Evaluating Environmental Impacts on Tropical Cyclone Rapid Intensification Predictability Utilizing Statistical Models

JOHN KAPLAN,* CHRISTOPHER M. ROZOFF,+ MARK DEMARIA,#,## CHARLES R. SAMPSON,@ JAMES P. KOSSIN,& CHRISTOPHER S. VELDEN,+ JOSEPH J. CIONE,* JASON P. DUNION,**,* JOHN A. KNAFF,# JUN A. ZHANG,**,* JOHN F. DOSTALEK,++ JEFFREY D. HAWKINS,@ THOMAS F. LEE,@,@@ AND JEREMY E. SOLBRIG@

* *NOAA/AOML/Hurricane Research Division, Miami, Florida*
+ *CIMSS/University of Wisconsin–Madison, Madison, Wisconsin*
# *NOAA/Center for Satellite Applications and Research, Fort Collins, Colorado*
@ *Naval Research Laboratory, Monterey, California*
& *NOAA/National Centers for Environmental Information, Asheville, North Carolina*
** *University of Miami/Cooperative Institute for Marine and Atmospheric Studies, Miami, Florida*
++ *Cooperative Institute for Atmospheres, Fort Collins, Colorado*

(Manuscript received 27 February 2015, in final form 10 June 2015)

ABSTRACT

New multi-lead-time versions of three statistical probabilistic tropical cyclone rapid intensification (RI) prediction models are developed for the Atlantic and eastern North Pacific basins. These are the linear-discriminant analysis–based Statistical Hurricane Intensity Prediction Scheme Rapid Intensification Index (SHIPS-RII), logistic regression, and Bayesian statistical RI models. Consensus RI models derived by averaging the three individual RI model probability forecasts are also generated. A verification of the cross-validated forecasts of the above RI models conducted for the 12-, 24-, 36-, and 48-h lead times indicates that these models generally exhibit skill relative to climatological forecasts, with the eastern Pacific models providing somewhat more skill than the Atlantic ones and the consensus versions providing more skill than the individual models. A verification of the deterministic RI model forecasts indicates that the operational intensity guidance exhibits some limited RI predictive skill, with the National Hurricane Center (NHC) official forecasts possessing the most skill within the first 24 h and the numerical models providing somewhat more skill at longer lead times. The Hurricane Weather Research and Forecasting Model (HWRF) generally provides the most skillful RI forecasts of any of the conventional intensity models while the new consensus RI model shows potential for providing increased skill over the existing operational intensity guidance. Finally, newly developed versions of the deterministic rapid intensification aid guidance that employ the new probabilistic consensus RI model forecasts along with the existing operational intensity model consensus produce lower mean errors and biases than the intensity consensus model alone.

1. Introduction

Although some improvements in tropical cyclone (TC) intensity forecasting have been achieved over the past few decades (DeMaria et al. 2014), predicting changes in TC intensity remains problematic. This is particularly true for identifying episodes of rapid intensification (RI), which remains the highest operational forecasting priority of the National Hurricane Center (NHC) (Rappaport et al. 2012). The difficulty of predicting RI is due in large part to the multiscale nature of the problem with environmental, oceanic, and inner-core processes all likely playing important roles in determining if and when a TC will undergo RI.

The importance of environmental influences in the RI process has been shown in a number of observational and numerical modeling studies. For example,

Current affiliation: NOAA/National Hurricane Center, Miami, Florida.
@@ Retired.

Corresponding author address: John Kaplan, NOAA/AOML/Hurricane Research Division, 4301 Rickenbacker Cswy., Miami, FL 33149.
E-mail: john.kaplan@noaa.gov

DOI: 10.1175/WAF-D-15-0032.1

© 2015 American Meteorological Society

Molinari and Vollaro (1990) showed that environmental forcing from an upper-level trough produced enhanced midlevel inflow, upper-level outflow, and increased upward motion in the near-storm environment surrounding Hurricane Elena (1989) and speculated that might have contributed to the RI of that system. However, a subsequent composite study by Hanley et al. (2001) showed that systems were more likely to undergo RI when there was no forcing from an upper-level trough or cold low. Thus, additional factors such as the strength and location of the upper-level trough or cold low relative to the TC likely play critical roles in determining if such interactions ultimately trigger an episode of RI, as the results of the Hanley et al. (2001) study also indicated.

More recently, Kaplan and DeMaria (2003, hereafter KD03) and Kaplan et al. (2010, hereafter KDK10) showed that statistically significant differences existed between the kinematic and thermodynamic environments of TCs that underwent RI and those that did not. Specifically, the aforementioned studies found that RI was more likely to occur for TCs that were situated over regions of higher than average sea surface temperature, oceanic heat content, and low- to midtropospheric moisture. Additionally, the results of KD03 and KDK10 indicated that systems were more likely to undergo RI if they were located in regions of weaker than average vertical wind shear, larger than average upper-level divergence, were farther from their maximum potential intensity, and were intensifying at the time of RI.

The modulating impact of the underlying oceanic conditions on the potential for a TC to undergo RI has also been demonstrated in several previous studies. For example, research conducted by Hong et al. (2000) and Shay et al. (2000) showed that a Gulf of Mexico warm-core eddy likely contributed positively to the observed RI of Hurricane Opal (1995), perhaps by minimizing the oceanic cooling beneath that system. The above findings are consistent with those of Cione and Uhlhorn (2003), who showed that systems that experienced the least amount of inner-core sea surface temperature reduction generally experienced the largest rates of intensification. It is worth noting, however, that Bosart et al. (2000) showed that Opal was embedded in an environment of low vertical wind shear and enhanced divergence generated by an upper-level trough during the period prior to that system's interaction with the ocean eddy so the presence of the eddy itself was likely only one of several factors that led to Hurricane Opal's RI.

In recent years, the role of inner-core processes on RI has been investigated in more detail. In one such study conducted by Kossin and Schubert (2001), it was shown that the development of eyewall mesovortices might induce RI provided that the vorticity of the parent vortex is initially sufficiently large. More recently, the potentially important role that convective-scale bursts may play in the RI process has been examined by a number of researchers (e.g., Guimond et al. 2010; Wang and Wang 2014; Rogers et al. 2013, 2015). The results of these studies suggest that a positive feedback between the axisymmetric vortex and convective bursts can develop, particularly if the bursts are located near and just inside the radius of maximum wind where the background symmetric component of the parent vortex's vorticity is largest, as was shown in Rogers et al. (2015). If such a superposition of the convective burst and parent vortex is observed, then the axisymmetrizaton (Montgomery and Kallenbach 1997) of the TC's inner-core vorticity (Reasor et al. 2009) may produce intensification and ultimately RI of the TC. Perhaps if the aforementioned axisymmetrizaton of the inner-core vorticity also coincides with the development of an axisymmetric ring of convection, then RI may prove more likely to occur since Willoughby (1990) found that the development of an isolated well-defined convective ring can lead to rapid strengthening if that ring contracts, particularly for TCs of hurricane strength. The importance of the existence of a low-level symmetric convective ring during the RI process is supported by both the recent observational results of Kieper and Jiang (2012) and the previous modeling results of Nolan and Grasso (2003), which demonstrated that symmetric heating around the vortex center is generally more conducive for TC intensification than is asymmetric heating. Finally, the results of Chen and Zhang (2013) suggest that convective bursts may be more likely to produce RI if the resultant warming that those bursts produce is concentrated in the upper levels where they might prove to be more efficient in reducing the surface pressure of the underlying TC as their modeling simulations of Hurricane Wilma (2005) suggested. It is worth noting, however, that Jiang (2012) found that while the existence of convective bursts increased the likelihood that a TC would undergo RI, the increases observed in their study were not that large, suggesting that the location and overall coverage, as well as other factors, likely play crucial roles in determining the degree to which the convective bursts contribute to RI.

The multiscale nature of RI described in the above studies has made the operational prediction of RI quite difficult (Elsberry et al. 2007). Thus, utilizing the observational findings of their study, KD03 derived the initial version of the Statistical Hurricane Intensity

Prediction Scheme Rapid Intensification Index (SHIPS-RII) that employed five large-scale predictors from the SHIPS model (DeMaria et al. 2005) to estimate the probability of RI (defined as a $\geq 15\,\mathrm{m\,s^{-1}}$ increase in maximum sustained wind in 24 h). In the original version of the SHIPS-RII, the probability of RI estimates was determined by assessing whether the magnitudes of various environmental and climatological predictors fell above or below previously determined thresholds values that represented the RI sample–mean magnitudes of each of the RI predictors. The SHIPS-RII was first employed operationally by the NHC commencing with the 2004 Atlantic hurricane season while a SHIPS-RII developed for the eastern North Pacific basin was first implemented for the 2006 hurricane season. Over the next few years, new versions of the SHIPS-RII that included additional RI predictors and more sophisticated statistical methods were developed and the current linear-discriminant analysis version of the SHIPS-RII described in KDK10 was implemented operationally at the NHC prior to the 2008 hurricane season.

Although the SHIPS-RII is currently an operational forecasting tool at the NHC, its utility is somewhat restricted since it was developed for a single (24 h) lead time. Furthermore, its skill is somewhat limited, particularly in the Atlantic basin (KDK10). This paper presents an effort to improve the overall usefulness of SHIPS-RII by enhancing the model in both the Atlantic and eastern North Pacific basins. First, an updated version of the SHIPS-RII that employs more near-storm information is derived. Second, new consensus-based RI model forecasts that employ the current SHIPS-discriminant RII as well as the Bayesian and logistic regression RI models described in Rozoff and Kossin (2011, hereafter RK11) are developed. To provide additional guidance for the critical 48-h period during which the NHC issues watches and warnings, these new models are developed for the current 24-h operational forecast lead time, as well as for the added lead times of 12, 36, and 48 h. Finally, new versions of the rapid intensification aid (Sampson et al. 2011) that provide deterministic intensity forecasts utilizing the combination of conventional operational intensity forecast models and the new multi-lead-time probabilistic RI model guidance are derived. It is worth noting that a parallel effort to develop a microwave imagery–based version of the RI model is also under way (Rozoff et al. 2015). Section 2 of this manuscript provides a description of the new multi-lead-time RI models while a probabilistic and deterministic verification of their skill is provided in section 3. Finally, a summary of the current study's findings as well as some concluding remarks are offered in section 4.

2. Development of new multi-lead-time RI models

a. Enhanced SHIPS-RII model

1) Predictor selection

As noted above, KDK10 indicate that while the current operational version of SHIPS-RII generally exhibits some skill, the amount of skill is rather limited. This is particularly true in the Atlantic basin. Thus, predictors derived from three new sources are tested for their potential to improve the skill of the existing operational SHIPS-RII. First, a multiyear dataset of total precipitable water (TPW) developed by Remote Sensing Systems (RSS) and the National Environmental Satellite, Data, and Information Service (NESDIS) (Kidder and Jones 2007) is tested for its potential to improve the SHIPS-RII. The RSS–NESDIS TPW dataset is derived from a unified, physically based algorithm that utilizes the 19-, 22-, and 37-GHz channels on the constellation of Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave Imager (SSM/I) and Special Sensor Microwave Imager/Sounder (SSM/IS) satellites, the NASA *Aqua* Advanced Microwave Scanning Radiometer for EOS (AMSR-E), and the NASA Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI). Next, principal-component-analysis-derived channel-4 (10.7 μm) infrared (IR) imagery (PCIR) predictors computed from Geostationary Operational Environmental Satellite (GOES-East and GOES-West) data are examined following the methods described in Knaff (2008). Finally, boundary layer predictors derived from the National Centers for Environmental Prediction (NCEP) global operational gridded 2° × 2° (2.5° × 2.5° prior to 1996) latitude–longitude low-level temperature and moisture fields are examined for their ability to improve the SHIPS-RII. A brief description of the enhanced Atlantic and eastern North Pacific versions of SHIPS-RII developed utilizing the above new data sources is provided below.

As a first step in screening predictors derived from the aforementioned three new data sources, each is subjected to statistical significance testing for a homogenous sample of cases obtained from the 1995–2012 SHIPS Atlantic and eastern North Pacific developmental databases (DeMaria et al. 2005). Following KDK10, predictors whose sample-mean RI and non-RI[1] values are found to be statistically different at the ≥99.9% level for at least one lead time based upon a standard two-sided Behrens–Fisher t' test (Dowdy and Wearden 1991) are

[1] The non-RI stratification includes slowly intensifying, weakening, and steady-state systems.

TABLE 1. The predictors used in the current operational SHIPS-RII. Predictors are averaged every 6 h over the specified lead time (time avg) unless denoted as being either evaluated at the initial time ($t = 0$ h) or during the previous 12 h. For a more complete description of the methods used to compute the predictors listed in Table 1, consult KDK10.

Predictor	Definition
PER	Previous 12-h intensity change
SHRD	850–200-hPa vertical shear within a 500-km radius after vortex removal (time avg)
D200	200-hPa divergence within a 1000-km radius (time avg)
RHLO	850–700-hPa relative humidity within a 200–800-km radius (time avg)
PX30	Percentage of area with −30°C GOES-IR brightness temp ($t = 0$ h) within a 50–200-km radius
SDBT	Std dev of GOES-IR brightness temp ($t = 0$ h) within a 50–200-km radius
POT	Potential intensity (current intensity − max potential intensity) (time avg)
OHC	Oceanic heat content (time avg)

examined for their ability to increase the skill of the operational RII. These predictors are then substituted for several of the existing operational SHIPS-RII predictors shown in Table 1 to determine if they improve the skill of SHIPS-RII.

Specifically, a set of new TPW predictors that are correlated with RI are tested as replacements for the 850–700-hPa relative humidity (RHLO) since both the TPW and RHLO predictors are measures of atmospheric moisture. Similarly, the PCIR predictors are tested as replacements for the percent area within a 50–200-km radius with GOES-IR brightness temperatures < -30°C (PX30), as each provides a measure of the TC inner-core convective organization. Finally, low-level NCEP temperature and moisture variables are tested as replacement predictors for the potential intensity (POT) and oceanic heat content (OHC), as each is related to boundary layer processes.

Sensitivity tests are then performed to determine if any of the new replacement predictors increase the skill of the operational SHIPS-RII when substituted for existing RI predictors. The change in skill of each of the new replacement predictors is assessed by comparing the average skill of the new enhanced version of SHIPS-RII to that obtained utilizing the current version for the three original operational RI thresholds [25, 30, and 35 knots (kt; where 1 kt = 0.51 m s^{-1}) in 24 h] described in KDK10, as well as an additional threshold of 40 kt in 24 h.[2]

[2] The RI threshold of 40 kt in 24 h for SHIPS-RII was developed and implemented at the request of the NHC prior to the 2010 hurricane season.

Based on the aforementioned sensitivity tests, three new predictors are selected for use in the new enhanced versions of SHIPS-RII. The first of these predictors is the percentage of the area within a 500-km radius of the storm center and within ±45° of the upshear SHIPS wind direction with TPW < 45 mm at time $t = 0$ h. This predictor replaces RHLO in the enhanced SHIPS-RII. RI is favored when this predictor is small and hence the amount of dry air that is being advected into the storm circulation is relatively low. The cutoff of 45 mm as a delineator for dry air is based on the results of Dunion (2011). Figure 1 shows the distribution of TPW on a select day during the 2003 hurricane season as well as an example of the real-time blended NESDIS TPW product (Kidder and Jones 2007). Note that the blue and green areas (TPW $<$ 45 mm) depict regions where the atmosphere is relatively dry between the surface and 500 hPa (from where 90%–95% of the contribution from TPW comes). The orange and red areas (TPW > 45 mm) indicate where the lower to midlevels of the atmosphere are relatively moist.

The second new RI predictor is the second of the nine PCIR predictors (PC2) derived as described in Knaff (2008). This predictor replaces the PX30 predictor employed in the current operational SHIPS-RII. Figure 2 depicts the favored overall pattern for PC2 as well as an example of what this pattern looked like just prior to Hurricane Wilma's record-setting period of RI. Figure 2 shows that convection tends to be enhanced in the left-front quadrant while being suppressed in the right-rear quadrant near the time RI commences. This pattern often precedes the axisymmetrizaton of the overall convective pattern (Knaff 2008). This suggests that systems that undergo RI are more likely to move into regions where the thermodynamic environment is relatively favorable since TCs in both basins generally move to the west-northwest (not shown), toward the region of enhanced convection shown in Fig. 2.

The third new predictor is the inner-core dry-air predictor (ICDA), which is given by

$$(q_{10_{\text{layer}}} - q_{10})\text{VMX0}, \tag{1}$$

where q_{10} is the inner-core specific humidity at 10 m derived from the 1000-hPa NCEP temperature and relative humidity (RH) from a 200–800-km radius; $q_{10_{\text{layer}}}$ is the 10-m specific humidity obtained using the ambient 200–800-km radius, 1000-hPa temperature, and the layer-mean RH from 1000 to 500 hPa; and VMX0 is the NHC maximum sustained wind at $t = 0$ h. The value of q_{10} is obtained by bringing down q at 1000 hPa to the surface (dry adiabatically if unsaturated at 1000 hPa and moist adiabatically if the air is saturated) and then allowing the air to cool assuming that the RH reaches 95%

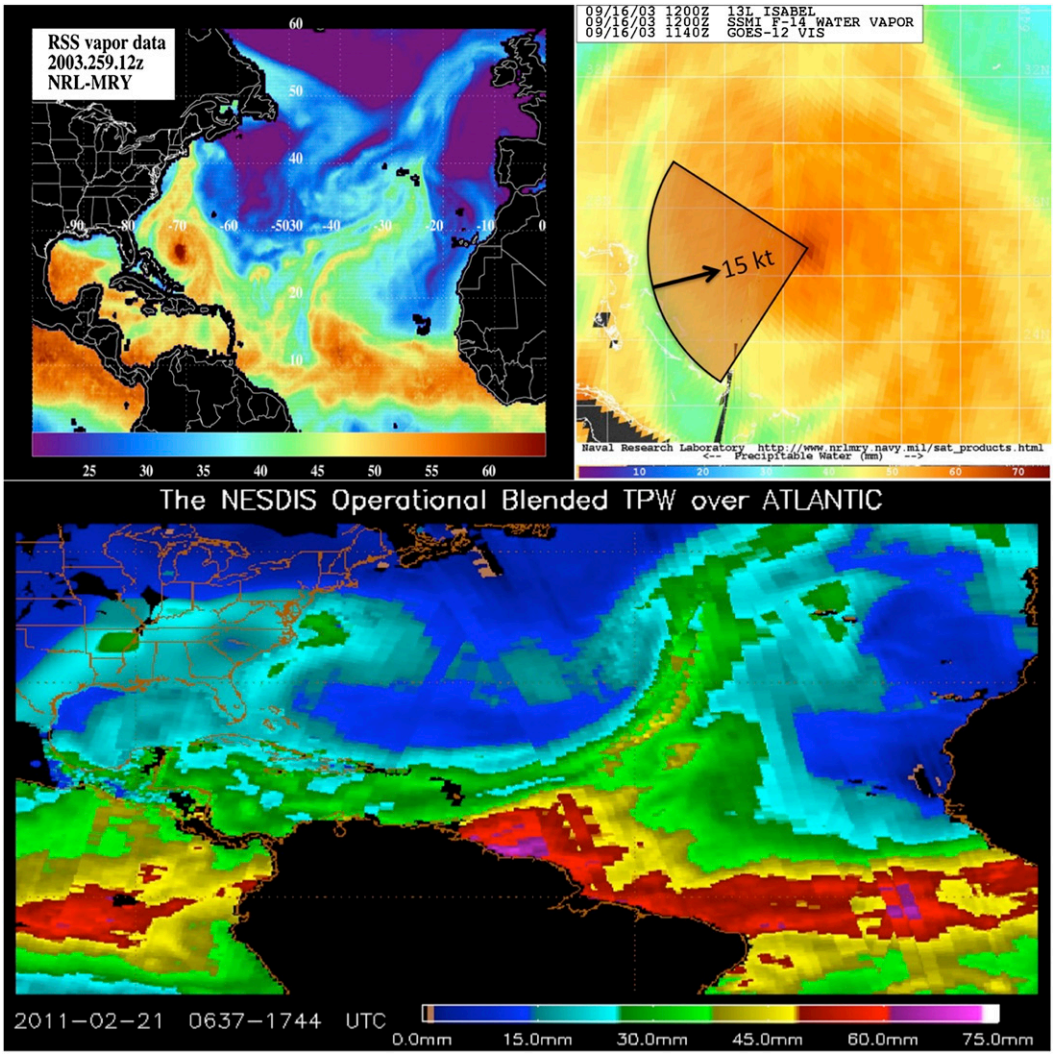

FIG. 1. (top left) TPW analysis at 1200 UTC 16 Sep 2003 (courtesy of the Naval Research Laboratory). (top right) A close-up view of the area over which the TPW predictor is computed for Hurricane Isabel (2003), several hundred kilometers east of Florida. (bottom) An example of the blended real-time NESDIS TPW analysis product.

as the parcel spirals into the storm core (Cione and Uhlhorn 2003). The value of $q_{10_{\text{layer}}}$ is obtained following the same methodology as described above except that the layer-mean RH between 1000 and 500 hPa is used in place of the 1000-hPa RH. It should be noted that RI is favored for small values of ICDA, which indicates less potential for dry air to mix down to the surface. Although initially tested as a replacement for POT and/or OHC, sensitivity tests show that the skill of the enhanced SHIPS-RII is maximized when all three predictors (i.e., OHC, ICDA, and POT) are employed. Finally, adding VMX0 as a 10th predictor improved the overall skill of the model even further. Table 2 shows the 10 predictors that are included in the new enhanced Atlantic and eastern North Pacific SHIPS-RII.

As noted above, the original version of the operational SHIPS-RII was developed exclusively for a 24-h lead time. In an effort to provide added RI guidance to NHC forecasters for the critical watch and warning period that was recently extended to 48 h, additional versions are developed for the 12-, 36-, and 48-h lead times. Following the methodology used to develop the current operational 24-h lead-time version of SHIPS-RII and using the 1995–2012 SHIPS developmental database, new SHIPS-RII models are developed for RI thresholds that correspond to approximately the 95th percentile of the overwater intensity change of all subtropical and tropical cyclones that formed during the 1995–2012 time period at the 12-, 36-, and 48-h lead times. Table 3 indicates that the 95th percentile intensity change levels of

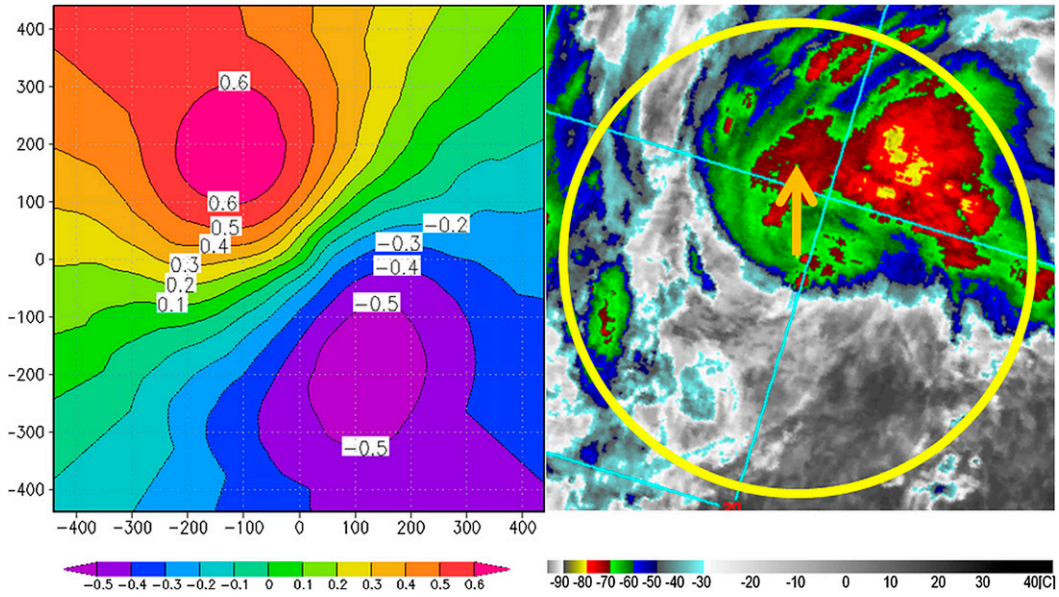

FIG. 2. Preferred pattern of (left) PC2 and (right) an example of the corresponding GOES-IR representation for Hurricane Wilma at 1800 UTC 17 Oct. The yellow circle denotes the 440-km radius region over which the PCIR predictors were evaluated. The magnitude of the PC2 is −1.32. Note that the direction of motion is to the upper center. The relative distances from the storm center (km) are shown along the *x* and *y* axes (left) while the GOES-IR brightness temperatures (°C) are depicted along the *x* axis (right).

TCs at the above lead times are quite similar in both the Atlantic and eastern North Pacific basins. Thus, for consistency, the same RI thresholds are employed when deriving the RI models for both basins. Specifically, new versions of SHIPS-RII are developed for the 20-kt RI threshold at 12-h lead time, the 45-kt threshold at 36-h lead time, and the 55-kt threshold at 48-h lead time in each ocean basin. Thus, versions of the SHIPS-RII are derived for a total of seven RI thresholds, since versions of that model had been previously developed for the four operational RI thresholds of 25, 30, 35, and 40 kt at the 24-h lead time.

Tables 4 and 5 show the mean and standard deviation of each RI predictor composing the new SHIPS-RII for the Atlantic and eastern North Pacific RI and non-RI samples at four RI thresholds [i.e., 20 kt $(12\,\text{h})^{-1}$, 30 kt $(24\,\text{h})^{-1}$, 45 kt $(36\,\text{h})^{-1}$, and 55 kt $(48\,\text{h})^{-1}$]. The sample statistics for only these specific RI thresholds are presented since each corresponds to approximately the same 95th percentile of intensity change at the four lead times. Note that these sample averages are obtained using only those cases for which the POT equals or exceeds that of the relevant RI threshold so as to better isolate the differences in the environmental conditions for systems that had an opportunity to undergo RI. It can be seen that each of the predictors is generally statistically significant at ≥95th percentile at most lead times with the majority exhibiting statistical significances at the ≥99th percent level.

2) MODEL DERIVATION

As described in KDK10, prior to deriving the discriminant SHIPS-RII, the magnitude of each of the RI predictors is first scaled between 0 and 1, with the latter representing conditions that are most conducive to RI. To illustrate, for the 30-kt Atlantic RI developmental

TABLE 2. As in Table1, but for the predictors used in the new enhanced SHIPS-RII model. Predictors that either replaced or have been added to those in those in the existing operational SHIPS-RII are shown in boldface.

Predictor	Definition
PER	Previous 12-h intensity change
SHRD	850–200-hPa vertical shear within a 500-km radius after vortex removal (time avg)
D200	200-hPa divergence within a 1000-km radius (time avg)
TPW	Percentage of area with TPW < 45 mm within a 500-km radius and ±45° of the upshear SHIPS wind direction (t = 0 h)
PC2	Second principal component of GOES-IR imagery within a 440-km radius (t = 0 h)
SDBT	Std dev of GOES-IR brightness temp (t = 0 h) within a 50–200-km radius
POT	Potential intensity (current intensity − max potential intensity) (time avg)
OHC	Oceanic heat content (time avg)
ICDA	Inner-core dry-air predictor (time avg)
VMX0	Max sustained wind (t = 0 h)

TABLE 3. Percentile thresholds associated with the corresponding intensity changes for the Atlantic basin, with the eastern North Pacific in parentheses. Values of N and N_{RI} in the 1995–2012 development samples from which each of the RI thresholds is derived is also provided.

RI threshold	N	N_{RI}	Probability of RI (%)
20 kt (12 h)$^{-1}$	4387 (3218)	255 (206)	94.2 (93.6)
25 kt (24 h)$^{-1}$	3768 (2729)	448 (357)	88.1 (86.9)
30 kt (24 h)$^{-1}$	3768 (2729)	286 (237)	92.4 (91.3)
35 kt (24 h)$^{-1}$	3768 (2729)	168 (165)	95.6 (94.0)
40 kt (24 h)$^{-1}$	3768 (2729)	113 (116)	97.0 (95.8)
45 kt (36 h)$^{-1}$	3224 (2293)	166 (147)	94.9 (93.6)
55 kt (48 h)$^{-1}$	2767 (1903)	148 (114)	94.7 (94.0)

sample, the SHRD of the RI cases ranges from 1.5 to 14.8 m s^{-1}. Since Tables 4 and 5 indicate that, on average, smaller values of SHRD are more conducive to RI than larger values, cases with SHRD $\leq$ 1.5 m s^{-1} are assigned a scaled value of 1 while those with SHRD $\geq$ 14.8 m s^{-1} are assigned a scaled value of 0 with a linear variation in the scaled magnitude of SHRD assumed for cases where 1.5 $\leq$ SHRD $\leq$ 14.8 m s^{-1}. It is worth noting that the above scaling methodology is employed for each of the 10 RI predictors except for VMX0. For VMX0, the scaling methodology differs somewhat from that described above and in KDK10. Specifically, cases for which VMX0 is equivalent to the mean value of all of the RI cases in the developmental sample are assigned a scaled value of 1 with values of 0 assigned when its magnitude is equivalent to the minimum and maximum VMX0 values of the RI sample. The above methodology is utilized when scaling VMX0 rather than that described previously in KDK10 as it yields higher overall skill of the enhanced SHIPS-RII based upon sensitivity tests performed on the developmental sample. Physically, the methodology found to provide optimal scaling of VMX0 suggests that weak systems are not sufficiently organized to undergo RI while strong systems are less likely to experience RI since they are closer to their maximum potential intensity, as was previously shown in KDK10.

The scaled predictors obtained through the above methods are used to determine the discriminant analysis weights for each of the 10 RI predictors. However, as noted in KDK10, only cases with predictor magnitudes within the range of values for which RI is observed to occur are employed to compute the discriminant weights, since sensitivity tests indicate this methodology improves the overall skill of the RI model. To illustrate, cases with SHRD $>$ 14.8 m s^{-1} are not employed when obtaining the discriminant weights as no RI cases had values greater than this magnitude. Similar screening procedures are employed for the other nine RI predictors. All 10 discriminant weights, which are simply the magnitude of Fisher's linear-discriminant function vector (Wilks 2011), are multiplied by their scaled values ranging between 0 and 1 to obtain the magnitudes of the discriminant function. This discriminant function value is then used

TABLE 4. Atlantic RI and non-RI sample mean and std dev for the 20-, 30-, 45-, and 55-kt RI thresholds with RI (non-RI) sample sizes of 251 (3654), 284 (2710), 164 (1944), and 146 (1419), respectively.

		RI		Non-RI		
Predictor	Unit	Mean	Std dev	Mean	Std dev	RI − non-RI (mean)
20 kt (12 h)$^{-1}$						
PER	m s^{-1}	5.2	5.0	1.2	4.6	4.0[a]
SHRD	m s^{-1}	5.6	3.0	8.1	4.5	−2.5[a]
PC2		0.0	0.8	0.3	1.1	−0.3[a]
OHC	KJ cm^{-2}	54.5	34.6	37.0	30.6	17.5[b]
SDBT	°C	12.4	6.2	16.6	6.8	−4.2[a]
D200	10^{-7} s^{-1}	44.4	32.5	30.7	34.3	13.7[a]
VMX0	m s^{-1}	34.0	11.6	29.4	11.3	5.6[a]
ICDA		293.3	165.1	307.1	167.3	−13.8
POT	m s^{-1}	36.5	12.4	35.4	14.4	1.1
TPW	%	11.9	22.3	26.1	31.4	−14.2[a]
30 kt (24 h)$^{-1}$						
PER	m s^{-1}	4.5	4.3	1.4	4.4	3.1[a]
SHRD	m s^{-1}	5.2	2.2	8.1	4.3	−2.9[a]
PC2		−0.1	0.9	0.3	1.1	−0.4[a]
OHC	KJ cm^{-2}	53.0	32.1	38.2	29.7	14.8[a]
SDBT	°C	13.5	5.9	17.1	6.8	−3.6[a]
D200	10^{-7} s^{-1}	44.1	31.5	29.2	32.8	14.9[a]
VMX0	m s^{-1}	29.8	9.9	27.9	9.9	1.9[b]
ICDA		255.9	147.5	294.9	152.3	−39.0[a]
POT	m s^{-1}	41.0	11.9	37.3	12.8	3.7[a]
TPW	%	8.8	18.8	25.2	30.5	−16.4[a]
45 kt (36 h)$^{-1}$						
PER	m s^{-1}	4.2	4.7	1.7	4.1	2.5[a]
SHRD	m s^{-1}	4.9	1.7	7.8	3.8	−2.9[a]
PC2		−0.1	0.9	0.4	1.1	−0.5[a]
OHC	KJ cm^{-2}	55.4	31.4	41.9	28.7	13.5[a]
SDBT	°C	14.3	6.3	17.6	6.9	−3.3[a]
D200	10^{-7} s^{-1}	47.2	30.7	29.3	30.9	17.9[a]
VMX0	m s^{-1}	27.8	7.5	25.8	8.2	2.0[b]
ICDA		235.5	121.8	268.3	132.9	−32.8[a]
POT	m s^{-1}	43.8	8.7	40.9	10.4	2.9[b]
TPW	%	7.5	15.4	21.9	28.1	−14.4[a]
55 kt (48 h)$^{-1}$						
PER	m s^{-1}	4.1	2.6	1.8	4.1	2.3[a]
SHRD	m s^{-1}	4.9	1.7	7.6	3.4	−2.7[a]
PC2		−0.2	1.0	0.4	1.1	−0.6[a]
OHC	KJ cm^{-2}	55.7	31.1	43.3	27.6	12.4[a]
SDBT	°C	16.2	5.5	17.9	6.9	−1.7[a]
D200	10^{-7} s^{-1}	47.1	30.5	29.9	29.2	17.2[a]
VMX0	m s^{-1}	24.3	6.0	24.8	7.4	−0.5
ICDA		205.4	103.3	255.3	119.5	−49.9[b]
POT	m s^{-1}	47.3	7.7	42.6	8.9	4.7[a]
TPW	%	8.9	17.9	19.1	25.9	−10.2[a]

[a] Statistical differences between sample means at the 99.9th percent level.

[b] Statistical differences between sample means at the 99th percent level.

TABLE 5. As in Table 4, but for the eastern North Pacific basin at the 20-, 30-, 45-, and 55-kt RI thresholds with RI (non-RI) sample sizes of 206 (2903), 237 (2216), 147 (1647), and 114 (1254), respectively.

		RI		Non-RI		
Predictor	Unit	Mean	Std dev	Mean	Std dev	RI − non-RI (mean)
20 kt (12 h)$^{-1}$						
PER	m s^{-1}	8.1	4.6	0.2	5.6	7.9[a]
SHRD	m s^{-1}	4.1	1.7	6.4	3.8	−2.3[a]
PC2		−0.1	0.5	0.0	0.9	−0.1[c]
OHC	KJ cm^{-2}	31.9	15.8	16.3	16.3	15.6[a]
SDBT	°C	9.4	4.0	14.9	6.6	−5.5[a]
D200	10^{-7} s^{-1}	59.5	30.2	32.8	31.9	26.7[a]
VMX0	m s^{-1}	34.5	11.2	27.7	12.0	6.8[a]
ICDA		176.6	129.7	197.3	144.6	−20.7
POT	m s^{-1}	44.4	11.8	41.7	14.9	2.7[b]
TPW	%	2.8	9.0	10.8	20.4	−8.0[a]
30 kt (24 h)$^{-1}$						
PER	m s^{-1}	6.7	4.1	0.8	5.6	5.9[a]
SHRD	m s^{-1}	4.0	1.6	6.4	3.7	−2.4[a]
PC2		−0.1	0.5	0.0	0.9	−0.1[c]
OHC	KJ cm^{-2}	32.6	15.3	16.4	15.3	16.2[a]
SDBT	°C	11.1	5.0	15.5	6.7	−4.4[a]
D200	10^{-7} s^{-1}	59.4	28.7	33.1	30.2	26.3[a]
VMX0	m s^{-1}	27.7	7.6	25.8	9.0	1.9[a]
ICDA		157.8	107.6	193.5	136.9	−35.7[a]
POT	m s^{-1}	48.5	9.9	42.4	13.7	6.1[a]
TPW	%	2.0	7.6	9.2	18.8	−7.2[a]
45 kt (36 h)$^{-1}$						
PER	m s^{-1}	6.4	3.5	1.9	5.1	4.5[a]
SHRD	m s^{-1}	4.0	1.4	6.4	3.6	−2.4[a]
PC2		−0.2	0.6	−0.2	0.9	0.0
OHC	KJ cm^{-2}	34.4	14.9	18.1	14.5	16.3[a]
SDBT	°C	11.1	5.0	15.5	6.7	−4.4[a]
D200	10^{-7} s^{-1}	61.7	26.6	36.7	28.7	25.0[a]
VMX0	m s^{-1}	27.7	7.6	25.8	9.0	1.9[c]
ICDA		146.8	101.4	168.3	101.6	−21.5[c]
POT	m s^{-1}	51.9	8.3	45.5	11.2	6.4[a]
TPW	%	1.8	7.1	7.4	17.5	−5.6[a]
55 kt (48 h)$^{-1}$						
PER	m s^{-1}	5.7	3.1	2.4	4.6	3.3[a]
SHRD	m s^{-1}	4.1	1.3	6.3	3.4	−2.2[a]
PC2		−0.4	0.6	−0.2	0.9	−0.2[c]
OHC	KJ cm^{-2}	34.2	14.8	19.0	13.8	15.2[a]
SDBT	°C	12.3	5.0	15.6	6.8	−3.3[a]
D200	10^{-7} s^{-1}	62.6	26.5	37.9	27.1	24.7[a]
VMX0	m s^{-1}	25.0	6.0	24.9	8.1	0.1
ICDA		133.5	91.5	159.2	90.9	−25.7[b]
POT	m s^{-1}	54.6	6.6	47.0	9.8	7.6[a]
TPW	%	1.6	4.6	6.6	16.8	−5.0[a]

[a] Statistical differences between sample means at the 99.9th percent level.
[b] Statistical differences between sample means at the 99th percent level.
[c] Statistical differences between sample means at the 95th percent level.

to estimate the probability of RI, as described in detail in KDK10.

Figure 3 shows the relative predictor weights of the new enhanced SHIPS-RII for the four RI thresholds employed in Table 4. These relative weights are obtained by dividing the absolute weight of each of the individual RI predictors by the sum of all 10 individual predictor weights at each lead time and for all four lead times combined. Thus, the sum of the relative weights is equivalent to 1 at each lead time. It can be seen that variations in the relative weights of the individual predictors are observed as a function of lead time and basin. However, when ranked by their mean weight for the four lead times, their importance is generally similar in both basins. The exceptions to this rule are PC2, which has a much larger weight in the Atlantic than in the eastern North Pacific, and POT, which has a much larger weight in the eastern North Pacific. Interestingly, at the longer lead times (i.e., 48 h) when PER becomes much less important, SHRD has the largest relative weight of any of the environmental predictors in the Atlantic, while POT exhibits the largest relative weight in the eastern North Pacific. This is perhaps an indication that kinematic factors play a more important role in determining if an episode of RI will occur in the Atlantic during longer forecast intervals, while thermodynamic factors may be more of a controlling factor in the eastern North Pacific.

It can be seen that the relative weight of each RI predictor is either positive or very small in the Atlantic and eastern North Pacific. However, the relative weight of the ICDA variable is, on average, somewhat negative in the eastern North Pacific, suggesting that the existence of too much moisture in the lower to midtroposphere may actually lessen the likelihood of RI. Nevertheless, the sample-mean ICDA of eastern North Pacific RI cases (Table 5) is still, on average, less than is found for non-RI cases. This suggests too much dry air still likely hinders RI.

By physical reasoning, the response of a TC to environmental forcing is expected to be the same regardless of TC location. Therefore, one might expect that the relative weights of the RI predictors in Fig. 3 should be equivalent. However, there are a variety of factors that could contribute to the differences in relative weights found between the two basins. First, differences in the mean TC structure in the two basins could be a factor in the weighting differences. To illustrate, numerical TC simulations performed by Riemer et al. (2013) indicate that the radial profile of a TC's symmetric wind field is an important modulator of TC resiliency and ultimately TC change. For instance, TCs in the eastern North Pacific are about a third smaller than those in the Atlantic

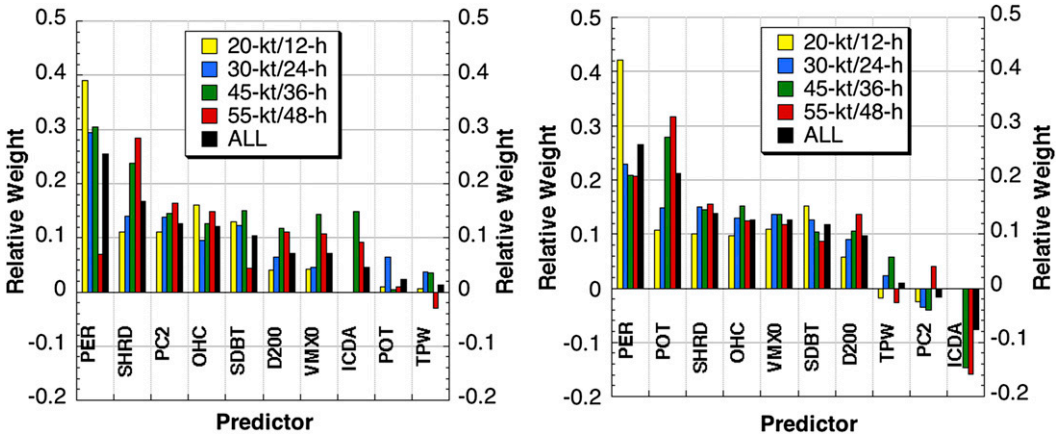

FIG. 3. Relative weights of the enhanced SHIPS-RII predictors (as defined in Table 2) for the (left) Atlantic and (right) eastern North Pacific basins for the 20- (yellow bars), 30- (blue bars), 45- (green bars), and 55-kt (red bars) RI thresholds at 12-, 24-, 36-, and 48-h lead times, respectively. The mean relative weight of each of the 10 RI predictors for all four RI thresholds (ALL) is also provided (black bars).

(Knaff et al. 2014). Thus, differences in TC structure that are not implicitly accounted for in SHIPS-RII could produce variations in the response of a TC to environmental forcing factors represented by SHIPS-RII predictors. Second, basinwide thermodynamic variations may also contribute to differences in the manner in which a TC responds to a given environmental forcing. Specifically, Tables 4 and 5 indicate that TCs in the eastern North Pacific are more often embedded in a favorable thermodynamic environment than Atlantic systems (i.e., higher POT, lower ICDA, and higher TPW). Thus, the intensity changes of TCs in each basin may be modulated by variations in the combined basinwide differences in TC structure and the thermodynamic conditions of the environment. Third, differences in the variability of certain conditions may contribute to the relative weighting differences in Fig. 3. For example, SHRD in the eastern Pacific is generally rather favorable since the SHRD of non-RI cases is only slightly higher than the SHRD in Atlantic RI cases. Consequently, other environmental factors such as variations in the thermodynamic environments and TC structures between the two basins may play an important role in differentiating between RI and non-RI cases for TCs experiencing similar kinematic forcing.

It is worth noting that DeMaria and Kaplan (1999) found substantial ocean basin differences in the relative importance of SHIPS model predictors such as vertical shear and potential intensity. Thus, while the precise reasons for the observed differences in coefficient magnitudes remain unclear, similar differences in predictor importance have been found in previous statistical TC intensity studies.

3) MODEL VERIFICATION

Figure 4 shows the skill of the existing operational and new enhanced versions of SHIPS-RII for the 1995–2012 developmental Atlantic and eastern North Pacific samples for all seven RI thresholds. Following KDK10, model performance is assessed using the Brier skill score (BSS) (Wilks 2011) for all overwater cases that remain tropical or subtropical during the forecast period as determined from the updated NHC best-track database (Landsea and Franklin 2013). In Fig. 4, both versions of SHIPS-RII exhibit skill relative to climatology, but the enhanced version generally has more skill than the current operational version. Specifically, the mean (maximum) absolute improvements for the seven RI thresholds of the enhanced SHIPS-RII over the current version are 1.3% (3.2%) and 0.8% (3.1%) for the Atlantic and eastern North Pacific, respectively, which correspond to relative mean (maximum) improvements of 3.2% (25.6%) and 3.1% (11.9%) for those same two basins.

Hamill (1999) suggested that a simple paired t test is an appropriate method for performing significance tests of probabilistic forecasts. Thus, in an attempt to gauge the relative significance of the above skill improvements, a paired two-sample t test (Wilks 2011) is conducted between the Brier skill scores of the enhanced and current operational versions of SHIPS-RII for each of the RI thresholds in each basin. The results of the paired t test indicate that the improvements of the enhanced SHIPS-RII over the current operational version are statistically significant at the $\geq$90% level for the 20-, 35-, and 45-kt RI thresholds in the Atlantic and the 40-kt RI threshold in the eastern

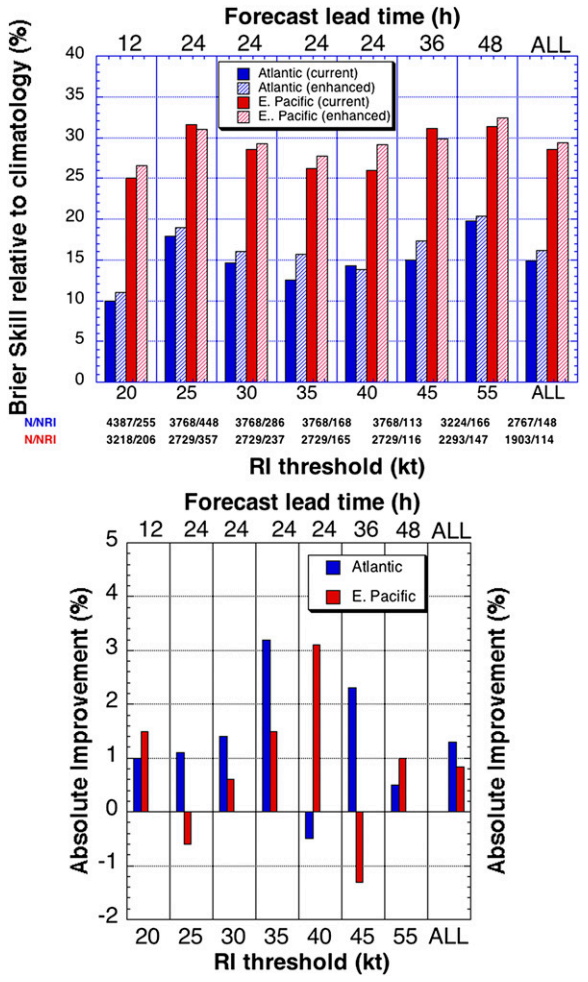

FIG. 4. (top) BSS relative to climatology of the current (solid bars) and enhanced (hatched bars) versions of the SHIPS-RII for the Atlantic (blue) and eastern North Pacific (red) basins at each lead time and RI threshold and for all seven of the RI thresholds combined (ALL) (upper x axis) for the 1995–2012 developmental sample. The total number of cases N and RI cases N_{RI} at each lead time are provided for the Atlantic (blue) and eastern Pacific basins (red) (lower x axis). (bottom) The improvements in skill of the enhanced SHIPS-RII over the current version are also provided.

North Pacific and that none of the degradations in skill of the enhanced version relative to the current version are statistically significant at that same level of significance in either basin. It should be noted that Wilks (2011) suggests that results of the paired t test tend to underestimate the significance of the differences of the relevant test statistic between the two samples; consequently, the results presented here likely represent somewhat of an underestimate of the significance of the improvements discussed above. Thus, while the overall improvements in skill of the new enhanced version of SHIPS-RII discussed above are somewhat modest, the new enhanced SHIPS-RII does generally provide some improvements in skill relative to the current operational version. Consequently, it will be used to obtain the results presented in the remainder of this study.

b. *Derivation of new consensus RI models*

In RK11, two new probabilistic RI models were developed based upon a logistic regression (LR) model and a naïve Bayesian classifier technique. Such models were derived for both the Atlantic and eastern North Pacific. These new RI models were then run in parallel with the discriminant analysis–based SHIPS-RII of KDK10 to produce a model forecast consensus. The consensus is simply an average of the LR, Bayesian, and SHIPS-RII model probabilities. RK11 found that the consensus RI model provides more skill than any of individual RI models.

Similar to the original SHIPS-RII, the models in RK11 were developed for just the 24-h lead time. To generate a new consensus RI model to accompany the updated SHIPS-RII, additional versions of the LR and Bayesian models are developed here for the added lead times of 12, 36, and 48 h. Also, the 24-h lead-time versions of the RK11 models are updated to take advantage of an expanded developmental dataset.

In RK11, the choice of each model's predictors was the same for all RI thresholds in a given ocean basin, although the model predictors differed between ocean basins. In the current study, the RI predictors are now allowed to also vary between RI thresholds and lead times in order to optimize cross-validated BSS further than is possible by fixing model predictors across the various definitions of RI.

In seeking optimal predictors for each model, the entire developmental dataset is included in the predictor search. Some of the predictors in Tables 1 and 2 that compose the SHIPS-RII are also chosen for the LR and Bayesian RI models. Ultimately, however, a number of new optimally chosen predictors from the developmental dataset are needed as well and are summarized in Table 6. The complete sets of model predictors for the Atlantic and eastern North Pacific basins are summarized in Tables 7 and 8. In these tables, a sense of the relative difference between mean RI and non-RI predictor magnitudes is also provided. For the predictors in common with SHIPS-RII, the relative differences are consistent with the results shown in Tables 4 and 5. The new predictors also behave in accordance with well-known climatological studies of TC intensification. In the succeeding section, an evaluation of the real-time performance of the new multi-lead-time consensus and constituent model forecasts is provided.

TABLE 6. As in Table1, but for new predictors used in the LR and Bayesian RI models.

Predictor	Definition
SHRD2	850–200-hPa vertical shear within a 200–800-km radius (time avg)
SHRG	Generalized 850–200-hPa vertical shear magnitude that takes all vertical levels into consideration (time avg)
DIVC	200-hPa divergence within a 1000-km radius, using a center based on the 850-hPa vortex center location (time avg)
U200	200-hPa zonal wind (time avg)
EPSS	Equivalent potential temp excess of a parcel lifted from the surface and the saturated equivalent potential temp of the environment within a 200–800-km radius (time avg)
ENSS	Vertical avg negative difference between the equivalent potential temp of a parcel lifted from the surface and the saturated equivalent potential temp of the environment within a 200–800-km radius (time avg)
TPWC	Azimuthally averaged TPW within a 200-km radius ($t = 0$ h)
TPWW	Percentage of area with TPW < 45 mm, within a 200-km radius, western quadrant of TC ($t = 0$ h)
TPWE	Percentage of area with TPW < 45 mm, within a 200-km radius, eastern quadrant of TC ($t = 0$ h)
TPWL	Percentage of area with TPW < 45 mm, within a 200-km radius, left quadrant relative to TC motion ($t = 0$ h)
TPWB	Percentage of area with TPW < 45 mm, within a 200-km radius, back quadrant relative to TC motion ($t = 0$ h)
TPWOR	Percentage of area with TPW < 45 mm, within a 400–600-km radius, right quadrant relative to TC motion ($t = 0$ h)
TPWA	Percentage of area with TPW < 45 mm, within a 500-km radius, quadrant centered upshear ($t = 0$ h)
PC1	First principal component of GOES-IR imagery within a 440-km radius ($t = 0$ h)
DPC2	Trend in the second principal component of GOES-IR imagery within a 440-km radius during the previous 6 h
AVBT	Avg GOES-IR brightness temp ($t = 0$ h) within a 200-km radius
AVBT2	Avg GOES-IR brightness temp ($t = 0$ h) within a 100–300-km radius
SDBT2	Std dev of GOES-IR brightness temp ($t = 0$ h) within a 100–300-km radius
PX10	Percentage of area covered by -10°C GOES-IR brightness temp ($t = 0$ h) within a 50–200-km radius
PX20	Percentage of area covered by -20°C GOES-IR brightness temp ($t = 0$ h) within a 50–200-km radius
PX50	Percentage of area covered by -50°C GOES-IR brightness temp ($t = 0$ h) within a 50–200-km radius
TBMX	Max brightness temp within a 30-km radius ($t = 0$ h)
RSST	Reynolds sea surface temp (time avg)

3. RI model rerun forecast verification

Although the new multi-lead-time RI models described above were run in parallel mode at the Cooperative Institute for Research in the Atmosphere (CIRA) in Fort Collins, Colorado, commencing in August 2013, the 2013 hurricane season was unusually quiet in terms of RI events in both the Atlantic and eastern North Pacific. Thus, the new RI models are rerun for the period 2004–13 using the operational NCEP global model forecast fields and initial storm data archived for that 10-yr period in an effort to better evaluate the forecasting capabilities of the newly developed RI models. Prior to performing the 2004–13 RI model reruns, leave-one-year-out cross-validated (Wilks 2011) versions of each of the three RI models (Bayesian, SHIPS-RII, and logistic regression) are derived for all seven RI thresholds for each of the years that compose the 10-yr (i.e., 2004–13) sample. For example, the developmental data from the periods 1995–2003 and 2005–12 are used to rederive the 2004 RI models. The RI models derived by the leave-one-year-out cross-validation technique are then rerun for each of the 10 yr in the 2004–13 sample period using the archived operational NCEP forecast fields, GOES imagery, and NHC initial storm intensity and location data for that period. A probabilistic and deterministic verification of those cross-validated RI rerun model forecasts is provided below.

a. Probabilistic verification

Figure 5 shows the skill of the three RI models and their consensus as a function of forecast lead time and RI threshold for the Atlantic and eastern North Pacific for a homogeneous sample of subtropical and tropical rerun forecasts from 2004 to 2013. It can be seen that each of the three individual RI models generally exhibit skill (i.e., positive BSS relative to climatology) at all lead times and RI thresholds (save for the LR model at the 40-kt threshold) and that the model skill is appreciably larger in the eastern North Pacific basin than in the Atlantic. Overall, the performance of the individual models tended to be similar at the shorter (i.e., 12 and 24 h) lead times with the LR version performing somewhat better at the longer (i.e., 36 and 48 h) lead times for this 10-yr sample. Consistent with the dependent 24-h lead-time results in RK11, the RI model consensus generally outperformed the three individual RI models at each lead time in both basins. Interestingly, it can be seen that the skill of the RI models tended to be somewhat larger at the 48-h lead time than the 12-h lead time in both basins (particularly in the Atlantic), suggesting that RI may be more predictable

TABLE 7. Predictors utilized in the Atlantic multi-lead-time LR and Bayesian RI models. See Tables 1, 2, and 6 for an explanation of the predictor. A minus (plus) sign denotes whether a lower (higher) value is found to be more conducive for RI.

Predictor	Tendency for RI storms	RI threshold
Logistic regression		
PER	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
SHRD	−	40 kt (24 h)$^{-1}$
SHRG	−	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
DIVC	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$
RHLO	+	40 kt (24 h)$^{-1}$, 55 kt (48 h)$^{-1}$
U200	−	40 kt (24 h)$^{-1}$
TPWW	−	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$
TPWE	−	55 kt (48 h)$^{-1}$
TPWB	−	45 kt (36 h)$^{-1}$
PC2	−	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$
DPC2	−	40 kt (24 h)$^{-1}$
SDBT2	−	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 55 kt (48 h)$^{-1}$
PX10	+	45 kt (36 h)$^{-1}$
PX20	+	55 kt (48 h)$^{-1}$
PX50	+	20 kt (12 h)$^{-1}$
TBMX	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$
POT	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
RSST	+	20 kt (24 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$
Bayesian		
PER	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
SHRD	−	20 kt (12 h)$^{-1}$
SHRD2	−	40 kt (24 h)$^{-1}$
SHRG	−	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
D200	+	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
DIVC	+	20 kt (12 h)$^{-1}$
RHLO	+	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
U200	−	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
TPWC	+	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$
TPWOR	−	35 kt (24 h)$^{-1}$
TPWA	−	55 kt (48 h)$^{-1}$
PC1	−	45 kt (36 h)$^{-1}$
PC2	−	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 55 kt (48 h)$^{-1}$
SDBT	−	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$
PX30	+	20 kt (12 h)$^{-1}$
PX50	+	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
TBMX	+	45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
POT	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
OHC	+	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$

at longer forecast intervals. This finding may be due in part to a couple of different factors. First, RI events that occur over short lead times (i.e., 12 h) are more likely to be controlled by internal processes (e.g., convection) as opposed to environmental factors and thus might be less predictable (Judt et al. 2015) by the RI models using predictors derived primarily from large-scale environmental information. Second, since RI events usually occur over time intervals of less than 48 h but greater than 12 h (Kieper and Jiang 2012), the timing of when RI commences might be less critical to the predictability of RI at the longer lead times.

It is worth noting that while the RI predictors employed in each of the three individual RI models are selected separately so as to maximize the BSS of each model as was noted previously, there likely exists some correlation between the forecasts produced by each as there is an overlap in some of the predictors utilized in the various RI models. Nevertheless, inspection of a cross section of the forecasts produced by each of the RI models (not shown) indicates that for the same case there can sometimes be fairly large differences in the RI probabilities that are forecasted by each of the individual RI models. Thus, despite the similarities in some of the predictors employed, an appreciable degree of independence likely exists between the individual model RI probability forecasts.

Figure 6 provides reliability diagrams (Wilks 2011) depicting the forecasted versus observed RI probabilities

TABLE 8. Predictors used in the new eastern North Pacific basin LR and Bayesian multi-lead-time RI models. A minus (plus) sign denotes whether a lower (higher) value is found to be more conducive for RI.

Predictor	Tendency for RI storms	RI threshold
Logistic regression		
PER	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
SHRD	−	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
SHRD2	−	20 kt (12 h)$^{-1}$
DIVC	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
ENSS	−	25 kt (24 h)$^{-1}$
AVBT	−	35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$
SDBT2	−	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$
PX20	+	55 kt (48 h)$^{-1}$
PX50	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$
TBMX	+	25 kt (24 h)$^{-1}$
POT	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
Bayesian		
PER	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
SHRD	−	20 kt (12 h)$^{-1}$
DIVC	+	20 kt (12 h)$^{-1}$
EPSS	+	25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
TPWL	−	55 kt (48 h)$^{-1}$
AVBT2	−	55 kt (48 h)$^{-1}$
SDBT2	−	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$
PX30	+	45 kt (36 h)$^{-1}$
PX50	+	20 kt (12 h)$^{-1}$
POT	+	20 kt (12 h)$^{-1}$, 25 kt (24 h)$^{-1}$, 30 kt (24 h)$^{-1}$, 35 kt (24 h)$^{-1}$, 40 kt (24 h)$^{-1}$, 45 kt (36 h)$^{-1}$, 55 kt (48 h)$^{-1}$
RSST	+	20 kt (12 h)$^{-1}$

for both the Atlantic and eastern North Pacific basins. Note that points below the diagonal line denote forecasts that are overconfident (i.e., the forecasted probabilities exceed the observed) while those above that line represent underconfident (i.e., probabilities forecast are less than those observed) forecasts. These diagrams are constructed utilizing the consensus RI model forecasts since Fig. 5 indicates that these generally provide the most skill. The number of bins employed to construct the reliability diagrams for each RI threshold are chosen so that each generally contained no less than five cases, as this was subjectively determined to be a reasonable minimum number of cases to obtain representative reliability statistics.

Figure 6 shows the Atlantic RI model forecasts are fairly well calibrated (i.e., the forecasted probabilities are similar to the observed probabilities). However, the 30-kt RI threshold forecasts are somewhat noisy for probabilities above ~30% while the 40-kt RI threshold forecasts exhibit a marked tendency to produce underconfident forecasts at all but the lowest probabilities. It is worth noting that while the consensus RI forecast probabilities generally do not exceed 40% very often in the Atlantic basin, RI is likely to occur (i.e., observed probabilities exceed 50%) when such probabilities are forecast. Figure 6 also indicates that the eastern North Pacific RI model consensus forecasts are generally fairly well calibrated up to probabilities of around 40%–50%, after which they exhibit underconfidence, particularly for the 40- and 55-kt RI threshold forecasts. The lone exception is the 25-kt RI threshold forecasts that exhibit underconfidence at all forecasted probabilities. Nevertheless, compared to the Atlantic RI model consensus, the eastern North Pacific consensus is more capable of correctly identifying forecast periods when the probability of RI is comparatively high. It can be seen that the maximum RI probabilities typically forecasted by the consensus model are approximately 65% and 85% for the Atlantic and eastern North Pacific basins, respectively, although probabilities exceeding those values are sometimes forecast in those basins (see Fig. 7). Thus, it is possible that the addition of some type of bias correction might improve the consensus model's capability to predict higher RI probabilities.

The aforementioned findings suggest that environmental conditions associated with higher probabilities of RI are more prevalent in the eastern Pacific than in the Atlantic. Consequently, the ability to analyze and predict changes in inner-core structure may be more crucial to the prediction of RI in the Atlantic basin, where environmental conditions most conducive to RI tend to be less prevalent.

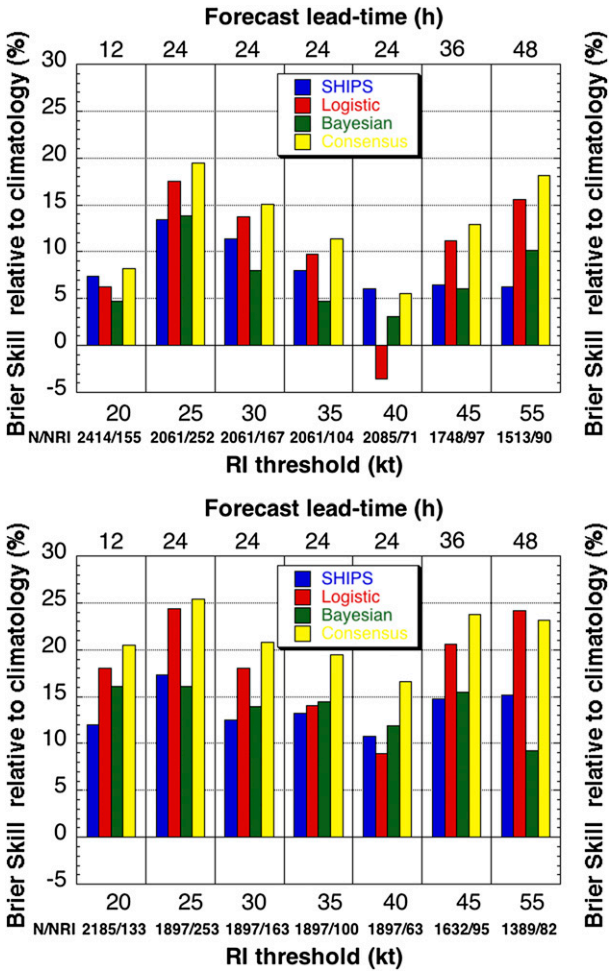

FIG. 5. BSSs of the (top) Atlantic and (bottom) eastern North Pacific 2004–13 RI rerun forecasts relative to the climatological probabilities of RI for each RI threshold. Skill is shown for the SHIPS (blue bars), logistic regression (red bars), Bayesian (green bars), and consensus (yellow bars) RI models. Values of N and N_{RI} are shown at each lead time (lower x axis). Note that sample sizes can vary at a given lead time as a result of the lack of availability of one or more of the RI models for a given lead time and RI threshold.

Figure 7 shows examples of the forecast performance of the consensus RI model for a few select systems that underwent RI. These systems are chosen since they represent cases for which the consensus model correctly forecasts relatively high RI probabilities as well as those for which the model failed to do so. It can be seen that the RI model correctly forecast relatively high RI probabilities for both Atlantic Hurricane Wilma and eastern North Pacific Hurricane Rick (2009) while failing to do so for Atlantic Hurricane Michael (2013) and eastern North Pacific Hurricane Flossie (2007). These examples underscore both the importance and limitations (Hendricks et al. 2010) of using environmental information to predict RI.

b. Deterministic evaluation

Although the RI models are formulated to produce probabilistic forecasts, the methodology described in KDK10 can be also be utilized to construct a deterministic forecast. To accomplish this, cutoff probabilistic RI thresholds that when exceeded would trigger an RI forecast need to be determined. In the KDK10 study, the cutoff probabilities were assumed to be equivalent to the climatological probability of false detection (also known as the false alarm rate; Wilks 2011) for each RI threshold in each basin as determined utilizing the SHIPS-RII developmental model probabilities. However, in the present study the cutoff probabilities are determined using the consensus RI model rerun forecasts since those were previously shown to provide the most skill (see Fig. 5). Additionally, the consensus model rerun forecasts from 2004 to 2007 (rather than those from the entire 2004–13 period) are employed when determining the probabilistic RI cutoffs to allow the consensus model to be used as an independent benchmark against which its performance, as well as that of the operational intensity guidance, can be evaluated for the period 2008–13.

Table 9 shows the RI forecast cutoff probabilities for each of the RI thresholds for both basins. For the conventional deterministic models, RI is forecast if the operational model-predicted intensity change determined from the Automated Tropical Cyclone Forecasting System (ATCF; Sampson and Schrader 2000) equals or exceeds the specified RI threshold for that basin. The accuracy of the RI forecasts is assessed utilizing the probability of detection (POD), the false alarm ratio (FAR), and the Peirce skill score (PSS), which are the same three metrics previously utilized in KDK10.

Briefly, the POD is the percentage of time that RI events are correctly identified so that a POD of 100% (0%) indicates that all (no) RI events are forecast correctly. The FAR is the percentage of time that RI is forecasted but is not observed so that an FAR of 0% (100%) indicates that all (no) forecasted RI events occurred. The PSS provides a means of quantifying the overall level of skill of contingency-type forecasts, particularly those made for infrequent events like RI, and ranges from -1 to 1 with perfect forecasts receiving a PSS of 1 and negative values not considered skillful. A more detailed description of the above three metrics can be found in KDK10 and Wilks (2011).

Using the above metrics for a homogenous set of Atlantic and eastern North Pacific forecasts from 2008 to 2013, the Atlantic and eastern North Pacific RI forecasts are evaluated (Figs. 8 and 9). The 2008–13 time

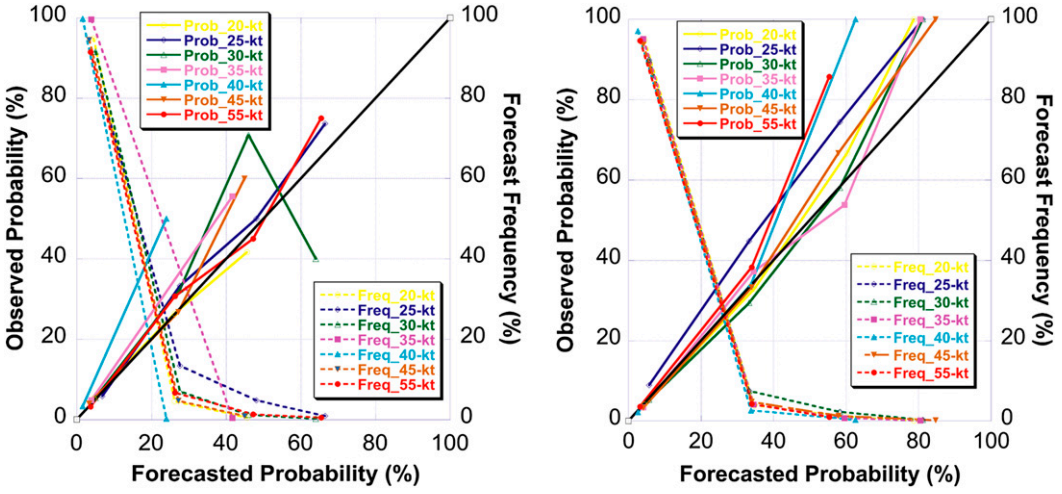

FIG. 6. Reliability diagrams for the (left) Atlantic and (right) eastern North Pacific consensus 2004–13 RI rerun forecasted (*x* axis) vs observed (left *y* axis) probability of RI for the 20- (yellow), 25- (dark blue), 30- (green), 35- (magenta), 40- (light blue), 45- (orange), and 55-kt (red) RI thresholds. The solid lines depict the probability of RI while the dashed lines of the same color show the frequency (right *y* axis) at which the RI forecasts are observed. The diagonal solid black line depicts forecasts with perfect reliability.

period is chosen since each of the intensity models described below is routinely available throughout that entire time period. The evaluations are performed for the 5-day SHIFOR model (SHF5; Knaff et al. 2003), the decay version of the SHIPS model (DSHP; DeMaria et al. 2005), the logistic growth equation model (LGEM; DeMaria 2009), the Geophysical Fluid Dynamical Laboratory (GFDL) hurricane prediction model early (GFDI) and late (GFDL) versions (Kurihara et al. 1998) and the Hurricane Weather Research and Forecasting Model early (HWFI) and late (HWRF) versions (Tallapragada et al. 2014), the NHC official forecast (OFCL), and the newly developed consensus RI Index model (CON-RII). Note that SHF5, DSHP, LGEM, HWFI, GFDI, and CON-RII are considered early model guidance since each typically would be available to the NHC forecasters at the time they are preparing their forecasts, while HWRF and GFDL are considered late models since they are not typically available until after the OFCL forecasts are issued. A more complete description of the distinctions between the late and early model guidance described above may be found online (http://www.nhc.noaa.gov/verification). Finally, it is important to note that although the CON-RII rerun forecasts were not available in real time to NHC forecasters, they are provided here to demonstrate their potential utility to operations as they employ the same early model input data as SHIPS while also serving as a baseline tool for evaluating the importance of large-scale environmental forcing on RI.

Figure 8 shows that the OFCL forecasts have higher PODs than any of the operational early models at the shorter (i.e., 12 and 24 h) lead times in the Atlantic basin while the HWFI forecasts exhibit the largest POD rates at the longer (i.e., 36 and 48 h) lead times. Encouragingly, the newly developed CON-RII model generally produces higher POD rates than any of the operational guidance. In terms of the FAR, the OFCL forecasts generally have the lowest values of any of the operational Atlantic early model guidance, particularly at the shorter lead times. It is worth noting that CON-RII produces FAR values that are generally comparable to those of the operational intensity guidance except for the 35- and 40-kt RI thresholds at 24-h lead time where the FAR is appreciably higher and for the 45- and 55-kt RI thresholds at 36- and 48-h lead times, respectively, where the CON-RII FAR values are slightly higher than those of the existing operational intensity guidance. This result likely demonstrates the limitations of employing mainly environmental information to predict the likelihood of the largest rates of intensification, particularly at shorter lead times. As noted earlier, inner-core processes that are not well represented by the RI models are likely more critical to the forecasting of the most rapid rates of intensification, especially at the shorter forecast time intervals.

When the operational Atlantic basin early model RI forecasts are evaluated according to PSS, which accounts for both the POD and FAR, the OFCL forecasts generally provide the most skill of any of the early model operational intensity guidance at 12- and 24-h lead times while the GFDI and HWFI forecasts exhibit the most skill at 36- and 48-h lead times. CON-RII generally

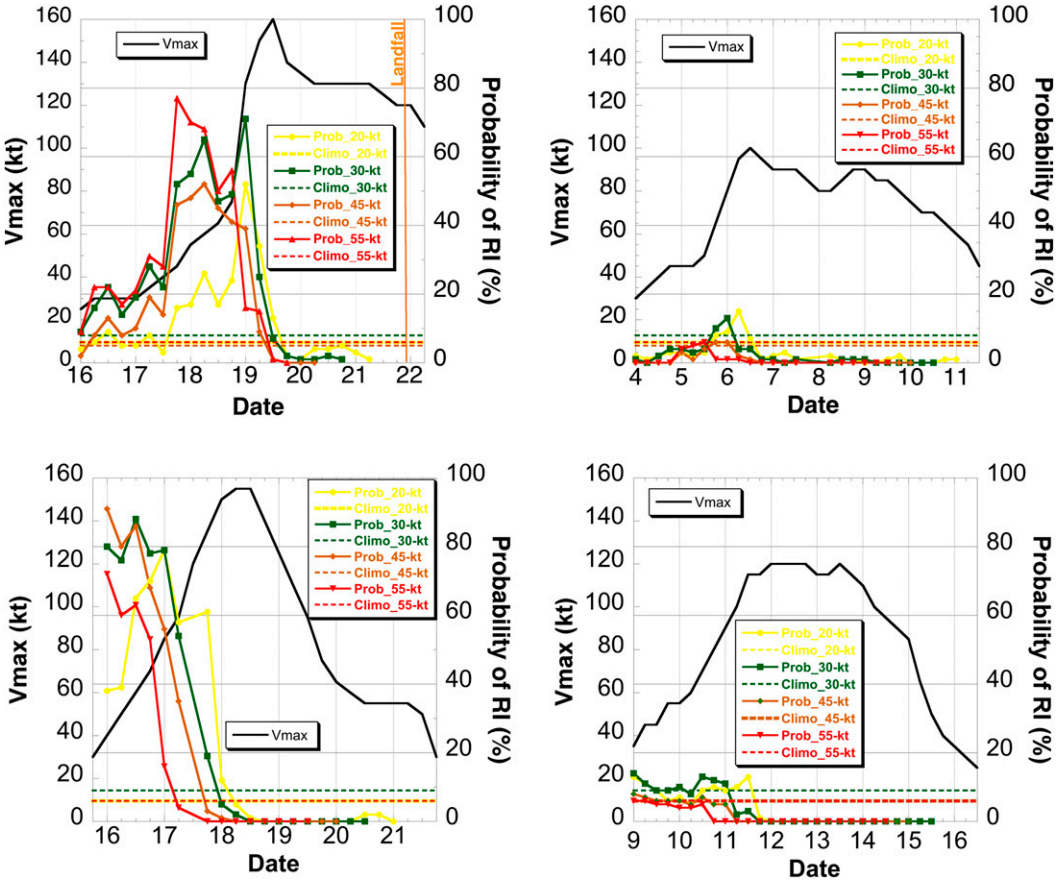

FIG. 7. The consensus 6-h RI rerun forecast probabilities for the 20- (solid yellow), 30- (solid green), 45- (solid orange), and 55-kt (solid red) RI thresholds (right *y* axis) for Atlantic hurricanes (top left) Wilma (2005) and (top right) Michael (2012) and eastern North Pacific hurricanes (bottom left) Rick (2009) and (bottom right) Flossie (2007). The horizontal dashed lines represent the climatological probability of RI for each threshold. The best-track NHC max sustained wind estimates (left *y* axis) are depicted by the solid black line. The forecasts are shown for all time periods when the systems remained over water and had a tropical or subtropical designation. The time at which Wilma made landfall along the Yucatan Peninsula is depicted by a vertical solid brown line.

exhibits somewhat higher PSS values than the operational early model guidance, particularly at the shorter lead times, suggesting that it has the potential to be a useful operational forecasting aid. It is also worth noting that the HWFI forecasts generally provide the most skill among the early operational intensity model guidance.

It can be seen that although the late models (HWRF and GFDL) generally produce higher POD rates than the early versions of those same models (HWFI and GFDI), they also tend to produce higher FAR values. Hence, the overall PSS of the late versions is, on average, only somewhat better than the early versions.

Inspection of the early eastern North Pacific operational guidance forecast verifications shown in Fig. 9 indicates that the OFCL forecasts generally provide the highest POD, particularly at the shorter lead times and lower RI thresholds, with the HWFI and OFCL forecasts exhibiting comparable POD values for the longer lead times and higher RI thresholds. Similar to the Atlantic, the CON-RII forecasts produce higher POD values than the operational intensity guidance, although the observed increase in POD of the CON-RII forecasts over that of the operational guidance is much more substantial for the eastern North Pacific than the Atlantic.

Generally, the OFCL forecasts produce the lowest FAR of any of the early operational guidance as was also observed in the Atlantic. The CON-RII FAR is comparable to or slightly higher than that of the operational intensity guidance except for the 35- and 40-kt RI thresholds for which it is observed to be appreciably higher, as was also seen in the Atlantic basin.

The PSSs of the OFCL forecasts generally exceed those of the other operational guidance, save for the

TABLE 9. Cutoff probability thresholds (%) for each of the seven RI thresholds determined from the Atlantic and eastern North Pacific consensus RI model 2004–07 rerun forecasts. The climatological probability of RI of the development sample is provided for reference.

RI threshold	Cutoff probability (%)	Climatological probability (%)
Atlantic		
20 kt (12 h)$^{-1}$	17	5.8
25 kt (24 h)$^{-1}$	26	12.0
30 kt (24 h)$^{-1}$	21	7.3
35 kt (24 h)$^{-1}$	17	4.3
40 kt (24 h)$^{-1}$	10	3.0
45 kt (36 h)$^{-1}$	22	5.1
55 kt (48 h)$^{-1}$	24	5.3
Eastern North Pacific		
20 kt (12 h)$^{-1}$	18	6.4
25 kt (24 h)$^{-1}$	18	12.8
30 kt (24 h)$^{-1}$	25	8.6
35 kt (24 h)$^{-1}$	21	6.0
40 kt (24 h)$^{-1}$	17	4.3
45 kt (36 h)$^{-1}$	21	6.4
55 kt (48 h)$^{-1}$	18	6.0

40- and 45-kt RI thresholds, where the HWFI exhibits the highest PSS values. However, consistent with the Atlantic basin the HWFI forecasts generally provide the most skill among the early model intensity guidance. Also, similar to the Atlantic findings, the PSSs of the CON-RII forecasts exceed those of the conventional operational intensity guidance. In the case of the eastern North Pacific, however, the CON-RII's PSS improvement over the operational guidance is significantly larger than in the Atlantic.

Finally, a comparison of the eastern North Pacific RI forecasts produced by the early and late numerical models shows results similar to those obtained for the Atlantic. Specifically, the late versions of the numerical models (GFDL and HWRF) possess higher POD and FAR values and, to a degree, higher PSS values, than the early (GFDI and HWFI) versions of those models.

The results of Figs. 8 and 9 indicate that the FAR values of both GFDL and HWRF are shown to be substantially larger for the late versions of those models when compared to the early versions particularly at the shorter lead times, perhaps suggesting that the numerical models may have a tendency to overintensify systems over the shorter forecast time intervals (i.e., 12–24 h). Inspection of these figures also indicates that the RI forecast guidance generally exhibits higher skill in the eastern Pacific than in the Atlantic. This is particularly true for CON-RII. The higher level of skill found in the eastern North Pacific relative to the Atlantic may indicate that RI is somewhat more predictable when the large-scale conditions are more favorable since Tables 4 and 5 indicate that the environmental conditions that form the basis of the RI models examined here are, on average, more conducive to RI in the eastern Pacific than in the Atlantic. In addition, the large-scale conditions themselves may be more predictable in the eastern North Pacific basin than in the Atlantic. Nevertheless, the overall finding of relatively low POD and a rather high FAR for all of the intensity guidance demonstrates the difficulty of predicting the timing and magnitude of the RI events.

Although the above results indicate that the predictive skill of RI remains rather limited, NHC's ability to predict such events appears to have increased somewhat in recent years, as is illustrated in Fig. 10, which depicts the changes in the PSSs of the OFCL forecasts between the periods 2004–07 and 2008–13. These two periods are chosen since the current study encompasses the aforementioned 10-yr time period and the HWRF and discriminant versions of SHIPS-RII both became operational and routinely available to NHC forecasters at the start of the 2008 hurricane season. Figure 10 indicates that based upon the PSS metric, there appears to have been some improvements in the ability of the NHC to forecast RI events in recent years, particularly in the eastern North Pacific. Specifically, the PSS of the OFCL forecasts improved by an average of about 6% (11%) in the Atlantic (eastern Pacific) with improvements of up to 10% (18%) observed in those same two basins over the above time period.

c. *RAPID guidance evaluation*

Sampson et al. (2011) developed a deterministic rapid intensification aid (IVRI; also known as RAPID) that employs probabilistic SHIPS-RII forecasts in conjunction with the existing intensity five-member consensus (IVCN). Sampson et al. (2011) suggested that the RAPID guidance be run when the SHIPS-RII forecasted probability of RI exceeds 40% for any of the 24-h lead-time operational RI thresholds. To generate the new deterministic IVRI forecast for such cases, the magnitude of the largest RI threshold for which the RI model forecast probability exceeds 40% is averaged together with the intensity changes forecasted by each of the five IVCN models. To illustrate, if the forecast 24-h SHIPS-RII probabilities for the 25-, 30-, 35-, and 40-kt RI thresholds are 50%, 45%, 35%, and 30%, respectively, then an intensity change of 30 kt is averaged together with the intensity change forecasts of the IVCN models to generate the new 24-h IVRI forecast since 30 kt represents the highest RI threshold for which the SHIPS-RII forecast probability exceeded 40%.

Results of the RAPID forecasts generated utilizing the 2008–13 operational IVCN and the multi-lead-time

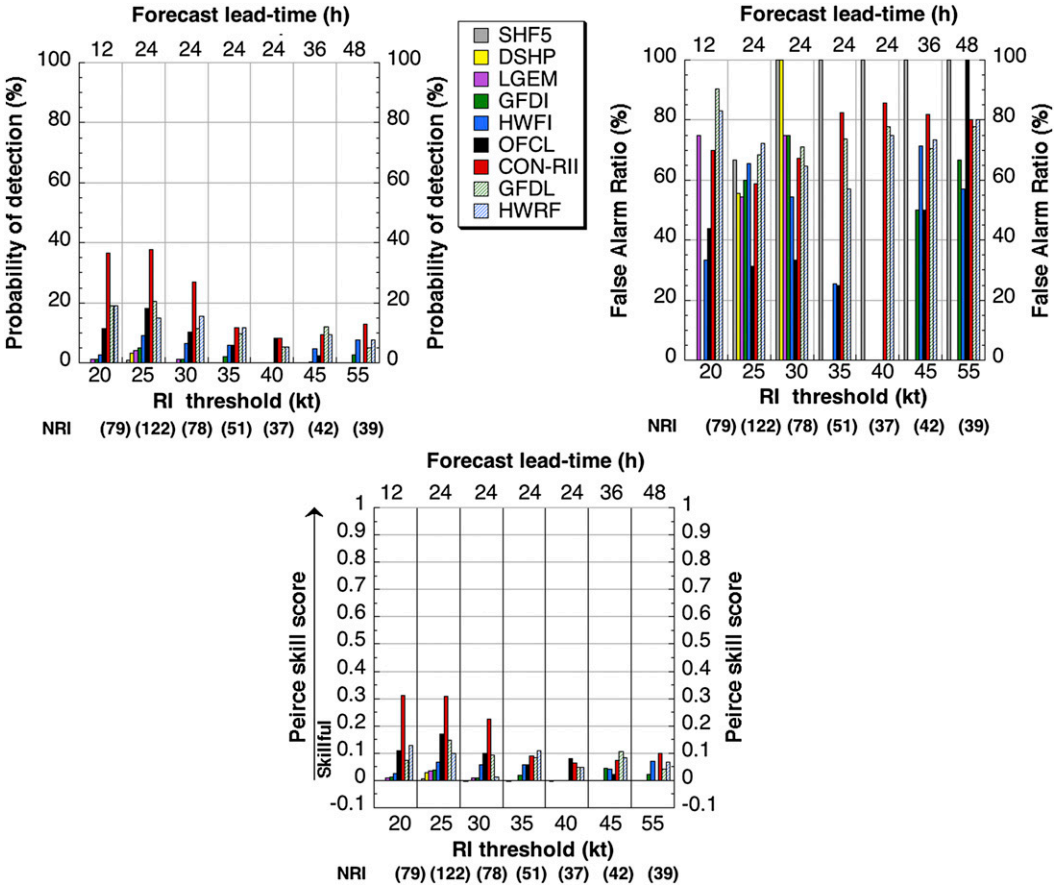

FIG. 8. (top left) POD, (top right) FAR, and (bottom) PSS of the Atlantic basin 2008–13 operational early model forecasts for SHF5 (gray bar), DSHP (yellow bar), LGEM (purple bar), GFDI (solid green bar), HWFI (solid blue bar), OFCL (black bar), and CON-RII (red bar) as a function of RI threshold (lower *x* axis) and forecast lead time (upper *x* axis). The results of the late model GFDL (hatched green bar) and HWRF (hatched blue bar) forecasts are also provided. The number of RI cases at each forecast lead time is provided (lower *x* axis).

CON-RII rerun forecasts are shown in Fig. 11. These results are obtained for all Atlantic and eastern North Pacific cases for which at least one CON-RII probability forecast exceeds 40% at one of the four (12, 24, 36, and 48 h) lead times. Although the number of cases is limited, indicating that the CON-RII forecast probabilities do not exceed 40% very often, it can be seen that when the probabilities exceed that cutoff probability threshold inclusion of the RAPID in the IVCN lowers the mean absolute forecast errors of the IVCN for the four lead times by about 8% and 6% in the Atlantic and eastern North Pacific, respectively, with reductions up to 13% and 8% observed for those same two basins. While these are modest reductions in the mean error, the IVRI produces more substantial decreases in the magnitude of forecast bias with mean reductions of 53% (18%) observed for the Atlantic and eastern North Pacific, respectively, and reductions of up to 94% and 26% observed for those same two basins.

It is notable that the OFCL forecasts outperform all of the objective intensity guidance for the first 24 h, indicating that the NHC forecasters made good use of the available operational intensity and RI guidance that existed for this time period. The improvements of the OFCL forecasts over the RI guidance shown here are consistent with recent basinwide operational intensity verification results conducted by the NHC for all (RI and non-RI) forecasts (see http://www.nhc.noaa.gov/verification) that indicate that the OFCL intensity forecasts typically provide more skill than do the individual intensity forecast models themselves. This result suggests that the human forecasting element typically yields added skill for operational intensity forecasts.

Nevertheless, it is encouraging that the aforementioned results indicate that the IVRI produces the smallest errors and biases of any of the guidance at the longer lead times (i.e., beyond 24 h) in the Atlantic basin and comparable or slightly higher errors in the eastern

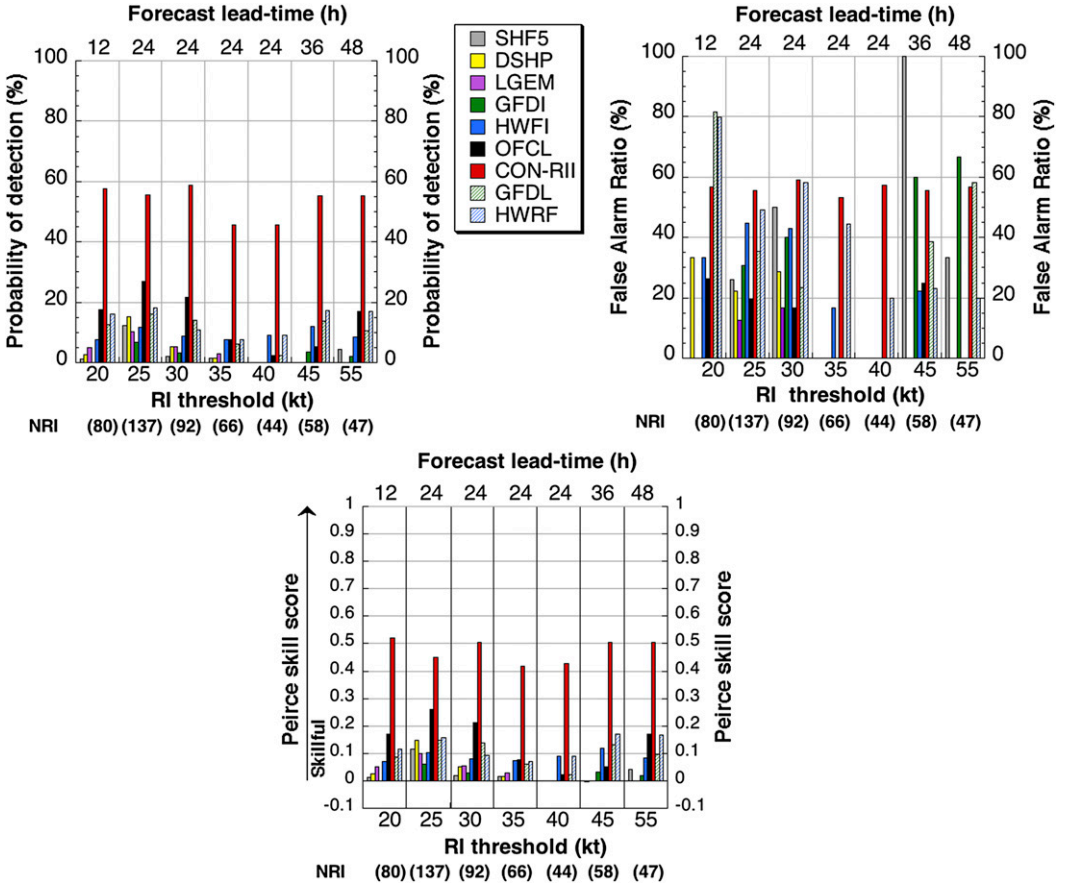

FIG. 9. As in Fig. 8, but for the eastern North Pacific basin.

Pacific even though the probability threshold employed to trigger the IVRI guidance is not tuned to the new multi-lead-time RI model guidance. Thus, further improvements in the overall performance of the IVRI are believed to be achievable after some additional refinements are made to this product. This result also indicates that the new multi-lead-time RI guidance should have a positive impact on the NHC official forecast when it becomes available operationally, since the multi-lead-time RI guidance improves the consensus when included in a very simple way in the IVRI technique.

4. Summary and conclusions

An enhanced version of the SHIPS-RII that includes several new and/or replacement predictors has been developed for the 24-h lead time for both the Atlantic and eastern North Pacific basins. These new predictors included the area within a 500-km radius of the storm center and within ±45° of the upshear wind direction with total precipitable water <45 mm, the second principal component of GOES-IR brightness temperature within a 440-km radius, and an inner-core dry-air predictor that provides a measure of the impact of the near-storm moisture from 1000 to 500 hPa. Finally, the maximum sustained wind at time $t = 0$ h was employed as an additional predictor in the new enhanced SHIPS-RII.

To provide added RI guidance to NHC forecasters for the critical watch and warning period that was recently extended to 48 h, additional versions of the newly developed enhanced SHIPS-RII were derived for 12-, 36-, and 48-h lead times. Following the methodology used to develop the current operational SHIPS-RII, new versions of the enhanced SHIPS-RII were formulated for RI thresholds representing approximately the 95th percentile of overwater intensity change of all subtropical and tropical cyclones. For the 1995–2012 developmental sample, the mean (maximum) absolute improvements of the new enhanced SHIPS-RII over the current operational version for all seven RI thresholds were 1.3% (3.2%) for the Atlantic and 0.8% (3.1%) eastern North Pacific basins, which corresponded to

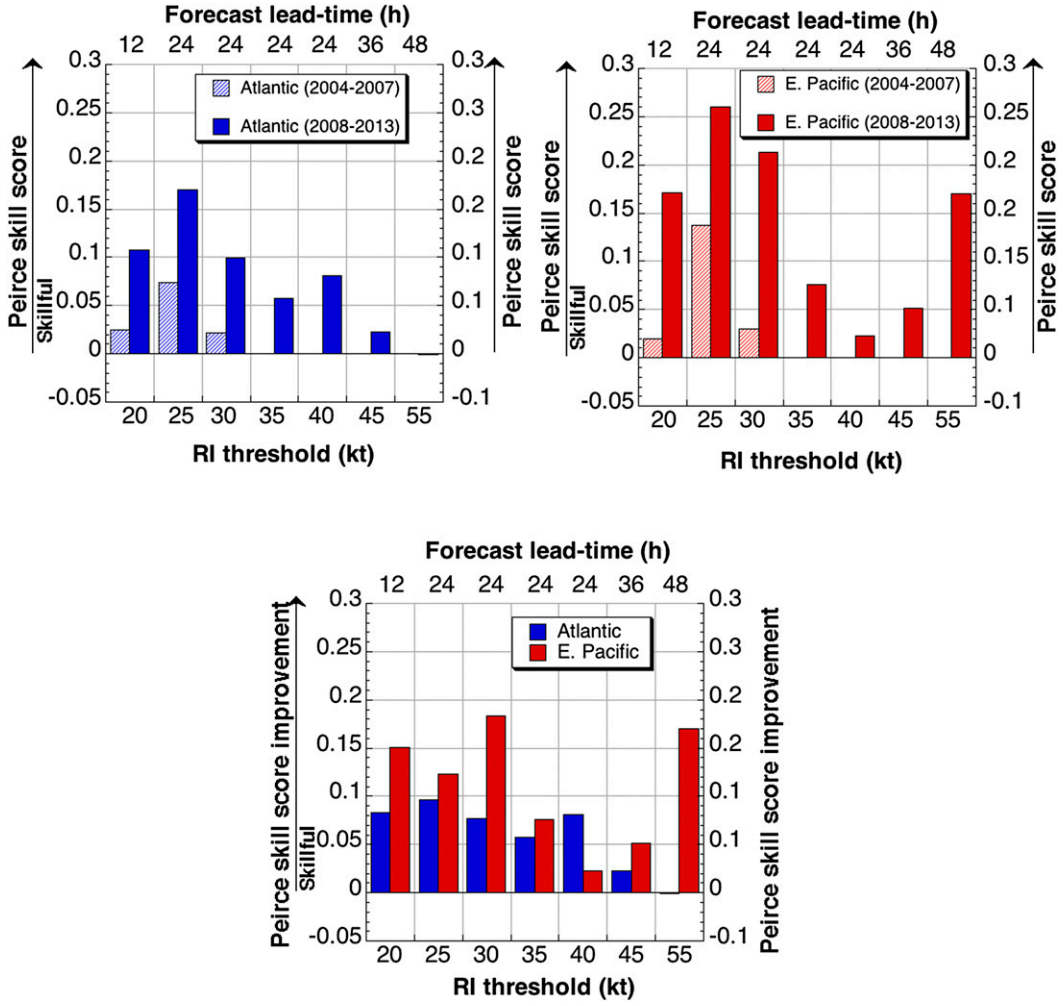

FIG. 10. PSS of the OFCL operational RI forecasts from 2004 to 2007 and 2008 to 2013 for the (left) Atlantic, (right) eastern North Pacific, and (bottom) improvement or degradation in forecast skill between these two periods.

mean (maximum) relative improvements of 8.3% (25.6%) and 2.9% (11.9%), respectively.

The order of the relative importance of the new enhanced SHIPS-RII predictors was found to be comparable in both the Atlantic and eastern North Pacific basins save for the second principal component of IR imagery and potential intensity variables that were observed to have much higher weights in the Atlantic and eastern North Pacific basins, respectively. It is also worth noting that at the longer lead times when persistence became less important, the vertical wind shear had the largest relative weight in the Atlantic while potential intensity had the largest weight in the eastern North Pacific. This may imply that kinematic factors play a more important role in determining the likelihood that an episode of RI will occur in the Atlantic and thermodynamic factors are more of a controlling factor in the eastern North Pacific.

New multi-lead-time versions of logistic regression (LR), Bayesian, and their consensus (the arithmetic average of the LR, Bayesian, and SHIPS-RII forecasts) RI model forecasts have also been developed for 12-, 24-, 36-, and 48-h lead times for both the Atlantic and eastern North Pacific basins. These new RI models generally exhibit skill at all lead times in both basins when evaluated for a cross-validated set of rerun forecasts using operational forecast data from the 2004–13 hurricane seasons. The consensus model generally provides the most skill. Interestingly, the skill of the RI models is typically higher at the longer lead times than the shorter ones, suggesting that predicting RI on shorter time scales may prove to be more difficult.

A verification of the deterministic RI forecasts shows that, while the operational intensity models and OFCL forecasts exhibit some skill in anticipating episodes of RI, the skill is quite limited, particularly in the Atlantic

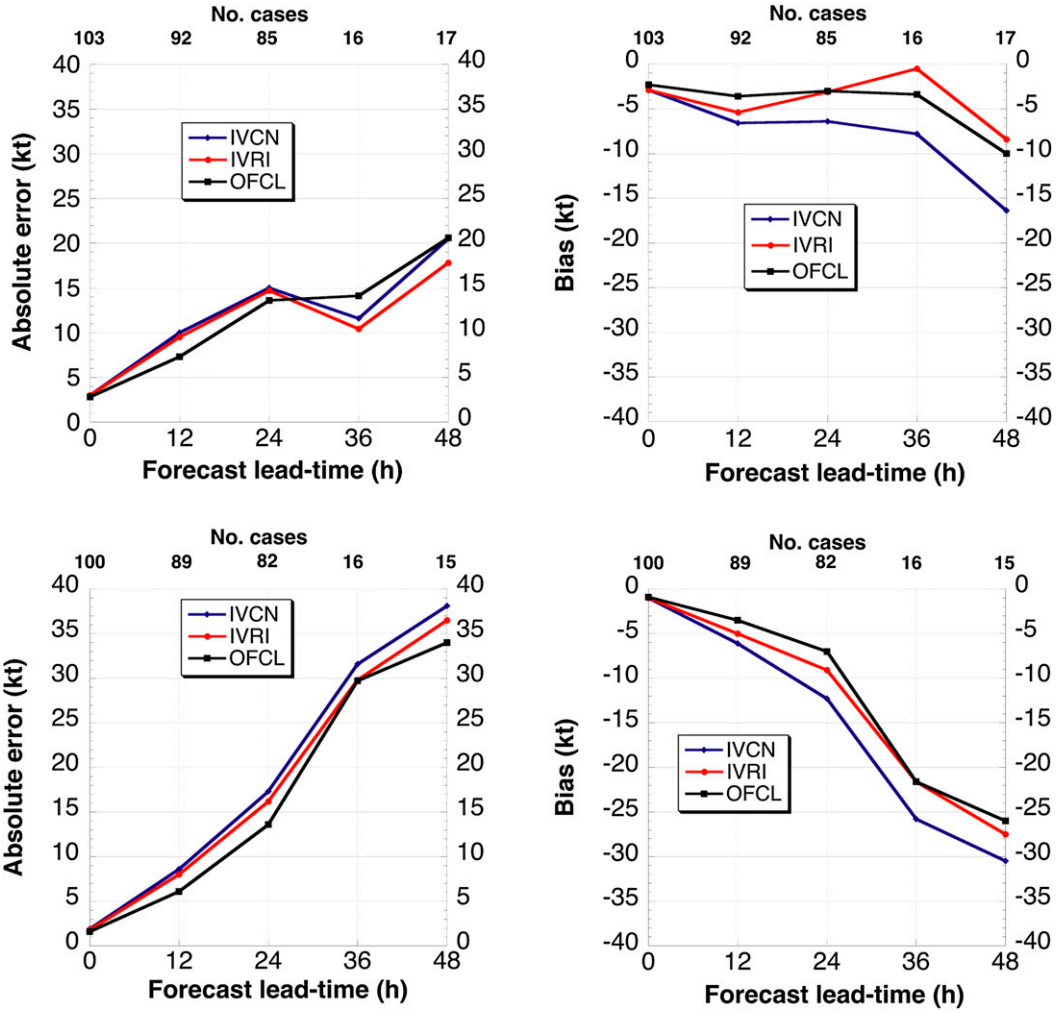

FIG. 11. (top left) Mean absolute forecast errors and (top right) forecast biases for the Atlantic basin for the IVCN, NHC OFCL, and IVRI rerun forecasts for the 2008–13 homogenous sample. (bottom) As in (top), but for the eastern North Pacific basin. The number of forecasts at each forecast time interval is also provided (upper *x* axis).

basin. The OFCL forecasts generally exhibit the most skill within the initial 24-h forecast interval with the numerical models providing somewhat more skill at the longer lead times. HWRF generally provides the most skillful RI forecasts of any of the conventional intensity model guidance. The newly developed consensus RI model rerun forecasts generally display an increase in skill over the existing operational intensity model forecasts. The above finding also indicates that environmental factors, upon which the RI models are based, play an important role in modulating the predictability of RI events. Although the skill of the operational RI forecasts remains limited, it was shown that the OFCL forecasting skill increased by an average of 6% (11%) in the Atlantic (eastern North Pacific) basin in recent years, suggesting that improvements in the operational intensity models and statistical RI guidance during that time period appear to have been at least somewhat helpful.

In addition to the development of improved probabilistic RI models, this paper described the development of a new deterministic intensity consensus forecast aid that combines the multi-lead-time probabilistic consensus RI forecasts with existing intensity model consensus products. Although the number of cases is limited, inclusion of the RI aid guidance in the existing intensity model consensus lowers the mean absolute forecast errors for the four lead times (12, 24, 36, and 48 h) by about 8% and 6% in the Atlantic and eastern North Pacific basins, respectively, with reductions of as much as 13% and 8% observed for those same two basins. The new RI aid guidance also produces very substantial reductions in the overall sample-mean bias with

reductions of 53% and 18% for the Atlantic and eastern North Pacific basins, respectively, with reductions in the forecast bias of as much as 94% and 26% obtained for those same two basins.

Future work will include investigating the potential for improving the RI models by incorporating structural information such as the radius of maximum wind, and other wind radii deduced in real time by the NHC, as well as the distribution of lightning surrounding the inner core as studies by Carrasco et al. (2014) and DeMaria et al. (2012) have shown that such information is correlated with the potential for a system to undergo RI. Additionally, the potential for stratifying RI model forecast skill by factors such as system location, track, and intensity will also be explored in an attempt to investigate the physical factors that might impact RI predictability and to provide forecasters with added information regarding the likely accuracy of the RI model forecasts.

Acknowledgments. This research was funded, in part, by grants from the NOAA/Office of Atmospheric Research (OAR) U.S. Weather Research Joint Hurricane Testbed (JHT) Project and the NOAA/NESDIS GOES-R Proving Ground Program. Financial support from the Office of Naval Research (ONR) is also appreciated. Evan Kalina of the University of Colorado assisted with the computations of the new inner-core dry-air predictor, while Paul Leighton of NOAA/HRD provided programming assistance during the SHIPS-RII development and verification. We thank Drs. Frank Marks (NOAA/HRD) and Chris Landsea (NOAA/NHC) and three anonymous reviewers for their constructive comments on an earlier version of this manuscript. The helpful suggestions of the JHT Project points of contact Eric Blake, Stacy Stewart, and Chris Landsea of the NHC are also appreciated. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. government position, policy, or decision.

REFERENCES

Bosart, L. F., C. S. Velden, W. E. Bracken, J. Molinari, and P. G. Black, 2000: Environmental influences on the rapid intensification of Hurricane Opal (1995) over the Gulf of Mexico. *Mon. Wea. Rev.*, **128**, 322–352, doi:10.1175/1520-0493(2000)128<0322:EIOTRI>2.0.CO;2.

Carrasco, C. A., C. W. Landsea, and Y. L. Lin, 2014: The influence of tropical cyclone size on its intensification. *Wea. Forecasting*, **29**, 582–590, doi:10.1175/WAF-D-13-00092.1.

Chen, H., and D. L. Zhang, 2013: On the rapid intensification of Hurricane Wilma (2005). Part II: Convective bursts and the upper-level warm core. *J. Atmos. Sci.*, **70**, 146–162, doi:10.1175/JAS-D-12-062.1.

Cione, J. J., and E. W. Uhlhorn, 2003: Sea surface temperature variability in hurricanes: Implications with respect to intensity change. *Mon. Wea. Rev.*, **131**, 1783–1796, doi:10.1175//2562.1.

DeMaria, M., 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Mon. Wea. Rev.*, **137**, 68–82, doi:10.1175/2008MWR2513.1.

——, and J. Kaplan, 1999: An updated Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **14**, 326–337, doi:10.1175/1520-0434(1999)014<0326:AUSHIP>2.0.CO;2.

——, M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, 2005: Further improvements to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20**, 531–543, doi:10.1175/WAF862.1.

——, R. T. DeMaria, J. A. Knaff, and D. Molenar, 2012: Tropical cyclone lightning and rapid intensity change. *Mon. Wea. Rev.*, **140**, 1828–1842, doi:10.1175/MWR-D-11-00236.1.

——, C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, doi:10.1175/BAMS-D-12-00240.1.

Dowdy, S., and S. Wearden, 1991: *Statistics for Research.* 2nd ed. Wiley-Interscience, 555 pp.

Dunion, J. D., 2011: Rewriting the climatology of the tropical North Atlantic and Caribbean Sea. *J. Climate*, **24**, 893–908, doi:10.1175/2010JCLI3496.1.

Elsberry, R. L., T. D. B. Lambert, and M. A. Boothe, 2007: Accuracy of Atlantic and eastern North Pacific tropical cyclone intensity forecast guidance. *Wea. Forecasting*, **22**, 747–762, doi:10.1175/WAF1015.1.

Guimond, S. R., G. M. Heymsfield, and F. J. Turk, 2010: Multiscale observations of Hurricane Dennis (2005): The effects of hot towers on rapid intensification. *J. Atmos. Sci.*, **67**, 633–654, doi:10.1175/2009JAS3119.1.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, doi:10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2.

Hanley, D., J. Molinari, and D. Keyser, 2001: A composite study of the interactions between tropical cyclones and upper-tropospheric troughs. *Mon. Wea. Rev.*, **129**, 2570–2584, doi:10.1175/1520-0493(2001)129<2570:ACSOTI>2.0.CO;2.

Hendricks, E. A., M. S. Peng, B. Fu, and T. Li, 2010: Quantifying environmental control on tropical cyclone intensity change. *Mon. Wea. Rev.*, **138**, 3243–3271, doi:10.1175/2010MWR3185.1.

Hong, X., S. W. Chang, S. Raman, L. K. Shay, and R. Hodur, 2000: The interaction between Hurricane Opal (1995) and a warm core ring in the Gulf of Mexico. *Mon. Wea. Rev.*, **128**, 1347–1365, doi:10.1175/1520-0493(2000)128<1347:TIBHOA>2.0.CO;2.

Jiang, H., 2012: The relationship between tropical cyclone intensity change and the strength of inner-core convection. *Mon. Wea. Rev.*, **140**, 1164–1176, doi:10.1175/MWR-D-11-00134.1.

Judt, F., S. S. Chen, and J. Berner, 2015: Predictability of tropical cyclone intensity: Scale-dependent forecast error growth in high-resolution stochastic kinetic-energy backscatter ensembles. *Quart. J. Roy. Meteor. Soc.*, doi:10.1002/qj.2626, in press.

Kaplan, J., and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea. Forecasting*, **18**, 1093–1108, doi:10.1175/1520-0434(2003)018<1093:LCORIT>2.0.CO;2.

——, ——, and J. A. Knaff, 2010: A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **25**, 220–241, doi:10.1175/2009WAF2222280.1.

Kidder, S. Q., and A. S. Jones, 2007: A blended satellite total precipitable water product for operational forecasting. *J. Atmos. Oceanic Technol.*, **24**, 74–81, doi:10.1175/JTECH1960.1.

Kieper, M., and H. Jiang, 2012: Predicting tropical cyclone rapid intensification using the 37 GHz ring pattern identified from passive microwave measurements. *Geophys. Res. Lett.*, **39**, L13804, doi:10.1029/2012GL052115.

Knaff, J. A., 2008: Rapid tropical cyclone transitions to major hurricane intensity: Structural evolution of infrared imagery. Preprints, *28th Conf. on Hurricanes and Tropical Meteorology*, Orlando, FL, Amer. Meteor. Soc., 15A.1. [Available online at http://ams.confex.com/ams/pdfpapers/137929.pdf.]

——, M. DeMaria, C. R. Sampson, and J. M. Gross, 2003: Statistical 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80–92, doi:10.1175/1520-0434(2003)018<0080:SDTCIF>2.0.CO;2.

——, S. P. Longmore, and D. A. Molenar, 2014: An objective satellite-based tropical cyclone size climatology. *J. Climate*, **27**, 455–476, doi:10.1175/JCLI-D-13-00096.1.

Kossin, J. P., and W. H. Schubert, 2001: Mesovortices, polygonal flow patterns, and rapid pressure falls in hurricane-like vortices. *J. Atmos. Sci.*, **58**, 1079–1090, doi:10.1175/1520-0469(2001)058<1079:TDRITK>2.0.CO;2.

Kurihara, Y., R. E. Tuleya, and M. A. Bender, 1998: The GFDL Hurricane Prediction System and its performance in the 1995 hurricane season. *Mon. Wea. Rev.*, **126**, 1306–1322, doi:10.1175/1520-0493(1998)126<1306:TGHPSA>2.0.CO;2.

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.

Molinari, J., and D. Vollaro, 1990: External influences on hurricane intensity. Part II: Vertical structure and response of the hurricane vortex. *J. Atmos. Sci.*, **47**, 1902–1918, doi:10.1175/1520-0469(1990)047<1902:EIOHIP>2.0.CO;2.

Montgomery, M. T., and R. J. Kallenbach, 1997: A theory for vortex Rossby waves and its application to spiral bands and intensity changes in hurricanes. *Quart. J. Roy. Meteor. Soc.*, **123**, 435–465, doi:10.1002/qj.49712353810.

Nolan, D. S., and L. D. Grasso, 2003: Three-dimensional perturbations to balanced, hurricane-like vortices. Part II: Symmetric response and nonlinear simulations. *J. Atmos. Sci.*, **60**, 2717–2745, doi:10.1175/1520-0469(2003)060<2717:NTPTBH>2.0.CO;2.

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

Reasor, P. D., M. D. Eastin, and J. F. Gamache, 2009: Rapidly intensifying Hurricane Guillermo (1997). Part I: Low-wavenumber structure and evolution. *Mon. Wea. Rev.*, **137**, 603–631, doi:10.1175/2008MWR2487.1.

Riemer, M., M. T. Montgomery, and M. E. Nicholls, 2013: Further examination of the thermodynamic modification of the inflow layer of tropical cyclones by vertical wind shear. *Atmos. Chem. Phys.*, **13**, 327–346, doi:10.5194/acp-13-327-2013.

Rogers, R. F., P. D. Reasor, and S. Lorsolo, 2013: Airborne Doppler observations of the inner-core structural differences between intensifying and steady-state tropical cyclones. *Mon. Wea. Rev.*, **141**, 2970–2991, doi:10.1175/MWR-D-12-00357.1.

——, ——, and J. Zhang, 2015: Multiscale structure and evolution of Hurricane Earl (2010) during rapid intensification. *Mon. Wea. Rev.*, **143**, 536–562, doi:10.1175/MWR-D-14-00175.1.

Rozoff, C. M., and J. P. Kossin, 2011: New probabilistic forecast models for the prediction of tropical cyclone rapid intensification. *Wea. Forecasting*, **26**, 677–689, doi:10.1175/WAF-D-10-05059.1.

——, C. S. Velden, J. Kaplan, J. P. Kossin, and A. J. Wimmers, 2015: Improvements in the probabilistic prediction of tropical cyclone rapid intensification with passive microwave observations. *Wea. Forecasting*, doi:10.1175/WAF-D-14-00109.1, in press.

Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (version 3.2). *Bull. Amer. Meteor. Soc.*, **81**, 1231–1240, doi:10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2.

——, J. Kaplan, J. A. Knaff, M. DeMaria, and C. A. Sisko, 2011: A deterministic rapid intensification aid. *Wea. Forecasting*, **26**, 579–585, doi:10.1175/WAF-D-10-05010.1.

Shay, L. K., G. J. Goni, and P. G. Black, 2000: Effects of a warm oceanic feature on Hurricane Opal. *Mon. Wea. Rev.,* **128**, 1366–1383, doi:10.1175/1520-0493(2000)128<1366:EOAWOF>2.0.CO;2.

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

Wang, H., and Y. Wang, 2014: A numerical study of Typhoon Megi (2010). Part I: Rapid intensification. *Mon. Wea. Rev.*, **142**, 29–48, doi:10.1175/MWR-D-13-00070.1.

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

Willoughby, H. E., 1990: Temporal changes of the primary circulation in tropical cyclones. *J. Atmos. Sci.*, **47**, 242–264, doi:10.1175/1520-0469(1990)047<0242:TCOTPC>2.0.CO;2.