Forecasting Tropical Cyclones in the Western North Pacific Basin Using the NCEP Operational HWRF Model: Model Upgrades and Evaluation of Real-Time Performance in 2013

VIJAY TALLAPRAGADA,[a] CHANH KIEU,[a,b,d] SAMUEL TRAHAN,[a,b] QINGFU LIU,[a] WEIGUO WANG,[a,b] ZHAN ZHANG,[a,b] MINGJING TONG,[a,b] BANGLIN ZHANG,[a,b] LIN ZHU,[a,b] AND BRIAN STRAHL[c]

[a] *NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland*
[b] *I. M. Systems Group, Rockville, Maryland*
[c] *Joint Typhoon Warning Center, Pearl Harbor, Hawaii*
[d] *Atmospheric Science, Department of Geological Sciences, Indiana University, Indiana*

(Manuscript received 22 October 2014, in final form 19 October 2015)

ABSTRACT

This study presents evaluation of real-time performance of the National Centers for Environmental Prediction (NCEP) operational Hurricane Weather Research and Forecast (HWRF) modeling system upgraded and implemented in 2013 in the western North Pacific basin (WPAC). Retrospective experiments with the 2013 version of the HWRF Model upgrades for 2012 WPAC tropical cyclones (TCs) show significant forecast improvement compared to the real-time forecasts from the 2012 version of HWRF. Despite a larger number of strong storms in the WPAC during 2013, real-time forecasts from the 2013 HWRF (H213) showed an overall reduction in intensity forecast errors, mostly at the 4–5-day lead times. Verification of the H213's skill against the climate persistence forecasts shows that although part of such improvements in 2013 is related to the different seasonal characteristics between the years 2012 and 2013, the new model upgrades implemented in 2013 could provide some further improvement that the 2012 version of HWRF could not achieve. Further examination of rapid intensification (RI) events demonstrates noticeable skill of H213 with the probability of detection (POD) index of 0.22 in 2013 compared to 0.09 in 2012, suggesting that H213 starts to show skill in predicting RI events in the WPAC.

1. Introduction

In support of the operational tropical cyclone (TC) forecasts in the western North Pacific basin (WPAC) for the U.S. Joint Typhoon Warning Center (JTWC), the Environmental Modeling Center (EMC) at the National Centers for Environmental Prediction (NCEP) began providing experimental real-time TC forecasts for the WPAC in 2012, using the operational high-resolution triple-nested Hurricane Weather Research and Forecast (HWRF) modeling system (Tallapragada et al. 2015, hereinafter T15). Evaluation of the HWRF model's performance in 2012 showed lower forecast errors from the HWRF model as compared to other operational regional models currently used by JTWC (T15; Evans and Falvey 2013). Specifically, statistics of the real-time track forecast errors showed that the 2012 HWRF (hereinafter H212) achieved the 3-, 4- and 5-day track errors of ~125, 220, and 290 n mi (1 n mi = 1.852 km), respectively, as reported in T15. Intensity forecasts from H212 also showed improved performance as compared to other regional models with much reduced forecast errors during the first 24 h owing to better vortex initialization (Tallapragada et al. 2014a; Gopalaknishnan et al. 2012).

Even though H212 showed reasonably good performance in its maiden effort, several issues were noted during the evaluation of real-time WPAC TC forecasts in 2012. First, H212 forecasts tend to overestimate storm intensity for initially strong storms, which are defined as storms with the initial maximum 10-m wind greater than 50 kt (1 kt = 0.51 m s^{-1}), and underestimate storm intensity for initially weak storms. This issue is seen not

Corresponding author address: Dr. Vijay Tallapragada, NOAA/NWS/NCEP/Environmental Modeling Center, College Park, MD 20740.
E-mail: vijay.tallapragada@noaa.gov

DOI: 10.1175/WAF-D-14-00139.1

only in the WPAC but also in the eastern North Pacific basin (EPAC) and the North Atlantic basin (NATL) (Tallapragada et al. 2014a; Goldenberg et al. 2015). While the overestimation of storm intensity for the stronger storms could be attributed partly to the uncoupled version of the HWRF model in the WPAC (Kim et al. 2014), H212's underestimation of the intensity of weaker storms is worth noting as the model sometimes did not capture the storm development even in favorable conditions for intensification, such as warm sea surface temperatures (SSTs) or areas of weak vertical wind shear. One possible reason for such underestimation of initially weak storms is related to the vortex tracking algorithm that determines how the model innermost domain (3-km nest) moves in time (Trahan et al. 2014). Retrospective evaluation of the HWRF forecasts for weak storms during the 2010–12 hurricane seasons in the NATL and EPAC revealed that H212 frequently tracked an incorrect vortex center if the model storms were too weak or there was another low pressure system nearby, as the vortex tracking algorithm based on the centroid method was not optimally designed to address such situations. Occasionally, the moving nest was also found to be latched on to a low pressure center over high topography near the vicinity of the storm (Trahan et al. 2014). As a result, the model's innermost domain could not follow the model storm properly, thus limiting the potential upscale growth of the disturbances during their early stages of formation.

The second issue with H212 is related to the representation of the large-scale environmental circulation in the WPAC (see T15). Analysis of the track and intensity error distribution over the WPAC showed that H212 appears to have both track and intensity biases related to the environmental steering flows associated with weaker strength of the western Pacific subtropical high (WPSH) in the H212 forecasts. In particular, there are some distinct patterns in the along-track and cross-track errors, with significant left (right) cross-track bias in the higher (lower) latitudes on the edge of the WPSH. In the East China Sea, H212 has slower translational speed (with respect to the observed tracks), which may explain the positive intensity bias due to possibly longer exposure of the model storms over the warm ocean in this area (Zeng et al. 2007; Kim et al. 2014).

Given the positive performance of the HWRF model in the first real-time experiment in the WPAC in 2012, the HWRF team at EMC continued their efforts to provide real-time experimental TC forecasts in the WPAC to JTWC in 2013, using the upgraded version of the HWRF model implemented operationally at NCEP in 2013. Motivated by the successful demonstration of forecast improvements from retrospective experiments of the 2013 HWRF configuration (hereinafter H213) for the 2010–12 seasons in the NATL and EPAC (Tallapragada et al. 2013), as well as strong support from National Oceanic and Atmospheric Administration's (NOAA) Hurricane Forecast Improvement Project (HFIP; Gall et al. 2013), H213 was customized and implemented in real time for the WPAC in July 2013. In this study, EMC's efforts in implementing the 2013 HWRF upgrades for a real-time forecasting system and the model performance during the period from July to December 2013 are documented. One of our main objectives is to identify the strengths and weaknesses of the 2013 HWRF model in the WPAC region as compared to the 2012 HWRF, and to compare the real-time forecast performance with other operational models used by JTWC, which could offer further insights into the behavior of the HWRF model and provide better understanding of the forecast characteristics of TCs in the WPAC region for future improvements.

This paper is organized as follows. In section 2, an overview of the 2013 upgrades of the operational HWRF modeling system (H213), and the real-time setup for the WPAC is provided. Section 3 presents detailed forecast verification of H213 for the WPAC with specific focus on the overall track and intensity error statistics, as well as the rapid intensification forecasts. Some concluding remarks and future work are given in the final section.

2. The 2013 HWRF model upgrades and real-time experimental configuration

Several major upgrades to the operational HWRF model were implemented at NCEP for the 2013 NATL and EPAC hurricane seasons (Tallapragada et al. 2013), following an extensive evaluation of retrospective forecasts for three NATL and EPAC hurricane seasons (2010–12). Table 1 lists the details of H212 and the 2013 HWRF upgrades. Similar to H212, the basic modeling framework H213 was based on the WRF-NMM dynamical core developed at NCEP (Janjić et al. 2010). The vertical resolution of H213 remained the same as H212 with 43 levels and a 50-hPa model top. The outermost domain of H213 was configured at a horizontal resolution of 0.18° (roughly 27 km) that covers an area of about 80° × 80° and the intermediate domain at a resolution of 0.06° (~9 km), but the innermost domain at a resolution of 0.02° (~3 km) had its horizontal domain size extended from about 5.5° × 5° to about 6.5° × 6°. This larger innermost domain size allowed H213 to capture more of the storm central region, especially for large storms for which the gale-force winds could be sometimes as far as 800 km from the storm center. The innermost domain was allowed to move after every

TABLE 1. The 2012 HWRF model configuration and the 2013 HWRF upgrades for the 2012–13 real-time experiments in WPAC.

Scheme	2012 HWRF (H212)	2013 HWRF (W213/H213)
Model horizontal resolution	27/9/3 km (E grid)	27/9/3 km (E grid)
Dynamic core	WRF-NMM (version 3.4a)	WRF-NMM (version 3.5)
Model domain	216 × 432 (27-km domain), 88 × 170 (9-km domain), and 154 × 272 (3-km domain)	216 × 432 (27-km domain), 88 × 170 (9-km domain), and 180 × 324 (3-km domain)
No. of vertical levels	43	43
Cumulus parameterization	GFS convection parameterization	GFS convection parameterization
Microphysics	Ferrier microphysics parameterization	Ferrier microphysics parameterization
Boundary layer	Modified GFS PBL scheme	Modified GFS PBL scheme with variable critical Richardson number (Vickers and Mahrt 2003)
Surface physics	Improved GFDL surface physics	Improved GFDL surface physics
Radiation	GFDL short- and longwave schemes	GFDL short- and longwave schemes
Lateral boundary conditions	GFS updated every 6 h	GFS updated every 6 h
Physics time step (except radiation)	3 min	30 s
Radiation time step	60 min	30 min
Vortex tracking algorithm	Mass-weighted centroid method	NCEP vortex tracker (Trahan et al. 2014)

three time steps so that its center could follow the storm's center at all times. Both the initial and boundary conditions were taken from the NCEP Global Forecast System (GFS) analysis generated by the global ensemble-based hybrid data assimilation system and its 126-h forecasts obtained directly from the spectral coefficients at T574L64 resolution (about 27 km near the equator and 64 vertical levels), and the boundary conditions are updated at 6-h intervals.

There were a number of other important infrastructural and scientific upgrades implemented in H213. First, the storm tracking algorithm was completely revised based on eight different variables to locate the storm center in the model (Trahan et al. 2014), similar to the NCEP TC vortex tracking algorithm, instead of using the centroid method implemented in H212. This helped eliminate several issues with the centroid approach where the innermost domain occasionally followed incorrect vortex centers when model storms were too weak, or when there was a nearby low pressure system or high topography. Second, the nest–parent interpolation was redesigned to improve the treatment of interpolation at nest boundaries and upscale feedback (Trahan et al. 2014). The new interpolation scheme allowed for more efficient computations at the nest–parent interface and for advection of different microphysics species that the old interpolation scheme could not handle. Third, the frequency of physics calls was increased from every 180 s in H212 to every 30 s in the H213 configuration. Such a need for higher frequency of the model physics calls has long been known to be critical, but it was not implemented in H212 because of operational computational resource constraints at the NCEP supercomputer. The 30-s frequency of the physics calls was made affordable in H213 mostly because the new nest–parent interpolation significantly sped up the model integration, and was considered to be one of the key enhancements for improving the realism of the model physics that has long been advocated by the research community.

Along with these major infrastructure changes, H213 also had several physics upgrades including modifications to the GFS planetary boundary layer (PBL) parameterization scheme based on observational findings (Gopalakrishnan et al. 2013; Zhang et al. 2015) to allow the critical Richardson number to vary with the PBL stability and wind speeds. Instead of using a constant number, this new modification allows the critical Richardson number to be a function of the 10-m wind according to the formulation by Vickers and Mahrt (2003). This change helped account for the spatial variation of the boundary layer height and is associated with a shallower boundary layer near the storm region, as compared to that over land or in the region with lower 10-m winds (Kwon et al. 2014). A bug fix related to the ice cloud water contents in the Geophysical Fluid Dynamics Laboratory (GFDL) radiation scheme was also made, although the impacts related to this bug in the model track and intensity forecast process were insignificant. Another change in the H213 configuration was an implementation of a new hybrid data assimilation component, but this data assimilation system was not applied in the real-time experiments in the WPAC basin during 2012–13 because of data insufficiency and the lack of adequate testing. In addition to disabling the GSI component, no ocean coupling was implemented in the WPAC basin because of a lack of real-time global oceanic forecasts. All experiments in the WPAC using the H213 configuration were performed on NOAA HFIP supercomputers (Jet systems) in Boulder,

Colorado, employing dedicated reservations, and forecast guidance products were disseminated to JTWC through specially established channels, accomplishing about 90% on-time delivery.

To reduce the issues with the rapid spinup/spindown of the model vortex during the first 6 h of model integration, the vortex initialization component was modified in H213 to take into account a more accurate storm size correction. Specifically, the filter domain was increased to smoothly blend the vortex-scale analysis with the environmental flow of the global analysis in the parent domain. In addition, the vortex initialization process was modified to use the initial vortex directly from the GFS analysis instead of using the bogus vortex structure when storms were weaker than 16 m s^{-1}. More details of the vortex initialization changes can be found in Tallapragada et al. (2013).

3. Performance of H213 in the WPAC during 2013

a. *Track and intensity forecast verification*

Before implementing H213 for real-time forecasts in the WPAC, retrospective forecast experiments for all 2012 WPAC TCs were conducted using the H213 configuration to demonstrate the performance of the 2013 HWRF upgrades relative to the real-time HWRF forecasts in 2012. Figure 1 shows homogeneous verification of the track, the absolute intensity, and the intensity bias for all cases during the real-time experimental period in 2012. The 2013 HWRF model (denoted by W213 in Fig. 1) showed some significant improvements in the track forecasts with the most noticeable error reductions at 4- and 5-day lead times, with the 5-day track forecast errors (Fig. 1a) reduced from 270 to ~185 n mi (~30% improvement). Although there is little improvement in the absolute intensity errors (Fig. 1b), note that the intensity biases shown in Fig. 1c have indicated most of the negative intensity bias has been reduced at 0–72 h forecast lead times. Except for the slight positive bias at 12-h lead time due to the model tendency to overintensify most of the weak systems, the overall intensity bias reduction at longer lead times is realized not only in the WPAC but also in the NATL and EPAC (Tallapragada et al. 2013), which is attributed mostly to the improved vortex initialization, a better vortex tracking algorithm, and model physics upgrades in the 2013 HWRF implementation (Kwon 2014; Trahan et al. 2014). Note that the improvements in both the track and intensity forecasts during the retrospective experiments for 2012 TCs are due solely to the HWRF model upgrades, because the same global GFS analysis and forecasts are used to provide initial and lateral boundary conditions for the retrospective experiments. In fact, the reduction of the track errors in the W213 experiments is significantly larger than what was obtained in either the NATL or EPAC, despite using the same 2013 HWRF model upgrades (the 5-day track error in the W213 experiment is ~205 n mi, whereas the HWRF 5-day track error using a similar model configuration is ~210 n mi in the NATL (about 10% reduction with respect to H212), and ~180 n mi in the EPAC basin (about 15% reduction with respect to H212; cf. Fig. 1). While the improvement is different in different basins, noticeable reductions in track and intensity bias errors could at least serve as justification for NCEP to continue supporting the execution of H213 experiments for real-time forecasts in the WPAC during 2013.

To see the relative performance improvements of the HWRF model forecasts from 2013 compared to 2012, nonhomogeneous statistics of the absolute TC track forecast errors, the absolute intensity errors, and the intensity bias obtained from real-time forecast experiments of H212 and H213 during the experimental period[1] were also provided in Fig. 1. Here, all forecast errors are computed with respect to the postseason best-track datasets provided by JTWC. One notices significantly smaller track and intensity forecast errors in H213 as compared to H212 at all forecast lead times. The track error reduction from H213 (Fig. 1a) is most apparent at the 5-day lead time, which decreases by ~15% from 275 n mi in 2012 to about 235 n mi in 2013. The decrease in intensity forecast errors (Fig. 1b) is even more noticeable with roughly a 41% reduction at 5-day lead time from 27 to 16 kt. The intensity bias (Fig. 1c) also shows that the H213 intensity biases (red columns) between the 24- and 96-h lead times were reduced relative to the H212 biases (blue columns), except at the 5-day lead time for which H213 possesses a positive intensity bias. Such a positive bias at the 5-day lead time could be related to several factors such as landfalling storms where a slight difference in the landfall time or location could have a significant influence on the intensity bias (cf. Fig. 6c for an example of a negative along-track error as well as the corresponding positive intensity errors in Fig. 6).

Of interest is that the intensity error reduction of the H213 configuration in the WPAC basin is more substantial in terms of intensity bias errors and intensity skill (Figs. 1c and 2b) than that reported for NATL and EPAC in 2013 (Tallapragada et al. 2014b). Given the

[1] In general, the WPAC has TCs year round, and there is no official definition for the TC season in this basin. However, for the sake of discussion, the experimental period referred to in this work is from May to December 2012 and from July to December 2013.

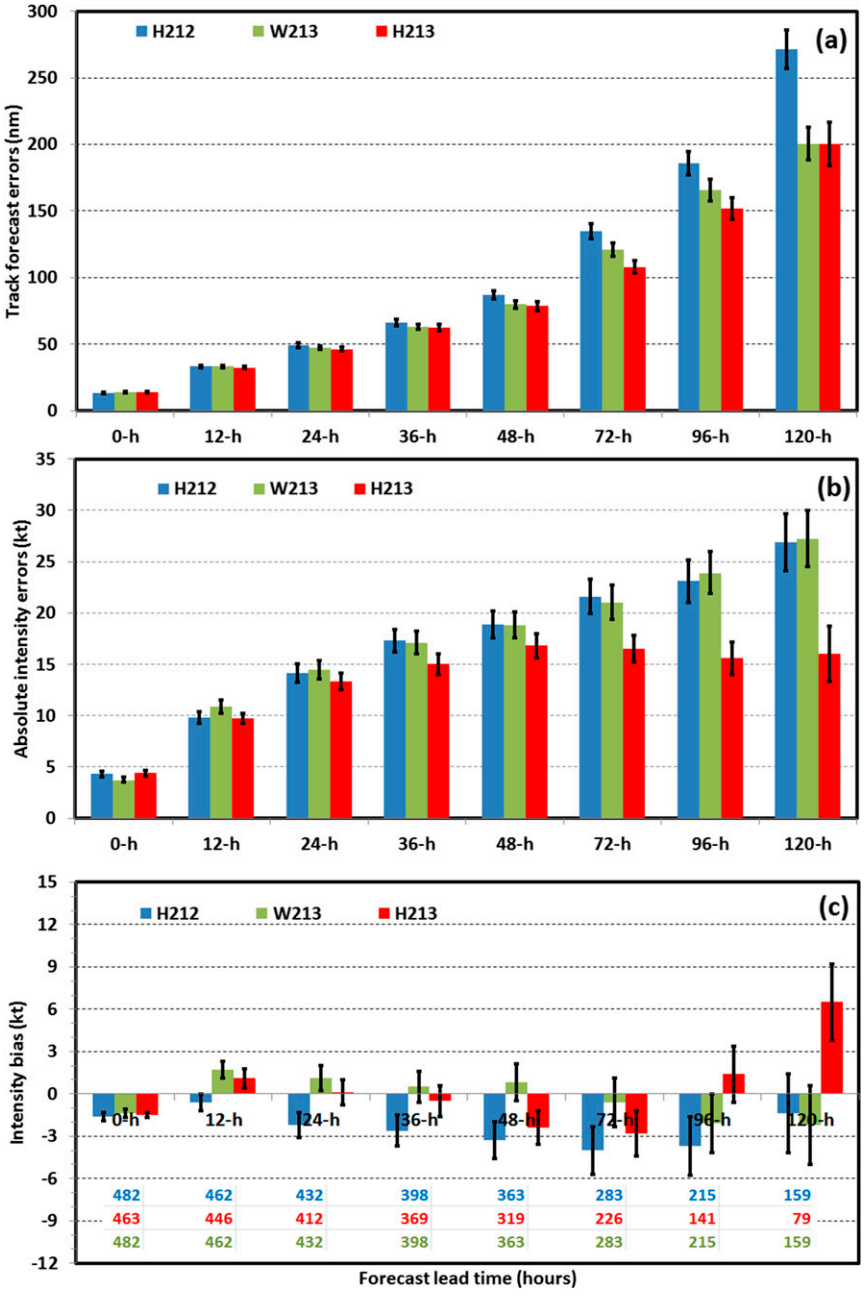

FIG. 1. Comparison between H212 during the 2012 season (blue) and H213 during the 2013 season (red), and W213 retrospective experiments (green) of the (a) absolute track forecast errors, (b) absolute intensity forecast errors, and (c) intensity bias. Comparison of H212 and W213 forecast errors are for the homogeneous sample from 2012 WPAC TCs while H213 forecast errors are obtained from 2013 WPAC TCs.

fact that the WPAC was very active in 2013 with 34 storms, of which 5 were supertyphoons, including Supertyphoon Haiyan, the most powerful landfalling storm ever recorded, the improvement in both the intensity and track forecast errors at the 3–5-day lead times indicates that the H213 was able to capture consistent development and structure of TCs in the WPAC. Because the model performance could depend on seasonal characteristics that the nonhomogeneous comparison may not fully reveal, Fig. 2 provides additional skill verification with respect to the Climatology and Persistence (CLIPER; Neumann 1992; denoted by C120) statistical model forecasts for track errors (Fig. 2a) and the 5-day Statistical Intensity Prediction

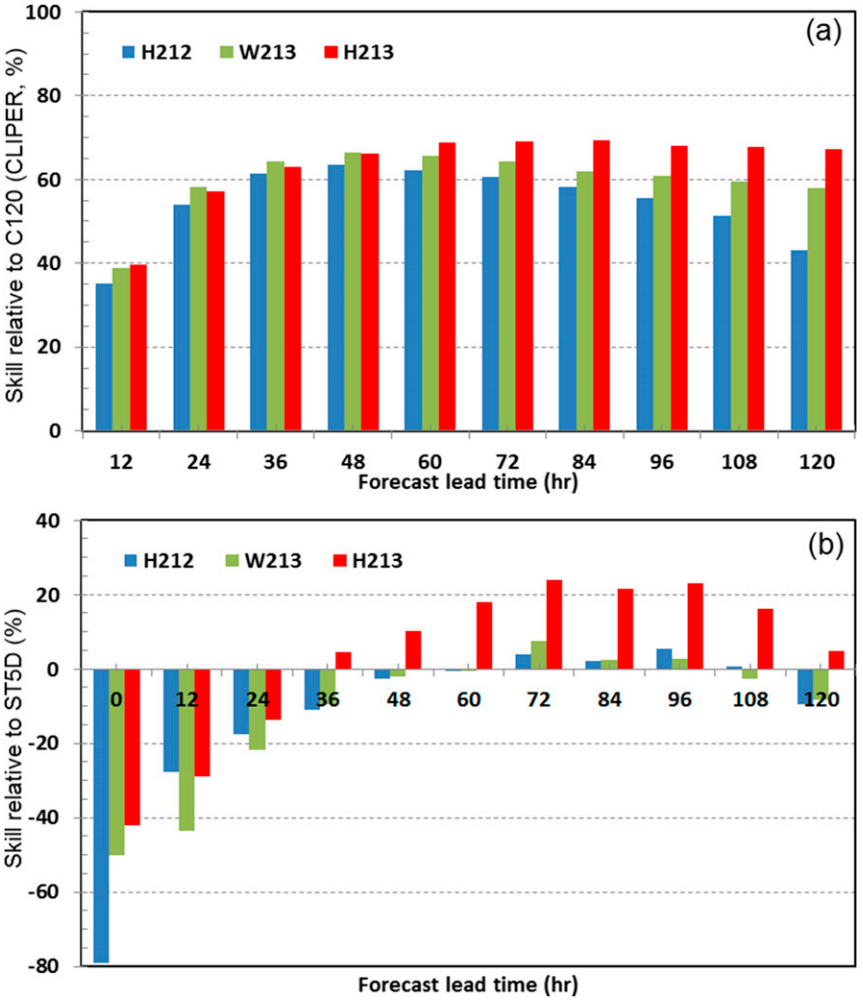

FIG. 2. Comparison of the HWRF forecast skill for (a) track errors and (b) intensity errors in the WPAC during 2012 and 2013. The model track forecast skill is relative to the CLIPER statistical forecasts, and the model intensity forecast skill is relative to the ST5D forecasts.

System (STIPS; Knaff et al. 2005; denoted by ST5D) forecasts for intensity errors (Fig. 2b). Here, the track skill is defined simply as the percentage of the HWRF forecast errors relative to the CLIPER forecast errors [i.e., (H213 − C210)/C120], and similarly for the intensity skill. By convention, positive (negative) skill means the model shows better (worse) performance by the model as compared to the statistical forecasts.

It can be seen from the skill verification that the 2013 version of HWRF shows consistent improvement in both W213 (for 2012 WPAC TCs) and H213 (for 2013 WPAC TCs) relative to H212, suggesting that the positive impacts of the 2013 model upgrades have been realized in 2013 for real-time performance in the WPAC. Particularly, W213 showed 5%–18% improvement in the intensity forecast skill from 2- to 5-day lead times compared to H212 for WPAC storms in 2012. Both the track and intensity forecast skills of H213 relative to the statistical forecasts shown in Fig. 2 apparently illustrate the significance of model upgrades in H213 as applied to real-time forecasts in 2013. Note also that the track and intensity forecast skill improvements take into account seasonal characteristics of the behavior of TCs (Neumann 1992; Knaff et al. 2005), which are indeed consistent with similar findings for the NATL and EPAC basins (Tallapragada et al. 2014b). It can be seen from this skill verification that H213 again shows consistent improvement for both W213 and H213 relative to H212, suggesting that the 2013 model upgrades have had positive impacts on the real-time performance in WPAC.

Because the performance of the HWRF model often depends on the initial storm strength (Tallapragada et al. 2014a, Goldenberg et al. 2015), Fig. 3 stratifies the statistics into strong storm and weak storm verifications

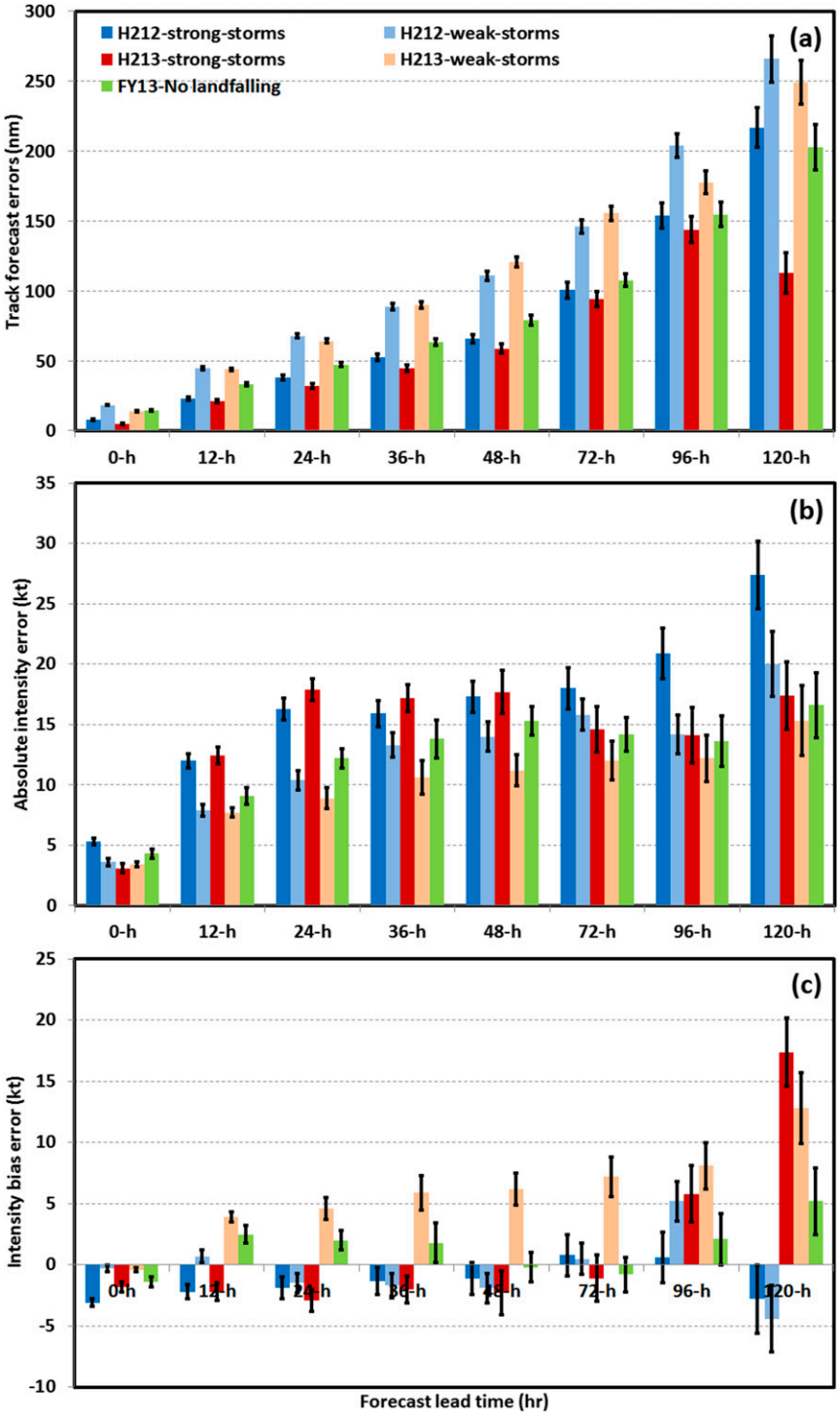

FIG. 3. (a) Comparison of the track forecast errors for strong storms (VMAX > 50 kt) during the 2012 (dark blue columns) and the 2013 experiments (dark red columns), for weak storms (VMAX < 50 kt) (orange and light blue columns), and for the 2013 statistics with land points removed (green columns). Similar to (a), shown are the (b) absolute intensity errors and (c) intensity bias. The stratification of the storm intensity is based on the best-track data values.

separately. To be explicit, a storm is considered initially strong (weak) if its best-track maximum 10-m wind (VMAX) at the initial time is greater (smaller) than or equal to 50 kt at each forecast lead time. It is of interest to see that for both 2012 and 2013, HWRF strong storms appeared to have lower track forecast errors (Fig. 3a), but they possessed somewhat larger absolute intensity errors (Fig. 3b). Despite smaller track forecast errors for the strong storms, H213 tends to have a large positive intensity bias for strong storms after 4-day lead time. Likewise, H213 also possesses a substantially positive intensity bias for weak storms at all lead times (Fig. 3c). While the positive intensity bias for weak storms is known to be caused by the overintensification tendency of H213, we notice that the positive intensity bias for strong storms is mostly related to storms making landfall, which could influence the intensity error statistics due to the forecast timing and location of these storms. Indeed, model performance statistics in which land points are removed from the verification show a significant reduction in the intensity bias at 4–5-day lead times (see Fig. 3c).

The HWRF model performance, as compared to a few other operational models used by JTWC, is shown in Fig. 4, where the homogeneous verification of H213 relative to COAMPS-TC (COTC; Doyle et al. 2011), the U.S. Navy's version of the GFDL model (GFDN; Dickerman 2006), NCEP GFS model (AVNO), and the official JTWC forecasts were included for all WPAC TCs in 2013. As seen in Fig. 4, HWRF outperformed all other regional models in terms of track and intensity forecasts. HWRF's track errors (Fig. 4a) are comparable to those for the GFS forecasts up to day 3 and slightly larger at 4–5-day lead times. In terms of the absolute intensity errors and bias (Figs. 4b,c), HWRF demonstrates better forecasts during most of the forecast range, except that it is not statistically significant at the 5-day lead time. It should be mentioned that although GFS could capture the TC tracks fairly well, its coarse resolution and model physics are not tailored to TC applications, rendering its intensity forecasts much weaker than observed, with a dominant negative intensity bias at all lead times. The improved performance of all regional models in forecasting storm intensity as compared to the GFS forecasts is apparent in both 2012 and 2013, indicating the importance of the higher resolution and appropriate model physics for TC forecasting.

Technically, the significant track improvements from H213 in both the absolute error and skill space can be attributed to all of the changes discussed in section 2. However, the larger innermost domain, the new nest motion algorithm, and the more frequent physical calls are the three main factors that have the most profound impacts to model forecast skill as demonstrated by the 3-yr retrospective experiments from 2010 to 2012 in the EPAC and NATL during our annual HWRF model development cycle. While such conclusions obtained from the large-scale sensitivity experiments in the EPAC and NATL may not be directly applicable for other basins, real-time and retrospective experiments of the HWRF model in the WPAC during both 2012 and 2013 seem to indicate that the improvements in the WPAC are more significant than in other basins, given the identical model upgrades across all ocean basins.

b. *Verification of dynamical constraints*

Similar to results shown in T15, an evaluation of the dynamical constraints between the mass and wind fields for H213 is presented in Fig. 5. Comparison of the pressure–wind relationship (PWR) between the H212 and H213 model configurations shows no significant improvement at the strong intensity limit (i.e., when VMAX $>$ 120 kt). Both H212 and H213 suffer from the same lower minimum sea level pressure (PMIN) at higher VMAX as compared to the observed PWR. Similar narrow spread of the pressure–wind relationship is observed in the retrospective experiment, W213, thus suggesting no substantial change to the dynamics of the model in the fiscal year (FY) 2013 upgrade. However, the slight difference between the observed PWR and the model-simulated PWR at low-intensity limits in H212 is alleviated in H213, which indicates that the 2013 upgrades to the HWRF model appear to correct some of the issues associated with the weak storms in 2013 (cf. Fig. 3). Note that PMIN data in the JTWC's best-track dataset is derived from a modified version of the Knaff–Zehr–Courtney (KZC) PWR proposed by Knaff and Zehr (2007) and Courtney and Knaff (2009). It is possible that the best-track PWR may not have fully captured the dynamical constraints in WPAC due to different TC characteristics in that basin, and hence the difference between the model and observed PWRs indicates that the relationship between PMIN and VMAX may not work well at higher-intensity limits. Such an issue, however, requires more systematic evaluation of the PWR in different basins (see, e.g., Velden et al. 2006), which is not the scope of this study.

c. *Geographical distribution of errors and evaluation of large-scale environment*

As speculated in T15, one of the main issues with H212 was the tendency to underestimate the WPSH, which resulted in weaker large-scale steering flows. Such underrepresentation of the large-scale environmental flow in H212 appeared to cause the model storms to have some specific across/along-track bias in different latitude regions with slow storm movement in the East

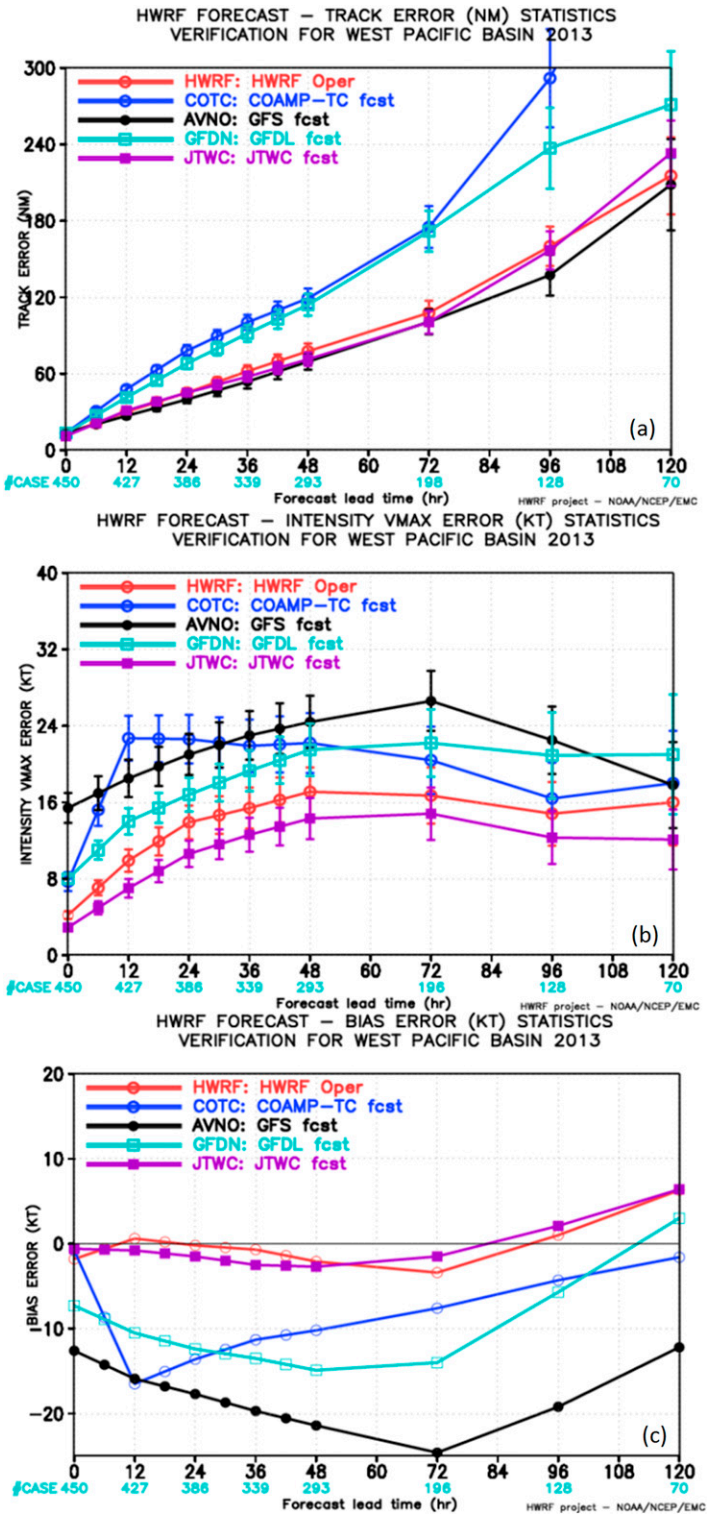

FIG. 4. Verification of (a) the track forecast errors, (b) absolute intensity forecast errors, and (c) intensity bias errors (n mi) during the 2013 tropical cyclones in the WPAC for the HWRF (red), COAMPS-TC (blue), AVNO (black), GFDN (cyan), and JTWC official (purple) forecasts. The number along the *x* axis denotes the number of cases (cycles) during the real-time experiments.

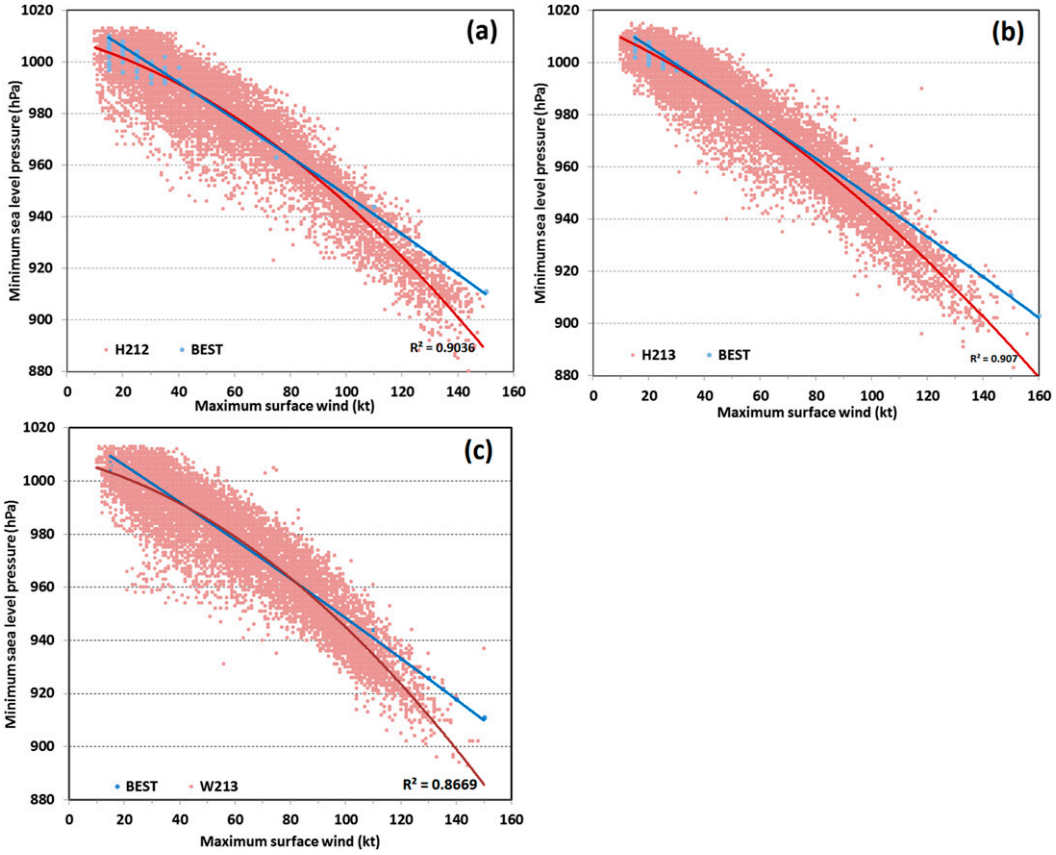

FIG. 5. Scatterplots of the VMAX and PMIN from (a) the 2012 real-time experiments with the H212 model, (b) the 2013 real-time experiments with H213, and (c) the retrospective experiment W213. Blue dots denote the observational forecast values, and red dots denote the model forecast values. Solid lines are the second-order polynomial best-fit curves.

China Sea (i.e., negative along-track errors; cf. Fig. 1b in T15). An indirect consequence of such a slow-moving bias in H212's forecasts is that the model storms may be exposed to the warm SSTs longer than the observed storms, causing systematically positive intensity bias in the same region. Recall that unlike in the NATL and EPAC, real-time experiments with the HWRF model in the WPAC have so far been run uncoupled, limiting the model's ability to accurately represent changes in the SSTs and associated air–sea interactions. As such, any systematic biases in the storm movement may have some further impact on the storm intensity in situations where the model storms incorrectly move into the warm SSTs for an extended period of time.

To examine this issue in the H213 forecasts, Fig. 6 shows a similar track and intensity error distribution as was analyzed by T15, but here it is applied to the WPAC storms in 2013. Here, the biases for both track and intensity at each box are the 48-h errors computed for a storm that is initially located within that box. Much like what has been seen in 2012, H213 showed several similar clusters of the positive/negative intensity bias in the WPAC in 2013, even though the storm density is not sufficiently large so that the clustering analysis is statistically conclusive. Specifically, there is a dominant cluster of positive intensity errors in the East China Sea, followed by an elongated negative intensity bias area in the Philippine Sea and a positive intensity bias area in the South China Sea. Although clusters of the cross-track and along-track errors corresponding to these intensity biases in the South China Sea did not show any clearly distinguishable patterns seen in 2012 (T15), we still notice some remnant cross-track errors at the higher latitudes to the east of the East China Sea (i.e., 20°–30°N, 130°–140°E, Fig. 6b). This left-track bias suggests that the model storms seem to stay over the warmer SSTs at lower latitudes (assuming that the general track direction is southeast to northwest in the WPAC basin), and so have more time to interact with the warm ocean beneath. As a result, these model storms tend to have a higher intensity than the observed overall intensity. Similar to 2012, the along-track errors in Fig. 6c appear

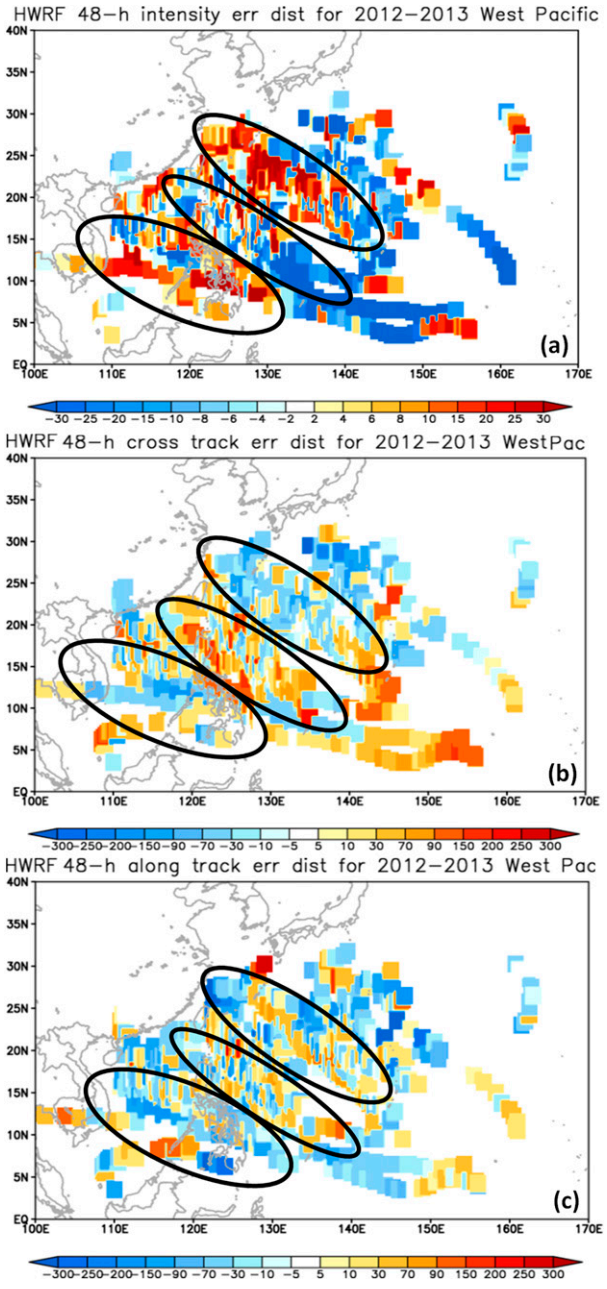

FIG. 6. Distribution of (a) the 48-h intensity bias (shaded; kt) in WPAC during 2013, (b) cross-track track forecast errors (n mi), and (c) along-track forecast errors (n mi). Black ovals denote the clusters of intensity/track errors that have connections to each other.

to show some signal of the slower storm movement to the southeast of the East China Sea (i.e., 16°–20°N, 135°–145°E), albeit the signal is not as clear as in the H212 configuration.

Figure 7 shows a composite analysis of the large-scale distribution of the 500-hPa geopotential height difference for 24-, 72-, and 120-h forecasts between H213 and the GFS analysis. Here, the composite analysis is computed across a large common domain, shown in Fig. 7a, where the shading denotes number of the forecasts that overlapped within the domain (the HWRF model outermost domain changes from cycle to cycle depending on the storms' initial locations), and the difference between the H213 and GFS analysis is averaged for all forecast cycles between July and November 2013. The negative 500-hPa bias in H213 relative to the GFS analysis, with the largest magnitude right around the area where WPSH is located, has grown in time from the 24- to the 72-h forecast period. Such a negative bias in the 500-hPa geopotential height was similar to the analysis of the H212 forecasts (T15) and is also observed in the NATL and EPAC. Although it is not known at this moment what caused the large-scale environment in the HWRF model to be weaker than the GFS analysis besides the potential impact due to lateral boundary conditions, our sensitivity experiments with different radiation and microphysics parameterization schemes, as well as different model configurations appear to have similar large-scale biases. In addition, differences in the large-scale flows between the HWRF forecasts and the GFS analysis are concentrated near the coastlines in WPAC as well, indicating that the HWRF model could have some issues with the land surface model that result in more errors along these coastlines. Because the large-scale flows in the vicinity of the TC circulation could have a large influence on the track and intensity forecasts, these large-scale errors suggest that the model physics in the far-field environment should be paid more attention along with the TC-tailored physics, which will be the target areas for improvement of the HWRF modeling system in the future.

d. RI forecast verification

While the absolute intensity forecast errors and the corresponding intensity bias could provide some useful information about the overall intensity forecast performance of the model, the magnitude of the absolute intensity errors is generally insufficient to exclusively quantify the model performance because of inherent uncertainties related to stochastic processes at small spatial and temporal scales that can significantly impact the model VMAX (Torn and Snyder 2012; Landsea and Franklin 2013). Because of such model-inherent uncertainties and the stochastic nature of the pointlike measurement metric, another valuable measure for evaluating the capability of a dynamical model in forecasting TC intensity is to see how the model predicts the phase of the storm development such as intensification, steady state, or weakening state rather than emphasizing entirely the absolute value of VMAX. In this regard,

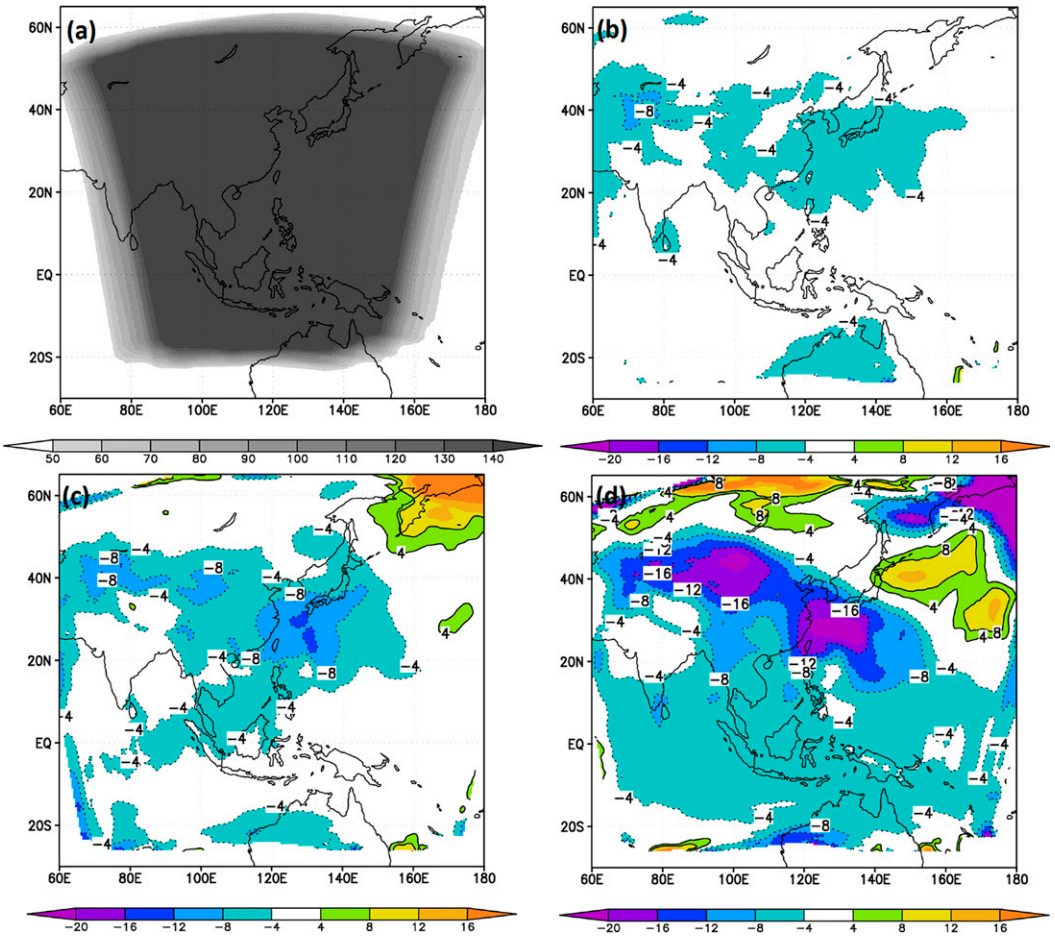

FIG. 7. (a) Number of overlapped outermost domains from the H213 forecasts from 2013 real-time experiments, and average of the 500-hPa geopotential height difference between the HWRF (b) 24-, (c) 72-, and (d) 120-h forecasts and the corresponding GFS analyses valid at the same time.

recent developmental efforts with the HWRF modeling system have also put weight on the capability of forecasting rapid intensification (RI) for the storms in the WPAC where such RI occurrences are more frequent.

Figure 8 shows scatterplots of the 24-h VMAX change in H213's real-time forecasts during the entire experimental period in 2013, as well as in the retrospective experiment W213 along with the ST5D statistical forecasts. If one defines an RI event as a period during which VMAX increases by 30 kt in 24 h, the verification of the model RI forecast can be formulated as a categorical verification problem in which a correct RI forecast can be quantified directly as a binary event (yes or no) when compared to the corresponding observed event during the same period. With this definition, all of the points within the black boxes in Fig. 8 indicate that the model captures RI events that are consistent with observations, which provides a specific probability of detection (POD) index for RI events. Likewise, all of the points within the gray boxes in Fig. 8 indicate that the model predicts incorrect RI events, which corresponds to a false alarm ratio (FAR) index. Table 2 lists all of the RI detection and false alarm events for H212 and H213. One notices in Fig. 8 that there is a significant improvement in the detection of RI events in the W213 experiment with a POD of ~0.17 as compared to the much lower POD index in H212 (~0.09). The real-time performance of the H213 configuration shows even higher results with POD ~ 0.22, while the statistical ST5D forecasts possess minimum skill in predicting RI, with a POD index of ~0.09. Note that in T15 we used a criteria of 25-kt VMAX change in 24 h as the definition for RI to allow for more RI forecasts. Use of similar criteria for RI also shows that the H212 POD for RI was much lower, which can be attributed to several deficiencies in the H212 model configuration discussed earlier in section 2. Along with a higher POD index, both W213 and H213 also have a lower FAR index (0.68 and 0.45) than H212

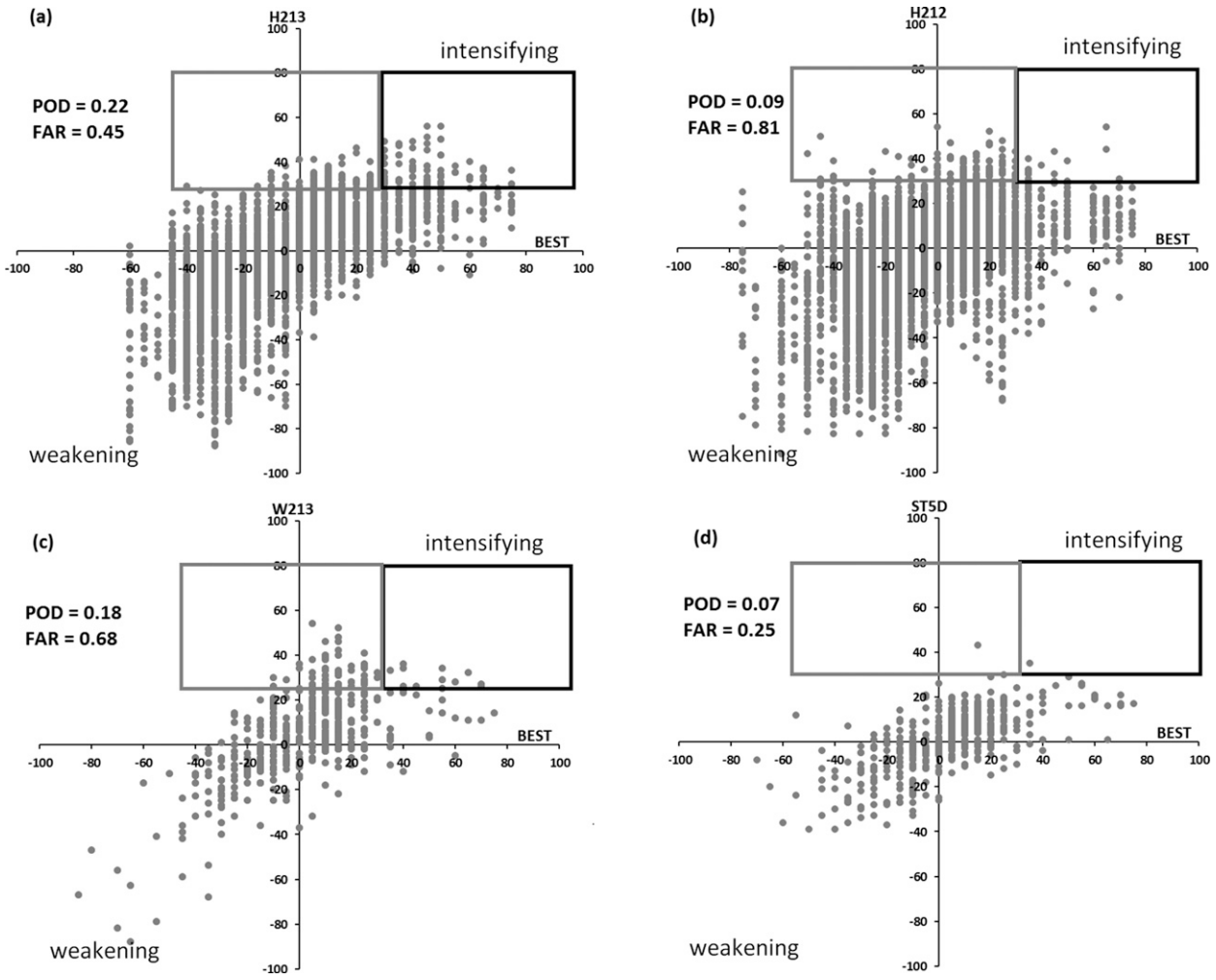

FIG. 8. Scatterplots of the 24-h change of the maximum 10-m wind from observation (BEST; *x* axis) and the real-time model forecasts (HWRF; *y* axis) for (a),(c) 2013 and (b),(d) 2012. Black boxes denote the points where both HWRF and observation capture RI, whereas gray boxes denote the points where HWRF forecasted RI events that were not observed in reality.

(0.81), confirming the 2013 HWRF upgrades have led to significant progress in reducing the spurious storm development despite the fact that H213 still has some issues with cases having very large intensification rates (Tallapragada and Kieu 2014).

Figure 9 illustrates two specific cases, Typhoon Soulik (07W) and Typhoon Usagi (17W), during 2013 for which H213 was able to forecast RI quite accurately and consistently, especially for Typhoon Soulik (Fig. 9a) during the seven consecutive cycles from 1800 UTC 7 July to 0000 UTC 9 July 2013. Note that the H213 forecasts overestimated Typhoon Soulik's intensity at 4–5-day lead times as a result of the slightly incorrect landfall location, which had a northward position bias and thus allowed the model storm to stay over warm SSTs for a longer period of time. While H213 could capture fairly well the RI events for major storms, as seen in the examples of Typhoons Soulik and Usagi in Figs. 9a,b, we notice that HWRF is not able to forecast the extreme RI, or the so-called explosive deepening event, during which the 24-h change of VMAX $>$ 50 kt (Tallapragada and Kieu 2014). The WPAC in 2013 experienced several instances of such extreme RI events for several

TABLE 2. Categorical forecast table of the real-time RI events in WPAC from H212 and H213.

	OBS	
	Yes	No
H213		
Yes	53	44
No	184	3356
H212		
Yes	16	71
No	169	3252

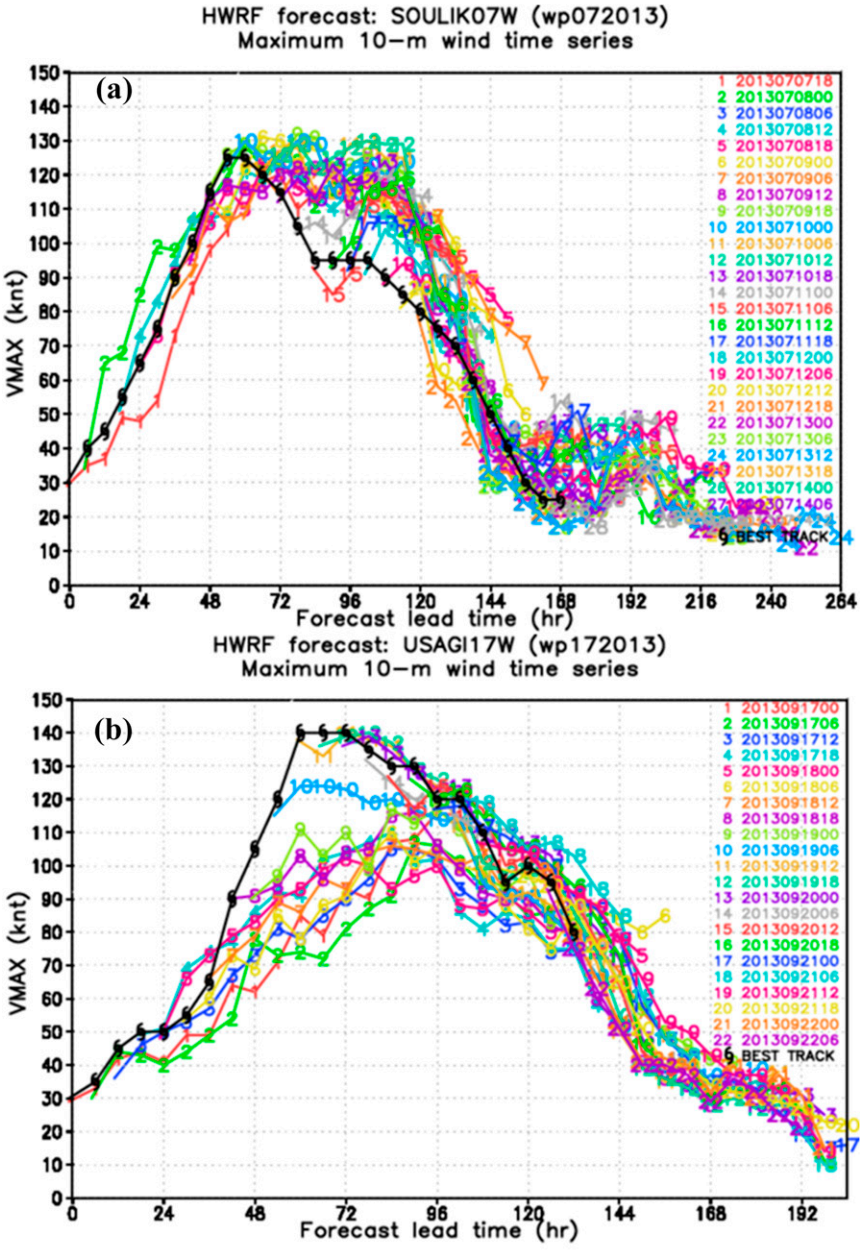

FIG. 9. Composite maximum 10-m wind forecasts of H213 during the entire life cycle of (a) Typhoon Soulik (2013) and (b) Supertyphoon Usagi (2013). Black solid curve denotes the observed intensity recorded in the JTWC best-track dataset.

supertyphoons. This inability to forecast the explosive deepening is not limited to H212 but also is found for the H213 version, which exposes our current inadequate understanding of TC dynamics and physics under extreme situations, which requires further exploration. Although causes for the inability of the model in capturing such extreme RI events are unknown at this moment, Kieu and Tallapragada (2014) have recently reported that these extreme RI cases are often associated with the development of a double-warm-core structure during a very high-intensity regime in the HWRF model, which is at present not fully documented or understood. Lack of specification of the upper-level warm core in HWRF's current vortex initialization scheme could be a factor that limits the model from experiencing the explosive deepening, which we will address in our upcoming detailed case study of Supertyphoon Francisco (26W).

TABLE 3. Table of 24-h intensity changes for the H213 forecast with a range of intensity thresholds from 1 to 25 kt during the 2013 real-time experimental period.

24-h intensity change threshold (kt)	POD	FAR
1	0.91	0.28
5	0.87	0.22
10	0.80	0.33
15	0.72	0.42
20	0.57	0.48
25	0.48	0.53

e. *Intensity change forecast verification*

From the forecaster's perspective, the ability to predict the general TC intensity trend is also very important. A statement that a storm will intensify, stay at the same intensity, or weaken in the next, for example, 12, 24, or 36 h is of similar value as forecasting RI events (Zelinsky 2014). In this regard, one can examine a simple intensity change forecast capability by verifying if the model can predict that storm intensification will happen or not at certain lead times regardless of the intensification rate or magnitude of the intensity change. This can be done by using a criterion of a range of intensity change thresholds for several forecast lead times, including 0–12, 0–24, 0–36, 0–48, and 0–72 h at which the model intensity forecasts are verified with the observed intensity change during these respective periods. Because forecasts of intensity changes for more than 3 days in advance are not reliable in general, we will limit the intensity change verifications to up to 72-h intervals.

As a first attempt to examine the predictive skill of the HWRF model in forecasting a general 24-h intensity change instead of the rapid intensification with criteria of 30 kt $(24\,h)^{-1}$, we relax the 30-kt threshold for RI to a range from 1- to 25-kt change of VMAX over a 24-h period as an indication of whether the model captures the intensification trend or not. Here, the lower limit of 1 kt is chosen instead of using 0-kt change (no change forecasts; often referred to as NCHG forecasts by NHC and JTWC in their intensity forecast verification reports) because the model output of the VMAX provides intensity change at every 1-kt interval. Note that the observed VMAX can only change by an increment of 5 kt (best-track intensity data increments). By introducing this range of intensity change thresholds, we can quantify the HWRF model performance in terms of the POD/FAR indices for different intensity change thresholds instead of a single RI criteria. By convention, a 0–24-h intensification event with an intensity change threshold of 5 kt is considered to be a ''yes'' if both the model forecast and observations show a change of at least 5 kt in the 0–24-h forecast period for any given cycle, or a false alarm if otherwise.

Table 3 provides the POD/FAR analysis for a range of the 24-h intensity change thresholds from 1 to 25 kt. One notices very high POD values (≥0.87) for the 1- and 5-kt thresholds, and the POD index gradually decreases for larger-intensity change thresholds. The high POD index and low FAR index for the smaller-intensity thresholds indicate that H213 provides reliable information about the phase of the storm development regarding whether a storm will intensify or not at the 24-h lead time. For the magnitude of the intensity change, the decrease (increase) of the POD (FAR) index at the larger-intensity change threshold suggests that the model may not capture fully the intensification rate for storms that undergo faster development, which is consistent with the relatively low POD index of ~0.22 for the RI forecasts examined in the previous section. Similar analysis for other models by Tallapragada and Kieu (2014) shows that despite having a decreasing POD index for larger-intensity change thresholds, H213 tends to possess a higher POD index than other regional and global model forecasts. Another noteworthy point is that the FAR index increases with greater-intensity threshold, with a FAR index at ~28% for the 1-kt threshold and reaching 53% for the 25-kt threshold. This indicates that the model seems to capture better the intensity tendency but has issues with the magnitude of the intensification. Specifically, the large FAR index for higher-intensity change thresholds means the model tends to overestimate the storm intensity during the first 24 h, which is somewhat consistent with the initial unbalanced adjustment related to the vortex initialization (Tallapragada et al. 2015).

Along with the change in the POD/FAR index with different intensity change thresholds at 24-h lead time, another aspect of the intensity change forecast capability is how these indices depend on the range of forecast lead times. To address this issue, Fig. 10 compares the POD and FAR indices for the intensity trend forecasts between H212 and H213 with different forecast intervals for a specific 5-kt intensity change threshold, which is sufficient to capture the tendency of the model intensity change. Because of the nonhomogeneous comparison between H212 and H213, the POD/FAR indices for a set of retrospective experiments (W213; see Fig. 1) for the 2012 TCs using the H213 configuration is also provided. It is seen in Fig. 10 that H213 possesses a significantly higher POD index than H212, with a higher POD index for longer forecast intervals (e.g., 0–72 h). Such improved performance may not reveal conclusively the improvement of the model due to both the model and the different sample periods considered in this evaluation. However, a further comparison of the H212 and

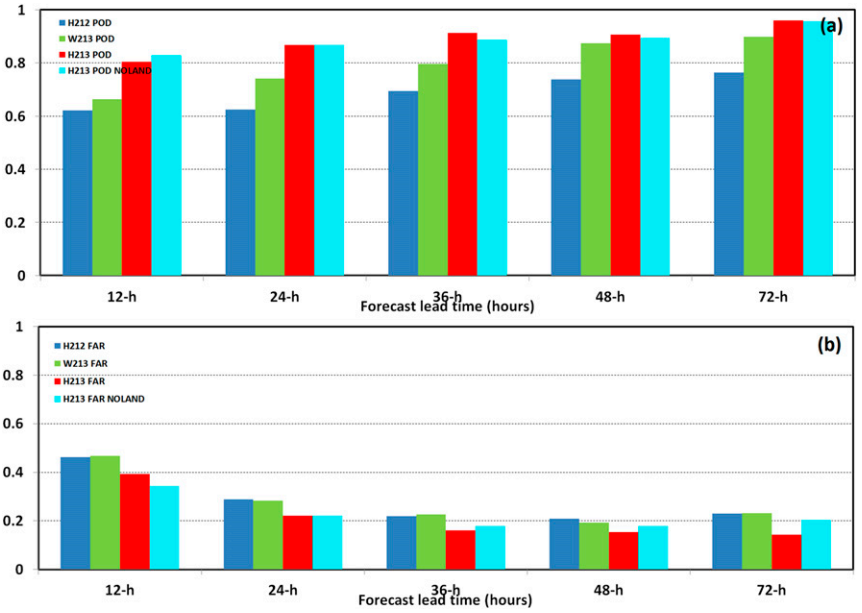

FIG. 10. (a) Comparison of the POD index for the 0–12-, 0–24-, 0–36-, 0–48-, and 0–72-h intensity change forecasts for H212 (blue columns), the 2012 retrospective experiments using W213 (green), and the 2013 real-time forecasts with H213 (red columns). (b) As in (a), but for the FAR index. Cyan columns denote statistics of the intensity changes in the 2013 season in which all land points are excluded.

W213 for 2012 TCs clearly demonstrates that the improvement in forecasting the intensity change could be attributed to the model upgrades implemented in 2013.

The intensity trend forecast showing better performance at the longer forecast lead times is to some extent expected from the regional modeling perspective, since regional models often take time to adjust to the ambient environment after the vortex initialization processes. This is consistent with a previous study by Tien et al. (2012), who showed that this 72-h threshold appears to be a limit at which the environmental conditions begin to dominate the vortex initial errors. As seen in Fig. 10, the POD index for 0–72-h intensity trend forecasts is ~0.65 for W213 and even higher (~0.80) for H213. The POD index for 72 h with the same 5-kt threshold is somewhat lower in the NATL and EPAC basins, but at least the relatively high POD index (>0.65) in all basins for the intensity change forecast at the 5-kt threshold is still useful, as it could give forecasters information about whether a model storm will intensify or not as far as 3 days in advance. This is consistent with a recent report by Zelinsky (2014), which showed that H213 had very good intensity trend forecasts in both NATL and EPAC as well.

Unlike the POD index, the FAR index for intensity trend forecasts does not show much improvement in H213's forecasts compared to H212 (Fig. 10b). Specifically, the FAR index for the real-time forecasts in 2012 and 2013, along with the retrospective experiments for 2012 from W213, all start roughly at ~0.4 for the 0–12-h interval and are reduced to ~0.20 at 0–36 h and beyond. The higher FAR index during the early forecast lead times could be attributed to the initial adjustments at the vortex scale, which takes about 12–24 h to attain a better balance with the ambient environment. Despite significant reduction of the FAR index at longer forecast intervals, the FAR index of H213 still stays relatively high (>0.20), suggesting that the model may provide false alarms of intensity change quite often. Further examination revealed that such a relatively high FAR index at longer forecast intervals is also associated with the incorrect timing of storm landfalls or their locations (Fig. 10b). For example, a delayed landfall of the model storm would allow the storm to stay over the ocean for a longer period of time and experience continuous growth, while the actual storm would have already made landfall and thus dissipated.

4. Summary, conclusions, and future plans

In this study, the performance of the HWRF model during 2013 real-time experiments in the western North Pacific basin has been presented. Forecast verifications showed that the 2013 version of the HWRF model (H213) achieved a substantial improvement as compared to the originally implemented 2012 version of the

HWRF model (H212) in both retrospective experiments and real-time forecasts with the 3-, 4-, and 5-day track forecast errors in the 2013 version reduced by about 14%, 11%, and 14%, respectively. Despite the dominance of strong storms in 2013, the H213 upgrades displayed more significant positive skill with respect to statistical persistence forecasts (ST5D) for the range from 36- to 96-h lead times. Verification of the H213's skill against the climate persistence forecasts shows that although part of such improvements in 2013 is related to the different seasonal characteristics between the years of 2012 and 2013, the new model upgrades implemented in 2013, including a larger innermost model domain size, an improved nest tracking algorithm, improved PBL physics, a higher frequency of physics calls, and a number of other bug fixes, could provide some further improvements that the 2012 version of HWRF could not achieve. Such improvements to the H213 were also realized in retrospective experiments for 2012 WPAC storms run with the 2013 version of the HWRF model (W213). Of course, it is difficult to isolate exactly what factors had contributed the most to the overall improvement of the W213/H213 as a result of the nonlinear feedback among different components; our sensitivity experiments during the annual upgrade of the HWRF model showed that the larger domain size, increased frequency of physics calls, and improved nest tracking algorithm are the three main factors that helped improve the model forecast performance most significantly in the 2013 version of the HWRF model (see, e.g., Trahan et al. 2014; Tallapragada et al. 2014b).

Further verification of the RI events in terms of the probability of detection (POD) and false alarm rate (FAR) demonstrated noticeable improvements by H213 compared to H212. Specifically, the POD index increased from 0.09 in 2012 to ~0.22 in 2013. Likewise, the FAR index was reduced by nearly half from 0.81 in 2012 to 0.45 in 2013. Categorical verification of the intensity trend forecasts showed an even more promising capability, with the POD index being ~0.8 and FAR index being ~0.3 for the forecast intervals from 0–12 to 0–72 h. This suggests that H213 could give forecasters some useful information about whether a model storm will intensify up to 3 days in advance.

While H213 showed promising forecast skill in predicting the WPAC TCs in 2013, it should be mentioned that the large-scale biases in the environmental fields showed similar characteristics, as noted in H212 (Tallapragada et al. 2015). Composite analysis of the differences between the HWRF forecasts and GFS analyses demonstrated that H213 tends to produce a weaker WPSH, which may render the steering flow associated with the WPSH weaker than the actual flow. As a result, both H212 and H213 possess systematic track biases, which could force model storms to stay over warm water for longer periods of time, resulting in a positive intensity bias in the East China Sea. In this regard, improving the large-scale environment is one of the aspects we would like to focus more on in our future model upgrades to HWRF.

Given the promising performance of the HWRF model as demonstrated in real time for WPAC in 2012 and 2013 and the endorsement by JTWC, further expansion of HWRF model forecasts for all global TC basins was accomplished in January 2014. Although the model forecasts are provided using nonoperational NOAA research and development supercomputers with support from HFIP, the reliability of the forecast delivery has increased from ~85% in 2012 to ~90% in 2013. The operational HWRF modeling system is currently being developed as a community model and is being supported by the Developmental Testbed Center (DTC; Bernardet et al. 2014). Performance of the HWRF model across all JTWC basins in 2014 will be summarized in a forthcoming paper, where we will also document new model upgrades implemented in 2014.

Acknowledgments. This work was supported by the Hurricane Forecast Improvement Project (HFIP) of NOAA. We thank JTWC, the China Shanghai Typhoon Institute, and the Taiwan Central Weather Bureau for their various feedback and suggestions that helped improve the performance of the HWRF model during the 2013 real-time experiments. We thank three anonymous reviewers for their constructive comments and suggestions.

REFERENCES

Bernardet, L., and Coauthors, 2014: Community support and transition of research to operations for the Hurricane Weather Research and Forecast (HWRF) model. *Bull. Amer. Meteor. Soc.*, **96**, 953–960, doi:10.1175/BAMS-D-13-00093.1.

Courtney, J., and J. A. Knaff, 2009: Adapting the Knaff and Zehr wind–pressure relationship for operational use in tropical cyclone warning centres. *Aust. Meteor. Oceanogr. J.*, **58**, 167–179.

Dickerman, C. L., 2006: A recent history of the GFDN Tropical Cyclone Forecast Model. Preprints, *27th Conf. on Hurricanes and Tropical Meteorology*, Monterey, CA, Amer. Meteor. Soc., P5.8. [Available online at https://ams.confex.com/ams/pdfpapers/108054.pdf.]

Doyle, J. D., and Coauthors, 2011: Real time tropical cyclone prediction using COAMPS-TC. *Advances in Geosciences*, Vol. 28, Academic Press, 15–28.

Evans, A. D., and R. J. Falvey, 2013: Annual Joint Typhoon Warning Center tropical cyclone report. JTWC, 118 pp. [Available online at http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/atcr/2012atcr.pdf.]

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. Quirio, 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 System (HWRF): Track and intensity forecast verifications. *Wea. Forecasting*, **30**, 710–729, doi:10.1175/WAF-D-14-00098.1.

Gopalakrishnan, S. G., 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 Jr., 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.

Janjić, Z., R. Gall, and M. E. Pyle, 2010: Scientific documentation for the NMM Solver. NCAR Tech. Note NCAR/TN-477+STR, 54 pp. [Available online at http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/NMM_scientific_2-2-10_final.pdf.]

Kieu, C. Q., and V. Tallapragada, 2014: On the development of double warm cores in intense tropical cyclones in the HWRF model. *Proc. 31st Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., 17D.4. [Available online at https://ams.confex.com/ams/31Hurr/webprogram/Paper243675.html.]

Kim, H.-S., C. Lozano, V. Tallapragada, D. Iredell, D. Sheinin, H. L. Tolman, V. M. Gerald, and J. Sims, 2014: Performance of ocean simulations in the coupled HWRF–HYCOM model. *J. Atmos. Oceanic Technol.*, **31**, 545–559, doi:10.1175/JTECH-D-13-00013.1.

Knaff, J. A., and R. M. Zehr, 2007: Reexamination of tropical cyclone wind–pressure relationships. *Wea. Forecasting*, **22**, 71–88, doi:10.1175/WAF965.1.

——, C. Sampson, and M. DeMaria, 2005: An operational statistical typhoon intensity prediction scheme for the western North Pacific. *Wea. Forecasting*, **20**, 688–688, doi:10.1175/WAF863.1.

Kwon, Y. C., 2014: 2014 upgrades of NOAA's operational hurricane model (HWRF) and its performance. *Proc. 31st Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., 11A.1. [Available online at https://ams.confex.com/ams/31Hurr/webprogram/Paper244004.html.]

——, V. Tallapragada, W. Wang, C. Kieu, E. Aligo, S. Trahan, Q. Liu, and Z. Zhang, 2014: The physics suite upgrades of the operational HWRF model for 2014 implementation. *Tropical Cyclone Research Forum (TCRF)/68th Interdepartmental Hurricane Conf.*, College Park, MD, Office of the Federal Coordinator for Meteorological Services and Supporting Research, S4-03. [Available online at http://www.ofcm.gov/ihc14/presentations/Session4/s04-03kwon.pdf.]

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

Neumann, C., 1992: Final report: Joint Typhoon Warning Center (JTWC92) model. SAIC Contract Rep. N00014-90-C-6042 (Part 2), 41 pp. [Available from Naval Research Laboratory, 7 Grace Hopper Ave., Monterey, CA 93943-5502.]

Tallapragada, V., and C. Kieu, 2014: Real-time forecasts of typhoon rapid intensification in the north western Pacific basin with the NCEP operational HWRF Model. *Tropical Cyclone Res. Rev.*, **3**, 63–77.

——, and Coauthors, 2013: Hurricane Weather Research and Forecasting (HWRF) model: 2013 scientific documentation. Developmental Testbed Center Rep., 99 pp. [Available online at http://www.emc.ncep.noaa.gov/gc_wmb/vxt/pubs/HWRFScientificDocumentation2013.pdf.]

——, C. Q. Kieu, Y. C. Kwon, S. G. Trahan, Q. Liu, Z. Zhang, and I.-H. Kwon, 2014a: 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.

——, and Coauthors, 2014b: Significant advances to the NCEP operational HWRF modeling system for improved hurricane forecasts. *Proc. 31st Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., 14D.1. [Available online at https://ams.confex.com/ams/31Hurr/webprogram/Paper244632.html.]

——, and Coauthors, 2015: Forecasting tropical cyclones in the western North Pacific basin using the NCEP operational HWRF model: Real-time implementation in 2012. *Wea. Forecasting*, **30**, 1355–1373, doi:10.1175/WAF-D-14-00138.1.

Tien, T. T., C. Thanh, H. T. Van, and C. Q. Kieu, 2012: Two-dimensional retrieval of typhoon tracks from an ensemble of multimodel outputs. *Wea. Forecasting*, **27**, 451–461, doi:10.1175/WAF-D-11-00068.1.

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

Trahan, S., and Coauthors, 2014: Improved Telescopic Nesting and Accurate Storm Tracking in the NCEP Operational HWRF Model. *Proc. 31st Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., 14D.2. [Available online at https://ams.confex.com/ams/31Hurr/webprogram/Paper244200.html.]

Velden, C., and Coauthors, 2006: The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bull. Amer. Meteor. Soc.*, **87**, 1195–1210, doi:10.1175/BAMS-87-9-1195.

Vickers, D., and L. J. Mahrt, 2003: The cospectral gap and turbulent flux calculations. *J. Atmos. Oceanic Technol.*, **20**, 660–672, doi:10.1175/1520-0426(2003)20<660:TCGATF>2.0.CO;2.

Zelinsky, D. A., 2014: Application of wind-pressure relationships to global model output for tropical cyclone intensity forecasting. *Proc. 31st Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., 17D.7. [Available online at https://ams.confex.com/ams/31Hurr/webprogram/Paper243748.html.]

Zeng, Z., Y. Wang, and C.-C. Wu, 2007: Environmental dynamical control of tropical cyclone intensity—An observational study. *Mon. Wea. Rev.*, **135**, 38–59, doi:10.1175/MWR3278.1.

Zhang, J. A., D. S. Nolan, R. F. Rogers, and V. Tallapragada, 2015: Evaluating the impact of improvements in the boundary layer parameterization on hurricane intensity and structure forecasts in HWRF. *Mon. Wea. Rev.*, **143**, 3136–3155, doi:10.1175/MWR-D-14-00339.1.