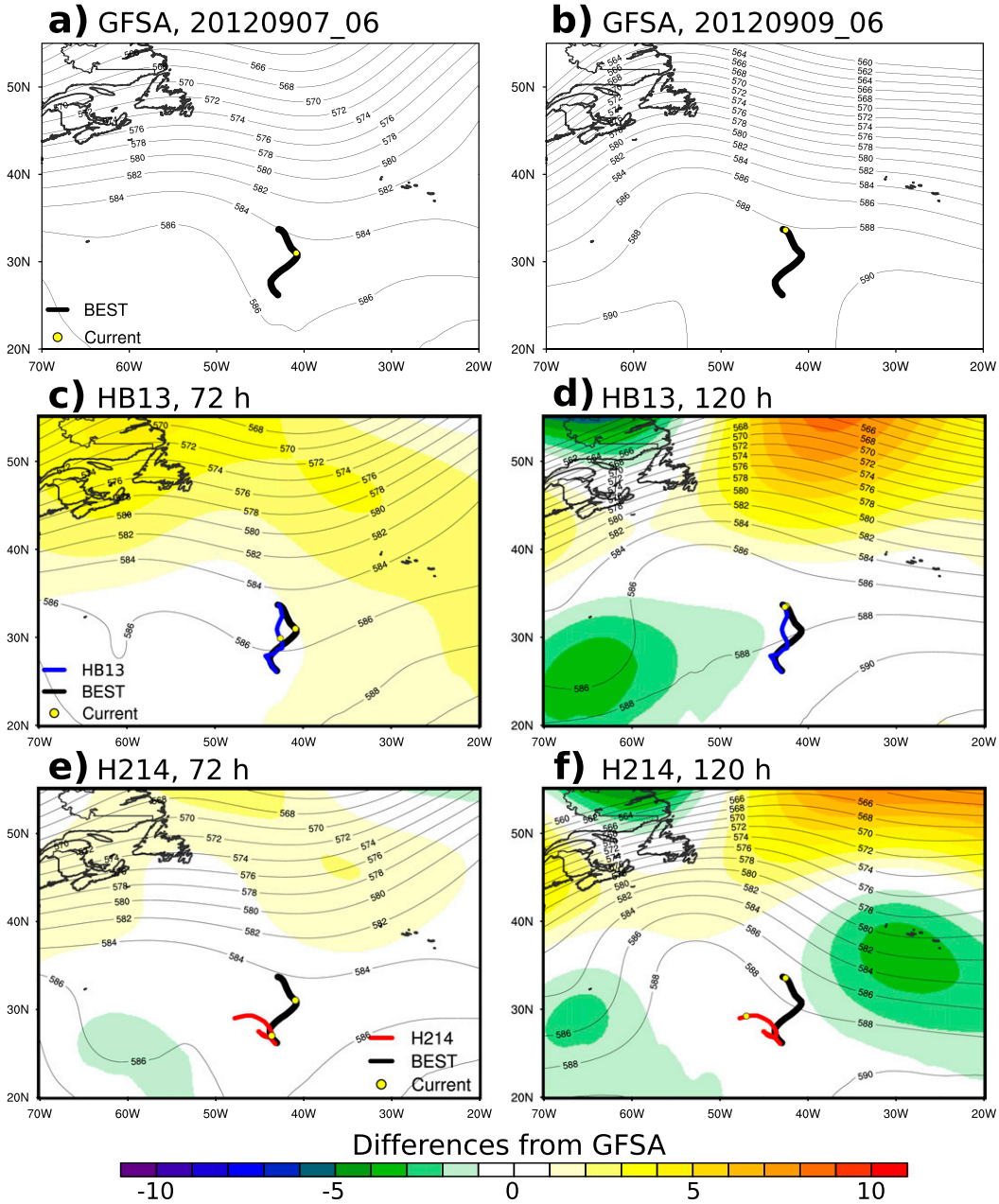

FIG. 12. GFSA HGT500 is contoured at (a) 0600 UTC 7 Sep 2012 (72 h) and (b) 0600 UTC 9 Sep 2012 (120 h). The HGT500 for HB13 is contoured and differences from GFSA are shaded (interval is 1 dam) at (c) 72 and (d) 120 h. The HGT500 for H214 is contoured and differences from GFSA are shaded at (e) 72 and (f) 120 h. HB13, H214, and BEST tracks are shown in blue, red, and black, respectively. The yellow circle corresponds to the verifying TC location at the given time.

(west of 60°W) portion of the Atlantic basin. These advantages suggest that the large outermost domain and multiple sets of MMLNs in HB13 are both important to TC track forecasts. A case study into Hurricane Michael on 0600 UTC 4 September 2012 (section 5) further supported these findings. Michael also demonstrated the importance of deep steering flow predictions, as supported HB13 and H214 forecast tracks for that particular TC. HB13 steering flows were generally consistent with GFSA, contributing to a track error of only 12 km at 120 h. Verification of steering flows in the operational HWRF and HWRF-B will be the focus of a future project. Hurricane Michael is one of several cases for which HB13 track forecasts were better than

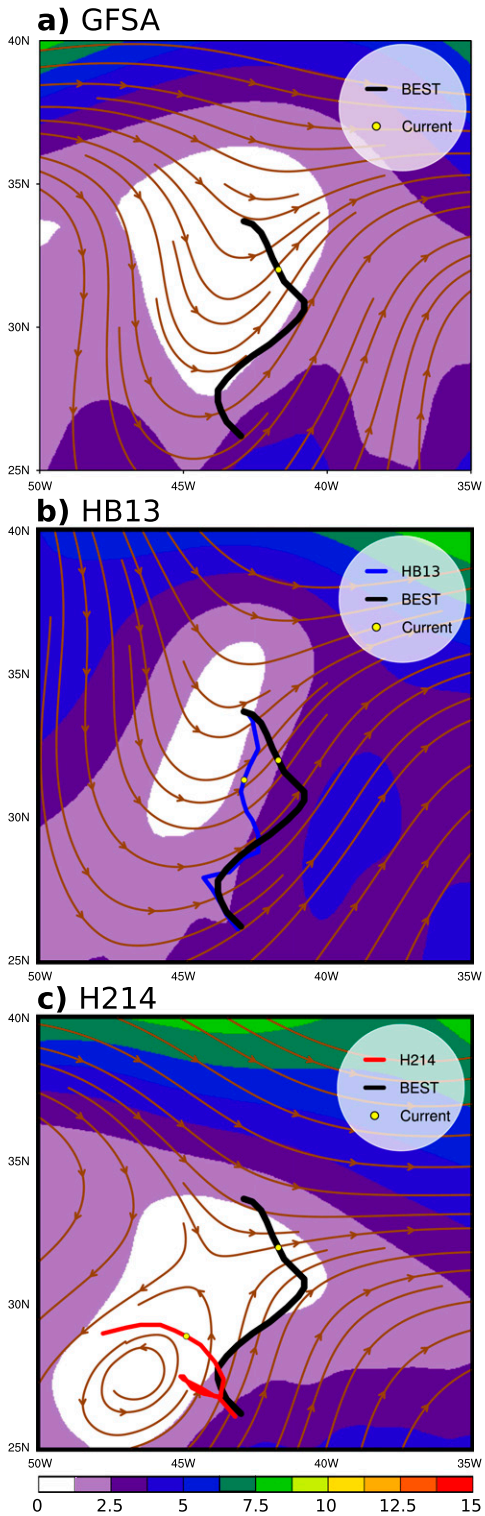

FIG. 13. Deep-layer (850–250 hPa) steering flow direction (streamlines) and amplitude (shading; interval is 1.25 m s^{-1}) at 0600 UTC 8 Sep 2012 (96 h) for (a) GFSA, (b) HB13, and (c) H214. H3HW, H214, and BEST tracks are shown in blue, red, and black, respectively. The yellow circle corresponds to the verifying TC location at the given time.

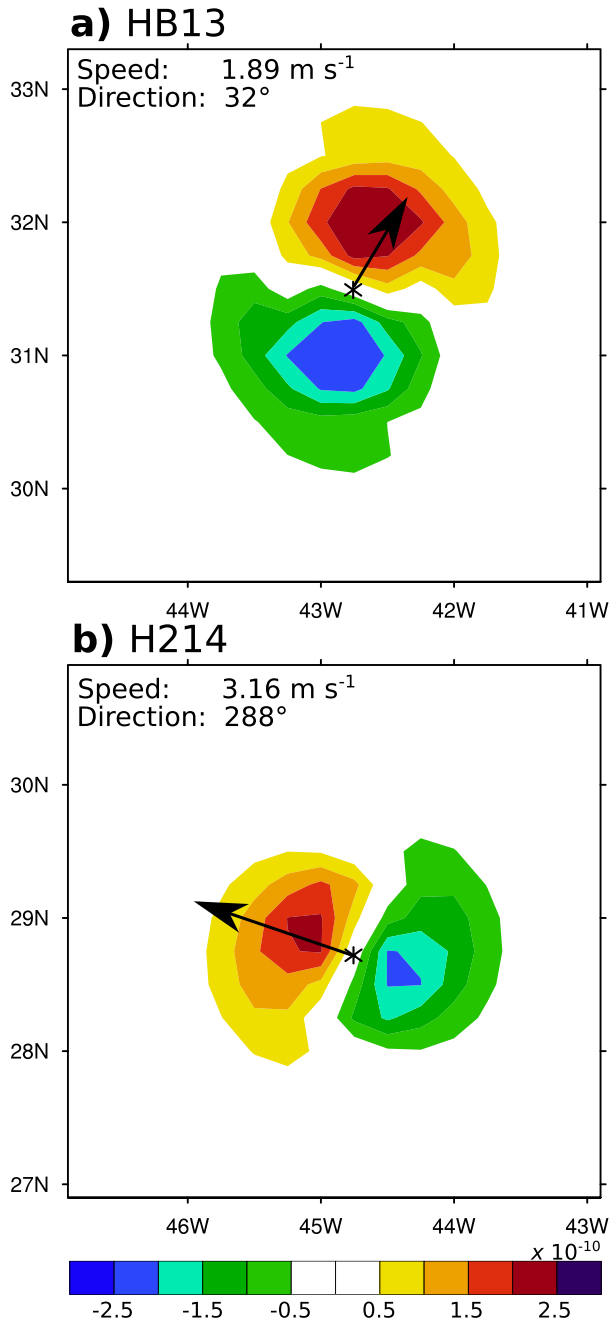

FIG. 14. Wavenumber 1 of 500-hPa PV tendency (shaded; interval is 0.5×10^{-10} m^2 s^{-2} K kg^{-1}) at 0600 UTC 8 Sep 2012 (96 h) for (a) HB13 and (b) H214. The 500-hPa center of Michael is marked by the asterisk. The black arrow represents the speed and direction of Michael, calculated as in Wu and Wang (2000).

H214 forecasts. For example, HB13 track forecast errors were lower than H214 for most Hurricane Nadine forecasts. Nadine was the longest-lived TC in the dataset and followed a very complex track. An exhaustive analysis on several case studies should be explored in a future study.

HB13 track improvement at 12 h was consistently significant at 95% and may be linked to poor intensity initialization stemming from H214 data assimilation. Additionally, HB13 forecasts demonstrated track improvement when the number of additional Atlantic/east Pacific tropical systems (nTC) ≥ 1 at the verification time. Most of this improvement was attributed to the presence of far-field tropical systems. Far-field tropical systems, which were quite common during the study period, were included (not included) in the HB13 (H214) outermost domain. Over ⅔ of all verifiable forecasts in the Atlantic basin featured at least one far-field tropical system at the verification time, and HB13 forecasts showed overall track improvement for those stratifications. Remarkably, HB13 track forecasts showed improvement of up to 14% in the presence of at least two far-field tropical systems (Fig. 8). These striking results strongly suggest an important role for remote TC–TC interactions that occur over synoptic to planetary scales and are captured by the large outermost domain of HB13. In a detailed review of TC–TC interactions, Carr et al. (1999) did not mention interactions over such large distances.

Since the results in section 4a are stratified by the verification time, there was no requirement that the far-field tropical systems existed at the initial time. Therefore, some of the far-field tropical systems were initialized with MMLN and some, since they did not exist at the initial time, were not. When stratifying by initial time, nTC does not include tropical systems that developed midforecast. Although not officially recognized at the initial time (i.e., by NHC), these tropical systems (or some signature of them) may have been apparent in the HWRF-B outermost domain during the forecast. On the other hand, any impact of these far-field tropical systems could have only been delivered to the operational HWRF via its GFS boundary conditions. The mere presence of a far-field tropical system within the outermost domain at some point during the forecast cycle appears to be responsible for HB13 track improvement versus H214.

If nTC is stratified at the initial time, an MMLN would be assigned to each of the additional tropical systems. Surprisingly, however, HB13 track improvement was near zero when considering additional MMLNs (i.e., nTC ≥ 1 at 0 h), which suggests that additional MMLNs are not necessarily important for HB13 improvement versus H214. Note that the presence of an additional tropical system at 0 h does not guarantee that it was present for the duration of the 120-h forecast. If a tropical system dissipates before the end of the forecast, the associated MMLN will disappear. Therefore, combined stratification by both initial *and* verification times was investigated, but case numbers were too small to draw conclusions, especially at long lead times. However, since the sensitivity of the HB13 track improvement to model configuration differences (see section 2) could not be independently measured, it might be premature to discount the importance of additional MMLNs to track forecasts at this time.

The precise mechanisms that govern remote TC–TC interactions are still unknown. Questions that need further research include the following: 1) How is information from a far-field tropical system exchanged over long distances during the forecast? 2) Why is this information lost or distorted at the operational HWRF lateral boundaries? 3) Do the extra sets of MMLNs make a difference, or is it more important to simply capture far-field tropical systems in the outer domain? One possible explanation for the impact of far-field tropical systems is that they cause small-amplitude fluctuations in mid- to upper-tropospheric geopotential height fields that can govern TC steering. The amplitude of these fluctuations would be similar to boundary-induced errors, which would effectively prevent a far-field tropical system from cleanly passing information through the H214 boundary. A high-resolution representation of these far-field tropical systems that develop midforecast may enhance the long-distance exchange of information. These results motivate the development of MMLNs that are able to initiate midforecast, which could further refine remote TC–TC interactions in the model.

The importance of a large outermost domain is also demonstrated by HB13 track improvement for TC forecasts initialized north of 25°N, where midlatitude flow is more likely to be a critical factor in TC track. In particular, midlatitude troughs and ridges impose rapidly evolving steering flows on TCs propagating out of the tropics in the North Atlantic Ocean. In H214, westerly flow can easily propagate through the outermost domain from the western boundary during a single 5-day forecast. However, the larger outermost domain of HB13 protects TCs from errors that originate from the model boundaries. Also, since most TCs accelerate as they approach the midlatitudes, the impact of ocean feedback, as captured only in a coupled model (i.e., H214), is likely small for TCs in the northern part of the Atlantic basin. HB13 track forecasts also generally show improvement compared with H214 when TCs were initialized in the western part of the Atlantic basin (i.e., west of 60°W), suggesting the importance of properly including the North American continent in the model domain to help track forecasts at long lead times. Since many TCs initialized in the western part of the Atlantic basin are recurving to the northeast, HB13 may have produced more realistic synoptic-scale flow for

these TCs, and, thus, more accurate track forecasts, in close proximity to North America.

Unfortunately, computing resources did not allow for the investigation into the sensitivity of HB13 track forecasts to the large outermost domain and the number of MMLNs independently. Thus, the impacts of configuration differences could not be explicitly separated in this study. However, the sensitivity of HWRF-B forecasts to these configuration options is crucial to guiding future model developments. Sensitivity experiments are under way at HRD with a later version of HWRF-B to assess the separate impacts of outermost domain size/location and the number of MMLNs. These HWRF-B control experiments will be the focus of a future study.

Although some intensity verifications for HB13 were presented in Zhang et al. (2016), more detailed intensity verifications will be the focus of future research. Four problem TCs (Gonzalo, Humberto, Leslie, and Ophelia), were identified that were associated with intensity forecast degradation at 12- or 24-h lead times and overall track degradation. For these problematic forecasts, HB13 intensity errors may have reflected vortex depth errors and TC structure errors, both of which can impact track errors. The forecasts from these four problem TCs should be examined more closely to help locate possible weaknesses (e.g., ocean coupling and data assimilation) in HB13 that could be corrected in a later version. A better understanding of model strengths (weaknesses) can also be used to communicate a priori to hurricane specialists certain situations in which the model is expected to perform optimally (poorly).

Since HB13 and H214, several upgrades were made to the HWRF system for the 2016 season (http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFv3.8a_ScientificDoc.pdf). The 2016 HWRF-B was configured with 1) data assimilation on the two innermost domains; 2) ocean coupling in both the Atlantic and eastern Pacific basins; 3) horizontal resolutions of 18, 6, and 2 km for the outermost domain and the MMLN, respectively; and 4) model physics upgrades based on HFIP community recommendations. These upgrades, combined with the core modifications elucidated in this study, are likely to further improve track and intensity forecasts in future versions of the HWRF-B system. Indeed, initial analysis of the 2016 version of HWRF-B revealed track improvements of 5%–10% over the concurrent operational HWRF.

While the eventual goal of NOAA's Next Generation Global Prediction System is to unify global models with high-resolution nests, it is imperative to understand if a larger domain may be able to improve track prediction without compromising HWRF intensity forecasts. HWRF-B is a more flexible and natural modeling system than the operational HWRF and represents a pathway forward for hurricane modeling. HWRF-B can serve as an effective bridge between the current operational HWRF system and these next-generation global model efforts at NCEP.

Acknowledgments. The authors acknowledge funding from NOAA's HFIP that supported this work. We thank EMC (especially Vijay Tallapragada, Samuel Trahan, and Zhan Zhang), which was responsible for running real-time and retrospective forecasts with H214, and DTC, which is responsible for maintaining the HWRF code. We also thank our colleagues Sim Aberson, John Kaplan, George Halliwell, and the HRD modeling team for helpful discussions that strengthened the manuscript. We are indebted to our colleagues Thiago Quirino and Russell St. Fleur for their invaluable expertise in running and managing the output for HB13 forecasts. We appreciate the comments from Morris Bender and two anonymous reviewers, which helped to strengthen this manuscript.

REFERENCES

Aberson, S. D., 2010: 10 years of hurricane synoptic surveillance (1997–2006). *Mon. Wea. Rev.*, **138**, 1536–1549, doi:10.1175/2009MWR3090.1.

——, 2011: The impact of dropwindsonde data from the THORPEX Pacific Area Regional Campaign and the NOAA Hurricane Field Program on tropical cyclone forecasts in the Global Forecast System. *Mon. Wea. Rev.*, **139**, 2689–2703, doi:10.1175/2011MWR3634.1.

——, and M. DeMaria, 1994: Verification of a nested barotropic hurricane track forecast model (VICBAR). *Mon. Wea. Rev.*, **122**, 2804–2815, doi:10.1175/1520-0493(1994)122<2804:VOANBH>2.0.CO;2.

Atlas, R., V. Tallapragada, and S. G. Gopalakrishnan, 2015: Advances in tropical cyclone intensity forecasts. *Mar. Technol. J.*, **49**, 149–160, doi:10.4031/MTSJ.49.6.2.

Bender, M. A., and I. Ginis, 2000: Real-case simulations of hurricane–ocean interaction using a high-resolution coupled model: Effects on hurricane intensity. *Mon. Wea. Rev.*, **128**, 917–946, doi:10.1175/1520-0493(2000)128<0917:RCSOHO>2.0.CO;2.

Brand, S., 1970: Interaction of binary cyclones of the western North Pacific Ocean. *J. Appl. Meteor.*, **9**, 433–441, doi:10.1175/1520-0450(1970)009<0433:IOBTCO>2.0.CO;2.

Cangialosi, J. P., and J. L. Franklin, 2016: 2015 hurricane season. National Hurricane Center Forecast Verification Rep., 69 pp. [Available online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2015.pdf.]

Carr, L. E., R. L. Elsberry, and M. A. Booth, 1999: Condensed and updated version of the Systematic Approach Meteorological Knowledge Base western North Pacific. Naval Postgraduate School Rep. NPS-MR-00-001, 140 pp. [Available online at http://hdl.handle.net/10945/28702.]

Fels, S. B., and M. D. Schwarzkopf, 1975: The simplified exchange approximation: A new method for radiative transfer calculations. *J. Atmos. Sci.*, **32**, 1475–1488, doi:10.1175/1520-0469(1975)032<1475:TSEAAN>2.0.CO;2.

Ferrier, B. S., 2005: An efficient mixed-phase cloud and precipitation scheme for use in operational NWP models. *Eos, Trans. Amer. Geophys. Union*, **86** (Spring Meeting Suppl.), Abstract A42A-02.

Franklin, J. L., S. E. Feuer, J. Kaplan, and S. D. Aberson, 1996: Tropical cyclone motion and surrounding flow relationships: Searching for beta gyres in omega dropwindsonde datasets. *Mon. Wea. Rev.*, **124**, 64–84, doi:10.1175/1520-0493(1996)124<0064:TCMASF>2.0.CO;2.

Fujiwhara, S., 1921: The mutual tendency towards symmetry of motion and its application as a principle in meteorology. *Quart. J. Roy. Meteor. Soc.*, **47**, 287–292, doi:10.1002/qj.49704720010.

Gall, R., J. Franklin, F. Marks, E. N. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project. *Bull. Amer. Meteor. Soc.*, **94**, 329–343, doi:10.1175/BAMS-D-12-00071.1.

Goldenberg, S. B., S. G. Gopalakrishnan, V. Tallapragada, T. Quirino, F. Marks, S. Trahan, X. Zhang, and R. Atlas, 2015: The 2012 triply nested, high-resolution operational version of the Hurricane Weather Research and Forecasting Model (HWRF): Track and intensity forecast verifications. *Wea. Forecasting*, **30**, 710–729, doi:10.1175/WAF-D-14-00098.1.

Gopalakrishnan, S. G., Q. Liu, T. Marchok, D. Sheinin, N. Surgi, R. Tuleya, R. Yablonsky, and X. Zhang, 2010: Hurricane Weather Research and Forecasting (HWRF) Model scientific documentation. NOAA/NCAR/Development Testbed Center Rep., 75 pp. [Available online at http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRF_final_2-2_cm.pdf.]

——, F. Marks, 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, doi:10.1175/2010MWR3535.1.

——, S. Goldenberg, T. Quirino, X. Zhang, F. Marks, K.-S. Yeh, R. Atlas, and V. Tallapragada, 2012: Toward improving high-resolution numerical hurricane forecasting: Influence of model horizontal grid resolution, initialization, and physics. *Wea. Forecasting*, **27**, 647–666, doi:10.1175/WAF-D-11-00055.1.

——, F. Marks, J. A. Zhang, X. Zhang, J.-W. Bao, and V. Tallapragada, 2013: A study of the impacts of vertical diffusion on the structure and intensity of the tropical cyclones using the high-resolution HWRF system. *J. Atmos. Sci.*, **70**, 524–541, doi:10.1175/JAS-D-11-0340.1.

——, and Coauthors, 2016: 2015 HFIP R&D activities summary: Recent results and operational implementation. HFIP Tech. Rep. HFIP2016-1, 44 pp. [Available online at http://www.hfip.org/documents/HFIP_AnnualReport_FY2015.pdf.]

Gray, W. M., 1998: The formation of tropical cyclones. *Meteor. Atmos. Phys.*, **67**, 37–69, doi:10.1007/BF01277501.

Halliwell, G. R., Jr., S. Gopalakrishnan, F. Marks, and D. Willey, 2015: Idealized study of ocean impacts on tropical cyclone intensity forecasts. *Mon. Wea. Rev.*, **143**, 1142–1165, doi:10.1175/MWR-D-14-00022.1.

Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2339, doi:10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2.

Kimberlain, T. B., and D. A. Zelinsky, 2012: Tropical cyclone report: Hurricane Michael (AL132012), 3–11 September 2012. National Hurricane Center, 13 pp. [Available online at http://www.nhc.noaa.gov/data/tcr/AL132012_Michael.pdf.]

Kurihara, Y., and R. E. Tuleya, 1974: Structure of a tropical cyclone developed in a three-dimensional numerical simulation model. *J. Atmos. Sci.*, **31**, 893–919, doi:10.1175/1520-0469(1974)031<0893:SOATCD>2.0.CO;2.

——, M. A. Bender, and R. J. Ross, 1993: An initialization scheme of hurricane models by vortex specification. *Mon. Wea. Rev.*, **121**, 2030–2045, doi:10.1175/1520-0493(1993)121<2030:AISOHM>2.0.CO;2.

Lacis, A. A., and J. E. Hansen, 1974: A parameterization for absorption of solar radiation in the earth's atmosphere. *J. Atmos. Sci.*, **31**, 118–133, doi:10.1175/1520-0469(1974)031<0118:APFTAO>2.0.CO;2.

Lander, M., and G. J. Holland, 1993: On the interaction of tropical-cyclone-scale vortices. I: Observations. *Quart. J. Roy. Meteor. Soc.*, **119**, 1347–1361, doi:10.1002/qj.49711951406.

Miller, B. I., 1958: The use of mean layer winds as hurricane steering mechanism. National Hurricane Research Project Rep. 18, 24 pp. [Available online at http://www.aoml.noaa.gov/general/lib/TM/NHRP_18_1958.pdf.]

Pan, H.-L., and W.-S. Wu, 1995: Implementing a mass flux convection parameterization package for the NMC medium-range forecast model. NMC Office Note 409, 43 pp. [Available online at http://www2.mmm.ucar.edu/wrf/users/phys_refs/CU_PHYS/Old_SAS.pdf.]

Pike, A. C., 1985: Geopotential heights and thicknesses as predictors of Atlantic tropical cyclone motion and intensity. *Mon. Wea. Rev.*, **113**, 931–940, doi:10.1175/1520-0493(1985)113<0931:GHATAP>2.0.CO;2.

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

Riehl, H., W. H. Haggard, and R. W. Sanborn, 1956: On the prediction of 24-hour hurricane motion. *J. Meteor.*, **13**, 415–420, doi:10.1175/1520-0469(1956)013<0415:OTPOHH>2.0.CO;2.

Schwarzkopf, M. D., and S. B. Fels, 1991: The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes. *J. Geophys. Res.*, **96**, 9075–9096, doi:10.1029/89JD01598.

Sirutis, J., and K. Miyakoda, 1990: Subgrid scale physics in 1-month forecasts. Part I: Experiment with four parameterization packages. *Mon. Wea. Rev.*, **118**, 1043–1064, doi:10.1175/1520-0493(1990)118<1043:SSPIMF>2.0.CO;2.

Susca-Lopata, G., J. Zawislak, E. J. Zipser, and R. F. Rogers, 2015: The role of observed environmental conditions and precipitation evolution in the rapid intensification of Hurricane Earl (2010). *Mon. Wea. Rev.*, **143**, 2207–2223, doi:10.1175/MWR-D-14-00283.1.

Tallapragada, V., and Coauthors, 2014a: Hurricane Weather Research and Forecasting (HWRF) model: 2014 scientific documentation. Developmental Testbed Center Rep., 105 pp. [Available online at http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFv3.6a_ScientificDoc.pdf.]

——, C. Kieu, Y. Kwon, S. Trahan, Q. Liu, Z. Zhang, and I.-H. Kwon, 2014b: Evaluation of storm structure from the operational HWRF Model during 2012 implementation. *Mon. Wea. Rev.*, **142**, 4308–4325, doi:10.1175/MWR-D-13-00010.1.

Tuleya, R. E., 1994: Tropical storm development and decay: Sensitivity to surface boundary conditions. *Mon. Wea. Rev.*, **122**, 291–304, doi:10.1175/1520-0493(1994)122<0291:TSDADS>2.0.CO;2.

Velden, C. S., and S. B. Goldenberg, 1987: The inclusion of high density satellite wind information in a barotropic hurricane-track forecast model. Preprints, *17th Conf. on Hurricanes and Tropical Meteorology*, Houston, TX, Amer. Meteor. Soc., 90–93.

Vukicevic, T., A. Aksoy, P. Reasor, S. D. Aberson, K. J. Sellwood, and F. Marks, 2013: Joint impact of forecast tendency and state

error biases in ensemble Kalman filter data assimilation of inner-core tropical cyclone observations. *Mon. Wea. Rev.*, **141**, 2992–3006, doi:10.1175/MWR-D-12-00211.1.

Wu, L., and B. Wang, 2000: A potential vorticity tendency diagnostic approach for tropical cyclone motion. *Mon. Wea. Rev.*, **128**, 1899–1911, doi:10.1175/1520-0493(2000)128<1899:APVTDA>2.0.CO;2.

Yablonsky, R. M., and I. Ginis, 2009: Limitation of one-dimensional ocean models for coupled hurricane–ocean model forecasts. *Mon. Wea. Rev.*, **137**, 4410–4419, doi:10.1175/2009MWR2863.1.

——, ——, B. Thomas, V. Tallapragada, D. Sheinin, and L. Bernardet, 2015: Description and analysis of the ocean component of NOAA's operational Hurricane Weather Research and Forecasting Model (HWRF). *J. Atmos. Oceanic Technol.*, **32**, 144–163, doi:10.1175/JTECH-D-14-00063.1.

Yeh, K., X. Zhang, S. Gopalakrishnan, S. Aberson, R. Rogers, F. Marks, and R. Atlas, 2012: Performance of the experimental HWRF in the 2008 hurricane season. *Nat. Hazards*, **63**, 1439–1449, doi:10.1007/s11069-011-9787-7.

Zhang, X., S. G. Gopalakrishnan, S. Trahan, T. S. Quirino, Q. Liu, Z. Zhang, G. J. Alaka, and V. Tallapragada, 2016: Representing multiple scales in the Hurricane Weather Research and Forecasting modeling system: Design of multiple sets of movable multilevel nesting and the basin-scale HWRF forecast application. *Wea. Forecasting*, **31**, 2019–2034, doi:10.1175/WAF-D-16-0087.1.

Zou, X., F. Weng, B. Zhang, L. Lin, Z. Qin, and V. Tallapragada, 2013: Impacts of assimilation of ATMS data in HWRF on track and intensity forecasts of 2012 four landfall hurricanes. *J. Geophys. Res. Atmos.*, **118**, 11 558–11 576, doi:10.1002/2013JD020405.