An operational rapid intensification prediction aid (RIPA) for the western North Pacific

John A. Knaff
NOAA Center for Satellite Applications and Research, Fort Collins, Colorado

Charles R. Sampson
Naval Research Laboratory, Monterey, California

and

Kate D. Musgrave
Cooperative Institute for Research in the Atmosphere, Fort Collins, Colorado

1. Introduction

This study develops intensification guidance designed to anticipate TC rapid intensification in the western North Pacific, a tool recently transitioned into operational use at JTWC. The work uses a statistical-dynamical methodology, where in application prognostic predictors from dynamic forecast models are used to make statistical forecasts, and uses a "perfect prog" development approach, where analyses of the prognostic data, considered perfect, are used for statistical inference (Neumann and Lawrence 1975). Section 2 describes the data and methods used to create the guidance and section 3 describes details of the statistical-dynamical models used and how those models provide the information to make deterministic forecasts. Section 4 shows results based on two years of independent data, and section 5 summarizes and provides comments on the utility of our algorithm and some ideas for improvement.

2. Data and methods

a. Independent and dependent variables

The intensity records from JTWC's best tracks provide the TC location and intensity based on a post-season reanalysis of all the data available and current operational practices. These data are in the native 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/). The files contain 6-hourly position, intensity, and wind radii information for each storm reaching tropical depression status in JTWC's warnings. In these files, the native units — operational units — are expressed in kt for intensities (i.e., 1-minute maximum sustained winds). For consistency with operational practices, this study uses kt throughout.

Table 1 contains the forecast parameters that are considered during the development of the rapid intensity change guidance. The parameters are grouped into three categories: 1) a subset of the environmental condition parameters available in the SHIPS (2017) developmental dataset, 2) storm-centered infrared (IR) imagery based initial conditions, and 3) real-time, best-track parameters. This limited set of predictors was chosen based on past research on RI forecasting, the authors' experience, and the many years of forecaster-derived insight on this topic.

Environmental condition parameters, top section of Table 1, used in this application come from the Statistical Hurricane Intensity Prediction Scheme (SHIPS; DeMaria and Kaplan 1999) developmental dataset (see SHIPS 2017). The developmental data consisted of western North Pacific TCs during the years 2000 to 2015. Most of the environmental predictors (e.g., RHMD, DIVC, and OHC) require little explanation. However, a brief description and/or justification of some of the less straightforward environmental predictors is provided. The 850- to 200-hPa layer shear (the vector difference between 200 hPa and 850 hPa) is the traditional measure of vertical wind shear, but it is probably less reliable in complicated vertical wind profiles. To capture more of the variability in the wind profile, we relying on the generalized vertical wind shear parameter (GSHR). The GSHR is the mass-weighted, root-mean-square deviations of the winds from 4 times the mass-weighted, deep-layer mean winds. The factor of four is used to make the values comparable to the more conventional measure of 200–850 hPa, and is equal to that scalar

difference for the case when shear is constant with respect to pressure. We also want to investigate a measure of upper level eddy fluxes to ascertain TC interactions with troughs. This is done using the relative eddy flux convergence (REFC, see Knaff et al. 2005). Finally, low-level temperature advection (TADV) can potentially be important to the TC intensification process (see Callaghan and Tory 2014), and is included for inspection.

IR imagery is a subjective analysis tool for predicting the onset of RI, and it is available almost instantaneously at operational centers, thus making the imagery ideal for an application such as ours. With this premise in mind, we use IR imagery to capture two characteristics related to intensification, see the middle section of parameters in Table 1. The first is convective vigor (PC50, PC60) and symmetry (SDO). The convective vigor predictors used here, PC50 and PC60, were chosen based on sensitivity work shown in Knaff et al. (2014a), where -50ºC and colder pixel counts improved discrimination or RI events in the Atlantic and East Pacific. The storm size and structure is the second characteristic provided by the IR imagery. Knaff et al. (2014a, 2016) developed a normalized IR-based TC size (FR5), where the TC size parameter (R5 from Knaff et al. 2014b) is divided by its intensity based climatological values. FR5, in essence, determines whether the TC is large or small relative to the global climatology. We include this parameter because studies (e.g., Xu and Wang 2015) find that small TCs have a tendency to intensify more quickly than larger TCs. In addition to the overall TC size, we also make use of the radius of minimum azimuthally averaged IR temperature (RMNT) — a crude measure of inner-core and eye size, the latter being inversely correlated to intensification (Musgrave 2011; Carrasco et al. 2014).

Due to the nature of operations, non-physical values can arise in the real-time, best-track parameters, bottom section of Table 1. Since large changes in operational intensity estimates are sometimes related to inspection of time-late information (e.g., microwave imagery and scatterometry) and/or restrictions mandated by diagnostic techniques (e.g., Dvorak intensity estimate), we limit our 12-hour intensification rate (DV) to physically realistic values using the following relationship (1).

$$DV = min(min(VMAX * 0.33, 17.5), DV) \quad (1)$$

This effectively limits DV to constraints used in Dvorak (1984) and provides a more stable predictor for the linear statistical techniques used in this work. Including this parameterized DV values improved dependent model fit, and sensitivity tests (not shown) indicated that capping DV in this physically consistent way resulted in better modeling of RI and answering the more basic question has the storm been intensifying.

Construction of the RI algorithm follows the same process as past RI algorithm development (e.g., Kaplan and DeMaria 2003; Kaplan et al. 2010; Kaplan et al. 2015) using a combination of current TC conditions, environmental conditions, and information about the current IR structure to forecast the probability of various intensification rates. In development, we use analyses (i.e. perfect prog) of environmental conditions. In application, environmental conditions are based on forecasts. We also use two statistical methods to create forecasts from which we construct a two-member consensus forecast. The two methods are a linear discriminant analysis and logistic regression.

b. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a classification method originally developed in Fisher (1936). In LDA, a linear combination of variables that best separates two or more groups is developed. We define just two groups for the LDA: Group 1 for when the intensification threshold is reached or exceeded, and Group 2 for when the intensification threshold is not reached. In the two-class LDA, the goal is to find the n-dimension vector of observations that best assigns a case to belonging to either Group 1 or Group 2. In our application, we assume both groups have the same covariance structure, so the vector has a direction in n-dimensional space that maximizes the distance between the means of Group 1 and Group 2 in standardized units. To estimate probabilities from the discriminant function provided by LDA, a windowing procedure relates prior probabilities (i.e., dependent data) to discriminant function values. In application, a cubic spline provides a probability given the discriminant function value.

c. Logistic regression

Logistic regression (LRE) is a model where the dependent variable is a defined category. In our case, “1” for reaching the intensification threshold and “0” for not having met the intensification threshold. LRE is a special case of the generalized linear model, where the natural

log of the odds ratio or logit based on categorical data is fit to a linear combination of independent predictors ($x_1, \ldots, x_n$) with intercept b_o and weights ($b_1 \ldots b_n$) that are determined via the method of maximum likelihood as shown in (2).

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

To perform variable selection and the model fit, we use FORTRAN 90 code from CSIRO (2017) that produces linear logistic models by iteratively reweighted least squares. Model fit is based maximum likelihood criteria. Once fitted, the probability of exceeding the intensification threshold takes the form (3).

$$p_{RI} = \frac{1}{\left(1+e^{-(b_o+b_1 x_1 \ldots + b_n x_n)}\right)} \qquad (3)$$

The measure of quality-of-fit for logistic regression is in terms of deviance — a generalization of the idea of using the sum of squares of residuals in ordinary least squares to cases, but where the model is fit using a maximum likelihood criterion. Deviance is 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).

3. Statistical-dynamical model formulations

For this work, we will examine several intensification thresholds. These include 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. These changes in the western North Pacific correspond to the 81.2, 87.0, 90.7, and 93.7 percentiles of 24-h intensity change, the 87.7 and 92.6 percentiles for the 36-h intensity change, and the 92.6 percentile of the 48-h intensity change. We will refer to the thresholds as RI25, RI30, RI35, RI40, RI45, RI55 and RI70, respectively. We also tried to predict the 85-kt increase of intensity in 72 h, but dependent fits were, in our opinion, not good enough to pursue real-time prediction. With these thresholds, we now describe the statistical-dynamical models for LDA and LRE methods for each of the intensity change threshold.

a. LDA-based models

For our intensification thresholds, we seek the model with the best developmental Brier Scores (BS; i.e., the mean square distance in probabilistic space) and Brier Skill Scores (BSS). The calculation of BSS is provided in (4), where BS_f is the BS of the forecasts and BS_r is the BS of the reference (Wilks 2006), In this case the reference is climatology.

$$BSS = 1 - \frac{BS_f}{BS_r} \qquad (4)$$

For fitting LDA models, we calculate time-averaged values of non-static predictors up until each forecast lead-time. In application, we use forecast values of these quantities for the calculation of averaged predictor values. The previous section provided the climatological rate of RI for each threshold and Table 2 provides statistics including climatological frequency of RI for each threshold, BS and BSS values, number of predictors used, and number of cases for each intensification threshold. BSS values decrease with increasing rates of intensity change and forecast difficulty.

The following set of eight predictors are used to make forecasts for RI25, RI30, and RI35 intensification rates: VMAX, DV, GSHR, OHC, PC50, SDO, POT, and DIVC. Forecasts for RI40 make use of the same predictors, save the PC50 predictor. In this case, PC60, the colder pixel count predictor, replaces PC50. The remaining intensification thresholds, RI45, RI55, and RI70, use the same predictors as RI40 with the addition of the inner core predictor RMNT.

Figure 1 shows the normalized (by their standard deviations) discriminant function weights [i.e., vector *a* in (2)] used for each intensification threshold; 24-, 36-, and 48-h lead times in blue, green and red hues, respectively. It is also important to note that VMAX and POT predictors appear to be inversely related to each other. In the absence of VMAX, one would expect POT to be positively correlated with RI, but in Fig. 1, we see the opposite. Upon further inspection, we found a much better discrimination when we used both of these variables despite their collinearity.

b. Logistic regression-based models

As described in section 2, we used logistic regression (LRE) to create probabilistic forecast modes for the intensity change thresholds. For the development of these models, we used the same potential predictors (Table 1), but here we seek to minimize the deviance (i.e., maximize the deviance explained) for each intensity change threshold. Table 3 presents number of predictors, the deviance explained, and BSS for each intensity change threshold. It is interesting to note that the static predictors DV, PC50, and

RMNT, and time-varying predictors GSHR and OHC are selected for every lead-time. In the LRE-based models, the importance of the SDO predictor is reduced compared to the LDA model. Also, the LRE predictor selection included FR5 (TC size) as a predictor for many of the lead times while it was not for the LDA model. RHMD and TADV become important specifically for the RI70 forecasts in the LRE model, and the reasons for this and other predictive relationships is now explored.

Since the coefficients of the logistic regression have essentially the same convention and meaning as linear regression coefficients, we show the normalized coefficients for these models in Fig. 2. The positively correlated RHMD (relative humidity in a 200-800 km annulus around the TC) and negatively correlated TADV (temperature advection within 500 km of the center) that are important specifically for the 48-h RI70 forecasts appear on the right of the figure. These are likely related to extra-tropical transition cases. Positive temperature advection and decreased humidity could both be effects from an approaching mid-latitude trough, and so should suppress RI. As is the case for LDA, the LRE predictors VMAX and POT are covariant. Both TC size (FR5) and inner-core size (RMNT) make significant contributions in most models suggesting that smaller storms and those with coldest brightness temperatures near the center are more likely to be rapid intensifiers.

The differences in weights and predictors in the LDA and LRE models suggest that the results of these two methods may be different. This difference implies some independence in the methods, making them ideal for use in a consensus forecast. Consensus forecasts, averages of forecasts from more than one method, have been shown to yield improvements over individual forecasts in a variety of fields. Furthermore, these studies all indicate that the degree of independence among consensus members is an important factor when combining forecasts; contributing to forecast improvements (e.g., Sampson et al. 2008, Appendix B). With the goal of creating the most skillful forecasts and useful guidance for rapid intensification, the next section discusses the preprocessing and combining the LDA- and LRE-based models and creation of deterministic intensity forecasts.

c. Preprocessing and combining forecasts and deterministic forecasts

The LDA- and LRE-based models described above provide probabilistic RI forecasts for distinct intensification thresholds and lead times. The models use slightly different predictors and therefore exhibit some degree of independence. As a result, the RI model forecasts can occasionally produce probabilities that are inconsistent, that is, a higher RI threshold (e.g., RI35) may have a higher probability than the lower RI thresholds for that lead time (RI25 and RI30). In such cases, probabilities of the lower RI thresholds are assigned the probability of the higher RI threshold. In our example above, the probabilities associated RI25 and RI30 are assigned the RI35 probability. This consistency check among RI thresholds is performed for both 24- and 36-h forecasts, and for each forecast methodology (LDA and LRE) independently. Following the consistency check, we use probabilities from the two forecast methodologies for each intensification threshold to create an equally weighted average (CON).

Traditionally the rapid TC intensification forecast problem led to categorical/binary (Mundell 1990) or probabilistic (i.e., DeMaria and Kaplan 2003) forecasts. Despite the reasoning for this decision, forecasters are still required to provide a deterministic forecast of intensity. To address the negative biases in RI forecasts, Sampson et al. (2011) demonstrates a method to provide deterministic intensity forecasts based on RI forecast probabilities. In this method, threshold values of the probabilistic forecast trigger deterministic forecasts for the valid forecast lead-time. For instance, when the RI35 forecast exceeds the threshold probability, the algorithm generates a twelve-hourly deterministic forecast of 35 kt in 24 h starting with the observed intensity at t=0 and ending 24 h later with an intensity of the initial intensity plus 35 kt. These deterministic forecasts are then added to the operational intensity consensus forecast at JTWC that is the most skillful intensity guidance (DeMaria et al. 2014). Both mean errors and biases are smaller when the intensity forecast aid contains deterministic RI aids (Sampson et al. 2011).

The threshold probability for triggering rapid aids is determined using past forecasts. For our purposes, and based on independent CON forecasts, we found the threshold of ~ 40% ± 10% for all the intensification thresholds. For this reason, the 40% probability triggers deterministic rapid intensification forecast aids for RI25, RI30, RI35, RI40, RI45, RI55, and RI70 based on the CON forecast probabilities. Only the deterministic

member with the highest intensification rate for each lead-time is triggered.

Above, we described probabilistic rapid intensification aids using LDA and LRE methods for seven intensification thresholds, RI25, RI30, RI35, RI40, RI45, RI55, and RI70. We now discuss the independent verification of LDA, LRE and CON forecasts in the next section.

4. Independent verification

To evaluate the performance of the models described in section 3, we now present verification statistics based on almost two years of independent forecasts produced as part of the vetting process. Verification data includes all forecasts from the 2016 season and forecasts through 27 October 2017. We present BSS with the climatology of the individual intensification thresholds, reliability, and the verification of the impact of the deterministic rapid intensification members triggered by the probabilistic models.

Table 4 shows the independent BSSs [%] for LDA, LRE, and CON forecasts for each intensification threshold. This is admittedly a limited sample with less than two years of verification. Brier Skill Scores show that the LRE methods are skillful for all the intensification threshold forecast models developed, and that the LDA method produces skillful forecasts for only the RI25 intensification threshold model. Consensus forecasts provided skill for all but the RI55 cases. These results are discouraging for the LDA-based methods, but suggest that the LRE and CON have delivered skillful guidance. Results are subject to change due to changes in final 2017 best tracks particularly because modification of best track intensities is likely when rapid intensity changes occur. Nonetheless, results are encouraging. Experience also suggests that at least three independent typhoon seasons are typically needed to make solid verification inferences.

Figure 3 shows reliability diagrams associated with our sample of independent forecasts. The LDA methods generally produce low biased reliability (i.e., below the 1:1 line) where as for the more rare intensification thresholds LRE are indicative of high bias. We wish we could report that the CON is the best method in this sample, but it appears that the LRE have better reliability for RI25, RI30, RI35, and RI45. The CON forecasts generally produce reliabilities in between the LDA and LRE, but this is not always the case; suggesting that there may be a fair bit of independence between LDA and LRE methods (e.g., RI45 for higher probabilities and frequencies). In general, these performance results are similar to rapid intensification aids in other basins. It is also interesting to note that the 40% CON forecasts in general would trigger deterministic forecasts for about 15-25% of the cases (i.e., over confidence).

The addition of deterministic forecast of rapid intensification based on the probabilistic models developed here should help with this issue by reducing the biases and possibly reducing the MAEs. Results based on our independent forecast sample used here have found this to be the case. Using the ideas presented in Sampson et al. (2011) and a 40% triggering probability, results in a dramatic decrease in the biases show a reduction in MAE for 24-, 36-, and 48-h forecasts. These results represent significant improvement of season intensity bias given the slow rate of change in intensity forecast performance. Table 5 shows the intensity verifications of the intensity consensus with the deterministic rapid intensification member versus those without. The 12- and 24-h periods have enough cases to get a sense of the performance. The improvement in MAE at 24 h is 0.4 kt and the expected low bias for RI cases is reduced by 1.9 kt. The improvements in MAE are very small or negative, but the reduction in bias is relatively large. The performance at the 36- and 48-h periods is promising, especially on inspection of the individual cases, but there are too few cases to make any firm conclusions other than the biases are reduced.

The improvements in MAE are small for these time periods, but this is partly a construct of availability. The RI aids at 24 h are available in 15% of all forecasts. As discussed in Sampson et al. (2011), raising the threshold above 40% improves the performance in terms of MAE and lessens the bias correction, but also reduces the availability. One can think of the 40% threshold as a tuning knob, turn it up and you get better MAEs, less bias correction, and fewer RI forecasts nudging the consensus. Turn it down and you typically increase the MAE, improve the bias correction, and more often nudge the consensus. In this work, we turned the knob, up to 50% and noticed the same behavior as discussed in detail in Sampson et al. (2011, their Figure 2), and decided that 40% would also work in this basin.

5. Summary

This manuscript describes the development of TC rapid intensification models for the western

North Pacific Basin. We chose seven intensification threshold including 25-, 30-, 35-, and 40-kt intensity changes in 24 h, 45- and 55-kt intensity changes in 36 h, and 70-kt intensity changes in 48 h, also referred to as RI25, RI30, RI35, RI40, RI45, RI55, and RI70, respectively. The models were developed as probabilistic algorithms following work in other basins and using two methodologies: Linear Discriminant Analysis (LDA) and Logistic Regression (LRE). The years 2000 to 2015 were used in the development, and then independent testing was performed with the years 2016 and 2017. Equally-weighted averages of the LDA and LRE probabilities are computed (CON), and those are then used to trigger deterministic forecasts for each of the seven intensification thresholds when the probabilities reach 40%.

Dependent LDA models had Brier Skill Scores ranging from 13-24% that indicated they could produce skillful (relative to climatology) probabilistic forecasts. Dependent LRE models, on the other hand, explained a little less than 30% of the deviance and had slightly better BSSs than LDA models. In independent tests, LDA models with the exception of RI25 failed to produce skillful forecasts, but both LRE and CON produced skillful probabilistic forecasts. We also examined reliability diagrams for independent forecasts of LDA, LRE, and CON probabilities. The reliability results are typical for the rapid intensification problem, resulting in forecasts that are over confident (e.g., Kaplan et al. 2015). The over confidence suggests that the 40% CON forecast threshold would ultimately trigger deterministic forecasts for about 15% of the cases.

In independent verification with the 2016 and 2017 data, the JTWC intensity consensus that includes deterministic RI guidance clearly shows reduced negative biases and somewhat improved MAEs (i.e., reduction in RMSE), indicating that some independence and skill are garnered by including them in the consensus. The deterministic RI aids should also give forecasters improved intensity guidance spread, though this is yet to be shown.

Acknowledgements: This work was funded by the Office of Naval Research through the Broad Area Announcement # N00173-17-S-BA01 and Program Elements 0602435N. The authors would like to thank Alan J. Miller for the software used to create the logistic regression models. 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:

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

CSIRO, 2017: Software from Alan J. Miller. [Available on-line at http://wp.csiro.au/alanmiller/]

DeMaria, M., and J. Kaplan, 1999: An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, 14, 326–337.

Dvorak, V. F., 1984: Tropical cyclone intensity analysis using satellite data. NOAA Tech. Rep. 11, 45 pp. [Available from NOAA/NESDIS, 5200 Auth Rd., Washington, DC 20333.]

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

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

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

Kaplan, J., 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: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, 2103 – 2116, doi:10.1175/WAF-D-17-0037.1

Knaff, J. A., M. DeMaria, S. P. Longmore, and R. T. DeMaria, 2014a: Improving tropical cyclone guidance tools by accounting for variations in size. *31st Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor.

Soc., 51. [Available online at https://ams.confex.com/ams/31Hurr/webprogram/Paper244165.html.]
Knaff, J. A., S. P. Longmore, and D. A. Molenar, 2014b: An objective satellite-based tropical cyclone size climatology. *J. Climate*, 27, 455–476. doi:10.1175/JCLI-D-13-00096.1
Knaff, J. A., C. J. Slocum, K. D. Musgrave, C. R. Sampson, and B. R. Strahl, 2016: Using routinely available information to estimate tropical cyclone wind structure. *Mon. Wea. Rev.*, 144, 1233-1247. doi:10.1175/MWR-D-15-0267.1
Knaff, J. A., C. R. Sampson, and M. DeMaria, 2005: An operational statisical typhoon intensity prediction scheme for the Western North Pacific.*Wea. Forecasting*, 20(4), 688-699.
Mundell, D. B., 1990: Prediction of tropical cyclone rapid intensification events. M.S. thesis, Dept. of Atmospheric Science, Colorado State University, 186 pp. [Available from Dept. of Atmospheric Science Colorado State University, Fort Collins, CO 80523.].
Musgrave, K. D., 2011: Tropical cyclone inner core structure and intensity change, PhD dissertation, Dept. of Atmospheric Science, Colorado State University, 103 pp. [Available from Dept. of Atmospheric Science Colorado State University, Fort Collins, CO 80523.].
Neumann, C.J. and M.B. Lawrence, 1975: An operational experiment in the statistical-dynamical prediction of tropical cyclone motion. *Mon. Wea. Rev.,* 103, 665–673,DOI: 10.1175/1520-0493(1975)103<0665:AOEITS>2.0.CO;2
Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (Version 3.2). *Bull. Amer. Meteor. Soc.*, 81, 1231–1240.
Sampson, C. R., J. Kaplan, J. A. Knaff, M. DeMaria, and C. Sisko, 2011: A deterministic rapid intensification aid. *Wea. Forecasting*, 26, 579–585.
Sampson, C. R., J. L. Franklin, J. A. Knaff, and M. DeMaria, 2008: Experiments with a Simple Tropical Cyclone Intensity Consensus. *Weather and Forecasting*, 23, 304–312.
SHIPS, 2017: “SHIPS developmental data.” [Available on-line at http://rammb.cira.colostate.edu/research/tropical_cyclones/ships/developmental_data.asp]
Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences: An Introduction*, Second Edition*,* 627 pp., Academic, San Diego, Calif.
Xu, J., and Y. Wang, 2015: A statistical analysis on the dependence of tropical cyclone intensification rate on the storm intensity and size in the North Atlantic. *Wea. Forecasting,* 30, 692–701, doi:10.1175/WAF-D-14-00141.1

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 26^{o}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 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 (m/sec/day) calculated in 100-600 km annulus vs. time
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^{o}C within a 50-200 km annulus*
PC60	*Percentage of IR pixels colder than -60^{o}C within a 50-200 km annulus*
SDO	*Standard Deviation of IR brightness temperatures 100-300 km*
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)*
DV	*12-hour change in TC intensity, which is limited by the following function,* $DV = min(min(VMAX * 0.33, 17.5), DV)$

Table 2. Statistics associated with the LDA models for various intensity change thresholds.

Intensification Threshold	Climatological Frequency (%)	BS	BSS	Number of Predictors	Number of Cases
RI25	18.8	1162.8	23.8	8	5447
RI30	13.0	890.4	21.1	8	5396
RI35	9.3	681.0	19.4	8	5301
RI40	6.5	519.7	15.0	8	5211
RI45	12.3	871.4	19.5	9	4432
RI55	7.4	576.4	15.8	9	4166
RI70	7.4	594.7	13.0	9	3290

Table 3. Number of predictors and percent deviance explained by dependent logistic regressions for the intensification thresholds used in this study. Brier Skill Scores are also provided for comparison with LDA-based models in the column labeled BSS.

Intensity Change Threshold	Number of Predictors	Percent Deviance Explained	BSS
RI25	10	27	26.5
RI30	8	28	24.4
RI35	9	29	21.4
RI40	9	28	22.2
RI45	10	26	21.5
RI55	9	28	18.0
RI70	9	24	12.4

Table 4. Probabilistic RI algorithm evaluation. Data set is independent and includes western North Pacific cases from January 1, 2016 through October 27, 2017.

	RI25	*RI30*	*RI35*	*RI40*	*RI45*	*RI55*	*RI70*
# Forecasts	678	678	678	678	601	601	531
% Climate	12	13	9	6	12	6	4
% Observed	14	10	7	5	10	6	5
LDA BSS (%)	2	-4	-32	-10	-19	-27	-5
LRE BSS (%)	18	13	10	9	11	4	9
CON BSS (%)	13	8	2	5	4	-2	5

Table 5. Deterministic intensity consensus evaluation. ICNW is the JTWC operational consensus that includes the deterministic RI aids while ICNC is the same consensus without RI aids. Cases include only those where at least one of the deterministic RI aids for a given forecast time was available. Data set is independent and includes western North Pacific cases from January 1, 2016 through October 27, 2017.

Consensus	*Statistic*	*12h*	*24h*	*36h*	*48h*
		127	125	58	14
ICNW	MAE (kt)/Bias (kt)	8.9/-3.7	13.9/-7.2	18.8/-12.8	16.5/-4.8
ICNC	MAE (kt)/Bias (kt)	9.0/-4.6	14.3/-9.1	18.7/-14.2	14.4/-6.9
Percent Improvement	MAE (%)/Bias (%)	1.0/19.6	2.7/21.0	-0.5/9.9	-14.6/30.4

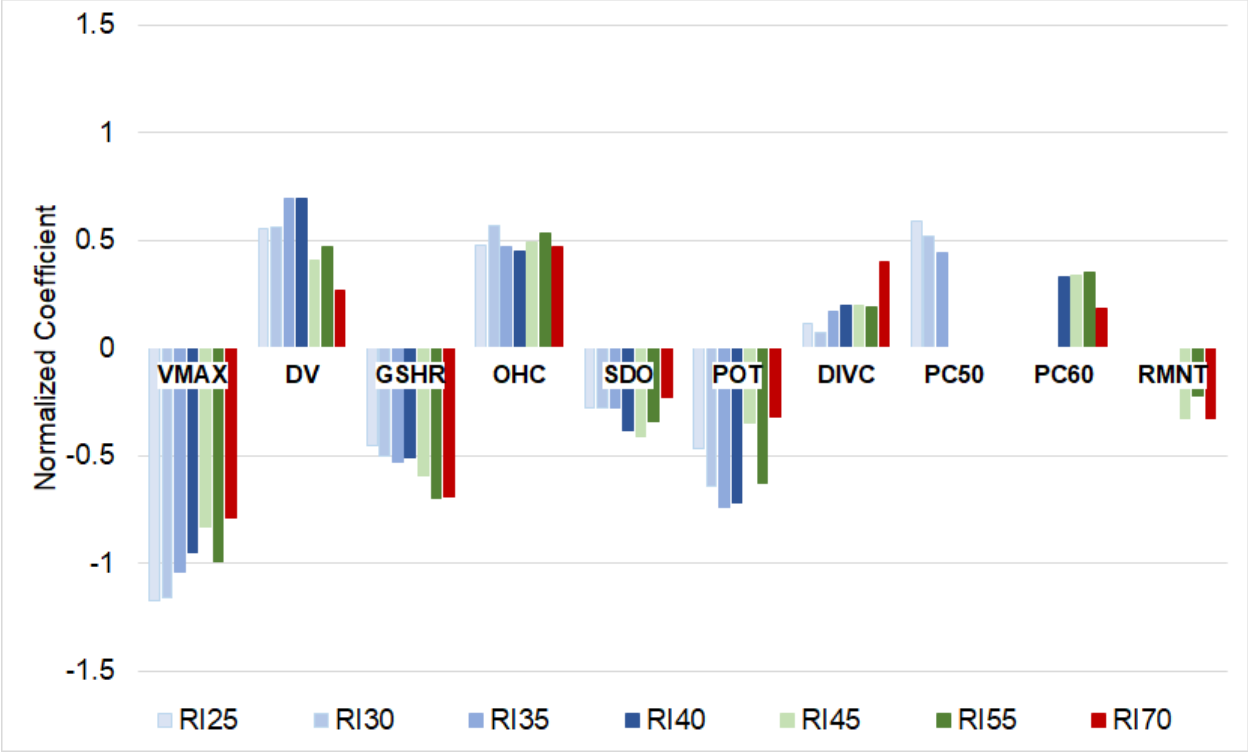

Figure 1. Normalized discriminant function coefficients for LDA models of (blue hues) RI25, RI30, RI35, and RI40, (green hues) RI45 and RI55, and (red) RI70. Predictors, which are listed just under the y=0 line, are described in Table 1.

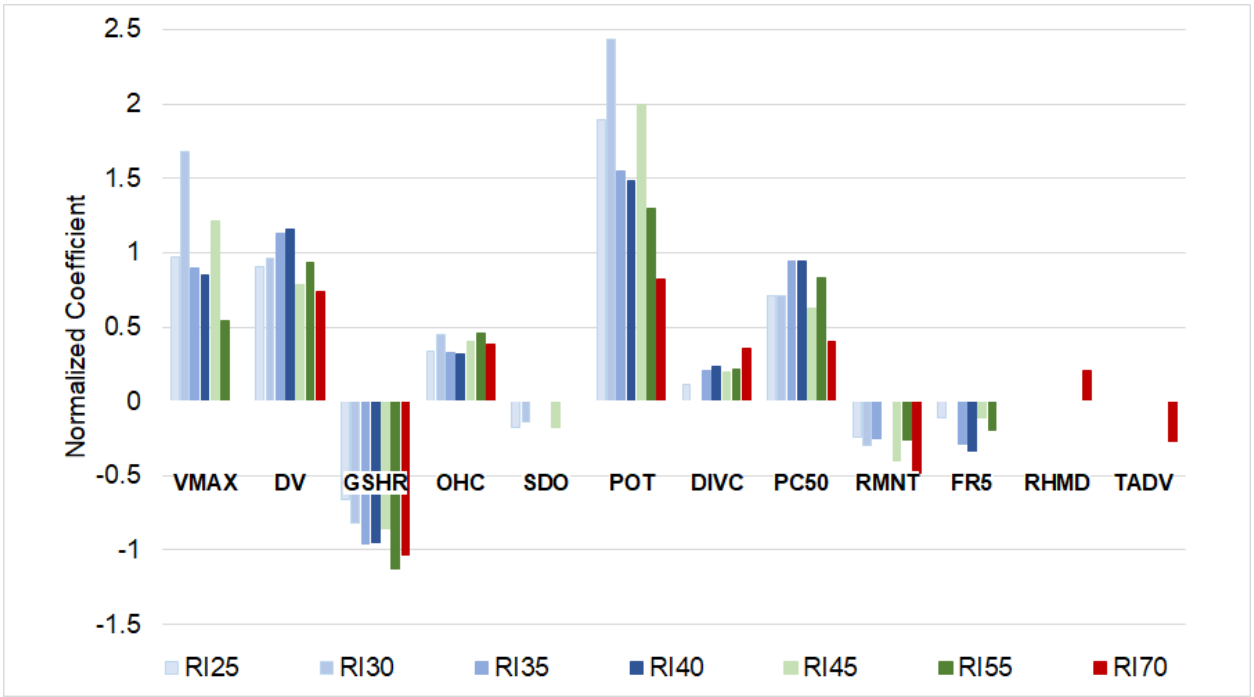

Figure 2. Normalized discriminant function coefficients for LRE models of (blue hues) RI25, RI30, RI35, and RI40, (green hues) RI45 and RI55, and (red) RI70. Predictors, which are listed just under the y=0 line, are described in Table 1.

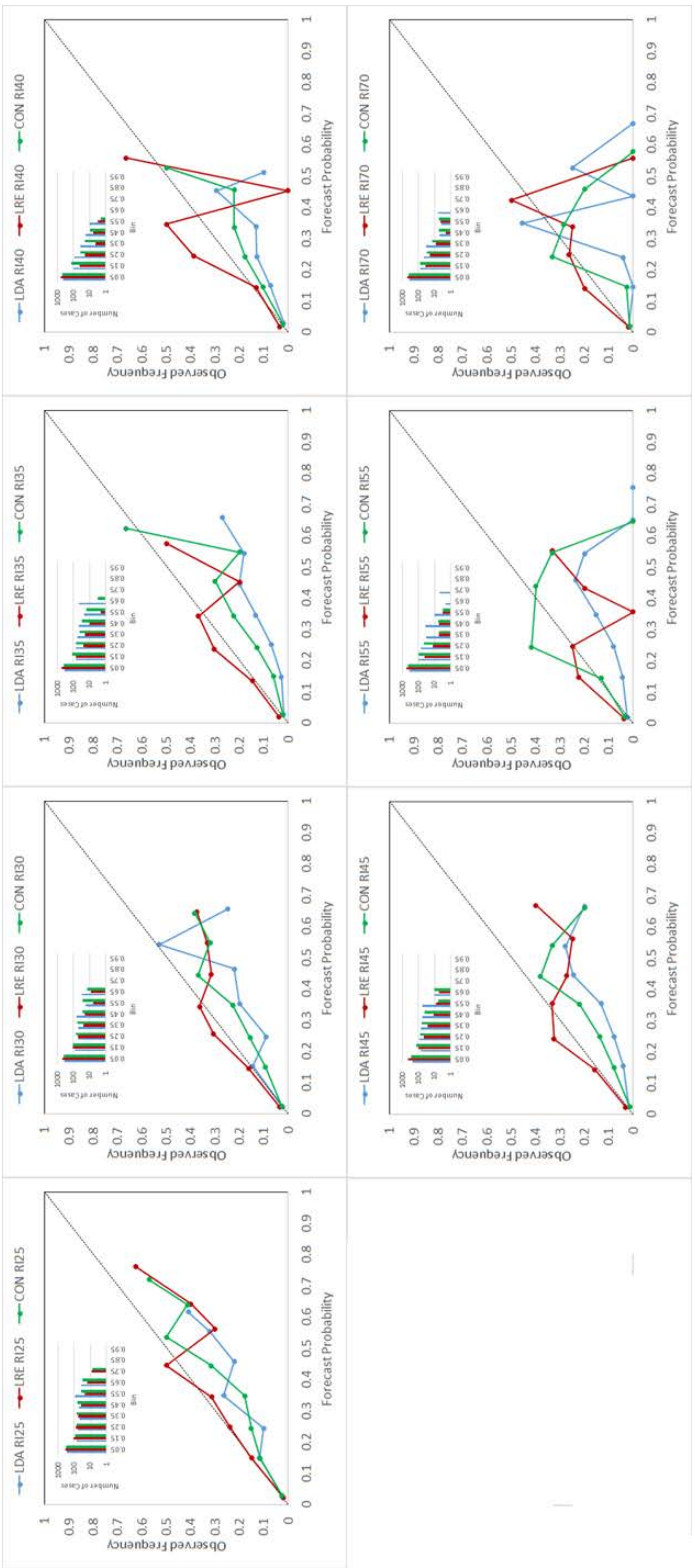

Figure 3. Reliability diagrams, (blue) LDA, (red) LRE and (green) CON, based on independent forecasts are shown. The top row shows reliability diagrams for RI25, RI30, RI35, and RI45, and the bottom row displays RI45, RI55, and RI70. Note the distributions of cases are shown in the upper left of each panel and have a log scale.