

OPERATIONAL PREDICTION SYSTEM NOTES

A Tropical Cyclone Rapid Intensification Prediction Aid for the Joint Typhoon Warning Center's Areas of Responsibility

JOHN A. KNAFF

NOAA/Center for Satellite Applications and Research, Fort Collins, Colorado

CHARLES R. SAMPSON

Naval Research Laboratory, Monterey, California

BRIAN R. STRAHL

Joint Typhoon Warning Center, Pearl Harbor, Hawaii

(Manuscript received 15 November 2019, in final form 24 March 2020)

ABSTRACT

In late 2017, the Rapid Intensification Prediction Aid (RIPA) was transitioned to operations at the Joint Typhoon Warning Center (JTWC). RIPA probabilistically predicts seven rapid intensification (RI) thresholds over three separate time periods: 25-, 30-, 35-, and 40-kt ($1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$) increases in 24 h (RI25, RI30, RI35, RI40); 45- and 55-kt increases in 36 h (RI45 and RI55); and 70-kt increases in 48 h (RI70). RIPA's probabilistic forecasts are also used to produce deterministic forecasts when probabilities exceed 40%, and the latter are included as members of the operational intensity consensus forecast aid. RIPA, initially designed for the western North Pacific, performed remarkably well in all JTWC areas of responsibility (AOR) and is now incorporated into JTWC's ever improving suite of intensity forecast guidance. Even so, making real-time operational RIPA forecasts exposed some methodological weaknesses such as overprediction of RI for weak/disorganized systems (i.e., systems with maximum winds less than 35 kt), prediction of RI during landfall, input data reliability, and statistical inconsistencies. Modifications to the deterministic forecasts that address these issues are presented, and newly derived and more statistically consistent models are developed using data from all of JTWC's AORs. The updated RIPA is tested as it would be run in operations and verified using a 2-yr (2018–19) independent sample. The performance from this test indicates the new RIPA—when compared to its predecessor—has improved probabilistic verification statistics, and similar deterministic skill while using fewer predictors to make forecasts.

1. Introduction

The Joint Typhoon Warning Center (JTWC) provides tactical tropical cyclone (TC) forecasts for U.S. Department of Defense installations operating in the western North Pacific, Indian, and South Pacific Oceans. These forecasts include position, intensity,¹ and the radii of 34-, 50-,

and 64-kt ($1 \text{ kt} = 0.514 \text{ m s}^{-1}$) winds through 5 days. The last decade or so has seen improvement in JTWC's intensity forecasts due to the availability of skillful intensity guidance coming from the Statistical Typhoon Intensity Prediction Scheme (Knaff et al. 2005), the Statistical Hurricane Prediction Scheme (DeMaria et al. 2005; transitioned to JTWC in 2013), the Hurricane Weather Research and Forecast Model (HWRF; Biswas et al. 2018), the Coupled Ocean–Atmosphere Model Prediction System-TC (COAMPS-TC; Doyle et al. 2014), and consensus aids (Sampson et al. 2008) as discussed in DeMaria et al. (2014). The improvements, while incremental, are evident in seasonal error statistics [IHC 2019; JTWC 2018 (cf. Figs. 6-4 and 6-8)].

¹Operational units for intensity are knots or nautical miles per hour. For that reason, those units are used throughout this manuscript.

Corresponding author: John Knaff, john.knaff@noaa.gov

DOI: 10.1175/WAF-D-19-0228.1

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](http://www.ametsoc.org/PUBSReuseLicenses).

The skillful deterministic statistical models, however, rarely produce forecasts with intensity changes associated with rapid intensification (RI) events due to the multiple time and space scales involved in the process (Kaplan et al. 2015, and references therein). As model resolutions have steadily increased, numerical weather predictions are increasingly capable of forecasting rapid changes in TC intensity (see Leroux et al. 2018; Courtney et al. 2019a,b and references therein). However, concerns such as false alarm rates and forecast timing of such events remain a barrier to operational reliability. To date, statistical guidance specifically designed to overcome the noted shortcomings of existing intensity guidance by forecasting probabilities associated with the occurrence of RI events has been developed on a basin-by-basin basis (see Kaplan and DeMaria 2003; Kaplan et al. 2010, 2015). These efforts have successfully provided guidance methods to anticipated RI events in the Atlantic and east Pacific that have helped forecasters make deterministic forecast decisions (see Gall et al. 2013; Rappaport et al. 2012). Statistics reveal that while the probability of detection ranges from 35% to 60%, the false alarm rates are 65%–60% for forecasts of 30-kt changes in 24 h in Atlantic and east Pacific (Kaplan et al. 2015), leaving forecasters with only difficult decisions. Following those efforts, Knaff et al. (2018) generated a set of tools to probabilistically predict several RI thresholds and trigger deterministic forecasts of those thresholds for use in JTWC's consensus intensity aids for the western North Pacific basin. This guidance, known as the Rapid Intensification Prediction Aid (RIPA), was transitioned to JTWC operations in late 2017.

Although there was a strong desire to test RIPA's capability and get forecaster feedback, this late installation provided only limited initial results in the western North Pacific. A decision was made to provide this guidance for the 2018 Southern Hemisphere and north Indian Ocean storms in addition to those that formed in the western North Pacific. The early verification during this real-time implementation revealed some performance issues, leading to updates to these tools and modifications to the way that RIPA tools are used in JTWC operations. Specifically, the deterministic aids generally worked well for TCs with initial intensities of 35 kt and above away from land; however, these aids had large errors for disturbances with initial intensities less than 35 kt and for landfall cases. Also, there appeared to be erratic forecast-to-forecast behavior for some predictions. Issues such as these were expected since this was the first attempt to apply these models to the western North Pacific in real time.

The purpose of this work is to: 1) evaluate the real-time runs of the methods presented in Knaff et al. (2018), 2) describe a few engineering solutions that make the

deterministic aids more plausible, and 3) present a re-derivation of the underlying statistical models using a dataset from the entire JTWC AOR (Knaff et al. 2018 used only western North Pacific data), using more improved statistical assumptions in the model construction. The next section reviews the expanded dataset and improved methods to derive the statistical models. This discussion is followed by our results, which demonstrate the performance of these new methods using the independent and real-time information collected during the 2018 and 2019 seasons. Finally, we close with a summary of the work and conclusions.

2. Data and methods

JTWC's historical best tracks provide quality controlled 6-hourly position, intensity, and radii information for each TC tracked by JTWC. These data are in the Automated Tropical Cyclone Forecast system (ATCF; Sampson and Schrader 2000) format (available at http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/best_tracks/). Due to issues such as latency of real-time data and operational resource protection considerations, “working” best track data may contain biases, whereas the “final” best tracks have been reanalyzed following the season using all available data and current operational practices. For the RI problem, working best tracks can underestimate intensity changes during RI events. Both working and final best tracks use units of knots and nautical miles (n mi; 1 n mi \approx 1.85 km) for intensity and distance, respectively, and these units will be used throughout to maintain consistency with JTWC operations.

For this work, we also use the SHIPS (2019) developmental dataset (2000–17) and large-scale diagnostic files (LSDFs). These datasets have the same format and contain the same predictors. The real time LSDFs are different than those in the developmental dataset in that they are based only on information available in real-time (i.e., an operational JTWC estimate of location and intensity, a 6-h-old² model forecast track, and corresponding environmental conditions) and so they provide slightly degraded information. For this work, we use the same set of predictors used in Knaff et al. (2018). For brevity, their descriptions and acronyms used in our discussion are provided in Table 1. Potential predictors are assembled into three groups: a subset of the environmental condition parameters in the LSDFs, storm-centered infrared (IR) imagery-based initial conditions,

²Current global model output data available for SHIPS are available after the tropical cyclone advisory warning package has been written and disseminated.

TABLE 1. Potential predictors for algorithms to predict the probabilities of rapid intensification at various intensification rate thresholds. Predictors include forecast parameters (environmental predictors) and initial conditions (IR predictors and best track/advisory-based predictors). Static predictors (i.e., those available only at $t = 0$) are italicized.

Acronym	Description
Environmental Predictors (time averaged from $t = 0$ to time of the forecast)	
GSHR	850–200-hPa generalized wind shear calculated as the mass-weighted root-mean-square deviations of the winds from the mass-weighted deep-layer mean winds times a factor of 4 calculated in a 200–800-km annulus (Knaff et al. 2005)
OHC	Oceanic heat content between the surface and the depth of the 26°C isotherm (Shay et al. 2000, and references within)
RHMD	700–500-hPa relative humidity averaged within a 200–800-km annulus
DIVC	200-hPa divergence following the storm, calculated in a 500-km circle centered on the TC
POT	Potential intensification calculated from the potential intensity as a function of SST at storm center and the current intensity (at $t = 0$)
REFC	Average relative eddy momentum flux convergence calculated in 100–600-km annulus ($\text{m s}^{-1} \text{day}^{-1}$)
TADV	The temperature advection between 850 and 700 hPa averaged from 0 to 500 km calculated from the geostrophic thermal wind
IR predictors	
PC50	Percentage of IR pixels colder than -50°C within a 50–200-km annulus
PC60	Percentage of IR pixels colder than -60°C within a 50–200-km annulus
SDO	Standard deviation of IR brightness temperatures 100–300-km annulus
RMNT	Radius of minimum brightness temperature (0–150 km)
FR5	The deviation of IR-based TC size ($R5$) from the climatological population as a function of TC intensity
Best track/advisory-based predictors	
VMAX	Current TC intensity ($t = 0$) up to 75 kt, fixed at 75 kt for all higher intensities
DV	12-h change in TC intensity, which is limited by the following function, $DV = \min[\min(VMAX \times 0.33, 17.5), DV]$

and real-time best track parameters. A full description of how these are calculated and their justifications are provided in Knaff et al. (2018) and SHIPS (2019). All TCs in best tracks with intensities greater than or equal to 25 kt that did not make landfall within the forecast period were used for development. Landfall was determined by a distance to land algorithm that contains continents, and moderate sized islands, and that has been used with SHIPS development. For instance, La Reunion, Bali, and Melville Island, Australia, are in the dataset, but Guam and Okinawa are not.

There is, however, one notable change in Table 1 and in this work. The initial intensity predictor (VMAX) is now capped at 75 kt for the following reason. Scatterplots of potential intensification (POT) versus VMAX reveal that the two variables strongly covary with $R^2 = 0.78$ (Fig. 1a). However, when POT’s contribution to intensification is removed via linear regression, the residual 24-h intensity change occurs in two separate regimes (Fig. 1b) one below 80 kt and before the eye typically forms in infrared imagery (Vigh et al. 2012) and another at intensities higher when an eye typically exists. For weaker TCs intensity change is positively correlated with VMAX, explaining about 5% of the remaining variance, but for stronger TCs VMAX is only slightly negatively correlated. Above 75-kt intensity, the intensification rate is mostly related to POT indicated by the limited scatter at high intensity shown in Fig. 1a. To address this dilemma simply, we limited the VMAX

term to 75 kt so that RI is more favored as VMAX approaches 75 kt.

As in Knaff et al. (2018), we use two statistical methods for making probabilistic forecasts. The first is linear discriminant analysis (LDA), which is a classification method originally developed in Fisher (1936), and second is logistic regression (LRE; Wilks 2006).

In LDA, a linear combination of variables that best separates two or more groups is developed. We assume two groups: those meeting the threshold intensity changes and those that do not. Additionally, both groups are assumed to have the same covariance structure. The discriminant vector \mathbf{a} has a direction in n -dimensional space that maximizes the distance between the means of Group 1 and Group 2 in standardized units or the discriminant function δ . Thus, δ is the scalar projection of the data vector \mathbf{x} in the direction of maximum separation along the discriminant vector, as shown in Eq. (1) to be

$$\delta = \mathbf{a}^T \mathbf{x}. \tag{1}$$

We used the International Mathematical and Statistical Libraries (IMSL 2019) to make these calculations. Prior probabilities are calculated from the matching dependent discriminant functions and a one-dimensional, single-pass Barnes (1964) analysis windowing procedure relates probabilities to discriminant function values. In application, a cubic spline provides a probability given the independent discriminant function value.

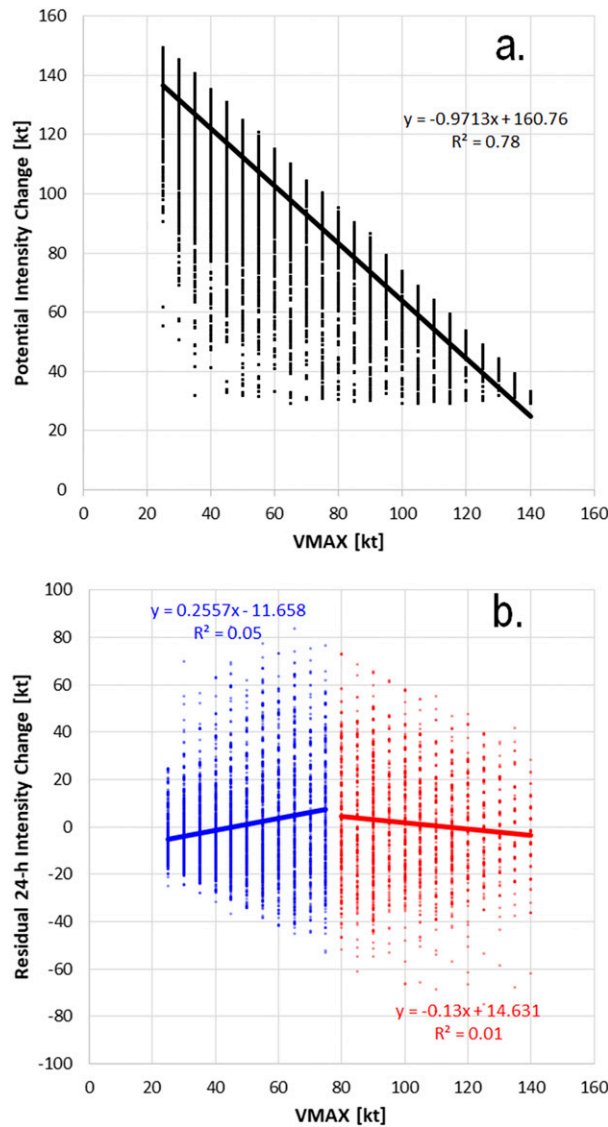


FIG. 1. (a) Scatterplots of current intensity (VMAX) vs potential intensification (POT) and (b) VMAX vs the residual of a POT-based 24-h intensity change, where points with VMAX values less than 80 kt are shown in blue and those 80 kt and above shown in red. The trend line and squared correlation coefficients (R^2) are provided.

In LRE, the dependent variable is a defined category; “1” for reaching the intensification threshold and “0” for not having met the intensification threshold. LRE is a special generalized linear model, where the logit (i.e., the log of the odds) based on categorical data are fit to a linear combination of independent predictors (x_1, \dots, x_n) with intercept b_o . The weights (b_1, \dots, b_n) in this case, are determined via the method of maximum likelihood. Equation (2) shows the logit (left-hand side of equation) as a linear combination of predictors:

$$\ln\left(\frac{p_e}{1-p_n}\right) = b_o + b_1x_1 + \dots + b_nx_n. \quad (2)$$

We use logistic regression code (CSIRO 2019) to fit the model by iteratively reweighted least squares. Forward variable selection with an occasional backward step was used with the knowledge of the LDA model predictors. For refined variable selection, we seek an LRE model with the best fit while trying to ensure physically consistent and statistically significant coefficients. The backward step removes predictors that had lost their statistical significance (99%, chi-squared). The LRE model has different assumptions about the relationship between dependent and independent variables when compared to linear regression. The two primary differences are 1) since the dependent variable is binary, the conditional distribution is a Bernoulli distribution (i.e., 1 or 0) rather than a Gaussian distribution, and 2) the predicted values are probabilities. Once fitted, the probability of exceeding the intensification threshold (p_{RI}) takes the form of Eq. (3):

$$p_{RI} = \frac{1}{1 + e^{-(b_o + b_1x_1 + \dots + b_nx_n)}}. \quad (3)$$

The quality of fit metric for logistic regression is called deviance—a generalization of the idea of using the sum of squares of residuals in ordinary least squares, but where the model is fit using a maximum likelihood criterion. Deviance is formally defined as -2 times the log-likelihood ratio of the fitted model compared to the full (i.e., perfect) model. One can also define the percent deviance explained as 1 minus the ratio of the fitted model deviance to the deviance of a model containing only the intercept b_o (Knaff and DeMaria 2017).

Using the LDA and LRE methods described above, we developed algorithms of weighted combinations of predictors to predict 25-, 30-, 35-, and 40-kt changes in 24 h; 45- and 55-kt changes in 36 h; and 70-kt changes in 48 h. We will refer to these as RI25, RI30, RI35, RI40, RI45, RI55, and RI70, respectively.

To determine how well models are fit to the dependent sample, we use the Brier score (Brier 1950), the Brier skill score and, for just the LRE models the percent deviance explained. The Brier score (BS) is essentially the mean squared error of the probabilistic forecast:

$$BS = \frac{1}{k} \sum_{k=1}^k (y_k - o_k)^2, \quad (4)$$

where y_k is the forecast probability and o_k is the observed probability (effectively 0 for no RI or 1 for RI), and k is the number of cases. A Brier skill score (BSS) is

formed by comparing the BS to the BS of a reference forecast [Eq. (5)]. For this study, climatology is used as a reference forecast (see, Wilks 2006). Here the BS_{ref} is equal to the climatological frequency of each RI threshold:

$$BSS = 1 - \frac{BS}{BS_{ref}}. \quad (5)$$

The BSS answers the question: “Is the forecast superior to the reference forecast?” The answer is “yes” for greater than zero, with a perfect score being equal to one. BSSs will also be calculated for independent forecasts and will again answer the same question. Engineering fixes related to weak storms are not applied to the probabilistic forecasts. There were very few landfalling cases during the period of verification. Nonetheless, BSSs and reliability diagrams were created without those cases that made landfall.

Since RIPA output also includes deterministic forecasts, which are produced when probabilities exceed a threshold of 40%, have initial intensities greater than or equal to 35 kt, and are located farther than 60 n mi from land, we will also verify these using mean absolute error (MAE) and biases. Sampson et al. (2011) found that deterministic forecasts produced with probabilities exceeding 40% were more optimal (lower MAE and smaller biases) than 30% and 50% when added to the intensity consensus. This value has been since revisited, but the 40% remains optimal for reducing biases and MAEs in the consensus intensity forecasts. One of the main motivations for developing a deterministic RIPA was to address negative biases in the intensity consensus during RI events, thereby producing improved guidance for operators. For this reason, the overall performance of the intensity consensus with and without RIPA deterministic forecasts will be examined to ensure that this is occurring.

3. Results

a. Operational performance of RIPA

Early subjective JTWC forecaster analysis of the RIPA tools developed in Knaff et al. (2018) indicates that collectively, the guidance worked as intended (JTWC 2018, personal communication). During the remainder of the 2017 western North Pacific season and the first few TCs in the Southern Hemisphere, we observed that RIPA deterministic forecasts were often triggered (when the probability of RI exceeded 40%) for weak and ill-formed disturbances in which TC formation was incomplete. As a proxy for ensuring a TC had formed, we implemented a change to require initial intensities of at least 35 kt to trigger the RIPA deterministic forecasts. We also observed that RIPA

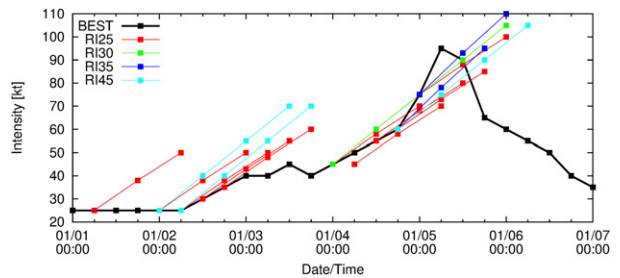


FIG. 2. Time series of the intensity of Tropical Cyclone Ava (sh032018) from the best track (BEST) and deterministic forecasts triggered by RIPA. Shown are RI25, RI30, RI35, and RI45, which were all triggered. Note the number of cases triggered when Ava’s estimated intensity was less than 35 kt. Landfall over Madagascar also occurred at approximately 1000 UTC 5 Jan, resulting in a rapid decay.

generated deterministic forecasts for TCs undergoing landfall, which is not only a distraction to forecasters but an inaccurate forecast as well. To eliminate this quandary, RIPA deterministic forecasts are now truncated when the TC center is forecasted to be within 60 n mi of land as a proxy for landfall processes in the RIPA deterministic forecast. Figure 2 shows the real-time deterministic forecasts for TC Ava, which shows both the RIPA deterministic forecasts from before its formation and intensification during landfall at (1000 UTC 5 January). Changes were made during the middle of the 2018 Southern Hemisphere season and resulted in deterministic forecasts that appeared more credible.³ These changes were not prescribed to the probabilistic guidance.

To examine RI forecast trends, we chose two RI thresholds, RI30 and RI45, and show seasonal MAEs, and Peirce skill scores (Peirce 1884; Wilks 2006) for all of JTWC’s forecast basins combined. The Peirce skill score answers the question of how well did the forecasts separate the “yes” events from the “no” events? The year 2005 was chosen for the start year because this is the first year STIPS was run operationally at JTWC. The results for this analysis are shown in Fig. 3. The number of cases for each year indicates that some years had very few RI cases, so the trends are noisy. For example, there were only 35 cases of RI30 in all of 2017. Still, some information can be readily gleaned from these plots. The red bars in Fig. 3 indicate the percent of time RI events were forecast by JTWC forecasters. The percentage has increased markedly in the last few years as guidance like

³ Forecasters were often looking at the deterministic forecasts of RI as a function of time. Removing the deterministic cases for very weak storms produced a time series where the deterministic RI forecasts often corresponded to real intensity changes, and thus appeared more credible.

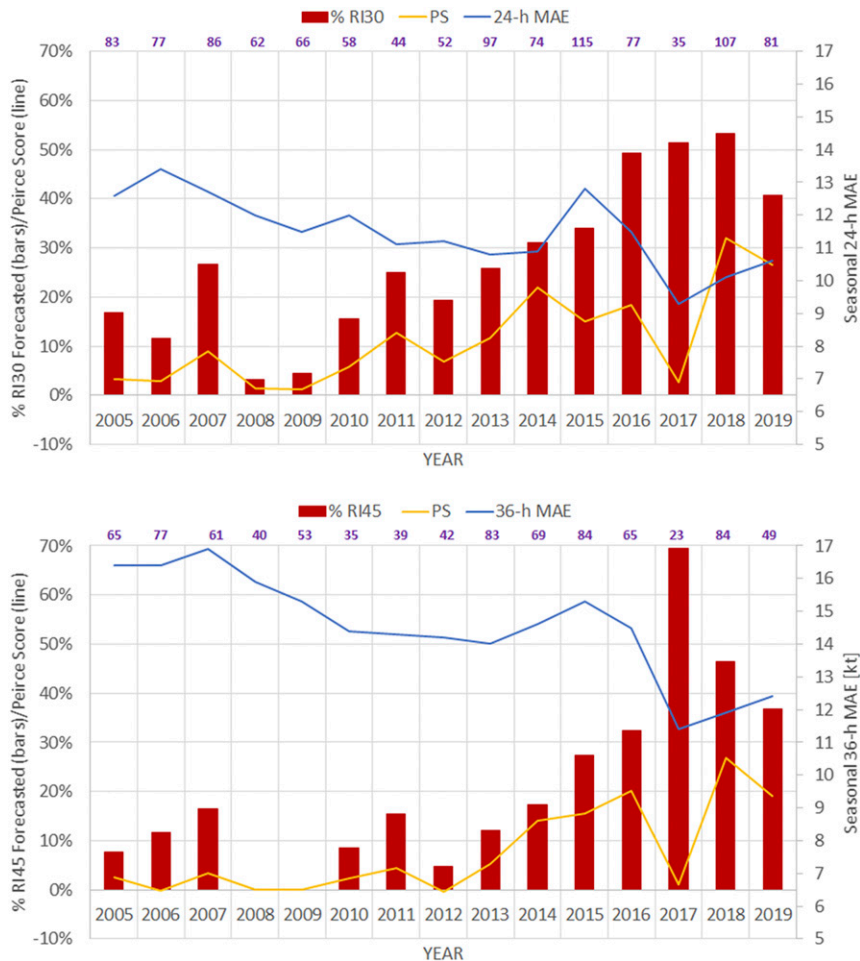


FIG. 3. JTWC forecast performance for observed RI of (top) 30 kt in 24 h (RI30) and (bottom) 45 kt in 36 h (RI45) for the years 2005–18 (all JTWC basins) and 2019 (Southern Hemisphere only). Red bars indicate percentage of observed RI cases for which JTWC predicted RI, blue lines indicate MAE for observed RI cases, and the yellow line indicates Peirce scores. The number of RI cases is listed across the top of each panel and is displayed with purple text.

SHIPS (DeMaria et al. 2005; Kaplan et al. 2015), HWRF (Biswas et al. 2018), COAMPS-TC (Doyle et al. 2014), and RIPA improve. Starting in 2017, the MAEs drop to less than 11 kt at 24 h and less than 13 kt at 36 h. The Peirce skill scores show trends similar to percent of time RI events were forecasted, and it is worth mentioning that the highest Peirce skill scores have been posted in 2018 and 2019.⁴ So, it appears that JTWC is forecasting RI events more frequently while also reducing their MAEs in those cases.

Other guidance that has been available since 2005 has less skill according to this metric. However, to adequately explain the numerous guidance techniques,

and their variations in availability over time is a study in its own and thus will be left to future research. It is, however, noteworthy that mesoscale hurricane models have steadily improved since 2014 and are catching up, particularly in statistics associated with 2018 and 2019.

Overall, the initial rollout of RIPA in JTWC operations was successful, and the engineering solutions led to more credible deterministic RIPA forecasts; however, there were still three issues that required further investigation. First, the forecasts for the 36- and 48-h lead times, as well as the higher 24-h intensity change thresholds (RI35 and RI40), varied widely from one 6-h forecast to the next. This issue was found to be related to differences in the high-frequency variations in the infrared brightness temperature standard deviation–based predictor (SDO) from one forecast to another, and SDO

⁴ Only Southern Hemisphere storms from 2019 were evaluated since the season runs July 2018–June 2019.

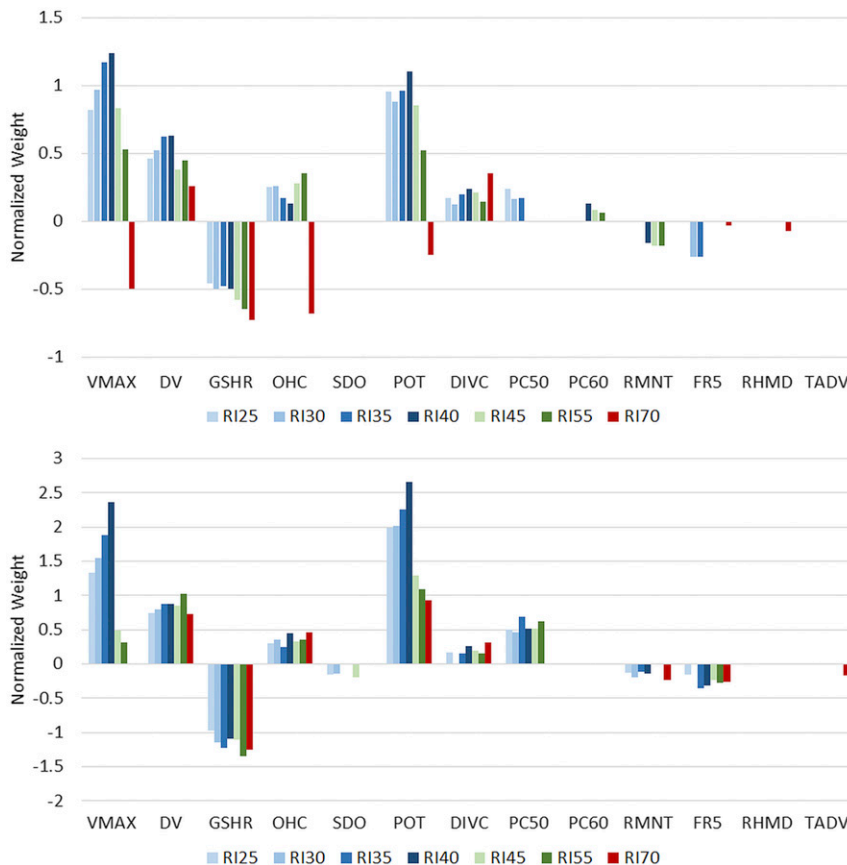


FIG. 4. (top) Normalized LDA and (bottom) LRE model coefficients. The magnitude of each coefficient provides an indication of that variable’s relative importance to predicting the various RI thresholds. The 24-, 36- and 48-h intensity change thresholds are shown in blue, green, and red tones, respectively.

had large weights in both the LDA and LRE prediction models. Second, these statistical models had difficulty forecasting RI once a TC had reached an intensity of about 85 kt. Further investigation showed that this effect was related to the colinearity between the VMAX and the POT predictors, which we alluded to earlier. Both of the issues outlined above are addressed in the new LDA and LRE models discussed in section 3b. And finally, the choice of IR satellite imagery for TC in the Indian Ocean was a dilemma. JMA’s *Himawari-8* images were available but were often from the limb and suffered from limb darkening/cooling and reduced resolution. This issue is addressed in the new operational RIPA implementation.

b. Dependent results (new models)

The individual LDA and LRE model components of RIPA were refit using developmental data from all of JTWC’s forecast basins (i.e., western North Pacific, north Indian Ocean, and Southern Hemisphere) for years 2000–17. Models were also refitted in such a way to remove where possible the dependence on the SDO

predictor that estimates the coherence of the brightness temperatures directly over the TC and to reduce the colinearity between the VMAX and POT predictors by capping VMAX at 75 kt (i.e., as shown in Fig. 1). The latter was thought to be related to storm organization, specifically the existence of an eyewall structure (cf. Vigh et al. 2012). Figure 4 shows the resulting normalized predictor weights; these can be compared to Figs. 1 and 2 in Knaff et al. (2018). It appears that the colinearity issues have been resolved except the LDA RI70 model, which still shows evidence of VMAX, POT, and OHC predictors having coefficients that have signs opposite of what would be expected in nature, suggesting colinearity among those three variables. The remaining models all have physically consistent coefficients.

The new models also have nearly identical dependent fits to the same components reported Tables 2 and 3 of Knaff et al. (2018). Figure 5 (top panel) shows the dependent BSSs for both the LDA and LRE models along with the climatological frequency of occurrence of each RI threshold. The bottom panel of Fig. 5 shows the

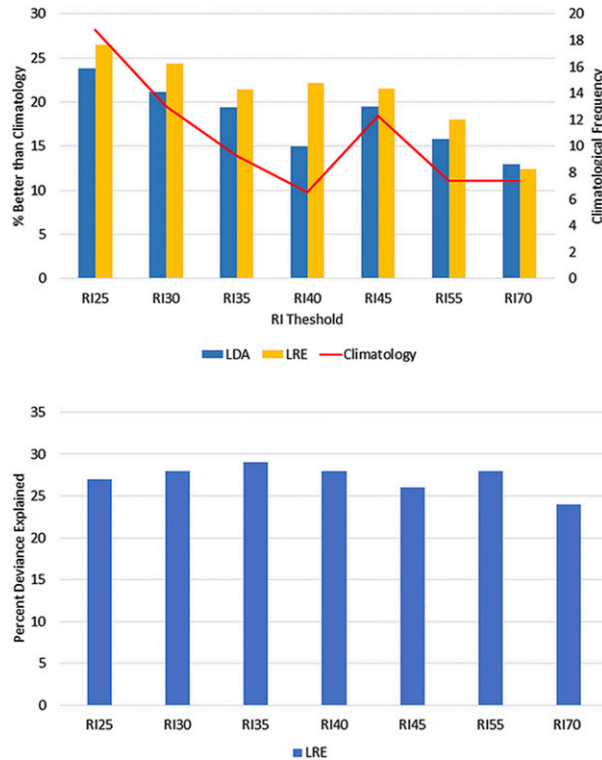


FIG. 5. (top) Brier skill scores for the LDA model (blue bars), the LRE model (yellow bars), and climatology (red line) for each RI threshold. (bottom) Percent deviance explained by the LRE model for each RI threshold. All results based on dependent data from the JTWC 2000–17 dataset.

goodness of fit provided for the individual LRE models in terms of percent deviance explained. Again, these fits are nearly identical as those reported in Knaff et al. (2018). These dependent statistics suggest that both the LDA and LRE models should outperform climatological frequency forecasts by 10% to nearly 30%. Additionally, with the new model issues with noisy variables (e.g., SDO) and collinearity (e.g., POT and VMAX) have been removed while the total number of predictors has been reduced. This is a preferred result because reducing the number of predictors has been shown to reduce artificial skill/ability (Mielke et al. 1996; Davis 1979; Knaff and Landsea 1997).

c. Independent results (new models)

1) PROBABILISTIC

Since the individual models were developed using analyses and best tracks from 2000 to 2017, independent analysis of the newly developed LDA and LRE models could only be accomplished using the tropical cyclones of 2018 and 2019. These reforecasts use the LSDFs produced at JTWC in real time during these times and thus

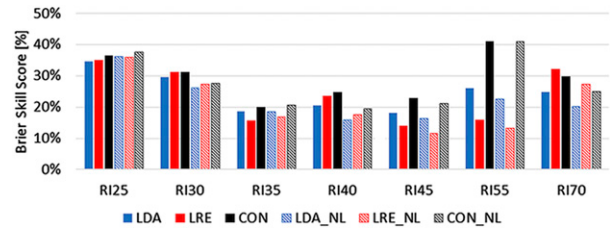


FIG. 6. BSSs associated with the 2018–19 independent verification of RI events associated with 25-, 30-, 35-, and 40-kt changes in 24 h (RI25, RI30, RI35, RI40); 45- and 55-kt changes in 36 h (RI45 and RI55); and 70-kt changes in 48 h (RI70). Note that final best tracks were available for 2018 and preliminary best tracks were used for 2019. There were 2177 forecasts made of which 1994, 1885, and 1788 cases were available for the no-landfall evaluation at 24-, 36-, and 48-h lead times, respectively.

introduce realistic errors caused by the use of working best tracks and 6-h global model forecasts. Probabilistic verification of results from the newly derived models run on this independent dataset is shown in Figs. 6–8.

Figure 6 shows the BSS associated with the individual models (the LDA and LRE models) and the equally weighted average of the two (consensus or CON). The verification is done with landfall cases included (LDA, LRE, and CON), and also removing landfall cases (LDA_NL, LRE_NL, and CON_NL). Removing the landfalling cases both improved and degraded BSSs, the former due to the effects of land on the intensity and the latter due to removing several correctly forecasted no-RI cases. In all but one of the RI thresholds (RI70, which is an extremely rare event), the consensus shows slightly higher skills. The highest BSS values are for RI55, but all the models are skillful compared with climatological forecasts (see top of Fig. 5). The independent forecasts have similar BSSs as dependent hindcasts based on the dependent data, which we would not expect to see if the models were over fitted to the dependent data (see Mielke et al. 1996; Davis 1979). In RI25, RI30, and RI70, the forecasts perform noticeably better for the independent forecasts than for the dependent forecasts. This result is likely due to the limited 2-yr sample and the infrequency of RI events in general (i.e., by chance).

The reliability diagrams associated with the RI thresholds are shown in Figs. 7 and 8. Reliability diagrams consist of two components that are the calibration function showing the relationship between binned forecast probabilities and observed frequencies, and the refinement distribution that shows the number of forecasts included in each bin and indicates the aggregate forecast model confidence. These results are comparable to the independent results presented in Fig. 4 of Knaff et al. (2018). It is important to note that there are very few cases at the highest forecast

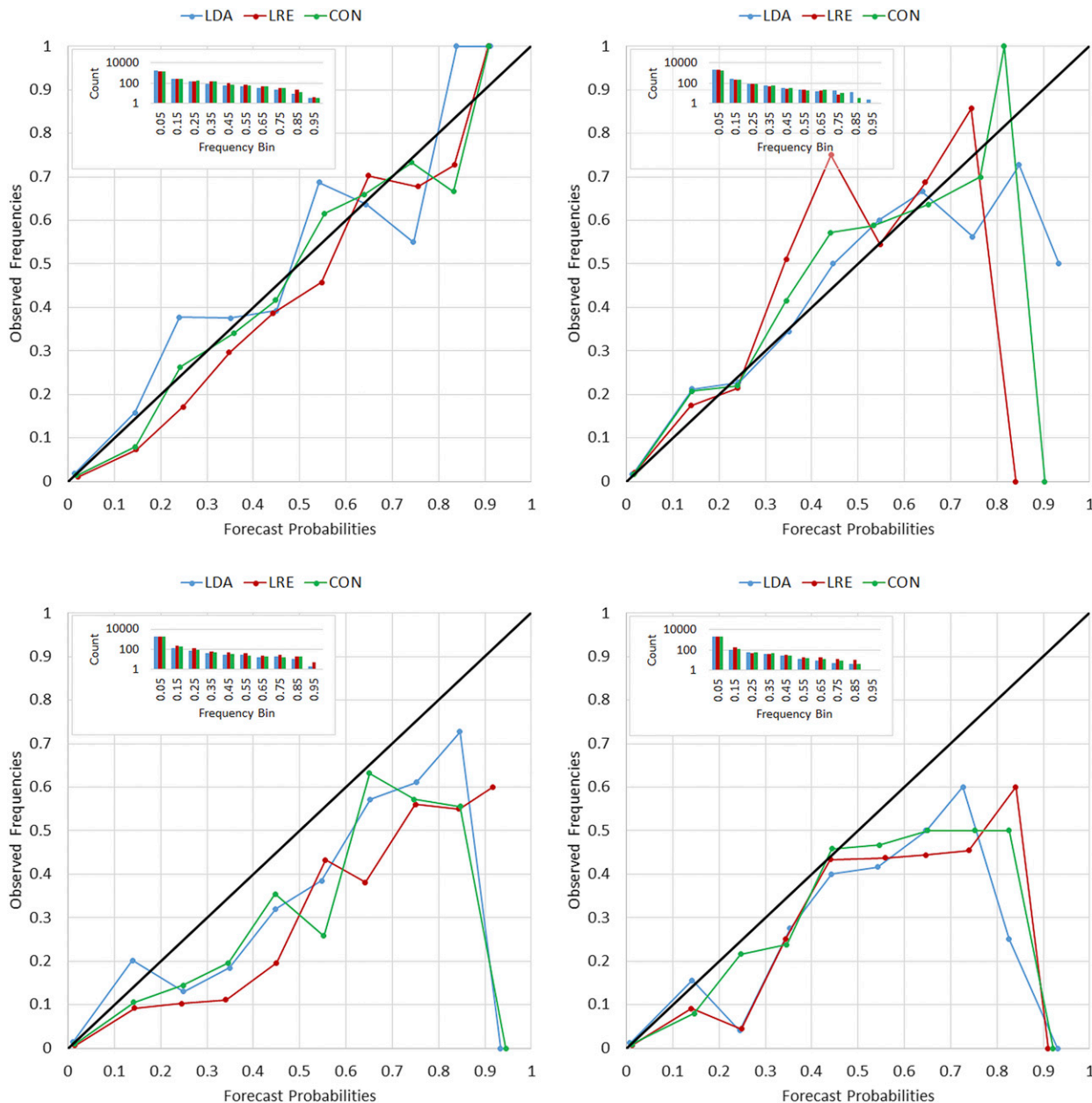


FIG. 7. Reliability diagrams for (top left) RI25, (top right) RI30, (bottom left) RI35, and (bottom right) RI40. Accompanying each reliability diagram is the refinement distributions that show the forecast frequency bins versus the log of the number of cases, which are inset to the upper left in each panel. Results are based on independent forecasts from 2018 and 2019 in the western North Pacific, Southern Hemisphere, and Indian Ocean basins, landfalling cases removed. Final and preliminary best tracks in 2018 and 2019, respectively, are used for verification.

probabilities and that the inset refinement distributions, which in general would be considered “high forecast model confidence” as described in Wilks (2006), are presented as a function of the log of the number of cases (count). In general, the independent results presented here show that the new models generally have better calibration or 1:1 correspondence with the observed frequency and tend to be less biased when compared to

their predecessors. Figure 7 shows that RI24 and RI30 have good calibration, and RI35 and RI40 forecasts tend to slightly overforecast. Figure 8 shows that RI45 and RI55 forecasts also slightly overforecast as well as how the calibration breaks down for the very rare event of RI70. The improvements in these new models resulting from capping VMAX at 75 kt and removing the SDO from the models are particularly evident at the higher-intensity and

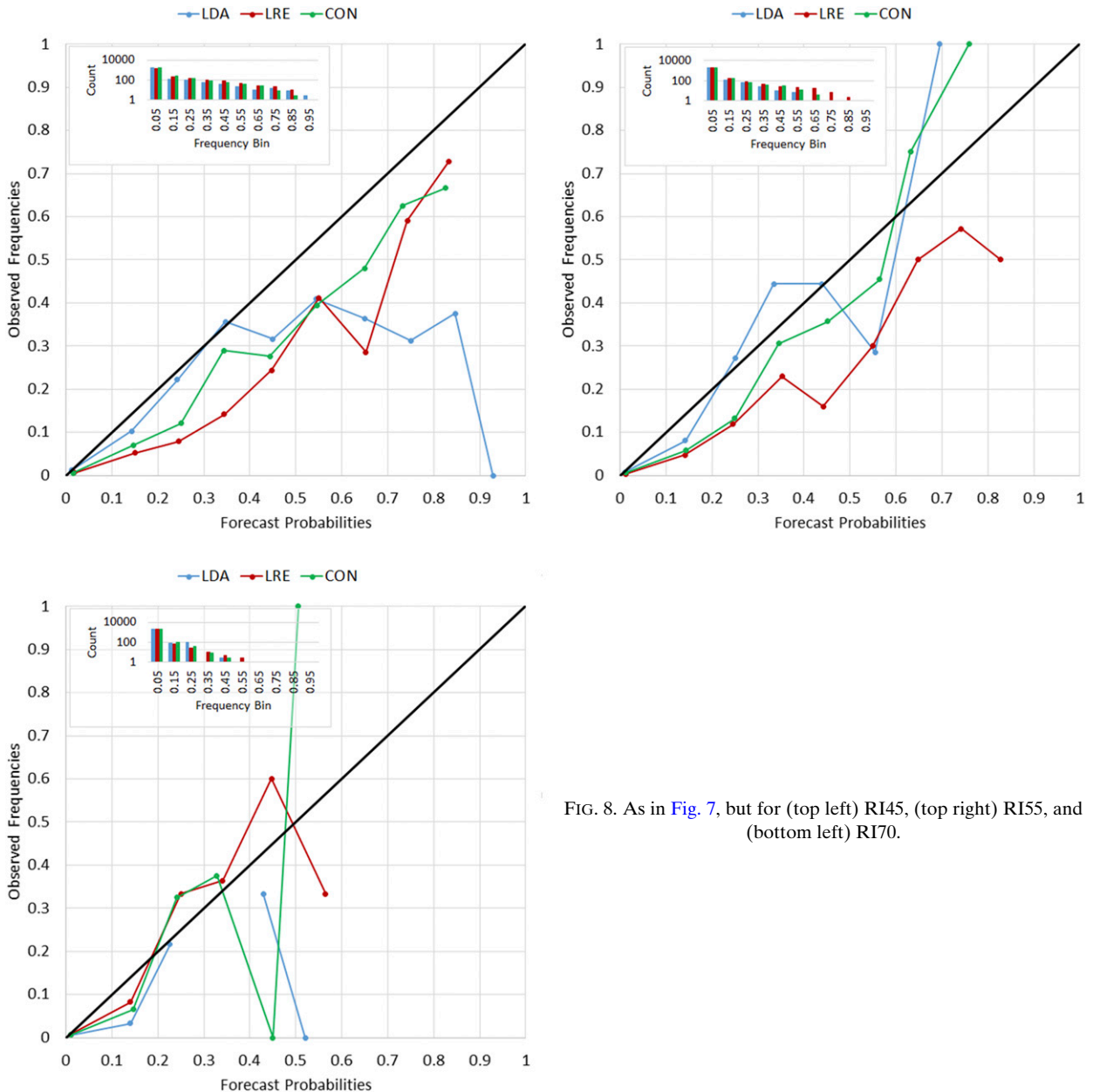


FIG. 8. As in Fig. 7, but for (top left) RI45, (top right) RI55, and (bottom left) RI70.

longer-lead RI thresholds. It is nonetheless noteworthy that some of the highest forecast probabilities were misses as seen in RI30, RI35, RI40, and RI45. All these misses were the same case. TC Dumazile (2018), while forecast to have very rapid/explosive intensification, only experienced a 25-kt increase in intensity.

2) DETERMINISTIC (CONSENSUS)

Deterministic RI guidance (RIPA) is evaluated in Fig. 9, which shows the National Hurricane Center’s RI thresholds of 20-kt intensification in 12 h, RI30, and RI45. Recall that the threshold used in Knaff et al.

(2018) to trigger a RIPA forecast was 40% probability or greater. In this effort, we combined the individual deterministic RI aids (RI25, RI30, RI35, RI40, RI45, RI55, RI70 in the ATCF format) so that the maximum RI rate for a given time interval is used as the RIPA forecast. This is done to simplify the construction of a consensus with (ICNW) and without RIPA (ICNC). As in Knaff et al. (2018), the aids that are used to construct the consensus ICNW are Decay SHIPS (see DeMaria et al. 2005) driven by two different global NWP models [the Global Forecasting System (GFS 2019) and the Navy Global Environmental Model (NAVGEM;

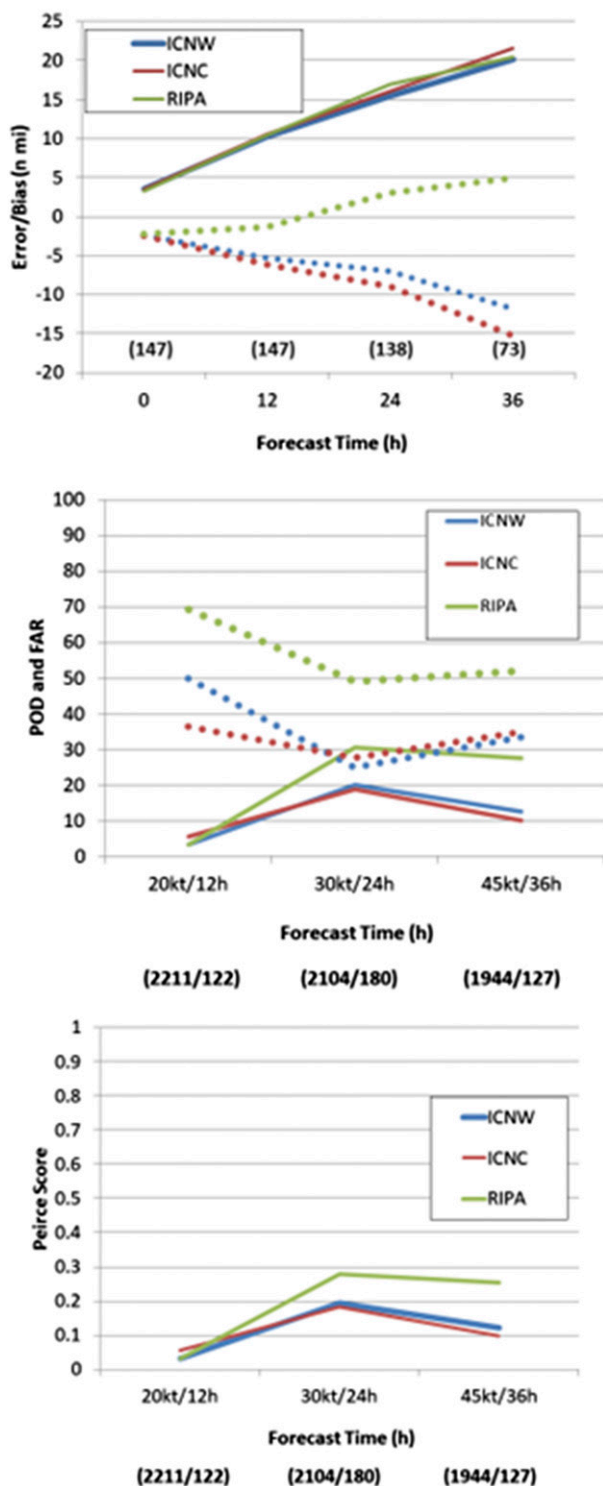


FIG. 9. (top) Mean absolute error (MAE; solid lines) and bias (dotted lines), (middle) POD (solid lines) and FAR (dotted lines), and (bottom) Peirce scores for RIPA and the consensus with and without RIPA (ICNW and ICNC, respectively). Evaluation includes the Southern Hemisphere, western North Pacific, and Indian Ocean 2018 seasons and Southern Hemisphere 2019 season. Total cases/RI cases are shown in parentheses.

Hogan et al. 2014); DSHA and DSHN in the ATCF format], COAMPS-TC (CTCI in the ATCF format, see Doyle et al. 2014), HWRF (HWFI in the ATCF format; see Biswas et al. 2018), a simplified intensity model (CHII in the ATCF format; see Emanuel et al. 2004), and RIPA. The available intensity forecasts for each individual aid are included in the ICNW average. These error/bias results (Fig. 9, top) are comparable to the independent results presented in Knaff et al. (2018) in that the RIPA bias is generally more positive than the consensus bias and that adding RIPA to the consensus reduces negative bias in RI cases. As seen in the negative biases in the consensus forecasts, RI is still a challenge for the consensus members as a group and RIPA is a positive contributor in that regard. In Fig. 9 (middle), and for RI30 and RI45, RIPA is shown to have a slightly higher probability of detection (POD) and lower false alarm rate (FAR) than consensus without RIPA. The corresponding Peirce skill scores shown in Fig. 9 (bottom) indicate some success at forecasting RI for RI30 and RI45 thresholds, and thus shows slight positive impact on the consensus in this measure of success.

4. Summary and conclusions

In late 2017, the Rapid Intensification Prediction Aid (RIPA) was transitioned to operations at the Joint Typhoon Warning Center. This was the first statistical guidance developed specifically for predicting the likelihood of RI in the western North Pacific. RIPA predicts several RI thresholds over three separate time periods and was described in Knaff et al. (2018). RIPA's probabilistic forecasts are also used to produce deterministic forecasts when probabilities exceed 40%. Deterministic forecasts are then incorporated into the operational intensity consensus, which effectively reduces negative biases while not increasing the MAE. The original RIPA worked surprisingly well and was incorporated into the operational forecast process. Nonetheless, running real-time operational RIPA forecasts exposed some weaknesses. These included over prediction of RI for weak and disorganized tropical systems (i.e., systems with maximum winds less than 35 kt), prediction of RI during landfall, input data reliability, and statistical inconsistencies within the models. All but the last of these were addressed by simple engineering solutions applied to just the deterministic forecasts triggered by RI probabilities exceeding 40%.

The last issue (statistical inconsistencies within the models) were traced to two specific issues: collinearity between the initial intensity (VMAX) and the potential intensity minus the initial intensity term (POT), and the noisy behavior of the IR brightness temperature standard deviation term (SDO). To remove the collinearity

between VMAX and POT, VMAX was capped at 75 kt. To address the noisy SDO behavior, the term was removed in the derivation of the new models. The dependent results for the new models were nearly identical to those presented in Knaff et al. (2018), and the independent results appeared to improve reliability and bias (Figs. 7 and 8). These updates were implemented in JTWC's operations in June 2019 and are now the operational basis for the RIPA forecasts.

One highlight of our analysis is that JTWC rapid intensification forecasts have become more frequent while the mean errors have remained near all-time lows (Fig. 2). The authors speculate that improvements in NWP and other guidance (e.g., RIPA) have enhanced JTWC's ability to forecast RI. Nonetheless, there is still plenty of room for improvement. Forecast busts are often a function of storm structure or unique environmental features (e.g., the cases discussed in Ryglicki et al. 2018) that are currently difficult to capture with existing NWP models and certainly statistical aids. Future work will involve studying false alarm and missed forecast cases for common features that could aid forecasters and algorithm developers and expanding RIPA capabilities to include other thresholds and other TC basins. The authors expect that the remaining RI forecast issues are going to be more difficult and time-consuming to address, but necessary to make further headway on this problem.

Acknowledgments. This work was funded by NOAA/NESDIS base funding, and the Office of Naval Research, Program Elements 0602435N. We thank Kate Musgrave at The Cooperative Institute for Research in the Atmosphere (CIRA) at Colorado State University (CSU) for preparing the large-scale diagnostics for JTWC's areas of responsibility; Dave Watson, CIRA/CSU for his invaluable help maintaining the infrared TC image archive; and Chris Slocum and Jack Dostalek for comments on the initial draft. The authors also would like to thank the editor Dr. Elizabeth Ritchie for assigning three excellent anonymous reviewers who provided very helpful and constructive comments. 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

- Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.*, **3**, 396–409, [https://doi.org/10.1175/1520-0450\(1964\)003<0396:ATFMDI>2.0.CO;2](https://doi.org/10.1175/1520-0450(1964)003<0396:ATFMDI>2.0.CO;2).
- Biswas, M. K., and Coauthors, 2018: Hurricane Weather Research and Forecasting (HWRF) model: 2018 scientific documentation. Developmental Testbed Center, accessed 22 April 2020, <https://dtcenter.org/HurrWRF/users/docs/index.php>.
- Brier, G., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).
- Courtney, J. B., and Coauthors, 2019a: Operational perspectives on tropical cyclone intensity change Part 1: Recent advances in intensity guidance. *Trop. Cyclone Res. Rev.*, **8**, 123–133, <https://doi.org/10.1016/j.tcr.2019.10.002>.
- , and Coauthors, 2019b: Operational perspectives on tropical cyclone intensity change Part 2: Forecasts by operational agencies. *Trop. Cyclone Res. Rev.*, **8**, 226–239, <https://doi.org/10.1016/j.tcr.2020.01.003>.
- CSIRO, 2019: Software from Alan J. Miller. CSIRO, accessed 22 April 2020, <https://wp.csiro.au/alanmiller/>.
- Davis, R. E., 1979: A search for short range climate predictability. *Dyn. Atmos. Oceans*, **3**, 485–497, [https://doi.org/10.1016/0377-0265\(79\)90027-7](https://doi.org/10.1016/0377-0265(79)90027-7).
- DeMaria, M., M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, 2005: Further improvement to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting*, **20**, 531–543, <https://doi.org/10.1175/WAF862.1>.
- , C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, <https://doi.org/10.1175/BAMS-D-12-00240.1>.
- Doyle, J., and Coauthors, 2014: Tropical cyclone prediction using COAMPS-TC. *Oceanography*, **27**, 104–115, <https://doi.org/10.5670/oceanog.2014.72>.
- Emanuel, K., C. Desautels, C. Holloway, and R. Korty, 2004: Environmental control of tropical cyclone intensity. *J. Atmos. Sci.*, **61**, 843–858, [https://doi.org/10.1175/1520-0469\(2004\)061<0843:ECOTCI>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<0843:ECOTCI>2.0.CO;2).
- Fisher, R. A., 1936: The use of multiple measurements in taxonomic problems. *Ann. Eugen.*, **7**, 179–188, <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>.
- Gall, R., J. Franklin, F. Marks, E. N. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project. *Bull. Amer. Meteor. Soc.*, **94**, 329–343, <https://doi.org/10.1175/BAMS-D-12-00071.1>.
- GFS, 2019: The Global Forecast System (GFS)—Global Spectral Model (GSM), GSM version 13.0.2. NOAA, accessed 22 April 2020, https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php.
- Hogan, T. F., and Coauthors, 2014: The Navy global environmental model. *Oceanography*, **27**, 116–125, <https://doi.org/10.5670/oceanog.2014.73>.
- IHC, 2019: JTWC—2018 year in review. OFCM, 14 pp., https://www.ofcm.gov/meetings/TCORF/ihc19/session_1/04-Cherrett.pdf.
- IMSL, 2019: IMSL Fortran numerical stat library. IMSL, accessed 22 April 2020, <http://docs.roguewave.com/imsl/fortran/7.0/stat/stat.htm>.
- JTWC, 2018: Joint Typhoon Warning Center annual tropical cyclone report 2017. JTWC, 132 pp., <https://www.metoc.navy.mil/jtwc/products/atcr/2017atcr.pdf>.
- Kaplan, J., and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea. Forecasting*, **18**, 1093–1108, [https://doi.org/10.1175/1520-0434\(2003\)018<1093:LCORIT>2.0.CO;2](https://doi.org/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 east Pacific basins. *Wea. Forecasting*, **25**, 220–241, <https://doi.org/10.1175/2009WAF2222280.1>.

- , and Coauthors, 2015: Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models. *Wea. Forecasting*, **30**, 1374–1396, <https://doi.org/10.1175/WAF-D-15-0032.1>.
- Knaff, J. A., and C. W. Landsea, 1997: An El Niño–Southern Oscillation Climatology and Persistence (CLIPER) forecasting scheme. *Wea. Forecasting*, **12**, 633–652, [https://doi.org/10.1175/1520-0434\(1997\)012<0633:AENOSO>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0633:AENOSO>2.0.CO;2).
- , and R. T. DeMaria, 2017: Forecasting tropical cyclone eye formation and dissipation in infrared imagery. *Wea. Forecasting*, **32**, 2103–2116, <https://doi.org/10.1175/WAF-D-17-0037.1>.
- , C. R. Sampson, and M. DeMaria, 2005: An operational statistical typhoon intensity prediction scheme for the western North Pacific. *Wea. Forecasting*, **20**, 688–699, <https://doi.org/10.1175/WAF863.1>.
- , —, and K. D. Musgrave, 2018: An operational rapid intensification prediction aid for the western North Pacific. *Wea. Forecasting*, **33**, 799–811, <https://doi.org/10.1175/WAF-D-18-0012.1>.
- Leroux, M.-D., and Coauthors, 2018: Recent advances in research and forecasting of tropical cyclone track, intensity, and structure at landfall. *Trop. Cyclone Res. Rev.*, **7**, 85–105, <https://doi.org/10.6057/2018TCRR02.02>.
- Mielke, P. W., K. J. Berry, C. W. Landsea, and W. M. Gray, 1996: Artificial skill and validation in meteorological forecasting. *Wea. Forecasting*, **11**, 153–169, [https://doi.org/10.1175/1520-0434\(1996\)011<0153:ASAVIM>2.0.CO;2](https://doi.org/10.1175/1520-0434(1996)011<0153:ASAVIM>2.0.CO;2).
- Peirce, C., 1884: The numerical measure of the success of predictions. *Science*, **4**, 453–454, <https://doi.org/10.1126/science.ns-4.93.453-a>.
- Rappaport, E. N., J. 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, <https://doi.org/10.1175/BAMS-D-11-00037.1>.
- Rygliski, D. R., J. H. Cossuth, D. Hodyss, and J. D. Doyle, 2018: The unexpected rapid intensification of tropical cyclones in moderate vertical wind shear. Part I: Overview and observations. *Mon. Wea. Rev.*, **146**, 3773–3800, <https://doi.org/10.1175/MWR-D-18-0020.1>.
- Sampson, C. R., and A. J. Schrader, 2000: The automated tropical cyclone forecasting system (version 3.2). *Bull. Amer. Meteor. Soc.*, **81**, 1231–1240, [https://doi.org/10.1175/1520-0477\(2000\)081<1231:TATCFS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2).
- , J. L. Franklin, J. A. Knaff, and M. DeMaria, 2008: Experiments with a simple tropical cyclone intensity consensus. *Wea. Forecasting*, **23**, 304–312, <https://doi.org/10.1175/2007WAF2007028.1>.
- , J. Kaplan, J. A. Knaff, M. DeMaria, and C. Sisko, 2011: A deterministic rapid intensification aid. *Wea. Forecasting*, **26**, 579–585, <https://doi.org/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, [https://doi.org/10.1175/1520-0493\(2000\)128<1366:EOAWOF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1366:EOAWOF>2.0.CO;2).
- SHIPS, 2019: SHIPS developmental data. CIRA, RAMMB, accessed 22 April 2020, http://rammb.cira.colostate.edu/research/tropical_cyclones/ships/developmental_data.asp.
- Vigh, J. L., J. A. Knaff, and W. H. Schubert, 2012: A climatology of hurricane eye formation. *Mon. Wea. Rev.*, **140**, 1405–1426, <https://doi.org/10.1175/MWR-D-11-00108.1>.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.