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
JTWC
Pearl Harbor, Hawaii

Prepared as a note to *Weather and Forecasting*
March 24, 2020

* Corresponding author: John Knaff, NOAA/Center for Satellite Applications and Research, CIRA/Colorado State University, Campus delivery 1375, Fort Collins, CO 80523-1375, email: john.knaff@noaa.gov

Early Online Release: This preliminary version has been accepted for publication in *Weather and Forecasting*, may be fully cited, and has been assigned DOI 10.1175/WAF-D-19-0228.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

© 2020 American Meteorological Society

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-knot increases in 24h (RI25, RI30, RI35, RI40), 45- and 55-knot increases in 36h (RI45 and RI55) and 70 knot increases in 48h (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 over prediction of RI for weak/disorganized (i.e., systems with maximum winds less than 35 knots), 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 two-year (2018 to 2019) independent sample. The performance from this test indicates the new RIPA—when compared to its predecessor—has improved

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

probabilistic verification statistics, and similar deterministic skill while using fewer predictors to make forecasts.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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-knot [kt; 1 kt =0.514 ms^{-1}] winds through five 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 Fig. 6-4 and 6-8)].

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, and Courtney et al. 2019a;2019b and references therein). However, concerns

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

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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; Kaplan et al. 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% to 60% for forecasts of 30kt changes in 24h 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

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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 data set 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 data set 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 six-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 re-analyzed following the season using all available data and current

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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 kt and nautical miles (n.mi; 1 n.mi = 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 (2000-2017) 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 old[2] 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, 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 which has been used with SHIPS development. For

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

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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. Scatter plots of potential intensification (POT) vs. 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, $\boldsymbol{a}$, has a direction in n-dimensional space that maximizes the distance between the means of Group 1 and Group 2 in

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

standardized units or the discriminant function, δ. Thus, δ is the scalar projection of the data vector, $\boldsymbol{x}$, in the direction of maximum separation along the discriminant vector, as shown in equation (1) to be:

$$\delta = \boldsymbol{a}^T \boldsymbol{x} \qquad (1)$$

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.

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 is fit to a linear combination of independent predictors ($x_1, \ldots, x_n$) with intercept b_o. The weights ($b_1 \ldots b_n$) 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:

$$\ln\left(\frac{p_e}{1-p_n}\right) = b_o + b_1\, x_1 + \cdots + b_n x_n \qquad (2)$$

We use logistic regression code (CSIRO 2019) to fit the model by iteratively reweighted least squares. Model fit is based on a maximum likelihood criterion. Forward variable selection with an occasional backward step was used with the knowledge of the LDA model predictors. For refined variable selection, we seek a LRE

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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 equation (3) :

$$p_{RI} = \frac{1}{\left(1+e^{-(b_o+b_1x_1\ldots+b_nx_n)}\right)} \qquad (3)$$

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_0 (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.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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 is 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 further than 60 n mi from land, we will also verify these using

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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 was value has been 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, 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 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.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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 (10:00 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 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 Pierce Skill Scores have been

[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.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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 & RI40), varied widely from one six-hour 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 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 co-linearity 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 the next subsection. And finally, choice of IR satellite imagery for TC in the Indian

[4] Only Southern Hemisphere storms from 2019 were evaluated since the season runs July of 2018 thru June 2019.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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.

a) 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 2000-2017. 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 co-linearity 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 co-linearity 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

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

frequency of occurrence of each RI threshold. The bottom panel of Fig. 5 shows the 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 percent. 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).

b) Independent results (new models)

i) 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 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 data set is shown in Fig. 6, Fig. 7, and Fig. 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

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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, & RI70, the forecasts perform noticeably better for the independent forecasts than for the dependent forecasts. This result is likely due to the limited two-year sample and the infrequency of RI events in general (i.e., by chance).

The reliability diagrams associated with the RI thresholds are shown in Fig. 7 and Fig. 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 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. Fig. 7 shows that RI24, RI30 have good

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

calibration, and RI35 and RI45 forecasts tend to slightly overforecast. Fig. 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 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.

ii) 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 12h, 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; Hogan et al. 2014)] DSHA and DSHN in the ATCF format), COAMPS-TC (CTCI in the ATCF format, see Doyle et al. 2014), H-WRF (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.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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.

2) Summary and conclusion

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

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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 (Fig. 7 and Fig. 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

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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 would like to 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, DOI: 10.1175/1520-0450(1964)003<0396:ATFMDI>2.0.CO;2

Biswas M. K., S. Abarca, L. Bernardet, I. Ginis, E. Grell, M. Iacono, E. Kalina, B. Liu, Q. Liu, T. Marchok, A. Mehra, K. Newman, J. Sippel, V. Tallapragada, B. Thomas, W. Wang, H. Winterbottom, and Z. Zhang, 2018: Hurricane Weather Research and

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

Forecasting (HWRF) Model: 2018 Scientific Documentation [Available online at 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. doi:10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2

Courtney, J. B., S. Langlade, C. R. Sampson, J. A. Knaff, T. Birchard, S. Barlow, S.D. Kotal, T. Kriat, W. Lee, R. Pasch, and U. Shimada 2019a: Operational perspectives on tropical cyclone intensity change Part 1: Recent advances in intensity guidance. *Tropical Cyclone Research and Review*, **8**(3), 123-133, doi: 10.6057/2019TCRR03.02.

Courtney, J. B., S. Langlade, C. R. Sampson, J. A. Knaff, T. Birchard, S. Barlow, S.D. Kotal, T. Kriat, W. Lee, R. Pasch, and U. Shimada , 2019b: Operational Perspectives on Tropical Cyclone Intensity Change Part 2: Forecasts by Operational Agencies. *Tropical Cyclone Research and Review*, **8**(4), 226-239. doi: 10.6057/2019TCRR04.03.

CSIRO, 2019: Software from Alan J. Miller. [Available online at http://wp.csiro.au/alanmiller/]

Davis, R. E., 1979: A search for short range climate productivity. *Dyn. Atmos. Oceans*, **3**, 485–497. 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.

___, C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc*., **95**, 387-398

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

Doyle, J., R. Hodur, S. Chen, Y. Jin, J Moskaitis, S. WANG, E. A. Hendricks, H Jin, and T. SMITH, 2014: Tropical cyclone prediction using COAMPS-TC. *Oceanography*, **27**(3), 104-115. Retrieved from http://www.jstor.org/stable/24862193

Emanuel, K., C. Desautels, C. Holloway and R. Korty, 2004: Environmental control of tropical cyclone intensity. *J. Atmos. Sci*., **61**, 843-858.

Fisher, R. A., 1936: The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179–188. doi: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. [available online at https://www.emc.ncep.noaa.gov/GFS/doc.php]

Hogan, T. F., M. LIU, J. A. Ridout, M. S. Peng, T. R. Whitcomb, B. C. Ruston, C. A. Reynolds, S. D. Eckermann, J. R. Moskaitis, N. L. BAKER, J. P. McCormack, K. C. Viner, J. G. McLay, M. K. Flatau, L. Xu, C. Chen, and S. W. Chang, 2014: The Navy global environmental model. *Oceanography,* **27**, 116-25. http://www.jstor.org/stable/24862194.

IHC, 2019: JTWC – 2018 Year in Review. [available online at https://www.ofcm.gov/meetings/TCORF/ihc19/session_1/04-Cherrett.pdf]

IMSL, cited 2019: IMSL® Fortran Numerical Stat Library. [available on-line at http://docs.roguewave.com/imsl/fortran/7.0/stat/stat.htm]

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

JTWC, 2018: Joint Typhoon Warning Center Annual Tropical Cyclone Report 2017. 132 pp [available online at 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(**6),1093-1108.

___, M. DeMaria, and J. A. Knaff, 2010: A revised tropical cyclone rapid intensification index for the Atlantic and east Pacific basins. *Wea. Forecasting*, **25**(1), 220-241.

___, C. M. Rozoff, M. DeMaria, C. R. Sampson, J. P. Kossin, C. S. Velden, J. J. Cione, J. P. Dunion, J. A. Knaff, J. A. Zhang, J. F. Dostalek, J. D. Hawkins, T. F. Lee, and J. E. Solbrig, 2015: Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models. *Wea. Forecasting*, **30**, 1374-1396. doi: http://dx.doi.org/10.1175/WAF-D-15-0032.1.

Knaff, J. A., and R. T. DeMaria, 2017: Forecasting tropical cyclone eye formation and dissipation in infrared imagery. *Wea. Forecasting*, **32**(6), 2103-2116, https://doi.org/10.1175/WAF-D-17-0037.1.

___. 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

___, C. R. Sampson, and M. DeMaria, 2005: An operational statistical typhoon intensity prediction scheme for the Western North Pacific. *Wea. Forecasting*, **20(**4), 688-699.

___, ___, and K. D. Musgrave, 2018: An operational rapid intensification prediction aid for the western North Pacific. *Wea. Forecasting,* **33**, 799–811, doi: /10.1175/WAF-D-18-0012.1.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

Leroux, M.-D., K. Wood, R. L. Elsberry, E. Cayanan, E.O. Hendricks, M. Kucas, P. Otto, P., R. Rogers, C.R. Sampson, and Z. Yu, 2018: Recent advances in research and forecasting of tropical cyclone track, intensity, and structure at landfall. *Tropical Cyclone Research and Review*, **7**(2), 85–105.

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

Peirce, C. 1884: The Numerical Measure of the Success of Predictions. *Science, **4***, 453-454. Retrieved from http://www.jstor.org/stable/1760565

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

Ryglicki, 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, DOI: 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.

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

___, J. Kaplan, J. A. Knaff, M. DeMaria, and C. Sisko, 2011: A deterministic rapid intensification aid. *Wea. Forecasting*, **26**, 579–585.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

SHIPS, 2019: “SHIPS developmental data.” [Available on-line at 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. doi: http://dx.doi.org/10.1175/MWR-D-11-00108.1.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences: An Introduction*, Second Edition, Academic, San Diego, Calif., 627 pp.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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 hPa to 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 (m s^{-1} 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	

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

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-hour change in TC intensity, which is limited by the following function,* $DV = min(min(VMAX * 0.33, 17.5), \mathrm{DV})$

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

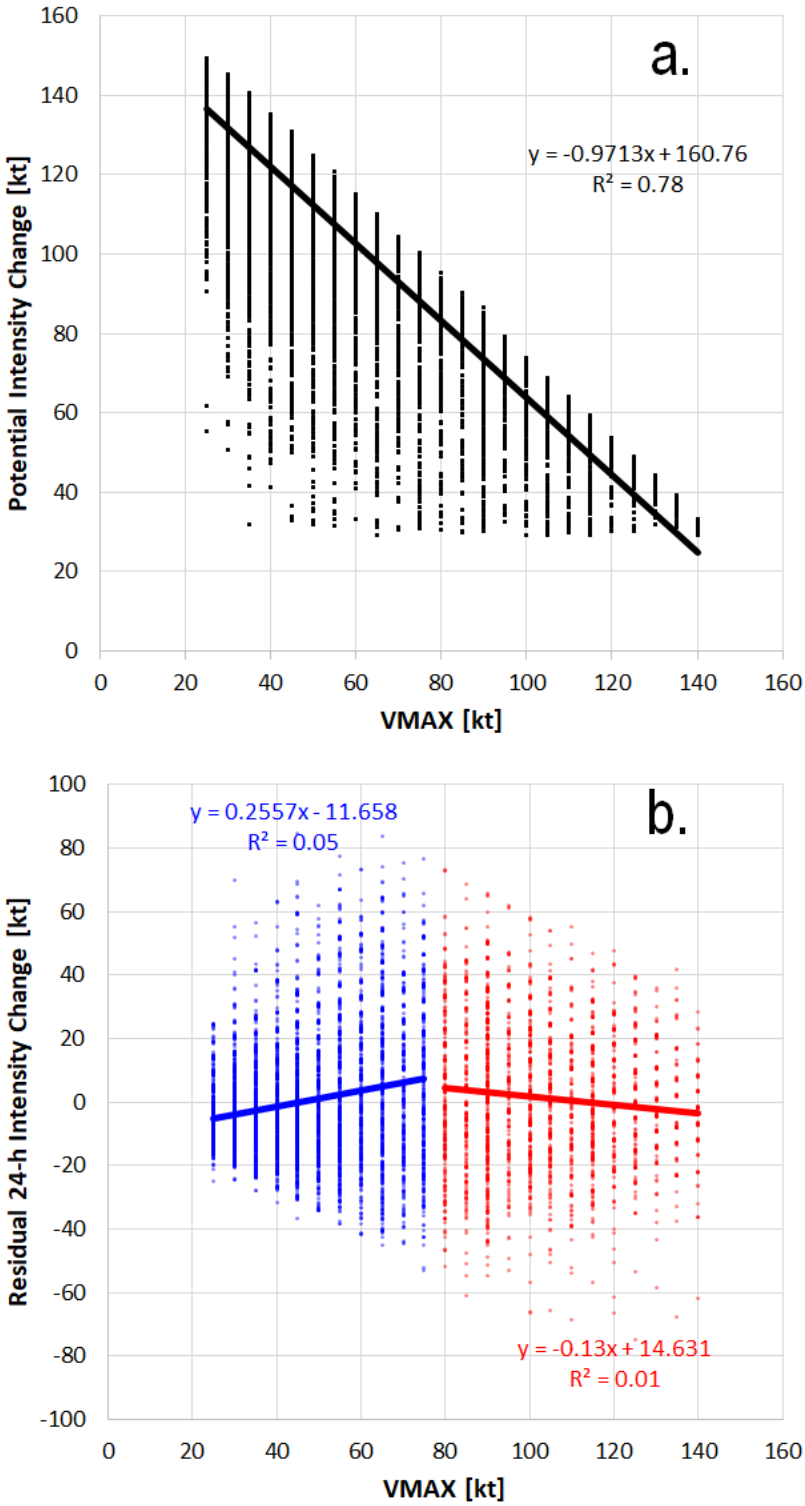

Figure 1. (a.) Scatter plots of current intensity (VMAX) vs. potential intensification (POT) and (b.) VMAX vs. the residual of a POT-based 24-h intensity change, where points with VMAX values less than 80 kt are shown in blue and those 80 kt and above shown in red. The trend line and squared correlation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

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

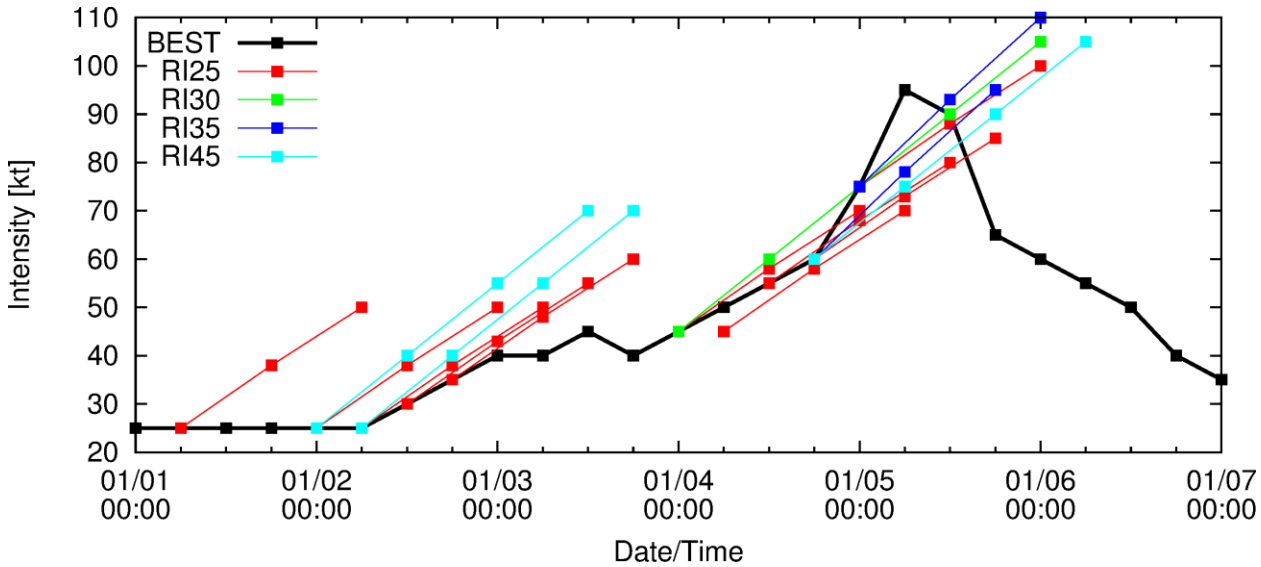

Figure 2. Time series of the intensity of Tropical Cyclone Ava (sh032018) from the best track (BEST) and deterministic forecasts triggered by RIPA. Shown are RI25, RI30, RI35 and RI45, which were all triggered. Note the number of cases triggered when Ava's estimated intensity was less than 35 knots. Landfall over Madagascar also occurred at approximately 10 UTC on 5 January, resulting in a rapid decay.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

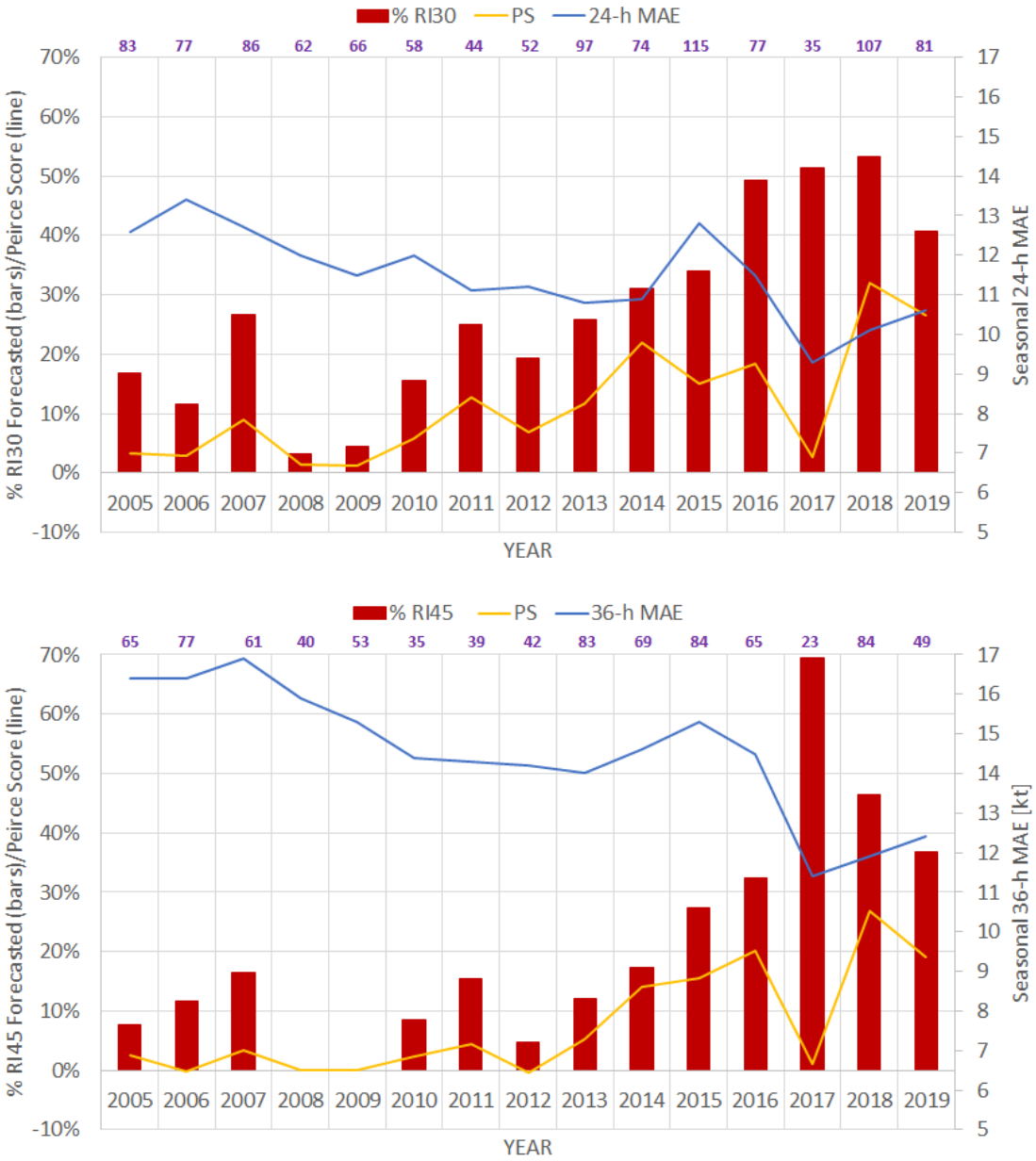

Figure 3. JTWC forecast performance for observed RI of 30 kt in 24 h (RI30; top) and 45 kt in 36 h (RI45; bottom) for the years 2005-2018 (All JTWC basins) and 2019 (Southern Hemisphere only). Red bars indicate percentage of observed RI cases for which JTWC predicted RI, blue lines indicate MAE for observed RI cases, and yellow line indicates Peirce Scores. The number of RI case is listed across the top of each panel and is displayed with purple text.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

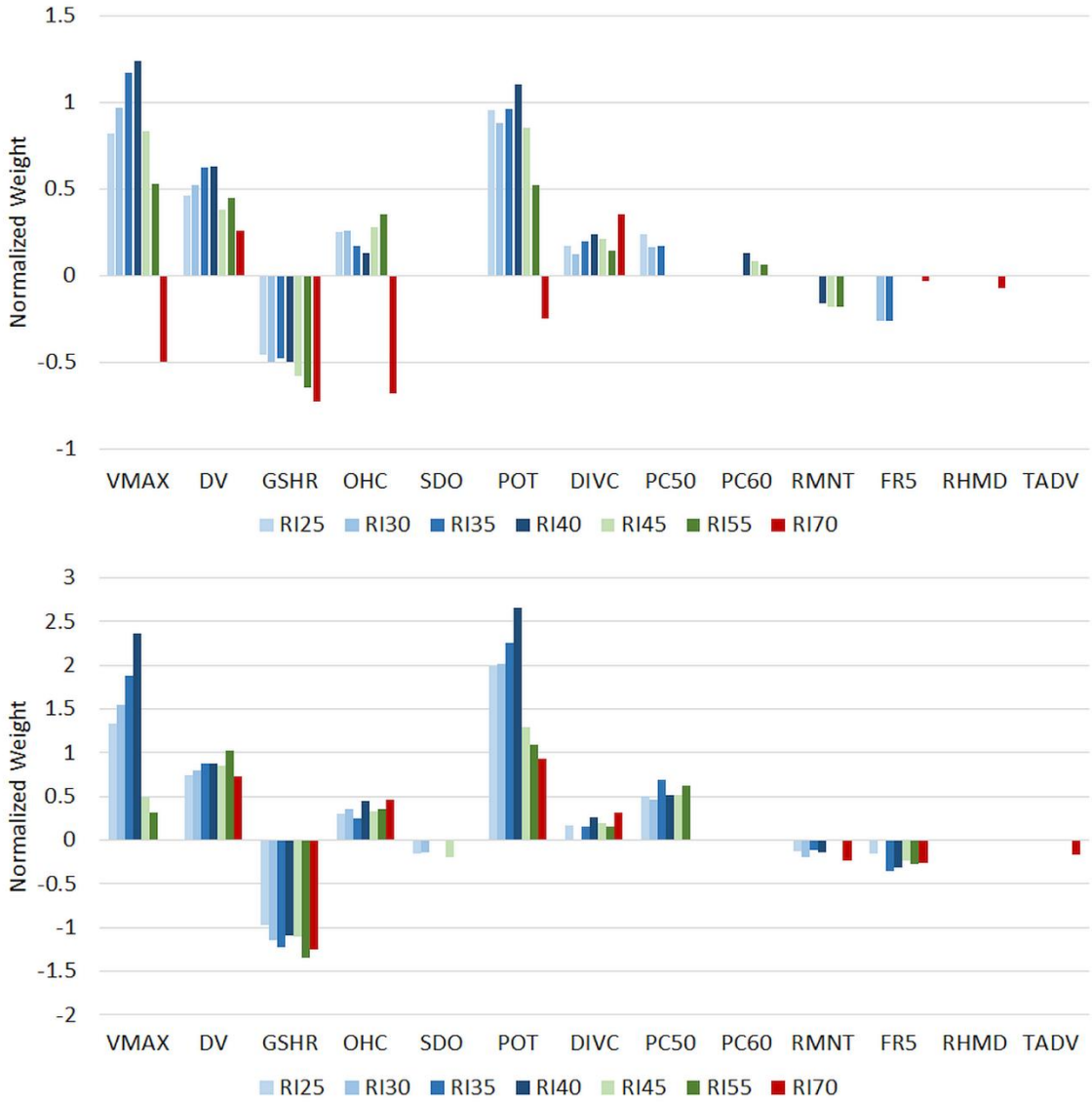

Figure 4. (Top) Normalized LDA and (Bottom) LRE model coefficients. The magnitude of each coefficient provides an indication of that variable's relative importance to predicting the various RI thresholds. The 24-h, 36-h and 48-h intensity change thresholds are shown in blue, green and red tones, respectively.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

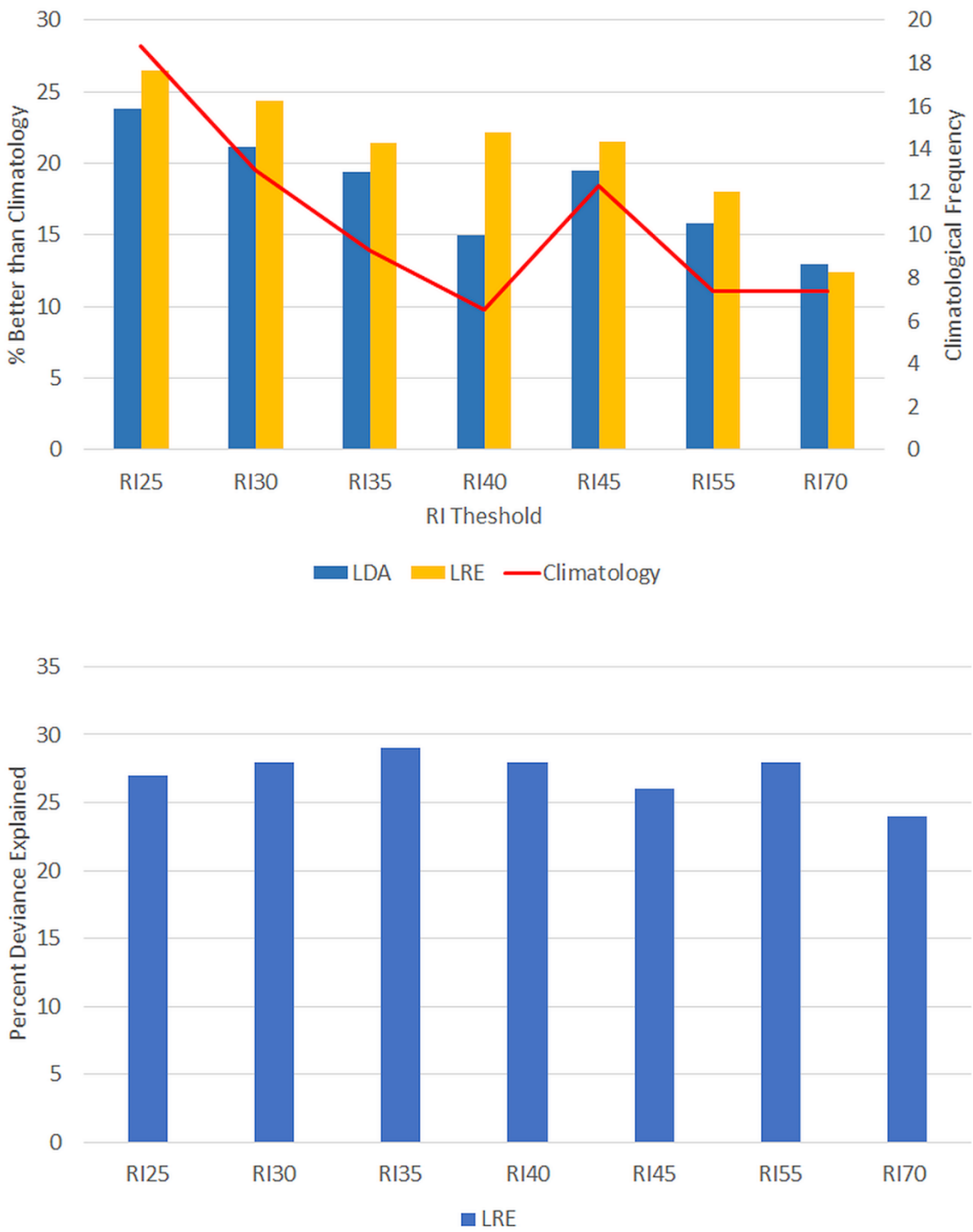

Figure 5. (Top) Brier Skill Scores for the LDA model (blue bars), the LRE model (yellow bars), and Climatology (red line) for each RI threshold. (Bottom) Percent deviance explained by the LRE model for each RI threshold. All results based on dependent data from the JTWC 2000-2017 dataset.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

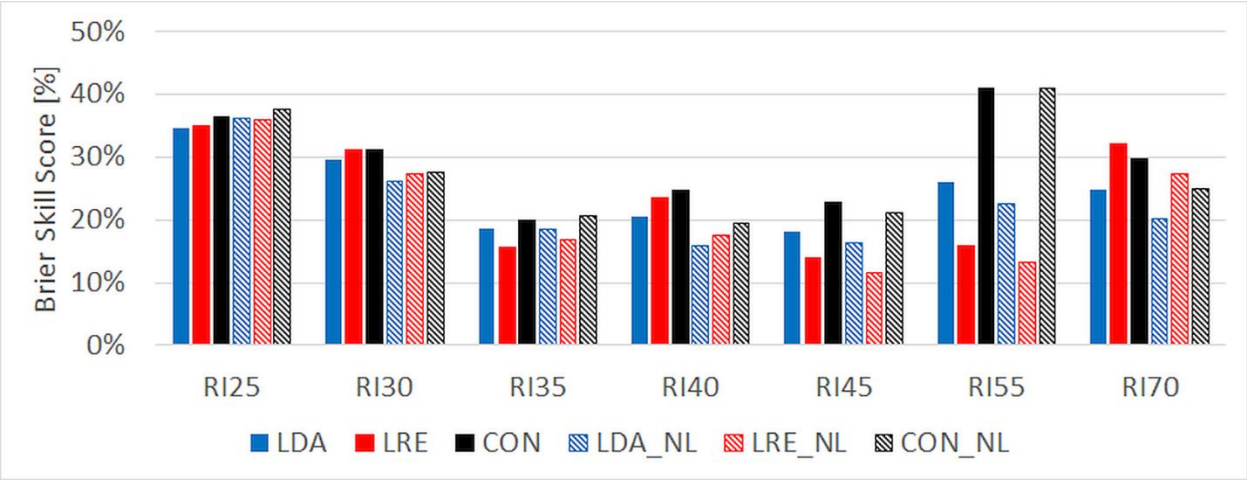

Figure 6. BSSs associated with the 2018-2019 independent verification of RI events associated with 25, 30, 35, and 40 kt changes in 24h (RI25, RI30, RI35, RI40), 45 and 55 kt changes in 36h (RI45 and RI55) and 70 kt changes in 48h (RI70). Note that final best tracks were available for 2018 and preliminary 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 48h lead times, respectively.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

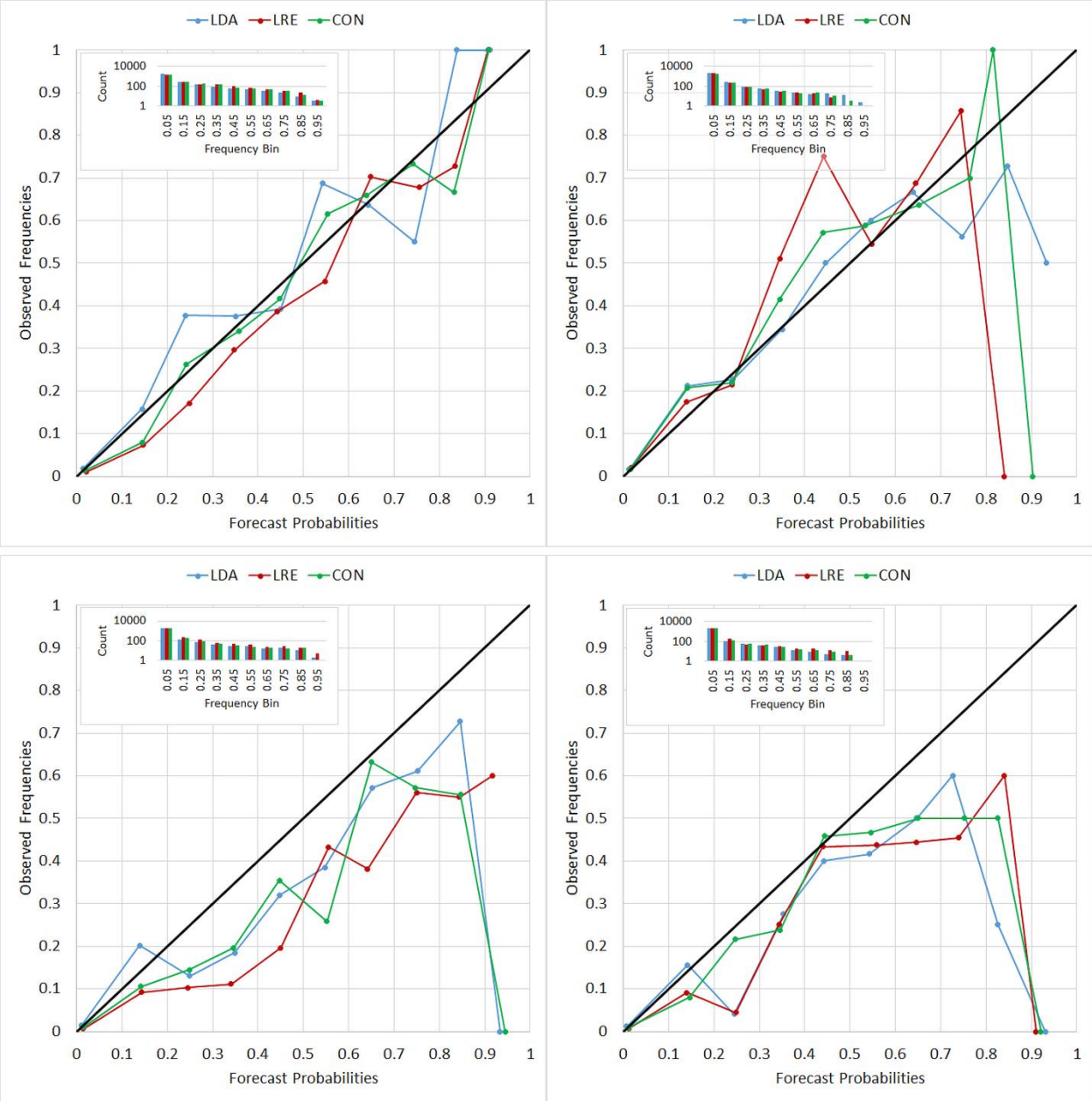

Figure 7. Reliability diagrams for (top left) RI25, (top right) RI30, (bottom left) RI35, (bottom right). Accompanying each reliability diagram is the refinement distributions that show the forecast frequency bins vesus the log of the number of case, which are inset to the upper left in each panel. Results are based on independent forecasts from 2018 and 2019 in WP, SH, and IO basins, landfalling cases removed. Final and preliminary best tracks in 2018 and 2019, respectively, are used for verification.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

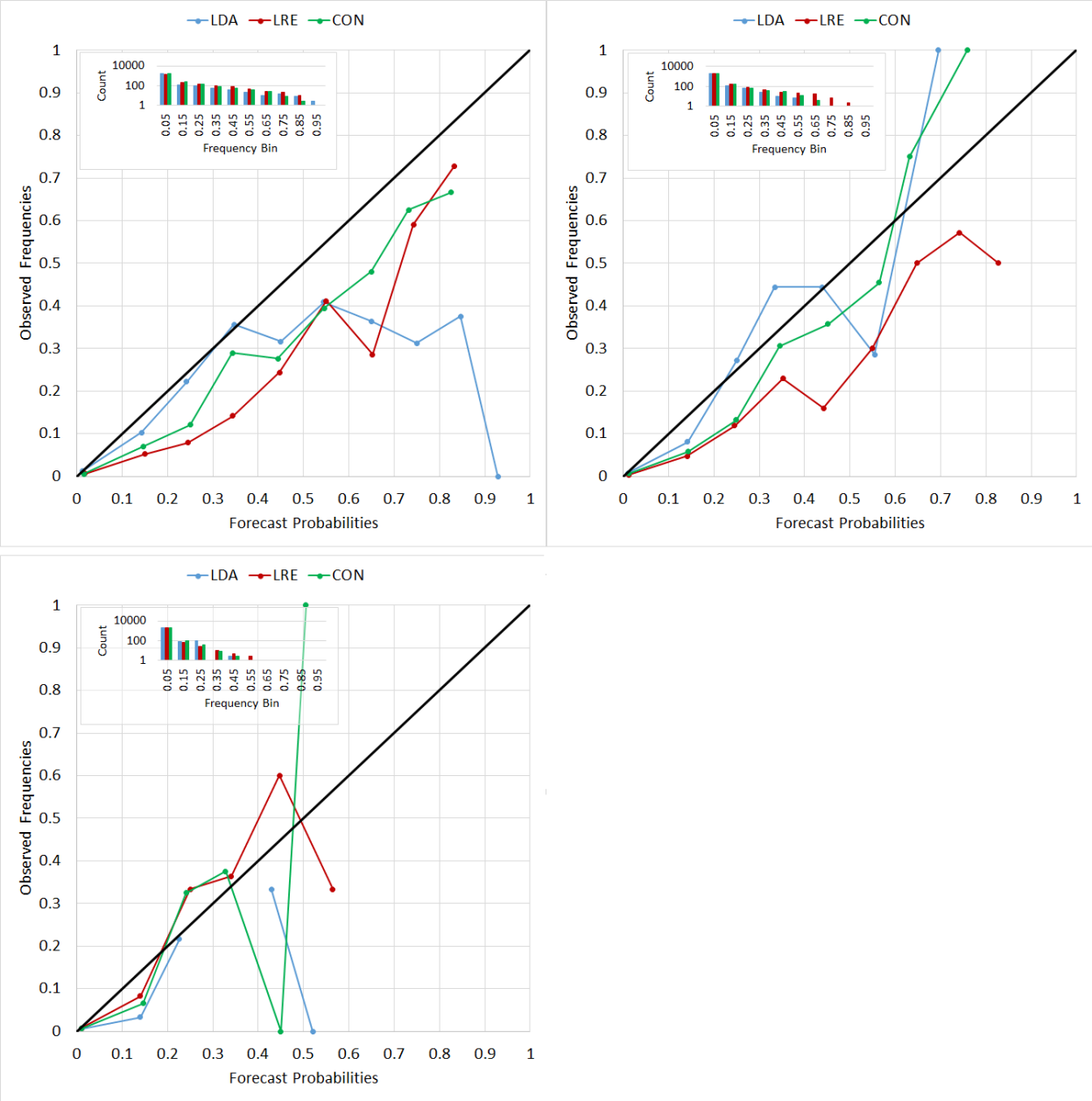

Figure 8. The same as Fig. 7, but for (upper left) RI45, (upper right) RI55, and (lower left) RI70.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.

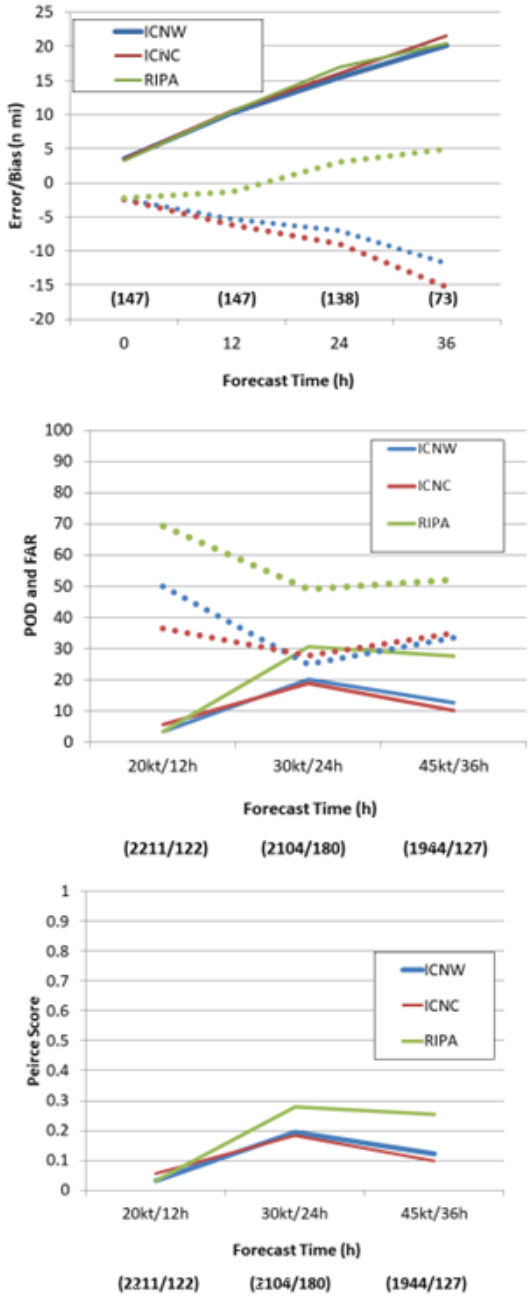

Figure 9. (Top) Mean Forecast Error (MAE, solid lines) and Bias (dotted lines), (Middle) POD (solid lines) and FAR (dotted lines), and (Bottom) Peirce Scores for RIPA, the consensus with and without RIPA (ICNW and ICNC, respectively). Evaluation includes Southern Hemisphere, western North Pacific and Indian Ocean 2018 seasons and Southern Hemisphere 2019 season. Total cases/RI cases are shown in parentheses.

Accepted for publication in *Weather and Forecasting.* DOI 10.1175/WAF-D-19-0228.1.