New Probabilistic Forecast Models for the Prediction of Tropical Cyclone Rapid Intensification

CHRISTOPHER M. ROZOFF

Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin—Madison, Madison, Wisconsin

JAMES P. KOSSIN

NOAA/National Climatic Data Center, Asheville, North Carolina, and Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin—Madison, Madison, Wisconsin

(Manuscript received 17 December 2010, in final form 19 February 2011)

ABSTRACT

The National Hurricane Center currently employs a skillful probabilistic rapid intensification index (RII) based on linear discriminant analysis of the environmental and satellite-derived features from the Statistical Hurricane Intensity Prediction Scheme (SHIPS) dataset. Probabilistic prediction of rapid intensity change in tropical cyclones is revisited here using two additional models: one based on logistic regression and the other on a naïve Bayesian framework. Each model incorporates data from the SHIPS dataset over both the North Atlantic and eastern North Pacific Ocean basins to provide the probability of exceeding the standard rapid intensification thresholds [25, 30, and 35 kt (24 h)$^{-1}$] for 24 h into the future. The optimal SHIPS and satellite-based predictors of rapid intensification differ slightly between each probabilistic model and ocean basin, but each set of optimal predictors incorporates thermodynamic and dynamic aspects of the tropical cyclone's environment (such as vertical wind shear) and its structure (such as departure from convective axisymmetry). Cross validation shows that both the logistic regression and Bayesian probabilistic models are skillful relative to climatology. Dependent testing indicates both models exhibit forecast skill that generally exceeds the skill of the present operational SHIPS-RII and a simple average of the probabilities provided by the logistic regression, Bayesian, and SHIPS-RII models provides greater skill than any individual model. For the rapid intensification threshold of 25 kt (24 h)$^{-1}$, the three-member ensemble mean improves the Brier skill scores of the current operational SHIPS-RII by 33% in the North Atlantic and 52% in the eastern North Pacific.

1. Introduction

Improving the prediction of rapid intensification (RI) in tropical cyclones (TCs) continues to be a top priority of the National Oceanic and Atmospheric Administration/National Hurricane Center (NOAA/NHC) (Jiing et al. 2011) and is a central focus area for modeling efforts in NOAA's Hurricane Forecast Improvement Program (HFIP) (Toepfer et al. 2010). RI is often defined as an increase in the 1-min maximum sustained surface wind speed beyond some predefined threshold (typically 25, 30, or 35 kt; where 1 kt = 0.514 m s^{-1}) over a 24-h period[1] (e.g., Kaplan et al. 2010, hereafter KDK10). Skillful prediction of RI remains one of the most challenging aspects of TC forecasting, but it is vitally important, particularly when storms are approaching land.

An incomplete physical understanding of the underlying processes of RI is part of the difficulty in accurate prediction. Nonetheless, statistical and theoretical studies

Corresponding author address: Christopher Rozoff, Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin—Madison, 1225 West Dayton St., Madison, WI 53706.
E-mail: chris.rozoff@ssec.wisc.edu

[1] These specific intensification thresholds are motivated by their representation of specific upper-level percentiles of climatological 24-h intensification rates. In KDK10, the threshold of 25, 30, and 35 kt (24 h)$^{-1}$ specifically represent the 90th (88th), 94th (92nd), and 97th (94th) percentiles of 24-h intensity changes of TCs in the Atlantic (eastern Pacific) Ocean basin from 1989 to 2006. These percentiles are very similar to those resulting from our period of study.

DOI: 10.1175/WAF-D-10-05059.1

© 2011 American Meteorological Society

allow us to confidently state that RI can be facilitated by both favorable environmental conditions and a TC's internal dynamics. Climatological analysis of the environments surrounding TCs suggests that RI events are favored in regions of higher sea surface temperatures (SSTs), ocean heat content, and low-level relative humidity, as well as lower vertical shear of the horizontal wind (e.g., Kaplan and DeMaria 2003, hereafter KD03; KDK10; Hendricks et al. 2010). Unresolved aspects of the associated internal TC dynamics complicate simple empirical relationships between RI and the environment. One likely internal mechanism for RI is the coupling of latent heating and enhanced inertial stability (e.g., Schubert and Hack 1982; Nolan et al. 2007; Vigh and Schubert 2009; Pendergrass and Willoughby 2009; Molinari and Vollaro 2010; Rodgers 2010). Increasing the inertial stability of an intensifying TC inner core allows for increasingly efficient intensification of the TC from the release of latent heat by precipitation processes occurring in that region. This important coupling suggests that predictive models should properly capture when a storm is acquiring sufficient inner-core organization of its winds and precipitating clouds. Additional factors influencing RI may include various asymmetric processes including rapid horizontal transport of angular momentum by mesovortices (e.g., Kossin and Schubert 2001; Eastin et al. 2005a,b; Sitkowski and Barnes 2009; Reasor et al. 2009). Hence, the most successful RI prediction schemes may require not only an accurate depiction of a TC's environment, but also some representation of the TC's system-scale and smaller-scale dynamics. Extant empirical–statistical models for predicting RI do not explicitly incorporate these dynamics, but limited information regarding inner-core convective structure can be deduced from infrared satellite data, and further progress is presently being made using microwave satellite data (Velden et al. 2010).

Statistical–empirical forecasting techniques have played a significant role in recent advances in RI prediction. The NHC utilizes an RI index (RII) derived from the Statistical Hurricane Intensity Prediction Scheme (SHIPS; DeMaria and Kaplan 1999) developmental dataset (known as SHIPS-RII). The SHIPS-RII was originally described in KD03 and recently improved in KDK10 to predict the probability of a TC intensifying at least 25, 30, and 35 kt $(24\ \mathrm{h})^{-1}$. The SHIPS-RII is based on linear discriminant analysis and uses a relatively small number of ocean-basin-dependent predictors describing both a TC's environment and some aspects of its internal structure. When compared to climatology, this operational model is skillful in both the North Atlantic and eastern North Pacific Ocean basins (hereafter abbreviated as the Atlantic and eastern Pacific, respectively).

Building on the success of existing statistical RI forecasting models, this paper presents two new prediction models for RI and evaluates their levels of skill with respect to the present SHIPS-RII model. In addition, the benefits of using an ensemble-mean forecast that combines different statistical techniques are examined. The probabilistic models and datasets are described in section 2, and section 3 provides a cross validation of these models. Finally, the forecasting skill levels of these two models are compared with the present SHIPS-RII model, and the value of averaging all three forecasts is explored in section 4.

2. Data and methods

The two statistical models used to predict the probability of RI in this study are both trained on optimally chosen environmental and satellite-based predictors. (The method for optimizing the predictors will be described below.) One model is based on logistic regression and the other is an empirical Bayesian probability model based on the naïve Bayesian framework. The logistic regression model is described in detail in Wilks (2006). Details of the Bayesian model are presented in Kossin and Sitkowski (2009) where the model is applied to forecasting secondary eyewalls in TCs. Although both the logistic regression and Bayesian models use similar environmental and satellite-based features, it will be shown below that each model provides independent information and that a simple average of the output from these two models and the SHIPS-RII yields superior forecast skill compared to any individual model.

All data used to train and test the logistic regression and Bayesian models are obtained directly or derived from the NHC's North Atlantic Hurricane Database (HURDAT; Jarvinen et al. 1984) and the SHIPS developmental dataset based on gridded operational global analyses data (DeMaria et al. 2005). The SHIPS dataset provides features describing the TC-ambient environmental conditions. Weekly gridded SSTs from the NOAA/National Climatic Data Center (Reynolds and Smith 1994, hereafter RS94; information online at http://www.ncdc.noaa.gov/oa/climate/research/sst/weekly-sst.php) and the oceanic heat content computed from the method of Mainelli et al. (2008) are included in the SHIPS dataset as well. The SHIPS dataset also provides measures of the 10.7-μm infrared (IR) satellite presentation of the storms [from Geostationary Operational Environmental Satellite (GOES) imagery]. These data are available at 0000 and 1200 UTC each day prior to the year 2000 and at 0000, 0600, 1200, and 1800 UTC from 2000 to the present. Data from 1995 to 2009 are utilized in this study to train and evaluate the probabilistic models over both the Atlantic and eastern Pacific.

To carry out objective comparisons between the logistic regression, Bayesian, and operational SHIPS-RII

(KDK10) models, the notation and predictor calculations follow KDK10 as closely as possible. As in KDK10, an Atlantic potential intensity predictor is obtained using an adjustment to the RS94 SSTs following the algorithm of Cione et al. (2010). This adjustment accounts for the upwelling of cooler water underneath slower-moving and/or higher-latitude TCs. On the other hand, the eastern Pacific potential intensity predictor is computed using unadjusted RS94 SSTs since the Cione et al. algorithm was developed only for the Atlantic. For each initial time ($t = 0$) analyzed in this study, the values of environmental predictors are averaged along the TC's track out to 24 h in the future, whereas satellite predictors are based only on the $t = 0$ h time.

To obtain the final training and testing dataset for the development of the new models and for their comparison to the SHIPS-RII, a small amount of data screening is carried out. To conform with KD03 and KDK10, data are only used when the center of the TC was not over land between $t = -12$ and 24 h. Also, any forecast time is dispensed with if any of the optimal features used by the logistic regression, Bayesian, or SHIPS-RII models are missing at a given forecast time. Unlike KDK10, no additional screening methods are employed, including the exclusion of cases where the difference between a TC's potential intensity and its current intensity is less than the RI threshold or where the environmental conditions are outside of the range of climatology. Including these types of screening techniques may improve the performance of the new models introduced here but such modifications are deferred to potential future refinements. When all of the data have been processed, the resulting sample sizes for the Atlantic and eastern Pacific are $N = 2572$ and 2614, respectively.

For each ocean basin and for each probabilistic model, optimal predictors must be chosen from the large number of predictors available in the SHIPS dataset and the variables derived from that data. In ordinary least squares regression models, this is generally accomplished through a forward or backward stepping procedure (e.g., DeMaria and Kaplan 1994). To choose optimal predictors for the logistic regression and Bayesian models, we employ a similar technique. First, all predictors whose sample-mean differences between RI and non-RI TCs are not significantly different at the 95% confidence level according to the two-sided Student's t test are discarded.

The selection of optimal predictors for the logistic regression model is accomplished via a stepwise algorithm (e.g., Cheng et al. 2006) that sequentially selects subsets of candidate predictors until the optimal predictors are found. In particular, optimal predictors are those that minimize the deviance of the logistic fit. Features are added to the set of optimal predictors as long as the decrease in deviance per added feature is statistically significant at the 95% level according to a χ^2 test.

Optimal predictors for the Bayesian model are chosen using the same methodology of Kossin and Sitkowski (2009), which is based on maximizing leave-one-year-out cross-validated forecast skill as measured by the Brier skill score (BSS). Here, the cross-validation period spans the years 1995–2009. Also, the BSS is defined here and throughout this paper with respect to the training data's baseline climatological probability of RI. An additional important constraint that is used in determining the set of optimal predictors for the Bayesian model is that significant cross correlation ($r = 0.7$ or greater) is not permitted between predictors. If such cross correlation is found between two or more predictors, only the predictor that contributes the most skill is kept.

The optimal predictors in the logistic regression and Bayesian models vary slightly between the Atlantic and the eastern Pacific but remain physically consistent with each other. The Atlantic predictors for the logistic regression model are summarized in Table 1.[2] These predictors include the previous 12-h change in intensity at $t = 0$ h (PER), the RS94 SSTs (RSST), the 200-hPa divergence averaged from 0- to 1000-km radius from the TC's center (D200), the 850–200-hPa vertical shear of the horizontal wind area averaged over a 0–500-km radius from the TC center after removing the vortex circulation (SHDC), the difference between the potential intensity and current intensity (POT), the standard deviation (from axisymmetry) of IR cloud-top brightness temperature over a 100–300-km radius from the TC center (SDBT2), and the mean IR cloud-top brightness temperature over a 0–30-km radius from the TC center (BTAV). The predictors PER, D200, and POT are also optimal Atlantic predictors in the Bayesian model, but the Bayesian model also includes the ocean heat content (OHC), the 850–700-hPa relative humidity averaged over the annular region 200–800-km radius from the TC's center (RHLO), the 850–200-hPa vertical shear of the horizontal wind averaged over a 200–800-km radius from the TC center (SHRD), the standard deviation of infrared (IR) cloud-top brightness temperatures over a 50–200-km radius from the TC center (SDBT1), and the percentage of the area within a 50–200-km radius from the TC center containing IR cloud-top brightness temperatures at least as cold as −30°C (PX30).

The predictors of the logistic regression model vary between the Atlantic and the eastern Pacific. The list of

[2] Here, we have adopted the same notation as the SHIPS Developmental Dataset to identify the predictors. Also, SHDC, as defined here, is actually the variable SHRD in KDK10.

TABLE 1. The predictors used for the Atlantic in the logistic regression (L) and Bayesian (B) models.

Predictor	Model	Definition
PER	L, B	Previous 12-h intensity change
RSST	L	Reynolds sea surface temperature
OHC	B	Ocean heat content
RHLO	B	850–700-hPa relative humidity (r = 200–800 km)
D200	L, B	200-hPa divergence (r = 0–1000 km)
SHDC	L	850–200-hPa vertical wind shear magnitude calculated with the vortex removed and relative to the 850-hPa center (r = 0–500 km)
SHRD	B	850–200-hPa vertical wind shear magnitude (r = 200–800 km)
POT	L, B	Departure from the tropical cyclone's maximum potential intensity
SDBT1	B	Std dev IR cloud-top brightness temperature (r = 50–200 km)
SDBT2	L	Std dev IR cloud-top brightness temperature (r =100–300 km)
PX30	B	Percent area covered by IR cloud-top brightness temperatures $\leq -30°C$ (r = 50–200 km)
BTAV	L	Mean IR cloud-top brightness temperature (r = 0–30 km)

optimal eastern Pacific predictors still includes PER, SHDC, and POT, but it also now contains four other predictors (Table 2). One of these four predictors, known as ENSS in the SHIPS developmental dataset, provides a measure of the moist convective inhibition. It is specifically computed by considering a surface-based parcel-lifted pseudo-adiabatically to its equilibrium layer and then computing the average of only the negative differences between the equivalent potential temperature of the parcel and the saturation equivalent potential temperature of the environment with height. The spatial average of this calculation over the annular region of 200–800-km radius from the TC center yields ENSS. The remaining three logistic regression predictors include IR cloud-top brightness temperatures averaged over a 100–300-km radius from the TC center (BTA), SDBT1, and the maximum IR cloud-top brightness temperatures over a 0–30-km radius from the TC center (BTMX). The optimal eastern Pacific predictors for the Bayesian model are identical to those in the Atlantic.

The composite mean values of the Atlantic predictors for the RI and non-RI samples are summarized in Table 3 for RI thresholds of 25, 30, and 35 kt $(24\ \mathrm{h})^{-1}$. These composites are comparable with the composites described in KDK10, in that PER, OHC, D200, RHLO, POT, and PX30 are larger and the vertical wind shear and SDBT are smaller for rapidly intensifying TCs. Interestingly, the mean inner-core cloud-top IR brightness temperatures (i.e., BTA) are colder in TCs about to undergo or continue RI. While one might expect a stronger signature of warm-core development in RI TCs as compared to non-RI TCs, vigorous convective activity with cold cloud-top temperatures near the center of RI TCs is evidently dominating the composite mean. A potential reason may be that initial eye formation often occurs underneath upper-level clouds and is therefore obscured in the IR imagery.

TABLE 2. The predictors used for the eastern Pacific in the logistic regression (L) and Bayesian (B) models.

Predictor	Model	Definition
PER	L, B	Previous 12-h intensity change
OHC	B	Ocean heat content
ENSS	L	Vertical average of the negative differences between the equivalent potential temperature of a parcel lifted from the surface and the environmental saturation equivalent potential temperature (r = 200–800 km)
RHLO	B	850–700-hPa relative humidity (r = 200–800 km)
D200	B	200-hPa divergence (r = 0–1000 km)
SHDC	L	850–200-hPa vertical wind shear magnitude calculated with the vortex removed and relative to the 850-hPa center (r = 0–500 km)
SHRD	B	850–200-hPa vertical wind shear magnitude (r = 200–800 km).
POT	L, B	Departure from the tropical cyclone's maximum potential intensity
BTA	L	IR cloud-top brightness temperature (r = 100–300 km)
SDBT1	L, B	Std dev of the IR cloud-top brightness temperature (r = 50–200 km)
PX30	B	Percent area covered by IR cloud-top brightness temperatures $\leq -30°C$ (r = 50–200 km)
BTMX	L	Maximum IR cloud-top brightness temperature (r = 0–30 km)

TABLE 3. The mean values of the predictors used in the Atlantic for the RI and non-RI samples for the 25, 30, and 35 kt (24 h)$^{-1}$ RI thresholds. All differences between the means are statistically significant at the 99.9th level. The sample sizes for the RI (non RI) samples with RI thresholds of 25, 30, and 35 kt (24 h)$^{-1}$ are N = 310 (2262), N = 194 (2378), and N = 113 (2459), respectively.

		RI mean			Non-RI mean			Difference		
Predictor	Units	25 kt	30 kt	35 kt	25 kt	30 kt	35 kt	25 kt	30 kt	35 kt
PER	m s^{-1}	4.5	4.9	5.3	1.3	1.5	1.6	3.2	3.4	3.7
RSST	°C	28.5	28.6	28.7	27.5	27.5	27.6	1.0	1.1	1.1
OHC	kJ cm^{-2}	51.2	53.8	57.0	32.4	33.1	33.7	18.8	20.7	23.4
RHLO	%	73.0	73.2	73.1	68.2	68.4	68.6	4.8	4.8	4.5
D200	10^{-7} s^{-1}	45.5	46.4	47.8	25.5	26.4	27.0	20.0	20.0	20.8
SHDC	m s^{-1}	5.3	5.1	4.8	8.4	8.3	8.2	−3.1	−3.2	−3.4
SHRD	m s^{-1}	6.3	6.1	5.8	9.3	9.2	9.1	−3.0	−3.1	−3.3
POT	m s^{-1}	40.6	40.3	40.4	29.7	30.2	30.5	10.9	10.1	9.9
SDBT1	°C	13.3	12.2	11.3	15.9	15.9	15.8	−2.7	−3.7	−4.5
SDBT2	°C	19.0	18.1	17.1	21.0	21.0	21.0	−2.0	−2.9	−3.9
PX30	%	77.8	81.0	84.0	61.7	62.3	62.7	16.1	18.7	21.4
BTAV	°C	−59.2	−61.4	−65.0	−37.4	−38.3	−38.9	−21.8	−23.1	−26.1

Overall, as indicated in Table 4, the composite means of eastern Pacific predictors for the RI and non-RI samples are consistent with the results in KDK10 and the results for the Atlantic above. It is worth noting that the predictor ENSS is less negative for rapidly intensifying storms, which indicates that the atmosphere surrounding rapidly intensifying storms is less statically stable. Also, consistent with BTA in the Atlantic, BTMX is more negative in TCs about to undergo or continue an RI episode in the eastern Pacific.

3. Validation

Both the logistic regression and Bayesian probabilistic models possess skill in predicting RI, as demonstrated through leave-one-season-out cross validation. Figure 1 provides an overview of the BSS for each model, RI threshold, and ocean basin. In both basins, the logistic regression model provides higher BSS values than the Bayesian model for all RI thresholds. In the Atlantic, the BSS ranges from 12% to 22% for the logistic regression model and from 8% to 15% for the Bayesian model. In the eastern Pacific, BSS ranges from 27% to 32% for the logistic regression model and 20% to 23% for the Bayesian model. As the specified RI threshold is increased from 25 to 35 kt (24 h)$^{-1}$, the forecast skill decreases as RI becomes even more of a rare event. Higher forecast skill was found in the eastern Pacific as compared to the Atlantic in the SHIPS-RII model as well (KDK10).

Reliability diagrams [also known as attributes diagrams; see Wilks (2006)] for each model, ocean basin, and RI threshold are provided in Figs. 2 and 3. In each of the reliability diagrams shown in the left-hand panels,

TABLE 4. The mean values of the predictors used in the eastern Pacific for the RI and non-RI samples for the 25, 30, and 35 kt (24 h)$^{-1}$ RI thresholds. All differences between the means are statistically significant at the 99.9th level. The sample sizes for the RI (non RI) samples with RI thresholds of 25, 30, and 35 kt (24 h)$^{-1}$ are N = 310 (2304), N = 210 (2404), and N = 150 (2464), respectively.

		RI mean			Non-RI mean			Difference		
Predictor	Units	25 kt	30 kt	35 kt	25 kt	30 kt	35 kt	25 kt	30 kt	35 kt
PER	m s^{-1}	5.9	6.6	7.3	0.4	0.6	0.7	5.5	6.0	6.6
OHC	kJ cm^{-2}	30.8	33.4	35.0	15.5	15.9	16.2	15.3	17.5	18.8
ENSS	°C	−23.1	−22.2	−21.2	−38.8	−38.3	−37.9	15.7	16.1	16.7
RHLO	%	77.8	77.7	77.8	73.4	73.6	73.7	4.4	4.1	4.1
D200	10^{-7} s^{-1}	54.4	58.2	61.1	29.8	30.5	31.0	24.6	27.7	30.1
SHDC	m s^{-1}	4.1	3.9	3.8	6.3	6.2	6.1	−2.2	−2.3	−2.3
SHRD	m s^{-1}	4.3	4.2	4.2	6.4	6.3	6.3	−2.1	−2.1	−2.1
POT	m s^{-1}	48.8	48.4	47.7	38.9	39.3	39.6	9.9	9.1	8.1
BTA	°C	−44.7	−45.8	−47.7	−27.5	−28.1	−28.4	−17.2	−17.7	−19.3
SDBT1	°C	11.0	10.6	9.8	14.8	14.7	14.6	−3.8	−4.1	−4.8
PX30	%	85.2	86.7	89.2	60.6	61.5	61.9	24.6	25.2	27.3
BTMX	°C	−63.7	−64.7	−66.2	−40.9	−41.8	−42.3	−22.8	−22.9	−23.9

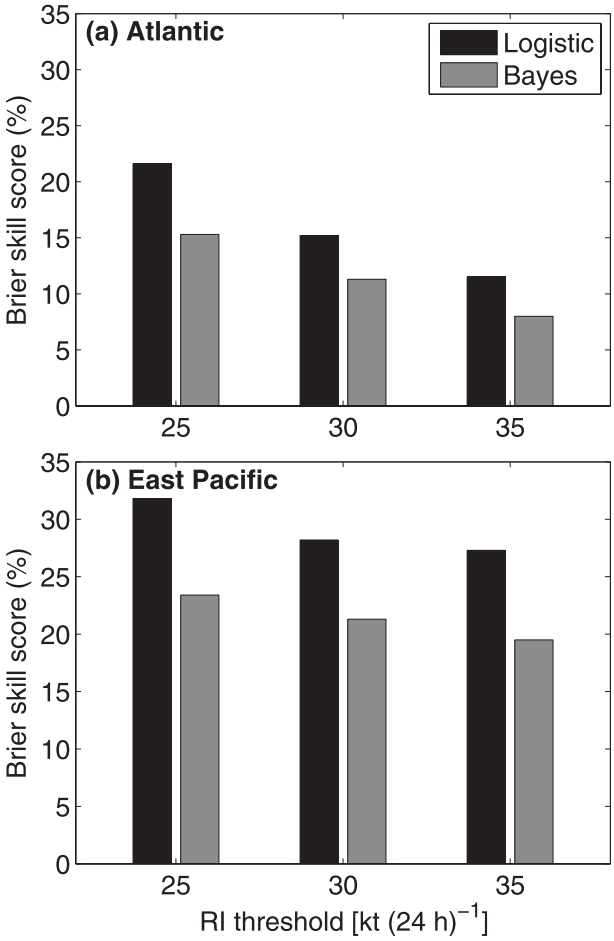

FIG. 1. BSSs (%) determined from leave-one-year-out cross validation (1995–2009) for the logistic regression (black) and Bayesian (dark gray) probabilistic models over the (a) Atlantic and (b) eastern Pacific and for the RI thresholds of 25, 30, and 35 kt (24 h)$^{-1}$. The sample sizes are $N = 2572$ for the Atlantic and $N = 2614$ for the eastern Pacific.

the 45° diagonal line indicates perfect reliability for all forecast probabilities. The horizontal and vertical dashed lines show the climatological probability of RI. Points within the shaded region indicate forecast probabilities that contribute positively to the BSS. Points below the 45° diagonal line indicate forecasted probabilities that are too high, whereas points above this line indicate that the forecasted probabilities are too low (i.e., these deviations provide the conditional bias). Overall, these figures indicate that forecasts are generally more reliable for lower RI thresholds and that forecasts for the eastern Pacific are more reliable than those made for the Atlantic. These aspects are consistent with the BSS results. Except for the marked reduction in reliability at higher forecasted probabilities, the logistic regression model tends to have greater reliability overall, which is reflected in the BSSs presented in Fig. 1. On the other hand, the Bayesian model provides greater reliability for forecasts of high RI probabilities. The corresponding histograms indicate that the sample sizes in the higher probability bins are typically larger for the Bayesian model (i.e., the Bayesian model is slightly more skewed toward higher probabilities of RI than the logistic regression model), which allows finite reliability for high-end probabilities. Still, model forecasts indicating a high probability of RI are quite rare. Therefore, there are little to no data to determine the reliability of higher-probability forecasts of RI. The histograms also help explain the superior reliability in the eastern Pacific, since both probabilistic models more commonly produce very high probabilities (80% or higher) of RI there.

Figure 4 provides an example of the performance of each model in predicting RI [using the 25 and 35 kt (24 h)$^{-1}$ thresholds] for Atlantic Hurricane Wilma (2005) during its time over the Caribbean Sea. Wilma experienced an astounding intensification of 95 kt between 1200 UTC 18 October and 1200 UTC 19 October. At the end of this 24-h period of intensification, Wilma achieved a new Atlantic record minimum sea level pressure of 882 hPa. Both models successfully show elevated probabilities of RI for forecast times that subsequently experience RI. The Bayesian model produces maximum probabilities of 90% and higher during Wilma's RI event. The logistic regression model's highest probabilities are smaller than are those of the Bayesian model, and the maximum occurs 6 h too late since the final 24-h increment of RI is already under way. The higher probabilities predicted by the Bayesian model are consistent with the reliability diagrams and the Bayesian model has better discernment of when to produce heightened probabilities of RI, although the logistic regression model is better at predicting the beginning of the period of RI. It is of interest to note that the timing of the observed onset of Wilma's RI was deduced in part from aircraft reconnaissance measurements. In the absence of such in situ data, intensity is strongly influenced by satellite-based techniques, which generally also lag the onset of RI because of prespecified rules–constraints (Knaff et al. 2010; Velden et al. 2006). In these cases, the statistical–empirical RI prediction models may in fact provide the first warning of an RI event.

The differences in the ability of the two models to capture the timing (onset, duration, and demise) of Wilma's RI event, and the difference in their ability to provide probabilities approaching 100% (Fig. 2), suggest that the two models carry independent information and could be combined to form a potentially more skillful forecast. The biases in forecasted probabilities computed from the entire cross-validated dataset suggest that the two schemes can indeed offer independent information. In the Atlantic,

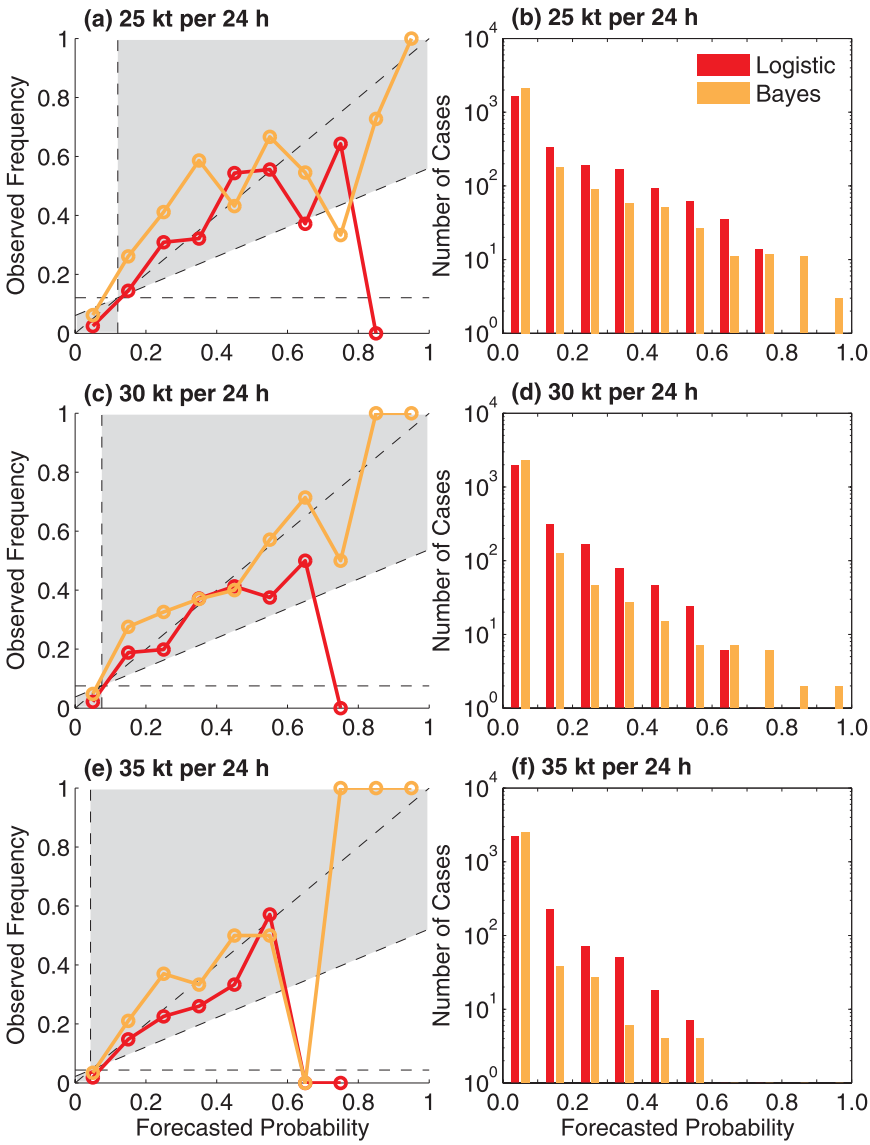

FIG. 2. Atlantic reliability diagrams resulting from independent testing of the logistic regression (red) and Bayesian (orange) models for RI thresholds of (a) 25, (c) 30, and (e) 35 kt (24 h)$^{-1}$, and corresponding figures for the number of forecasted probabilities falling between 0–0.1, 0.1–0.2, . . . , and 0.9–1.0 for (b) 25, (d) 30, and (f) 35 kt (24 h)$^{-1}$.

it is found that the logistic regression and Bayesian models have positive and negative biases, respectively, for all RI thresholds. Sampson et al. (2008) demonstrate how using a consensus of independent TC intensity forecasts can result in a forecast with reduced mean error.

4. Combining the models to improve skill

Given multiple forecast models, it is worth examining whether an ensemble mean of forecasted probabilities of RI can improve skill. To this end, the logistic regression and Bayesian models are utilized in an ensemble-mean forecast. In addition, the 6-hourly probabilities from the SHIPS-RII dataset described in KDK10 and trained with the operational SHIPS developmental dataset over the years 1995–2009 are used as a third ensemble member. We did not have ready access to the SHIPS-RII model in order to derive training subset model parameters for the independent testing method of leave-one-year-out cross validation. Consequently, including the SHIPS-RII model in the ensemble-mean constrains our analyses of skill to be based on dependent testing in which forecasted

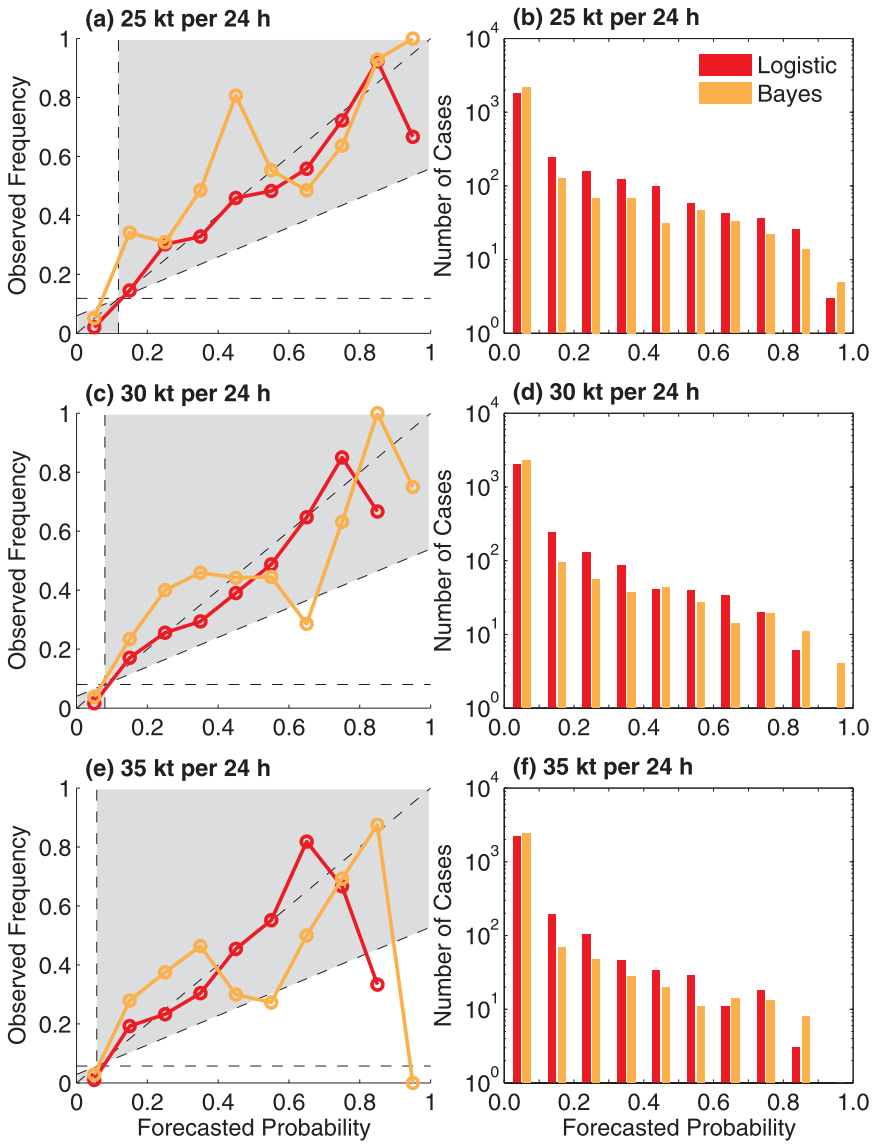

FIG. 3. As in Fig. 2, but for the eastern Pacific.

probabilities are computed for each time in the training dataset using model parameters derived from the same dataset. Compared to independent testing methods, dependent testing will inflate measures of skill, and a comparison of model biases (as discussed above) is not as meaningful, but the dependent analysis is adequate for the purpose of comparing the relative performances of the three models and their ensemble mean.

Figure 5 compares the BSS computed from the dependent dataset for each model and the three-model ensemble-mean forecast. The logistic regression model possesses higher forecast skill than does the Bayesian model. Except for the RI threshold of 35 kt $(24\ \mathrm{h})^{-1}$ in the Atlantic, both the logistic regression and Bayesian models perform somewhat better than the SHIPS-RII. More importantly, the three models evidently bring enough independent information so that the ensemble-mean model skill is higher than any other individual ensemble member. Intermodel correlation coefficients, which may provide an informal sense of model independence, range anywhere from 0.63 to 0.89 (Table 5). The BSS of the ensemble mean in the Atlantic is 33% greater than the existing SHIPS-RII scores for predictions of RI at the 25 kt $(24\ \mathrm{h})^{-1}$ threshold. Similar gains are seen at the 30 and 35 kt $(24\ \mathrm{h})^{-1}$ thresholds. In the eastern Pacific, the ensemble-mean BSS at the 25 kt $(24\ \mathrm{h})^{-1}$ threshold is 52% greater than the BSS of the SHIPS-RII alone. Again, similar gains are seen for the remaining thresholds. These improvements are particularly significant when considered within the context of improving intensity forecasting,

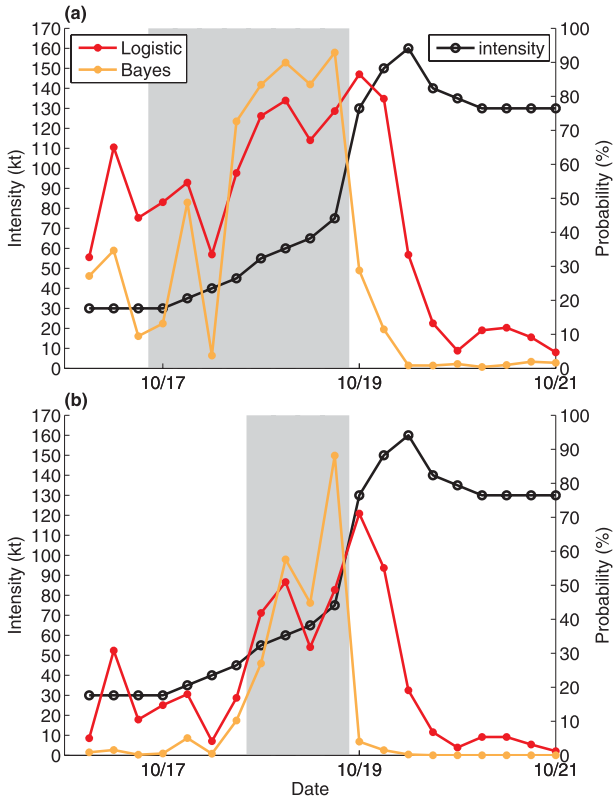

FIG. 4. Evolution of observed best-track intensity (solid black line, kt) and the probability of (a) 25 kt (24 h)$^{-1}$ or greater and (b) 35 kt (24 h)$^{-1}$ or greater intensification rates as predicted by the logistic regression (orange) and Bayesian (red) models for Hurricane Wilma (2005). The gray-shaded region indicates forecast times where RI was observed over the subsequent 24 h.

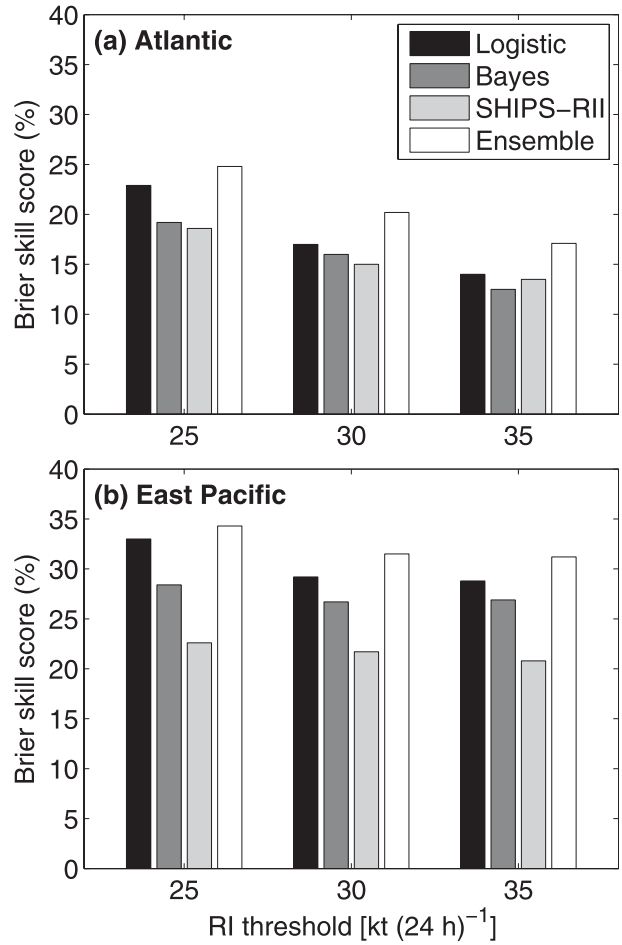

FIG. 5. BSSs (%) determined from dependent testing (1995–2009) for the logistic regression (black), Bayesian (dark gray), SHIPS-RII (light gray), and consensus (white) models over the (a) Atlantic and (b) eastern Pacific and for the RI thresholds of 25, 30, and 35 kt (24 h)$^{-1}$.

as the SHIPS-RII is presently a key operational forecasting tool at the NHC.

The reliability diagrams for the three probabilistic models and their ensemble mean are shown in Figs. 6 and 7. Because these diagrams are constructed using probabilities from the dependent testing framework, the reliability of the models is artificially inflated but the general pattern of behavior of the logistic regression and Bayesian models seen in Figs. 2 and 3 is still exhibited here. Namely, the logistic regression model has superior reliability at lower forecasted probabilities of RI, whereas the Bayesian model is more likely to successfully produce high probabilities of RI. Also, the reliability is once again superior for all models in the eastern Pacific.

The reliability of the SHIPS-RII is comparable to the other models, but exhibits a weakness similar to the logistic regression model in its inability to produce probabilities above around 80%. The reliability of the ensemble-mean model is best at lower- to midlevel probabilities as well. In the eastern Pacific (Fig. 7), the reliability of the ensemble mean is particularly high at the 25 kt (24 h)$^{-1}$ RI threshold, but its improvement over the individual members is less consistent at the remaining thresholds. Still, the ensemble mean is consistent in significantly increasing the skill of the SHIPS-RII in both basins and at all rapid intensification thresholds.

Following the Hurricane Wilma (2005) example shown in Fig. 4, probabilities of RI are shown in Fig. 8 for the

TABLE 5. Correlation coefficients between each of the model forecasted probabilities with the other models for the Atlantic and eastern Pacific and for the 25, 30, and 35 kt (24 h)$^{-1}$ RI thresholds.

Models	Atlantic			Eastern Pacific		
	25 kt	30 kt	35 kt	25 kt	30 kt	35 kt
Logistic/Bayesian	0.78	0.71	0.63	0.81	0.80	0.79
Logistic/SHIPS-RII	0.89	0.83	0.82	0.87	0.84	0.82
Bayesian/SHIPS-RII	0.77	0.74	0.69	0.78	0.77	0.76

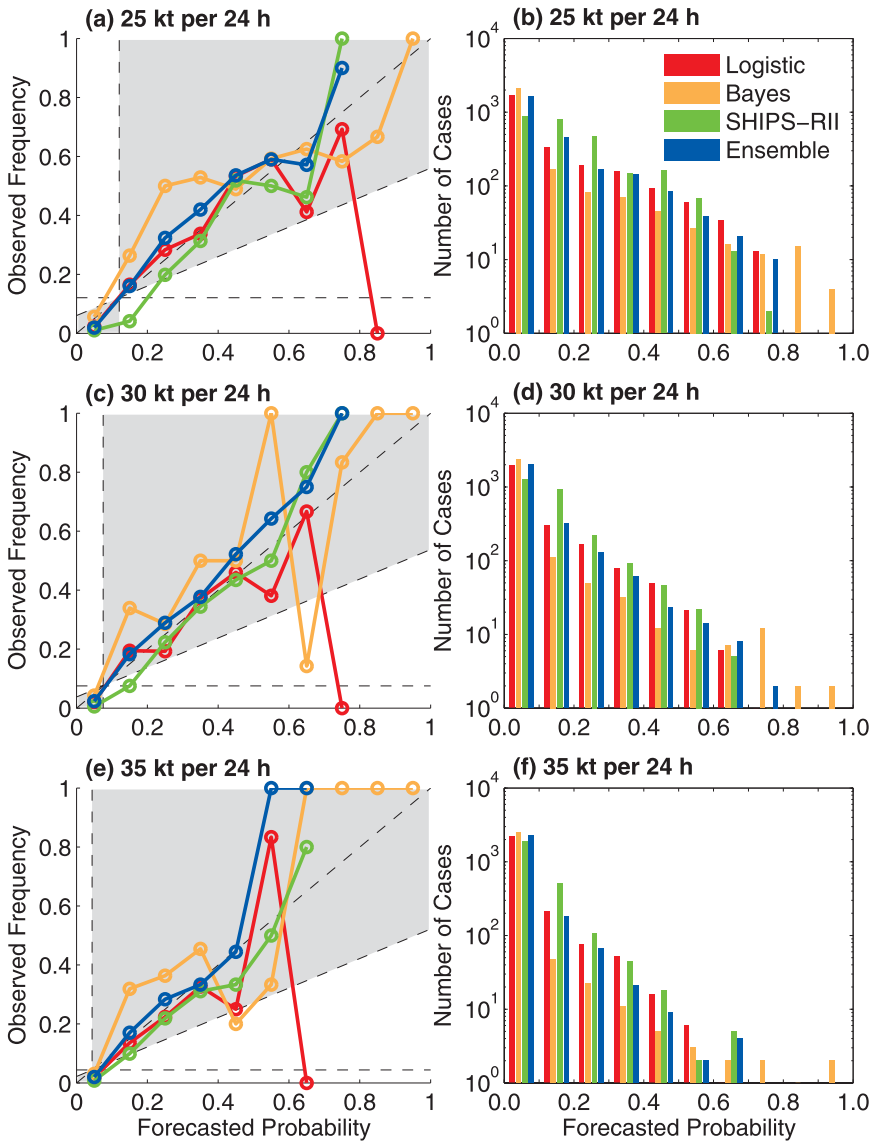

FIG. 6. As in Fig. 2, but for dependent testing of the logistic regression (red), Bayesian (orange), SHIPS-RII (green), and the ensemble (blue) models.

logistic regression, Bayesian, and SHIPS-RII models, as well as the three-model ensemble mean. Although these probabilities are now obtained within the dependent testing framework, the logistic regression and Bayesian models possess values that are qualitatively similar to the results found in the independent testing results throughout the course of Wilma's lifetime over the Caribbean. As noted previously, the Bayesian model provides the lowest probabilities during non-RI periods and the greatest maximum probabilities during the RI event. However, the Bayesian model is also slowest to recognize the onset of RI. Similar to the Bayesian model, the SHIPS-RII model successfully recognizes the end of the RI event, but the logistic regression model performs poorly in this regard. Both the SHIPS-RII and logistic regression models give overly high probabilities prior to the onset of RI, while the Bayesian model performs better during this period. As expected by its construction, the three-model ensemble mean tends toward a smoother evolution during Wilma's passage through the Caribbean Sea and generally captures the better patterns of behavior of each model.

5. Conclusions

In this paper, empirical prediction of rapid intensity change in tropical cyclones was revisited using two new probabilistic models based on logistic regression and Bayesian principles. Each model incorporated data from

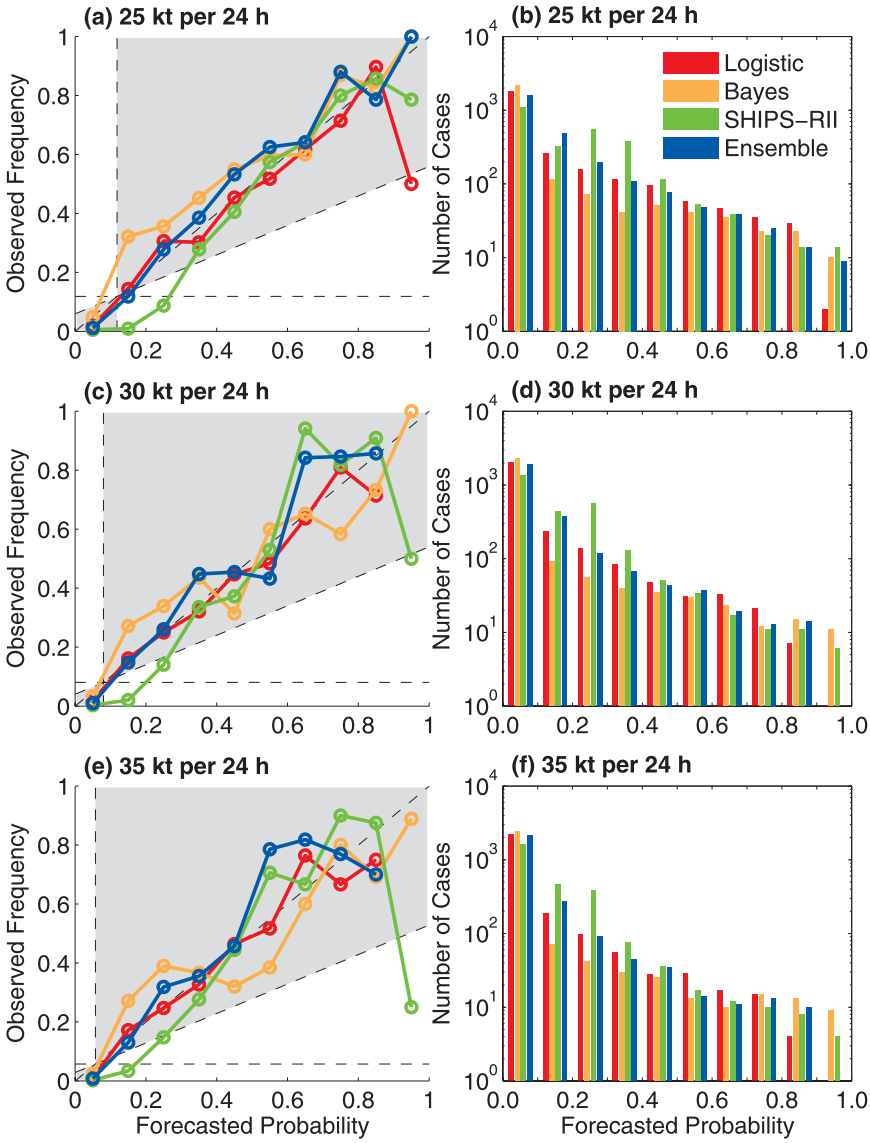

FIG. 7. As in Fig. 6, but for the eastern Pacific.

the SHIPS developmental database over both the Atlantic and eastern Pacific to provide the probability of exceeding the standard rapid intensification thresholds [25, 30, and 35 kt (24 h)$^{-1}$] for 24 h into the future. The optimal SHIPS and satellite-based predictors of RI differed slightly between each probabilistic model and ocean basin, but each set of optimal predictors incorporated aspects of the tropical cyclone's environment and its structure.

Cross validation demonstrated that both the logistic regression and Bayesian probabilistic models are skillful relative to climatology. Dependent testing indicated that both models exhibit forecast skill that is similar but superior to the operational SHIPS-Rapid Intensification Index (RII) presently employed at the NHC. This comparison was found to be consistent for the 25 and 30 kt (24 h)$^{-1}$ RI thresholds in the Atlantic and for all RI thresholds in the eastern Pacific. A simple three-member ensemble mean combining the SHIPS-RII with the logistic regression and Bayesian models showed superior skill compared to each individual ensemble member.

Unique patterns of behavior for each model during the life cycles of tropical cyclones were identified. As the logistic regression model's frequency of high probabilities of RI is lower than the other two models in situations of observed RI, the logistic regression model often displays longer lead times of enhanced RI probabilities than the Bayesian or SHIPS-RII models. Overall, the

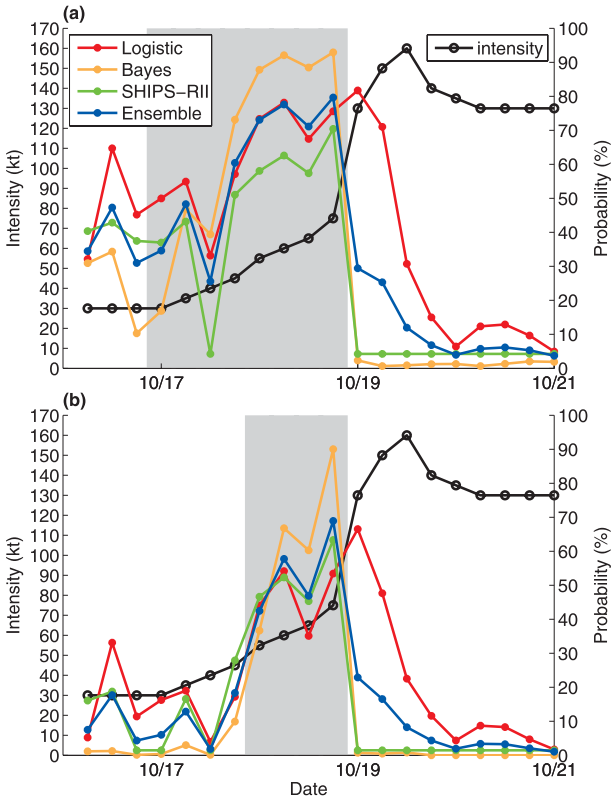

FIG. 8. As in Fig. 4, but for the logistic regression (orange), Bayesian (red), and SHIPS-RII (green) models and the three-member ensemble mean (blue). Model output is based on dependent training–testing.

ensemble mean of the three models provides additional reliability than is demonstrated by any individual model.

The logistic regression and Bayesian models can likely be improved in future research and operational forecasting. It has been found by KDK10 and in this study that coarse IR-based predictors add significant skill to probabilistic models. Using predictors developed from other individual geostationary satellite channels, multichannel predictors, and/or further innovations to IR predictors may further improve forecast skill. In addition, it has been shown that predictors from passive microwave imagery captured aboard low-earth-orbiting satellites improve the statistical intensity forecasting (Jones et al. 2006; Jones and Cecil 2007). Preliminary results with the logistic regression RI model have indicated that microwave-based predictors also add substantial skill to the prediction of RI (Velden et al. 2010). Thus, while improvements in RI forecasting from numerical modeling and data assimilation techniques should continue to increase substantially in coming years, there are still a number of rather inexpensive ways to achieve enhanced operational predictability of RI through the use of simple statistical techniques.

Improvement of rapid intensity change forecasts is the highest operational need and a topmost priority of HFIP. The new models introduced here have the potential to substantially increase the skill of RI forecasts and can be readily transitioned to operations at the NHC with reasonably minimal effort. All required model input data are available through the operational SHIPS model, and model output could be easily appended to the existing SHIPS-RII output. While efforts toward improving numerical guidance and data collection–assimilation are expected to ultimately lead to better RI forecasts, the simpler empirical–statistical models described here continue to provide the greatest skill and should be exploited to their maximum operational potential.

Acknowledgments. We are grateful to John Kaplan, Mark DeMaria, and John Knaff for their helpful discussions and for kindly sharing essential data from their rapid intensification prediction model. Careful reviews by Mark DeMaria, John Knaff, John Bates, and Edward Kearns have also led to substantial improvements in this manuscript. Discussions with Matt Sitkowski and Chris Velden have benefited this research as well. This research was supported by NOAA Grant NA06NES4400002 and NOAA's National Climatic Data Center.

REFERENCES

Cheng, Q., P. K. Varshney, and M. K. Arora, 2006: Logistic regression for feature selection and soft classification of remote sensing data. *IEEE Geosci. Remote Sens. Lett.,* **3,** 491–494.

Cione, J. J., J. Kaplan, C. Gentemann, and M. DeMaria, cited 2010: Developing an inner-core SST cooling algorithm for use in SHIPS. National Hurricane Center. [Available online at http://www.nhc.noaa.gov/jht/03-05_proj.shtml.]

DeMaria, M., and J. Kaplan, 1994: A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting,* **9,** 209–220.

——, and ——, 1999: An updated Statistical Hurricane Intensity Prediction Scheme for the Atlantic and eastern North Pacific basins. *Wea. Forecasting,* **14,** 326–337.

——, 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.

Eastin, M., W. M. Gray, and P. G. Black, 2005a: Buoyancy of convective vertical motions in the inner core of intense hurricanes. Part I: General statistics. *Mon. Wea. Rev.,* **133,** 188–208.

——, ——, and ——, 2005b: Buoyancy of convective vertical motions in the inner core of intense hurricanes. Part II: Case studies. *Mon. Wea. Rev.,* **133,** 209–227.

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.

Jarvinen, B. R., C. J. Neumann, and M. A. S. Davis, 1984: A tropical cyclone data tape for the North Atlantic basin, 1886–1983: Contents, limitations, and uses. NOAA Tech. Memo. NWS NHC 22, 21 pp.

Jiing, J. -G., C. Landsea, and S. Murillo, cited 2011: Joint Hurricane Testbed (JHT) 2011 update: Transition from research to operations. [Available online at http://www.nhc.noaa.gov/jht/ihc_11/s12-01Jiing_IHC2011.pdf.]

Jones, T. A., and D. J. Cecil, 2007: SHIPS-MI forecast analysis of Hurricanes Claudette (2003), Isabel (2003), and Dora (1999). *Wea. Forecasting,* **22,** 689–707.

——, D. Cecil, and M. DeMaria, 2006: Passive-microwave-enhanced Statistical Hurricane Intensity Prediction Scheme. *Wea. Forecasting,* **21,** 613–635.

Kaplan, J., and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea. Forecasting,* **18,** 1093–1108.

——, ——, 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.

Knaff, J. A., D. P. Brown, J. Courtney, G. M. Gallina, and J. L. Beven, 2010: An evaluation of Dvorak technique–based tropical cyclone intensity estimates. *Wea. Forecasting,* **25,** 1362–1379.

Kossin, J. P., and W. H. Schubert, 2001: Mesovortices, polygonal flow patterns, and rapid pressure falls in hurricane-like vortices. *J. Atmos. Sci.,* **58,** 80–92.

——, and M. Sitkowski, 2009: An objective model for identifying secondary eyewall formation in hurricanes. *Mon. Wea. Rev.,* **137,** 876–892.

Mainelli, M., M. DeMaria, L. K. Shay, and G. Goni, 2008: Application of oceanic heat content estimation to operational forecasting of recent Atlantic category 5 hurricanes. *Wea. Forecasting,* **23,** 3–16.

Molinari, J., and D. Vollaro, 2010: Rapid intensification of a sheared tropical storm. *Mon. Wea. Rev.,* **138,** 3869–3885.

Nolan, D. S., Y. Moon, and D. P. Stern, 2007: Tropical cyclone intensification from asymmetric convection: Energetics and efficiency. *J. Atmos. Sci.,* **64,** 3377–3405.

Pendergrass, A. G., and H. E. Willoughby, 2009: Diabatically induced secondary flows in tropical cyclones. Part I: Quasi-steady forcing. *Mon. Wea. Rev.,* **137,** 805–821.

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.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimal interpolation. *J. Climate,* **7,** 929–948.

Rodgers, R., 2010: Convective-scale structure and evolution during a high-resolution simulation of tropical cyclone rapid intensification. *J. Atmos. Sci.,* **67,** 44–70.

Sampson, C. R., J. L. Franklin, J. A. Knaff, and M. DeMaria, 2008: Experiments with a simple tropical cyclone intensity consensus. *Wea. Forecasting,* **23,** 304–312.

Schubert, W. H., and J. J. Hack, 1982: Inertial stability and tropical cyclone development. *J. Atmos. Sci.,* **39,** 1687–1697.

Sitkowski, M., and G. M. Barnes, 2009: Low-level thermodynamic, kinematic, and reflectivity fields of Hurricane Guillermo (1997) during rapid intensification. *Mon. Wea. Rev.,* **137,** 645–663.

Toepfer, F., R. Gall, F. Marks, and E. Rappaport, 2010: Hurricane Forecast Improvement Program five year strategic plan. HFIP Doc., 59 pp. [Available online at http://www.hfip.org/documents/.]

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

——, C. Rozoff, A. Wimmers, M. Sitkowski, M. E. Kieper, J. Kossin, J. Hawkins, and J. Knaff, 2010: An objective method to predict near real time rapid intensification of tropical cyclones using satellite passive microwave observations. Preprints, *29th Conf. on Hurricanes and Tropical Meteorology,* Tucson, AZ, Amer. Meteor. Soc., P1.53. [Available online at http://ams.confex.com/ams/pdfpapers/167742.pdf.]

Vigh, J. L., and W. H. Schubert, 2009: Rapid development of the tropical cyclone warm core. *J. Atmos. Sci.,* **66,** 3335–3350.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 91, Academic Press, 627 pp.