

## OPERATIONAL PREDICTION SYSTEM NOTES

### ③Deterministic Rapid Intensity Forecast Guidance for the Joint Typhoon Warning Center's Area of Responsibility

C. R. SAMPSON<sup>a</sup>, J. A. KNAFF,<sup>b</sup> C. J. SLOCUM,<sup>b</sup> M. J. Onderlinde,<sup>c</sup> A. BRAMMER,<sup>d</sup> M. FROST,<sup>a</sup> AND B. STRAHL<sup>e</sup>

<sup>a</sup> Naval Research Laboratory, Monterey, California

<sup>b</sup> NOAA/Center for Satellite Applications and Research, Fort Collins, Colorado

<sup>c</sup> NOAA/National Hurricane Center, Miami, Florida

<sup>d</sup> Cooperative Institute for Research in the Atmosphere, Colorado State University Fort Collins, Fort Collins, Colorado

<sup>e</sup> Joint Typhoon Warning Center, Pearl Harbor, Hawaii

(Manuscript received 17 May 2023, in final form 13 September 2023, accepted 25 September 2023)

**ABSTRACT:** Intensity consensus forecasts can provide skillful overall guidance for intensity forecasting at the Joint Typhoon Warning Center as they provide among the lowest mean absolute errors; however, these forecasts are far less useful for periods of rapid intensification (RI) as guidance provided is generally low biased. One way to address this issue is to construct a consensus that also includes deterministic RI forecast guidance in order to increase intensification rates during RI. While this approach increases skill and eliminates some bias, consensus forecasts from this approach generally remain low biased during RI events. Another approach is to construct a consensus forecast using an equally weighted average of deterministic RI forecasts. This yields a forecast that is generally among the top performing RI guidance, but suffers from false alarms and a high bias due to those false alarms. Neither approach described here is a prescription for forecast success, but both have qualities that merit consideration for operational centers tasked with the difficult task of RI prediction.

**SIGNIFICANCE STATEMENT:** Forecasters at the Joint Typhoon Warning Center are required to make intensity forecasts every watch. Skillful guidance is available to make these forecasts, yielding lower mean absolute errors and biases; however, errors are higher for tropical cyclones either undergoing rapid intensification or with the potential to do so. This effort is an attempt to mitigate higher errors associated with rapid intensification forecasts using existing guidance and consensus techniques. Resultant rapid intensification guidance can be used to reduce operational forecast intensity forecast errors and provide advanced warning to customers for these difficult cases.

**KEYWORDS:** Forecast verification/skill; Forecasting techniques; Operational forecasting; Postprocessing

### 1. Introduction

The Joint Typhoon Warning Center (JTWC; see [appendix A](#) for acronyms used within this work) in Pearl Harbor, Hawaii, is tasked with forecasting tropical cyclone (TC) track, intensity, and wind radii for the United States west of 180° in the Northern Hemisphere and for the entire Pacific and Indian Oceans in the Southern Hemisphere, commonly referred to as JTWC's Area of Responsibility or AOR. Overall, JTWC intensity forecast errors have been improving in the last 20 years ([Francis and Strahl 2022](#); [JTWC 2021](#)), with seasonal mean absolute intensity errors at longer forecast periods (e.g., 72 h) dropping from approximately 20 kt (1 kt  $\approx$  0.51 m s $^{-1}$ ; knots are used for the remainder of this work because they are the operational unit reported by JTWC) to approximately 15 kt. For

the shorter forecasts (e.g., 24 h) the trend is more subtle, averaging just over 10 kt in the 2000s and under 10 kt in the 2010s and 2020s.

Reduced mean absolute intensity forecast errors are a success story for the entire TC community and should be commended; however, issues remain. Individual forecasts can have errors in excess of 30 kt, even at 24-h lead times. Many of these larger error forecasts occur during TC rapid intensification (RI). These are not only some of the largest intensity errors in operations, they also represent some of the more important forecasts as they could result in preparedness and ship-routing issues that leave people and assets in dangerous locations.

The TC research community devoted significant effort addressing RI forecasts for the last 20 years, and JTWC has seen some gains ([Knaff et al. 2020](#)). Numerical weather prediction efforts (e.g., the U.S. Navy's Coupled Ocean–Atmosphere Mesoscale Prediction System for Tropical Cyclones or COAMPS-TC, see [Doyle et al. 2014](#)); and the Hurricane Weather Research and Forecast (HWRF) Model (see [Biswas et al. 2018](#)) have recently led to RI forecast improvements using metrics such as Peirce scores (also called Peirce skill

③ Denotes content that is immediately available upon publication as open access.

Corresponding author: C. R. Sampson, buck.sampson@nrlmry.navy.mil

scores; Peirce 1884; Manzato 2007; Ebert and Milne 2022) and threat scores (Wilks 2006). JTWC also has access to deterministic forecasts from algorithms designed specifically to address RI prediction. These include the Rapid Intensity Prediction Aid (RIPA; Knaff et al. 2018, 2020), the Forest-based Rapid Intensification Aid (FRIA; Slocum 2021), the Rapid Intensification Deterministic Ensemble (RIDE; Knaff et al. 2023), a deterministic algorithm based on the Deterministic To Probabilistic Statistical Model (DTOPS; DeMaria et al. 2021) converted to run in JTWC basins on the Automated Tropical Cyclone Forecast System (ATCF; Sampson and Schrader 2000), a deterministic forecast based on the COAMPS-TC Ensemble (Komaromi et al. 2021), and a deterministic algorithm based on the Coupled Hurricane Intensity Prediction System (CHIPS Ensemble; Emanuel et al. 2004). As a result, JTWC's TC intensity forecasts have shown slow and steady improvement from 2018 to 2021 (Zhang et al. 2022). However, glaring issues still exist for individual cases (Francis and Strahl 2022) and so efforts to improve RI forecasting need to continue. The effort described here is one attempt to do so.

The purpose of this work is to develop skillful deterministic consensus approaches to assist JTWC operational forecasters in RI forecasting, noting that JTWC intensity forecasts are deterministic even though many RI aids are developed as probabilistic. The first approach is to construct a consensus based solely on the six deterministic RI algorithms to investigate whether independence can improve RI skill scores as it does intensity forecast means (see Sampson et al. 2008 for discussion of independence in intensity forecasts). The second approach involves adding deterministic RI forecasts to the existing operational consensus in an attempt to improve Peirce scores and remove negative biases apparent during RI events. In section 2 we review four previously documented deterministic RI forecast algorithms and introduce two RI algorithms constructed from existing ensembles as examples of relatively simple ways to construct deterministic RI forecasts. We then freeze the algorithms and consensus algorithms, so they can be tested on independent data. In section 3 we run our deterministic RI forecasts and consensus on 2021 and 2022 TCs in JTWC AOR as independent data and evaluate results. In section 4, we discuss potential use and improvement of the newly developed consensus forecasts along with follow-on efforts that would likely improve both the guidance and operational forecasts of RI.

## 2. Data and methods

### a. Forecast guidance

As most NWP models require one forecast cycle (a 6-h period) to complete their forecast, the resultant forecast tracks and intensities produced by a tracker algorithm (Marchok 2021) are “late” for the current operational forecast, and thus post-processed or “interpolated” to the current operational forecast time using current TC conditions (Goerss and Sampson 2014). These 6-h late forecasts are assigned an “I” as the last character of the four-character ATCF identifier and are called “early” models as they are the deterministic forecasts

available in time for the current operational forecast. For example, the early HWRF forecast is HWFI, which is a modification of its “late” model identifier (HWRF). One other aspect of the “interpolated” forecasts is that there is an option to gradually eliminate or “phase out” the difference between the initial intensity specified by the forecaster and model over time with linear interpolation. For COAMPS-TC (CTCI), the phase out time is set at 36 h. For HWFI (used in many forecasts discussed below) there is no phase out time. And for HHFI (HWRF again), the phase out time is 18 h so the evaluations discussed herein (at 24 h and beyond) are based on model intensity change from the forecaster initial intensity. NWP model phase out times are periodically adjusted based on postseason analysis or test datasets.

Other models can be run directly on the ATCF as they require less computational resources than the NWP models discussed above. These include the RI, statistical-dynamical, and consensus forecast algorithms. These are also considered “early” models as they are run immediately after the current track position, intensity, central pressure and wind radii are specified and immediately prior to when the JTWC forecast is initiated.

Real-time runs for these forecast algorithms are all available in the JTWC or NHC ATCF “aids” or “adeck” files. Statistical Hurricane Intensity Prediction Scheme (SHIPS; DeMaria et al. 2005) diagnostics files are archived on the ATCF, as are the best tracks. The 2021 and 2022 seasons were cordoned off as independent data, whereas development data vary for the different algorithms. The JTWC version of DTOPS used in this effort, hereinafter called “DTOP” as they are labeled in the ATCF adeck, are reruns using solely ATCF data from JTWC.

### b. Evaluation metrics

Choice of metrics to measure RI performance varies, and there is likely no perfect single metric that suffices for the algorithms discussed below. Peirce scores are chosen over threat scores because they are well suited for rare events (e.g., Ebert and Milne 2022), and RI is a relatively rare but operationally significant event. An example highlighting issues using data from our independent sample evaluated in Fig. 1 and Table 1 follows. Using standard definitions for contingency table entries (hits, false alarms, misses, and correct negatives, represented by  $a$ ,  $b$ ,  $c$ , and  $d$ , respectively), the commonly used threat score =  $a/(a + b + c)$  and rare-event Peirce score =  $(ad - bc)/[(a + c)(b + d)]$  for the algorithm RICN (discussed below), and the RI threshold of 30 kt (24 h) $^{-1}$  yields  $a = 73$ ,  $b = 114$ ,  $c = 54$ ,  $d = 1399$ , threat score = 0.3, and Peirce score = 0.5. If we set  $d = 200$  while keeping the other values in the contingency table constant [this fictitious case is for an event that would not be considered rare since the number of RI events ( $a + c$ ) is 127 and the number of nonevents ( $b + d$ ) is 314], we get  $a = 73$ ,  $b = 114$ ,  $c = 54$ ,  $d = 200$ , threat score = 0.3, and Peirce score = 0.21. So while the threat score remains constant for different values of  $d$  the Peirce score does not and is dependent on the entire contingency table.

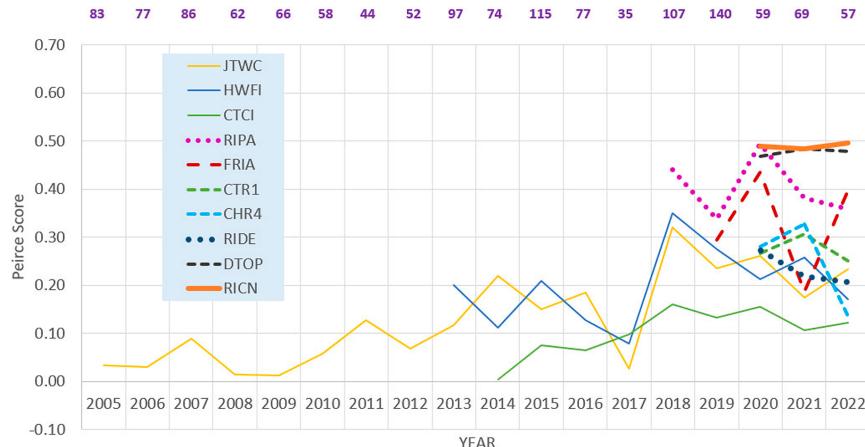


FIG. 1. Peirce score for RI30 through time for JTWC, NWP models (COAMPS-TC and HWRF), deterministic early models CTCI and HWFI, the deterministic RI forecasts (RIPA, FRIA, CTR1, CHR4, RIDE, and DTOP), and an RI consensus or average of two or more deterministic RI forecasts (RICN). The number of actual RI cases for each year is listed at the top of the graph.

Peirce score is the primary metric used in this work, but other metrics are considered. These include hit rate, false alarm rate, mean absolute errors, mean bias, and availability. Many of these metrics can be optimized by degrading others. for example, it is relatively easy to increase hit rate by sacrificing availability. But reducing availability likely decreases the number of forecasts prior to an RI event. These precursor forecasts do not verify as hits but are likely useful in operations. The authors present these metrics but acknowledge that there is no ideal single metric. The one metric that comes closest to being ideal is Peirce Score (Ebert and Milne 2022).

#### c. Deterministic RI guidance

RIPA is the first RI algorithm developed specifically for JTWC's AOR. Based on the earlier work of Kaplan et al. (2015) and prior efforts for the Atlantic and eastern North Pacific basins, RIPA uses current TC information (e.g., intensity and intensity change), atmospheric diagnostic information

along the TC track from either the Navy Global Environmental Model (NAVGEN; Hogan et al. 2014) or NOAA Global Forecast System (GFS 2021) computed from SHIPS, infrared temperatures in the vicinity of the TC from geostationary satellites such as *Himawari-8* and *Himawari-9*, and ocean heat content (Sampson et al. 2022) from the Naval Coupled Ocean Data Assimilation System (NCODA; Cummings 2005). Two statistical methods, linear discriminant analysis and logistic regression, are combined to create probabilistic forecasts for seven intensification thresholds including 25-, 30-, 35-, and 40-kt changes in 24 h; 45- and 55-kt changes in 36 h; and 55- and 70-kt changes in 48 h (RI25, RI30, RI40, RI45, RI55, RI56, and RI70, respectively). These two sets of forecast probabilities are then averaged, and that average is used to prescribe deterministic forecasts. The deterministic forecasts are equal to the intensification thresholds (RI25, RI30, RI35, RI40, RI45, RI55, RI56, and RI70) once the average probability reaches 40%. If two or more thresholds are reached, the highest intensification rate available is prescribed to the deterministic forecast. The terms “trigger” and “triggered” refer to cases when RI probabilities reach prescribed thresholds, and a deterministic forecast is generated. RIPA was introduced in 2018 and is among the best performers at predicting RI30. RI30 is highlighted throughout this work as it is one of the most discussed metrics in RI. RI30 also has an extra benefit in that it provides the most cases for evaluation of relatively rare RI events (Fig. 1). Even so, the number of cases even for RI30, the most common of our RI rates, is small and the development and analysis of performance within this effort is done with this in mind.

FRIA (Slocum 2021) makes use of the same TC, atmospheric diagnostic information, IR temperatures, and ocean heat content as RIPA but avoids use of intensity maximums used in RIPA. FRIA also applies random forest classification instead of the two classification methods used in RIPA. FRIA employs 100 trees per forest, which means that it generates an ensemble of 100 yes/no forecasts to compute its probabilities. Those probabilities are then applied as in RIPA, using 40%

TABLE 1. Hit rate and false alarm rate (separated by a slash) for the 2021 and 2022 western North Pacific, northern Indian Ocean, and Southern Hemisphere seasons. RI30 is 30 kt (24 h)<sup>-1</sup>, RI45 is 45 kt (36 h)<sup>-1</sup>, and RI56 is 55 kt (48 h)<sup>-1</sup>. The number of observed RI cases is in parentheses. Here, ID indicates identifier (e.g., CHR4).

ATCF ID	RI30 (127)	RI45 (47)	RI56 (42)
JTWC	22.8/50	19.1/55	23.8/50
CHR4	27.6/59.8	27.7/66.7	11.9/79.2
CTR1	37/71.3	36.2/71.2	23.8/71.4
HHFI	27.6/62	23.8/70.3	21.4/66.7
FRIA	33.1/63.8	42.6/61.5	38.1/61.9
RIPA	44.9/63.5	53.2/68.8	50/57.1
RIDE	22.8/46.3	23.4/50	16.7/53.3
DTOP	54.3/57.9	48.9/47.7	38.1/36
RICN	57.5/61	46.8/63.9	38.1/33.3
ICNC	5.6/36.4	5.6/25	4.3/33.3
ICNE	18.1/46.5	19.1/59.1	14.3/62.5

to trigger deterministic forecasts of RI. FRIA has been in operations at JTWC since 2021 with respectable Peirce scores for RI30 (Fig. 1), and it retains some independence from RIPA with routine differences of approximately 20% in RI probabilities for individual cases.

DTOPS (DeMaria et al. 2021) was developed for the National Hurricane Center (NHC) basins starting about 2015. NHC DTOPS applies binomial logistic regression to deterministic model forecasts, along with basic vortex and geographic parameters, to produce a probabilistic forecast of RI. NHC DTOPS uses AVNI ("early" GFS forecast), HWFI (the early HWRF forecast), LGEM (Logistic Growth Equation Model; DeMaria 2009), and SHIPS intensity forecasts, EMXI (the early European Center for Medium-Range Weather Forecasts global model) central pressure forecasts, initial TC intensity, latitude, and statistically combined values of these parameters. The version developed for JTWC basins (DTOP; Fig. 1) is similar, using AVNI, DSHA and LGEA (versions of SHIPS and LGEM that employ GFS tracks and atmospheric environmental parameters), HWFI, and EMXI wind speed intensities. The EMXI intensities are converted to TC central pressure using regression from Knaff and Zehr (2007) as central pressure is not reported in EMXI for JTWC basins in the ATCF files. DTOP, like many other RI forecast algorithms, was developed to provide probabilities. For a deterministic algorithm, the authors prescribed a 40% threshold like other algorithms used in this work. Performance on independent data from 2021 and 2022 seasons indicates that it is a top performer at predicting RI30 (Fig. 1).

RIDE (Knaff et al. 2023) uses intensity forecast output from seven deterministic models that were routinely available (and skillful) from 2018 to 2020: one statistical model—the Trajectory Climatology and Persistence Model (TCLP; DeMaria 2009; DeMaria et al. 2021), two early global model forecasts (AVNI and NVGI, the early model forecast from NAVGEM), two early mesoscale NWP model forecasts (CTCI and HWFI), and the two statistical-dynamical model forecasts (DSHA and LGEA). RIDE generates probabilities and deterministic RI forecasts with 40% thresholds as discussed in Knaff et al. (2023). As with DTOP, all of the guidance is found consistently in the ATCF at JTWC as RIDE is dependent on most being available for every operational forecast. Peirce scores for RI30 are not as high for RIDE as they are for RIPA and DTOP, as RIDE suffers from conservatively forecasting RI (see number of cases and mean bias in Fig. 2). However, this also indicates that RIDE has fewer false alarms and likely independence from other RI algorithms. Independence is a key attribute for improved performance in forming a consensus (Sampson et al. 2008).

CTR1 is a deterministic RI algorithm developed from the COAMPS-TC Ensemble. Development of CTR1 was done prior to the 2021 season and is described in appendix B. An effort was made to avoid fine tuning as the development dataset is small and the ensemble used as input to this algorithm changes each year as the COAMPS-TC and its ensemble evolve. Processing reruns prior to 2021 season indicated that using an equally weighted average of one or more forecasts exhibiting RI30 (only select and average forecasts that exhibit

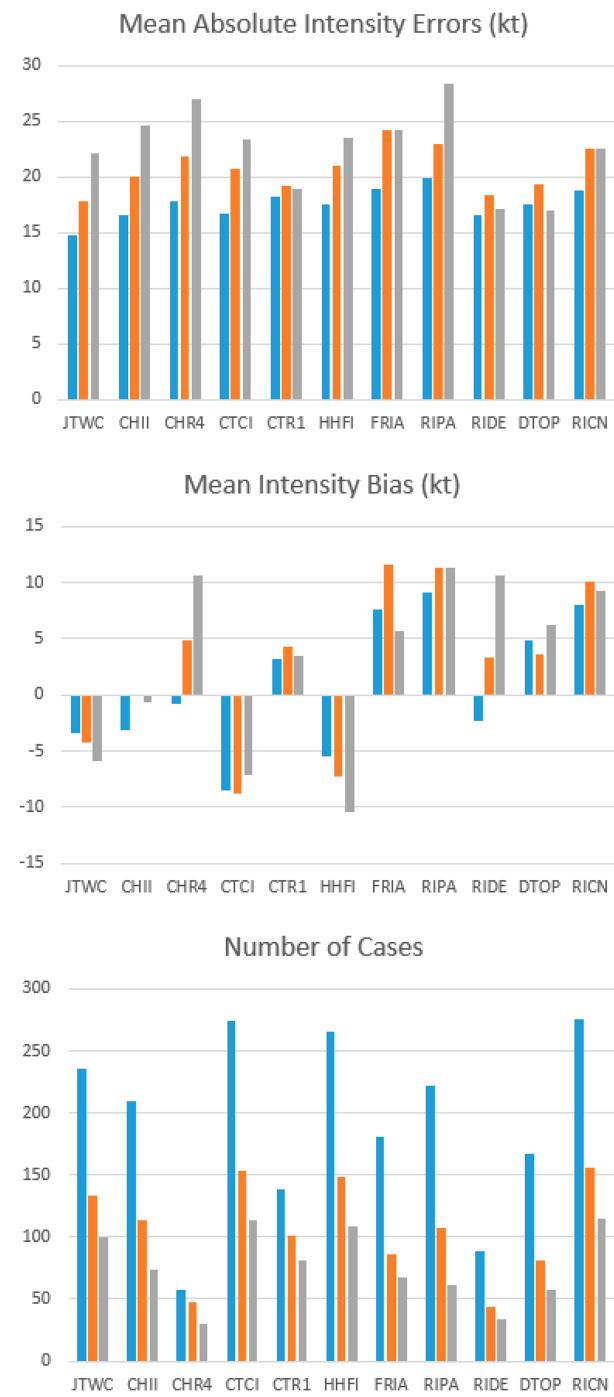


FIG. 2. (top) Mean absolute error, (middle) mean bias, and (bottom) number of cases for JTWC, CHIPS, the COAMPS-TC and HWRF deterministic early models (CHII, CTCI, and HHFI, respectively), the deterministic RI forecasts (RIPA, FRIA, CTR1, CHR4, RIDE, and DTOP), and an RI consensus that is the average of two or more deterministic RI forecasts (RICN) for independent data from JTWC for the 2021 and 2022 seasons. Blue, orange, and gray indicate forecasts of RI30, RI45, and RI56 events, respectively. Cases are limited to head-to-head comparisons with RICN.

RI of 30 kt per 24 h) in the deterministic algorithm gave skillful results, so this was implemented for the 2021 and 2022 seasons. Since then, the COAMPS-TC Ensemble has been upgraded, and processing reruns from 2021 indicate that a consensus of two or more forecasts now yields performance comparable to that of using just one. Starting in 2023, the minimum number of ensemble members required to surpass RI30 to trigger this deterministic forecast is two. Peirce scores for CTR1 (Fig. 1) are reasonably high and marginally higher than those of the COAMPS-TC deterministic early model (CTCI).

CHR4 is a deterministic RI algorithm developed from the CHIPS Ensemble. As with CTR1, development for CHR4 was done prior to the 2021 Northern Hemisphere summer season and is described in [appendix C](#). The number of ensemble members to provide an optimal deterministic RI forecast using Peirce scores and biases was not readily apparent, but it was clear that using an average of ensemble members exceeding RI thresholds provided forecasts that were extremely high biased. The authors elected to assign RI rates to account for high biases, and to therefore select a minimum of four ensemble members exceeding thresholds to trigger forecasts (e.g., a deterministic RI forecast of 30 kt in 24 h is provided when four or more ensemble members exceed 30 kt in 24 h). As with CTR1, tuning was avoided. Still, with little effort and little tuning, the deterministic RI forecast is shown to be skillful as demonstrated by Peirce scores (Fig. 1).

### 3. Consensus results

There are many ways to form a deterministic consensus using deterministic RI forecasts and we explore two simple constructs in this work: 1) form a consensus (an unweighted average of two or more forecasts) from only the deterministic RI forecasts and 2) form a consensus of the intensity forecasts that are routinely available (e.g., [Sampson et al. 2008](#)) including the deterministic RI forecasts when they are available.

#### a. RICN

The consensus of deterministic RI forecasts discussed above we name the RI Consensus (RICN):

$$\text{RICN} = (\text{RIPA} + \text{FRIA} + \text{CTR1} + \text{CHR4} + \text{RIDE} + \text{DTOP})/N,$$

where  $N$  represents the number of intensity forecasts available at the given forecast time, up to the maximum of six when all forecasts are present. RICN performance is shown in [Figs. 1](#) and [2](#) along with its individual members. Peirce scores for this limited independent dataset indicate that RICN is among the top performers at RI30, and retains high availability that is desirable for operational forecasting (Fig. 2). [Figure 2](#) also shows much of the RI guidance has large (15 kt and greater) mean absolute errors for these cases as compared with errors for the entire dataset that are about 5 kt lower. Biases for the deterministic RI forecasts are generally positive, reflecting the high number of false alarms for predicted RI events. The NWP models and JTWC also have elevated mean absolute errors for these difficult potential RI cases, but negative biases that indicate underforecasting of RI events.

Biases are a concern in general, but positive bias for deterministic RI forecasts are an artifact of the high false alarm rates (e.g., predicting RI30 when no RI occurs). Note that RI algorithms with higher hit rate in [Table 1](#) (FRIA, RIPA, DTOP, and RICN) also have high Peirce scores (Fig. 1). RIDE has small bias, relatively low false alarm rate, but also lower hit rate. Operational forecasters at the JTWC excel in terms of mean absolute error and bias, but have lower Peirce scores than the top deterministic RI algorithms.

#### b. ICNE

The second method discussed in this work is intended to serve as a consistently available consensus (ICNE). This is unlike RICN, which only runs when two or more deterministic RI forecasts are available. ICNE contains the same routinely available skillful forecasts as the consensus without deterministic RI forecasts (ICNC), but replaces and adds deterministic RI forecasts when those are available. The intent of doing so is to raise intensity consensus forecasts, which are low (negatively) biased during RI events, toward more realistic RI rates. For example, the two deterministic RI forecasts based on SHIPS (RIPA and FRIA) replace the two SHIPS forecasts (DSHA and DSHN, which is like DSHA but relies on NAVGEM instead of GFS) when the deterministic RI forecasts are triggered. Likewise, the deterministic RI forecast based on the COAMPS-TC Ensemble (CTR1) replaces CTCI when triggered. For the other deterministic RI forecasts there is no direct replacement, so those forecasts are just added to the consensus forecast. Consensus forecasts are computed when two or more member forecasts exist, and the two specified above can be summarized as follows:

$$\begin{aligned} \text{ICNE} = & [(\text{RIPA} \text{ or } \text{DSHA}) + (\text{FRIA} \text{ or } \text{DSHN}) \\ & + (\text{CTR1} \text{ or } \text{CTCI}) + \text{AVNI} + \text{HHFI} \\ & + \text{RIDE} + \text{CHR4} + \text{DTOP}]/N \quad \text{and} \end{aligned}$$

$$\text{ICNC} = (\text{DSHA} + \text{DSHN} + \text{CTCI} + \text{AVNI} + \text{HHFI})/N,$$

where  $N$  represents the number of intensity forecasts available at the given forecast time, up to the maximum of eight for ICNE and five for ICNC when all forecasts are present for the synoptic time (0000, 0600, 1200, 1800 UTC), and the “or” designates that the forecast right of “or” is used when the forecast left of “or” is unavailable. ICNC can serve as a baseline for ICNE in our work since it has no deterministic RI forecasts.

The consensus with deterministic RI forecasts (ICNE) and without (ICNC) were recomputed for the entire 2021 and 2022 JTWC seasons using real-time processing and results indicate overall similar performance, which is expected as the RI cases are only a small percentage of the entire set of forecasts. However, addition of the deterministic RI forecasts does increase the hit rates and false alarm rates for RI30, RI45, and RI56 ([Table 1](#)) and the Peirce scores (not shown) from approximately 0.05 to 0.15. The ICNE Peirce scores are also similar to the JTWC Peirce scores.

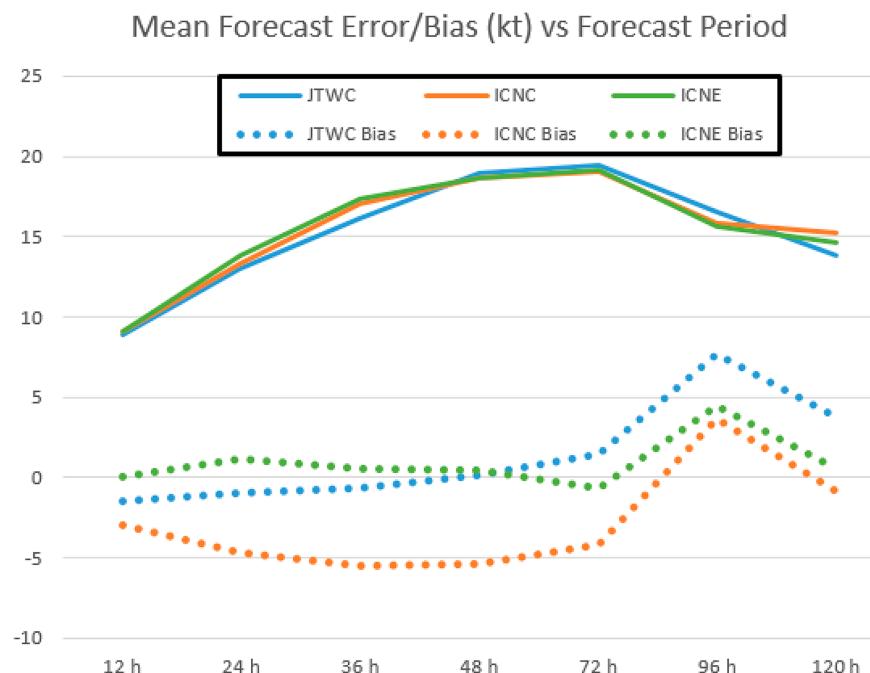


FIG. 3. Mean absolute forecast error (kt) and bias (kt) for consensus and JTWC forecasts when at least one deterministic RI forecast is available. Data are from the 2021 and 2022 western North Pacific, northern Indian Ocean, and Southern Hemisphere seasons. Numbers of cases are 454, 448, 319, 269, 183, 96, and 68 for 12, 24, 36, 48, 72, 96, and 120 h, respectively.

Evaluation of differences in consensus forecasts when one or more of the deterministic RI forecasts is available (i.e., when the ICNC and ICNE differ since now at least one deterministic RI forecast is added to the ICNE consensus) is shown in Fig. 3. The mean errors of the intensity consensus forecasts and JTWC forecasts are similar for these cases, but the biases for the consensus with deterministic RI forecasts (ICNE) are significantly closer to zero (one-tailed *t*-test; Wilks 2006) at 24, 36, 48, and 72 h than biases for the consensus without (ICNC). Since the long-term goal is to reduce errors and biases in RI forecasts to near those of seasonal averages, the small improvements seen here are only a step in the right direction with plenty of room for improvement.

#### 4. Summary and discussion of future work

This work describes and evaluates six deterministic RI forecast algorithms specifically designed for use in operational forecasting, all of which are operationally run on the ATCF at JTWC. The six deterministic RI forecasts are disparate in their construction and performance, yet all provided some skill as measured by Peirce scores in independent evaluation of the 2021 and 2022 JTWC seasons. Two different methods for combining these new deterministic forecasts into a consensus were then developed and evaluated. Neither is optimal for forecasters, but each has benefit over what is currently available. The RI CoNsensus (RICN) is among the top performing guidance for RI, and has reasonably high availability. It suffers from false alarms and related positive intensity forecast

bias. The intensity consensus that includes six deterministic RI forecasts (ICNE) provides performance similar to that of the intensity consensus without (ICNC) in terms of mean errors, but provides forecasts with approximately 5 kt less negative bias. ICNE performance measured by Peirce scores is below that of top-performing deterministic RI forecasts (e.g., DTOP, RIPA, FRIA) and the deterministic RI consensus (RICN), so operational forecast challenges remain (e.g., an operational forecaster cannot blindly select one of these forecasts to use in all cases).

On the bright side, development of RI forecast guidance will continue, and the methodology employed here requires little effort to adjust to updated guidance. The authors refrained from tuning deterministic RI forecast guidance and consensus to the dataset (e.g., defining the RI rates to be the RI rates found in the best tracks) in an effort to limit effects of model upgrades and changes on performance. This approach to addressing model upgrades was tested earlier this year when the HWRF was replaced by the Hurricane Forecast Analysis System (HAFS; Alaka et al. 2022). As of this writing, the RI guidance is still performing well with HAFS substituted for HWRF. A further benefit is that forecasters will be better able to understand and use algorithms developed.

The authors are also optimistic about the future of RI forecasting. There are many statistical efforts to develop better performing RI forecast guidance, employing machine learning, higher resolution shear products, satellite data, and more. NWP model ensembles continue to improve (e.g., Komaromi et al. 2021), and new deterministic NWP models such as HAFS

will continue to have direct impact on our ability to predict RI. There is still a requirement for more independent and skillful RI forecast methods, and as these become available and existing methods progress, the consensus methods described in this work will improve. For example, with many skillful RI forecasts in the future, the thresholds to trigger RI could be raised to minimize false alarms. Although raising the thresholds would degrade Peirce scores for individual deterministic RI forecast algorithms, consensus Peirce scores may improve.

Detection of RI events should also improve. New TC-specific wind speed algorithms such as those based on Synthetic Aperture Radar (SAR; [Mouche et al. 2019](#); [Jackson et al. 2021](#)), Soil Moisture Active Passive (SMAP; [Meissner et al. 2021](#)), Soil Moisture Operational Sensor (SMOS; [Reul et al. 2017](#)), and Advanced Microwave Scanning Radiometer (AMSR; e.g., [Meissner et al. 2021](#); [Alsweiss et al. 2023](#)) have enhanced JTWC capabilities to estimate high winds near the TC core ([Howell et al. 2022](#)). The Stepped Frequency Microwave Radiometer (SFMR; [Sapp et al. 2019](#)) and aircraft observations commonly used in the Atlantic do not directly impact the JTWC basins, but are absolutely essential for ground truth matchups of the remotely sensed data. As these Atlantic Ocean observation platforms improve, so will the remotely sensed algorithms. Also, the future will bring more advanced sensing capabilities such as those discussed in [Knaff et al. \(2021\)](#) and [Hauser et al. \(2023\)](#), enabling more frequent and hopefully more accurate detection of inner core winds critical for development, forecasting, and evaluation of RI forecast guidance.

**Acknowledgments.** The authors acknowledge funding provided by the Office of Naval Research under ONR Award N0001420WX00517 to NRL Monterey and a grant to CIRA (N00173-21-1-G008). We also thank all of the people at NRL, NHC, and JTWC who keep the ATCF running in operations, and we thank two anonymous reviewers. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of NOAA or the Department of Commerce. ATCF and COAMPS are registered trademarks of the Naval Research Laboratory.

**Data availability statement.** The best-track data used in this study are freely available from the Joint Typhoon Warning Center (<https://www.metoc.navy.mil/jtvc/jtvc.html?best-tracks>). Forecasts and SHIPS diagnostic files used are in the Joint Typhoon Warning Center's forecast database (a-deck) and are available online upon request (<https://pzal.metoc.navy.mil/php/rds/login.php>).

## APPENDIX A

### List of Acronyms

$a, b, c, d$	Hits, misses, false alarms, and correct negatives, respectively	Peirce score	$= (ad - bc)/[(a + c)(b + d)]$
AMSR	Advanced Microwave Scanning Radiometer	RI	Rapid intensification
AOR	Area of Responsibility	RI25, RI30, RI35, RI40	25-, 30-, 35-, and 40-kt change in 24 h, respectively
ATCF	Automated Tropical Cyclone Forecast System	RI45, RI55	45- and 55-kt change in 36 h, respectively
		RI56, RI70	55- and 70-kt change in 48 h, respectively
		RICN	RI Consensus
		RIDE	Rapid Intensification Deterministic Ensemble
		RIPA	Rapid Intensification Prediction Aid
		SAR	Synthetic Aperture Radar
		SFMR	Stepped Frequency Microwave Radiometer
		SHIPS	Statistical Hurricane Intensity Prediction Scheme
		SMAP	Soil Moisture Active/Passive
		SMOS	Soil Moisture Operational Sensor
		TC	Tropical cyclone
		Threat score	$= a/(a + b + c)$

AVNI	GFS forecast track and intensity, early model
CHII	CHIPS early model (interpolated)
CHIPS	Coupled Hurricane Intensity Prediction System
CHR4	Deterministic RI forecast based on CHIPS Ensemble
COAMPS-TC	Coupled Ocean–Atmosphere Mesoscale Prediction System for Tropical Cyclones
CTCI	COAMPS-TC forecast track and intensity, early model
CTCX	COAMPS-TC forecast track and intensity, late model
CTR1	Deterministic RI forecast based on COAMPS-TC Ensemble
DSHA	SHIPS intensity model, GFS model fields
DSHN	SHIPS intensity model, NAVGEM model fields
DTOP	JTWC version of DTOPS
DTOPS	Deterministic To Probabilistic Statistical Model
EMXI	ECMWF forecast track and intensity, early model
FRIA	Forest-based Rapid Intensification Aid
GFS	Global Forecast System (National Weather Service)
HHFI	Like HWFI, but without adjustment for current intensity
HWFI	HWRF forecast track and intensity, early model
HWRF	HWRF forecast track and intensity, late model
ICNC	Intensity forecast consensus with no RI forecasts
ICNE	Intensity forecast consensus including deterministic RI forecasts
JTWC	Joint Typhoon Warning Center
LGEA	LGEM using GFS model fields
LGEM	Logistic Growth Equation Model
NAVGEM	Navy Global Environment Model
NCODA	Navy Coupled Ocean Data Assimilation System
Peirce score	$= (ad - bc)/[(a + c)(b + d)]$
RI	Rapid intensification
RI25, RI30, RI35, RI40	25-, 30-, 35-, and 40-kt change in 24 h, respectively
RI45, RI55	45- and 55-kt change in 36 h, respectively
RI56, RI70	55- and 70-kt change in 48 h, respectively
RICN	RI Consensus
RIDE	Rapid Intensification Deterministic Ensemble
RIPA	Rapid Intensification Prediction Aid
SAR	Synthetic Aperture Radar
SFMR	Stepped Frequency Microwave Radiometer
SHIPS	Statistical Hurricane Intensity Prediction Scheme
SMAP	Soil Moisture Active/Passive
SMOS	Soil Moisture Operational Sensor
TC	Tropical cyclone
Threat score	$= a/(a + b + c)$

## APPENDIX B

**Development of COAMPS-TC Ensemble Deterministic RI Forecast**

The COAMPS-TC Ensemble (Komaromi et al. 2021) is an 11-member ensemble developed by the Naval Research Laboratory (NRL) to produce probabilistic forecasts of tropical cyclone (TC) track, intensity, and structure. Members run with a storm-following inner grid at 4-km horizontal resolution. A total of 10 ensemble members are perturbed (one is the control) through initial and boundary conditions, the initial vortex, and model physics to account for a variety of sources of uncertainty that affect track and intensity forecasts. The ensemble intensity forecast spread is correlated with intensity forecast

error, but ensemble intensity forecast spread underestimates uncertainty.

One method to turn an ensemble into RI guidance is to calibrate number of members in the ensemble to RI prediction. In a well-calibrated probabilistic system, we would expect probabilities near 40% (see Sampson et al. 2011) to be appropriate to trigger a deterministic RI forecast, but the early model COAMPS-TC (CTCI) is conservative about predicting RI and so is its ensemble. Experiments using anywhere from a minimum of one to five ensemble members to construct the deterministic RI forecast (an average of the ensemble members exhibiting RI) show that a minimum of just one yields reasonably high Peirce scores (Fig. B1).

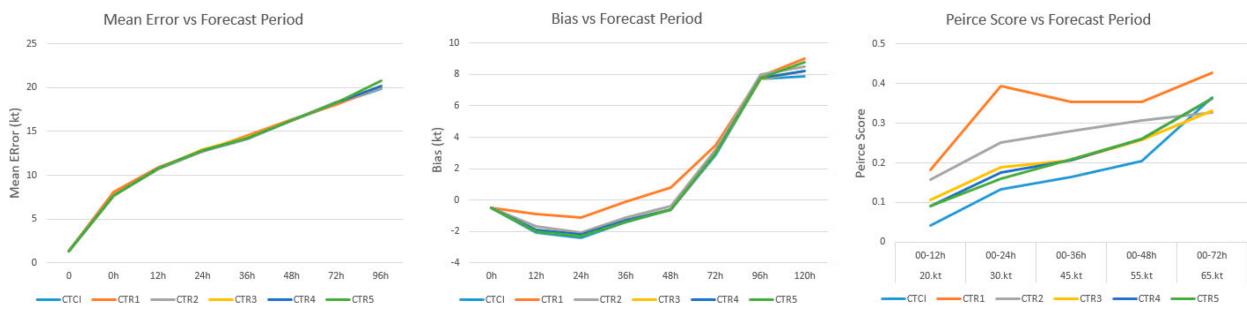


FIG. B1. (left) Mean absolute error, (center) bias, and (right) Peirce score for different ensemble member trials. CTCI is the deterministic COAMPS-TC forecast, and CTR1 is an average of one or more ensemble members that achieve RI30. CTR2, CTR3, CTR4, and CTR5 are averages of ensemble members that achieve RI30 with minimum numbers of 2, 3, 4, and 5, respectively. Development data are from the entire 2020 season and 2021 Southern Hemisphere. Peirce scores are for the homogeneous set.

## APPENDIX C

**Development of CHIPS Ensemble Deterministic RI Forecast**

CHIPS (Emanuel et al. 2004) has been generating real-time intensity forecasts since the mid-2000s. An excellent summary of CHIPS (Emanuel 2023) describes the CHIPS Ensemble as a seven-member ensemble:

- 1) Ensemble member 1 (control) uses unperturbed official JTWC track and GFS model fields.
- 2) In ensemble member 2, the initial intensity is enhanced by  $3 \text{ m s}^{-1}$  (during the initialization, the intensity increment is slowly ramped up over the previous 24 h.)
- 3) In ensemble member 3, the initial intensity is weakened by  $3 \text{ m s}^{-1}$  (during the initialization, the intensity increment is slowly ramped down over the previous 24 h.)
- 4) In ensemble member 4, the initial intensity is as reported but the intensity 12 h before is enhanced by  $1.5 \text{ m s}^{-1}$  so as to produce a negative intensification anomaly at the initial time.
- 5) Ensemble member 5 is the same as ensemble member 4 except that the initial intensification rate is enhanced rather than diminished.
- 6) In ensemble member 6, the initial intensity is enhanced as in ensemble member 2 and the environmental wind shear

is set to zero at all forecast times. This is intended to give an upper bound on forecast intensity.

- 7) In ensemble member 7, the initial intensity is diminished as in ensemble member 3 and wind shear is enhanced by  $10 \text{ m s}^{-1}$ . This is intended to give a plausible lower bound on forecast intensity.

Unlike the COAMPS-TC Ensemble described in appendix B, an average of CHIPS Ensemble member forecasts exhibiting RI is extremely high biased (not shown). Given this bias issue, the deterministic RI forecast algorithm was developed similar to what was done with RIPA. So a certain number of CHIPS Ensemble members exhibiting RI would trigger a deterministic RI forecast at that RI rate. For example, four CHIPS Ensemble members exceeding  $30 \text{ kt}$  ( $24 \text{ h}$ ) $^{-1}$  might trigger a deterministic RI30 forecast (i.e., a forecast of  $30 \text{ kt}$  in  $24 \text{ h}$ ).

To select the number of members to trigger the deterministic RI forecast, a sensitivity analysis was performed on the development set. Figure C1 shows sensitivity of the RI30 forecast to the number of ensemble members exceeding  $30 \text{ kt}$  ( $24 \text{ h}$ ) $^{-1}$ . There is no obvious choice for number of members. Using five members to trigger the deterministic RI forecast yields less high (positive) bias but low Peirce scores. With no obvious choice, the authors selected the algorithm with four ensemble members exceeding the RI threshold (50%; CHR4 in Fig. C1) as a compromise

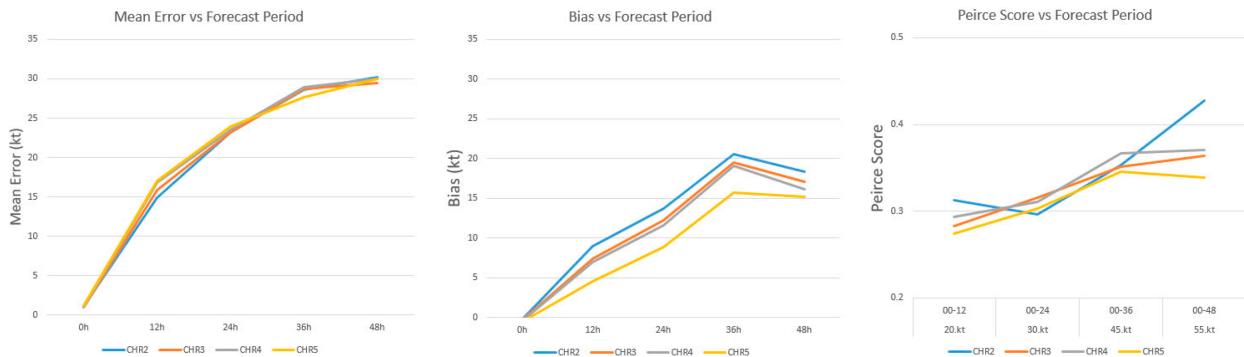


FIG. C1. (left) Mean absolute error, (center) bias, and (right) Peirce score for different ensemble member trials. CHR2, CHR3, CHR4, and CHR5 are prescribed RI30, RI45, and RI56 intensification rates with minimum numbers of 2, 3, 4, and 5, respectively. Development data are from 2020 and 2021 Southern Hemisphere. Peirce scores are for the homogeneous set.

between biases, Peirce scores, and number of deterministic RI forecast cases available (not shown).

#### REFERENCES

Alaka, G. J., Jr., X. Zhang, and S. G. Gopalakrishnan, 2022: High-definition hurricanes: Improving forecasts with storm-following nests. *Bull. Amer. Meteor. Soc.*, **103**, E680–E703, <https://doi.org/10.1175/BAMS-D-20-0134.1>.

Alsweiss, S., Z. Jelenak, and P. S. Chang, 2023: Extending the usability of radiometer ocean surface wind measurements to all-weather conditions for NOAA operations: Application to AMSR2. *IEEE Trans. Geosci. Remote Sens.*, **61**, 1–12, <https://doi.org/10.1109/TGRS.2023.3266772>.

Biswas, M. K., and Coauthors, 2018: Hurricane Weather Research and Forecasting (HWRF) Model: 2018 scientific documentation. Developmental Testbed Center Doc., 112 pp., [https://dtcenter.org/sites/default/files/community-code/hwrf/docs/scientific\\_documents/HWRFv4.0a\\_ScientificDoc.pdf](https://dtcenter.org/sites/default/files/community-code/hwrf/docs/scientific_documents/HWRFv4.0a_ScientificDoc.pdf).

Cummings, J. A., 2005: Operational multivariate ocean data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3583–3604, <https://doi.org/10.1256/qj.05.105>.

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

—, 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>.

—, J. L. Franklin, M. J. Onderlinde, and J. Kaplan, 2021: Operational forecasting of tropical cyclone rapid intensification at the National Hurricane Center. *Atmosphere*, **12**, 683, <https://doi.org/10.3390/atmos12060683>.

Doyle, J. D., and Coauthors, 2014: Tropical cyclone prediction using COAMPS-TC. *Oceanography*, **27**, 104–115, <https://doi.org/10.5670/oceanog.2014.72>.

Ebert, P. A., and P. Milne, 2022: Methodological and conceptual challenges in rare and severe event forecast verification. *Nat. Hazards Earth Syst. Sci.*, **22**, 539–557, <https://doi.org/10.5194/nhess-22-539-2022>.

Emanuel, K., 2023: The Coupled Hurricane Intensity Prediction System (CHIPs). MIT, 5 pp., <https://wind.mit.edu/~emanuel/CHIPs.pdf>.

—, 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).

Francis, A. S., and B. R. Strahl, 2022: Joint Typhoon Warning Center annual tropical cyclone report 2020. JTWC Rep., 145 pp., <https://www.metoc.navy.mil/jtwc/products/atcr/2020atcr.pdf>.

GFS, 2021: GFS Global Forecast System. GFS, accessed 4 January 2022, [https://www.emc.ncep.noaa.gov/emc/pages/numerical\\_forecast\\_systems/gfs.php](https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs.php).

Goerss, J. S., and C. R. Sampson, 2014: Prediction of consensus tropical cyclone intensity forecast error. *Wea. Forecasting*, **29**, 750–762, <https://doi.org/10.1175/WAF-D-13-00058.1>.

Hauser, D., and Coauthors, 2023: Satellite remote sensing of surface wind, waves, and currents: Where are we now? *Surv. Geophys.*, **44**, 1357–1446, <https://doi.org/10.1007/s10712-023-09771-2>.

Hogan, T. F., and Coauthors, 2014: The Navy global environmental model. *Oceanography*, **27**, 116–125, <https://doi.org/10.5670/oceanog.2014.73>.

Howell, B., S. Egan, and C. Fine, 2022: Application of microwave Space-Based Environmental Monitoring (SBEM) data for operational tropical cyclone intensity estimation at the Joint Typhoon Warning Center. *Bull. Amer. Meteor. Soc.*, **103**, E2315–E2322, <https://doi.org/10.1175/BAMS-D-21-0180.1>.

Jackson, C. R., T. W. Ruff, J. A. Knaff, A. Mouche, and C. R. Sampson, 2021: Chasing cyclones from space. *Eos*, **102**, <https://doi.org/10.1029/2021EO159148>.

JTWC, 2021: JTWC 2020 operational highlights, challenges, and future changes. *75th Interdepartmental Hurricane Conf.*, Miami, FL, JTWC, 16 pp., [https://www.icams-portal.gov/meetings/TCORF/TCORF21/01-Session/s1-03\\_jtwc.pdf](https://www.icams-portal.gov/meetings/TCORF/TCORF21/01-Session/s1-03_jtwc.pdf).

Kaplan, J., 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 R. M. Zehr, 2007: Reexamination of tropical cyclone wind–pressure relationships. *Wea. Forecasting*, **22**, 71–88, <https://doi.org/10.1175/WAF965.1>.

—, C. R. Sampson, 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>.

—, —, and B. R. Strahl, 2020: A tropical cyclone rapid intensification prediction aid for the Joint Typhoon Warning Center's

areas of responsibility. *Wea. Forecasting*, **35**, 1173–1185, <https://doi.org/10.1175/WAF-D-19-0228.1>.

—, and Coauthors, 2021: Estimating tropical cyclone surface winds: Current status, emerging technologies, historical evolution, and a look to the future. *Trop. Cyclone Res. Rev.*, **10**, 125–150, <https://doi.org/10.1016/j.tcr.2021.09.002>.

—, C. R. Sampson, A. Brammer, and C. J. Slocum, 2023: A Rapid Intensification Deterministic Ensemble (RIDE) for the Joint Typhoon Warning Center's area of responsibility. *Wea. Forecasting*, **38**, 1229–1238, <https://doi.org/10.1175/WAF-D-23-0012.1>.

Komaromi, W. A., P. A. Reinecke, J. D. Doyle, and J. R. Moskaitis, 2021: The Naval Research Laboratory's Coupled Ocean-Atmosphere Mesoscale Prediction System-Tropical Cyclone Ensemble (COAMPS-TC Ensemble). *Wea. Forecasting*, **36**, 499–517, <https://doi.org/10.1175/WAF-D-20-0038.1>.

Manzano, A., 2007: A note on the maximum Peirce skill score. *Wea. Forecasting*, **22**, 1148–1154, <https://doi.org/10.1175/WAF1041.1>.

Marchok, T., 2021: Important factors in the tracking of tropical cyclones in operational models. *J. Appl. Meteor. Climatol.*, **60**, 1265–1284, <https://doi.org/10.1175/JAMC-D-20-0175.1>.

Meissner, T., L. Ricciardulli, and A. Manaster, 2021: Tropical cyclone wind speeds from WindSat, AMSR and SMAP: Algorithm development and testing. *Remote Sens.*, **13**, 1641, <https://doi.org/10.3390/rs13091641>.

Mouche, A., B. Chapron, J. Knaff, Y. Zhao, B. Zhang, and C. Combot, 2019: Copolarized and cross- polarized SAR measurements for high-resolution description of major hurricane wind-structures: Application to Irma category-5 hurricane. *J. Geophys. Res. Oceans*, **124**, 3905–3922, <https://doi.org/10.1029/2019JC015056>.

Peirce, C. S., 1884: The numerical measure of the success of predictions. *Science*, **4**, 453–454, <https://doi.org/10.1126/science.ns-4.93.453.b>.

Reul, N., and Coauthors, 2017: A new generation of tropical cyclone size measurements from space. *Bull. Amer. Meteor. Soc.*, **98**, 2367–2385, <https://doi.org/10.1175/BAMS-D-15-00291.1>.

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>.

—, J. Cummings, J. A. Knaff, M. DeMaria, and E. A. Serra, 2022: An upper ocean thermal field metrics dataset. *Meteorology*, **1**, 327–340, <https://doi.org/10.3390/meteorology1030021>.

Sapp, J. W., S. O. Alsweiss, Z. Jelenak, P. S. Chang, and J. Carswell, 2019: Stepped Frequency Microwave Radiometer wind-speed retrieval improvements. *Remote Sens.*, **11**, 214, <https://doi.org/10.3390/rs11030214>.

Slocum, C. J., 2021: What can we learn from random forest in the context of the tropical cyclone rapid intensification problem? *Second NOAA Workshop on Leveraging AI in Environmental Sciences*, College Park, MD, NOAA, 13 pp., [https://www.star.nesdis.noaa.gov/star/documents/meetings/2020AI/presentations/202101/20210128\\_Slocum.pptx](https://www.star.nesdis.noaa.gov/star/documents/meetings/2020AI/presentations/202101/20210128_Slocum.pptx).

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

Zhang, Z., and Coauthors, 2022: Topic 2.3: Intensity change: Operational perspectives. *10th WMO International Workshop on Tropical Cyclones*, Bali, Indonesia, WMO, 2.3, 44 pp., <https://community.wmo.int/en/iwtc-10-reports>.