A Deterministic Rapid Intensification Aid

CHARLES R. SAMPSON

Naval Research Laboratory, Monterey, California

JOHN KAPLAN

NOAA/AOML/Hurricane Research Division, Miami, Florida

JOHN A. KNAFF AND MARK DEMARIA

NOAA/NESDIS, Fort Collins, Colorado

CHRIS A. SISKO

National Hurricane Center, Miami, Florida

ABSTRACT

Rapid intensification (RI) is difficult to forecast, but some progress has been made in developing probabilistic guidance for predicting these events. One such method is the RI index. The RI index is a probabilistic text product available to National Hurricane Center (NHC) forecasters in real time. The RI index gives the probabilities of three intensification rates [25, 30, and 35 kt $(24 \text{ h})^{-1}$; or 12.9, 15.4, and 18.0 m s^{-1} $(24 \text{ h})^{-1}$] for the 24-h period commencing at the initial forecast time. In this study the authors attempt to develop a deterministic intensity forecast aid from the RI index and, then, implement it as part of a consensus intensity forecast (arithmetic mean of several deterministic intensity forecasts used in operations) that has been shown to generally have lower mean forecast errors than any of its members. The RI aid is constructed using the highest available RI index intensification rate available for probabilities at or above a given probability (i.e., a probability threshold). Results indicate that the higher the probability threshold is, the better the RI aid performs. The RI aid appears to outperform the consensus aids at about the 50% probability threshold. The RI aid also improves forecast errors of operational consensus aids starting with a probability threshold of 30% and reduces negative biases in the forecasts. The authors suggest a 40% threshold for producing the RI aid initially. The 40% threshold is available for approximately 8% of all verifying forecasts, produces approximately 4% reduction in mean forecast errors for the intensity consensus aids, and corrects the negative biases by approximately 15%–20%. In operations, the threshold could be moved up to maximize gains in skill (reducing availability) or moved down to maximize availability (reducing gains in skill).

1. Introduction

There are a handful of tropical cyclone intensity (maximum 1-min mean wind at 10-m elevation) forecast objective aids available in real time at the National Hurricane Center (NHC), each with its own strengths and weaknesses. Statistical aids like the 5-Day Statistical Hurricane Intensity Forecast (SHF5; Knaff et al. 2003) have been in operations for approximately two decades.

SHF5 is a climatology and persistence statistical model that is frequently used as a skill baseline for other models. SHF5 is a poor predictor of rapid intensification and rapid decay, but its seasonal mean forecast errors are still quite competitive (DeMaria et al. 2007). For convenience, Table 1 provides a short summary of all the objective aids used in this study.

Dynamical models also produce forecasts of tropical cyclone intensity and have been available for approximately 15 years; however, many are handicapped by resolution constraints, poor initialization, and insufficient parameterizations of the smaller-scale processes (Knaff et al. 2006), and thus cannot adequately simulate the inner

Corresponding author address: Charles R. Sampson, NRL, 7 Grace Hopper Ave., Stop 2, Monterey, CA 93943-5502.
E-mail: sampson@nrlmry.navy.mil

TABLE 1. A list of tropical cyclone intensity forecast aids used in this study. The first column gives the name of the aid, the second column gives the name of the interpolated version of that forecast aid, and the final column gives a description of the numerical or statistical model that is the basis for those forecast aids.

Forecast aid ID	Interpolated version	Basis model
SHF5		5-day Statistical Hurricane Forecast Model
DSHP		Decay SHIPS
GFDL	GFDI	GFDL hurricane model
GFDN	GFNI	Navy version of the GFDL hurricane model
INT2		Two-member consensus (GFDI + DSHP)
HWRF	HWFI	Hurricane WRF model
ICON		Four-member consensus (GFDI + DSHP + LGEM + HWFI)
IVCN		Five-member consensus (GFDI + DSHP + LGEM + HWFI + GFNI)
LGEM		Logistic Growth Equation Model
RAPID		Deterministic RI aid
IVCN+RAPID		IVCN members + RI aid

core of a tropical cyclone where the highest wind (intensity) is usually located. Consequently, the dynamical models showing skill in intensity forecasting are high-resolution schemes designed specifically for tropical cyclone forecasting. The Geophysical Fluid Dynamics Laboratory (GFDL) Hurricane Prediction System (Bender et al. 2007) has been available to NHC forecasters since 1995. A version of GFDL, run with Navy Operational Global Atmospheric Prediction System (NOGAPS; Hogan and Rosmond 1991) initial and boundary conditions (GFDN; Rennick 1999; Bender et al. 2007), was also made available to forecasters in late 1998. More recently, the Hurricane Weather Research and Forecasting Model (HWRF; Bernardet et al. 2010) was made available in operations for the 2007 season.

The Statistical Hurricane Intensity Prediction Scheme (SHIPS; DeMaria et al. 2005; DeMaria et al. 2006; Kaplan and DeMaria 1995, 2001) and the Logistic Regression Equation Model (LGEM; DeMaria 2009) also produce intensity forecasts for the Atlantic and eastern North Pacific. These are statistical–dynamical models that do not resolve the tropical cyclone inner core, but perform surprisingly well using regression equations with large-scale environmental parameters. One of the main weaknesses of SHIPS is that even though it generally performs well when measured by mean absolute error statistics, it is not as skillful in forecasting rapid intensification (RI) events (Knaff et al. 2006; Kaplan et al. 2010). The LGEM outperforms SHIPS for longer-range forecasts (48–120 h), but still has difficulty predicting RI.

There are a number of intensity forecast aids specifically designed to forecast RI currently under development. Of these, only the RI index (Kaplan et al. 2010) is available to NHC forecasters in operations. Instead of providing a quantitative prediction of the intensity change, the original version estimated the probability that the maximum winds will increase by 30 kt [where 1 kt (or nautical miles per hour) = 0.514 m s^{-1}] or more in the following 24 h. More recently, the probabilities of additional intensification thresholds were added (25 and 35 kt). Using the Peirce skill score (probability of detection − false alarm rate), Kaplan et al. (2010) found that their RI index performance exceeded that of the SHIPS and GFDL models, as well as the statistical model SHF5. Optimum probability thresholds were found to be between 20% and 35%, and varied by basin and the forecasted rate of intensification (25, 30, or 35 kt). In this work, the authors construct a deterministic RI aid based upon RI index probability thresholds, evaluate its performance, and explore its inclusion in an operational consensus forecast run at NHC.

2. Data

The data used for this study are taken from the Automated Tropical Cyclone Forecasting System's (ATCF; Sampson and Schrader 2000) operational archives at the NHC. To provide stable statistics, the sample size should be as large as possible. However, the sample was limited to the years from 2006 to 2008 for development, and 2009 for independent evaluation since the authors wanted to employ both recent additions to the suite of operationally available objective aids (LGEM and HWRF) as well as upgraded versions of existing model guidance (GFDL, DSHP, and GFDN; model acronyms are defined in Table 1). All intensities (wind speeds) are reported in units of knots, since that is the unit used in operations.

The RI index deterministic forecasts are computed from the probabilistic text messages contained in the SHIPS text messages. These text messages are provided to NHC forecasters in real time for almost every forecast and are an integral part of the official forecast process. For consistency, the text messages used for this study are reconstructed ex post facto from the latest version of the RI index (Kaplan et al. 2010). The RI index forecasts are also limited to those verifying over water following the methodology used in Kaplan et al. (2010).

3. Methods

Forecast aids are characterized as either early or late, depending on whether or not they are available to the forecaster during the forecast cycle. The GFDL, HWRF, and GFDN forecasts are all late, so their forecasts from

the previous cycle are postprocessed to adjust the times and intensities with the initial time and intensity of the current forecast cycle via an algorithm named the *interpolator* (see Sampson et al. 2008 for further details). The postprocessed GFDL objective aid is called GHMI, the postprocessed HWRF is named HWFI, and the postprocessed GFDN is named GFNI. The DSHP, LGEM, and SHF5 forecasts are all made available for the current forecast cycle and are therefore considered early. This report focuses on verification of early forecast aids and the interpolated late aids since those are the relevant aids for the forecast process.

The deterministic RI aid constructed for this study (hereafter referred to as RAPID) is derived from the RI index probability text product available to the forecasters in operations. It is defined as the maximum rate available for probabilities at or above a given RI probability threshold (0%, 10%, 20%, 30%, 40%, 50%, 60%, or 70%). We chose to verify the various intensity aids at 10% increments from 0%–70% since the number of verifying cases is less than 250 above 30%, and we did not feel that using a smaller analysis increment was warranted at this time. If we choose a prescribed 40% probability threshold, RAPID is assigned a 24-h intensity forecast of +35 kt, when the probability of 35 kt in the RI index is at least 40%. If the 35-kt probability is less than the 40% threshold, RAPID is then assigned an initial 24-h intensity forecast of +30 kt. If the 30-kt probability is less than 40%, RAPID is then assigned a 24-h intensity of +25 kt. Finally, RAPID is undefined if none of the RI index probabilities reach the prescribed thresholds. At the 12-h forecast period, the intensity is defined as the average of the initial synoptic time intensity and 24-h forecasted intensity. The 24-h RAPID value is also used for forecasts through 72 h for evaluation, even though the RI index is not specifically designed to predict RI beyond 24 h. To accomplish this, the RAPID consensus is computed through 24 h, and then the 36-, 48-, and 72-h values are set to be the 24-h intensity. This is done to test the feasibility of developing an RI index that extends beyond 24 h.

The consensus forecasts described in this paper (i.e., INT2, IVCN, and ICON; see Table 1 for definitions of model names) are equally weighted averages of the consensus members. An attempt is made to compute a consensus forecast at each forecast period (12, 24, 36, 48, 72, 96, and 120 h). A variable consensus is computed if two or more consensus members exist for a given forecast period. If fewer than two members exist, the variable consensus is aborted for this and subsequent time periods. The exception to this rule is ICON, which is a fixed consensus and requires all ensemble members to be present. Table 1 provides descriptions of the intensity models and consensus aids used in this study.

Results presented are from recomputed interpolated aids and consensus forecasts using methods described above as well as operational input. The purpose of this is to ensure that all of the results are computed using the same version of the interpolator. Average differences in performance between recomputed interpolations and those produced in operations are generally on the order of 1%.

Forecasts are verified only when the best-track intensity is greater than 20 kt (10.3 m s^{-1}) and only when the system is tropical or subtropical. Interpolated forecast aids are used as described above. If 6-h interpolated forecast aids are not available, then 12-h interpolated forecast aids are used. The 12-h interpolations occur approximately 15% of the time or less for models that are available every 6 h. The dataset is further restricted in that there must be a verifying official forecast. Performance is discussed through the use of percent improvement over a given forecast aid. Graphs of this type are also called skill charts. The measure of skill in these charts is defined as $100 \times$ (baseline error $-$ model error)/baseline error. Thus, skill is positive when the forecast aid error is less than that of the baseline forecast aid (e.g., SHF5 or the variable consensus IVCN). A one-tailed Student's *t* test at the 95% level is also employed as a method of testing the significance in forecast error differences between individual intensity forecast aids. Serial correlation (30 h) was removed from the independent data for significance testing. Removing serial correlation is an attempt to ensure that the individual forecasts are independent from each other, thereby increasing the likelihood that significant results hold up over time.

4. Results

Figure 1 presents an overview of the performance of the deterministic intensity forecast guidance available to forecasters at NHC relative to SHF5 (a skill baseline commonly used in the Atlantic and eastern North Pacific basins) for the entire 2006–09 time period. Since the RI index is available in both the eastern North Pacific and Atlantic basins, these regions are combined in the Fig. 1 evaluation. The top-performing aids in terms of mean forecast errors are the three consensus aids: ICON, IVCN, and INT2 (see Table 1 for definitions). It is somewhat disappointing that the two-model consensus proposed in Sampson et al. (2008) and implemented in the operational suite in 2006 (INT2) demonstrates similar skill as the four- (ICON) and five- (IVCN) model consensus aids implemented in operations in 2008. However, this is not entirely a surprise since the large gain in skill for INT2 is due to the relative independence of the member models, and the addition of more skillful intensity forecast aids does not
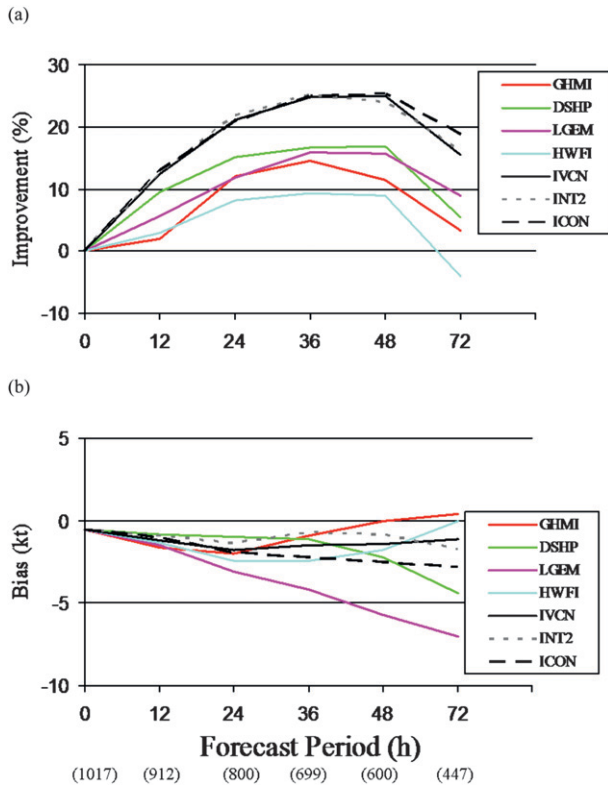
FIG. 1. (a) Intensity forecast skill (%) relative to SHF5 and (b) bias for selected intensity forecast aids, for the 2006–08 Atlantic and eastern North Pacific basins combined. Acronyms are defined in Table 1 and numbers of cases are shown in parentheses.
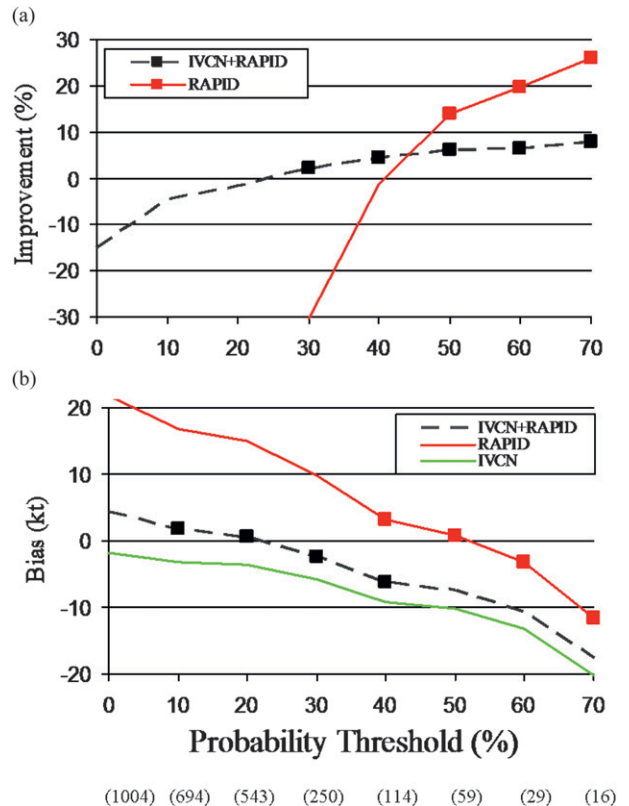

FIG. 2. (a) The 24-h forecast improvement (%) over IVCN and (b) bias for the deterministic RI aid (RAPID) and an intensity consensus aid that includes IVCN members and the RI aid (IVCN+RAPID) at selected probability thresholds. Each threshold defines subsets of the combined 2006–08 Atlantic and eastern North Pacific seasons. Significant improvement is shown with filled squares. Acronyms are defined in Table 1 and numbers of cases are shown in parentheses.

necessarily provide additional independence (see Sampson et al. 2006; Sampson et al. 2008). The level of skill for ICON is slightly higher than those of INT2 and IVCN at the longer forecast periods, but it has the largest negative biases of the consensus aids (Fig. 1b). Negative bias can indicate that an objective aid has difficulties forecasting rapid intensification and this will be discussed below.

An evaluation of RAPID at 24 h for RI index thresholds is shown in Fig. 2. The skill chart (Fig. 2a) indicates that RAPID becomes more skillful than IVCN in terms of mean forecast errors at the 40% threshold. The modified consensus that includes the members of IVCN and the RAPID aid (IVCN+RAPID) outperforms RAPID through the 40% threshold, but RAPID is the best performer at 50% and higher. One detriment to using 50% and higher thresholds for a deterministic aid is that those forecasts represent 4% or less of all the cases during 2006–08. Using a 40% probability threshold yields forecasts for approximately 8% of all cases, and RAPID performance is still comparable to IVCN.

Biases for the selected aids at 24 h are shown in Fig. 2b. Biases for all aids become more negative as the probability thresholds increase. The RAPID biases are nearly

zero at the 40% threshold, and the IVCN+RAPID biases are less negative at all thresholds than IVCN. Near-neutral bias should be a particularly desirable feature for objective aids in RI cases. A large negative bias in RI cases indicates that the aid is generally not forecasting enough rapid intensification. These results are consistent with those of Kaplan et al. (2010), which showed that the operational intensity guidance exhibits a low probability of detection of RI.

Selecting a threshold for the RAPID is subjective because any RI thresholds greater than 20% can be used to produce an objective aid (IVCN+RAPID) that outperforms IVCN in mean forecast error and bias performance. To minimize the negative bias and produce many RI aid forecasts, a 30% probability threshold could be selected. To maximize the skill improvement, a 50% probability threshold could be selected. For this study, the authors selected the 40% threshold for inclusion in the consensus aids. The 40% threshold produces a consensus (IVCN+RAPID) that outperforms

IVCN by approximately 4% in mean forecast errors while reducing the negative bias by a few knots. Both of these desired qualities are obtained while producing forecasts for a reasonable 8% of all cases This threshold could be lowered to increase availability (but reduce the gain in skill) or raised to increase the skill (but lower the aid availability), depending upon what verification characteristic the operational forecasters wish to optimize.

Figure 3 is an evaluation of IVCN, RAPID, and IVCN+RAPID for the 40% threshold. The IVCN+RAPID consensus has improved skill through about 36 h (though the improvement to IVCN is not statistically significant at 36 h), and the negative biases are reduced through the entire forecast. This is evidence that there could be some value in developing an RI aid that provides forecasts beyond 24 h. It should also be noted that the official NHC forecast (OFCL) outperforms both consensus aids in both skill and bias. The forecaster does have the advantage of using these aids in real time and can adjust the initial intensity after the guidance has been run (as seen in the initial estimate skill and bias); however, the forecasters clearly demonstrate skill over all their guidance in RI cases. The RAPID biases are near zero at all forecast lengths out to 48 h and are significantly less biased than the IVCN forecasts.

The 2009 season provides the opportunity to apply RAPID and IVCN+RAPID to independent data. Results for the 2009 data confirm that the inclusion of RAPID in IVCN reduces both the mean forecast error and negative bias (Fig. 4). The RAPID aid itself performs well for this small independent sample. Also, the official forecasts retain their superiority over the guidance both at analysis and all-forecast leads. Similar positive results were found for an objective aid formed with ICON members and RAPID, but the ICON aid by definition does not allow the inclusion of aids that only appear intermittently so it would be problematic to include the RI aid in the current implementation of ICON.

To be thorough, tests similar to those discussed above were run with a 35% threshold. This threshold may also be of interest to operations due to its increased availability. The availability fell between the 30% and 40% thresholds with only a small reduction in skill on dependent data; however, RAPID skill using this threshold was less for the independent 2009 cases and produced IVCN+RAPID forecasts that were not significantly improved over IVCN.

5. Summary and conclusions

A deterministic RI aid (RAPID) was developed using RI index text files for the 2006–08 Atlantic and eastern North Pacific seasons. RAPID outperforms the consensus aids at about 50% probability; however, the sample size is quite limited. When added as a member to
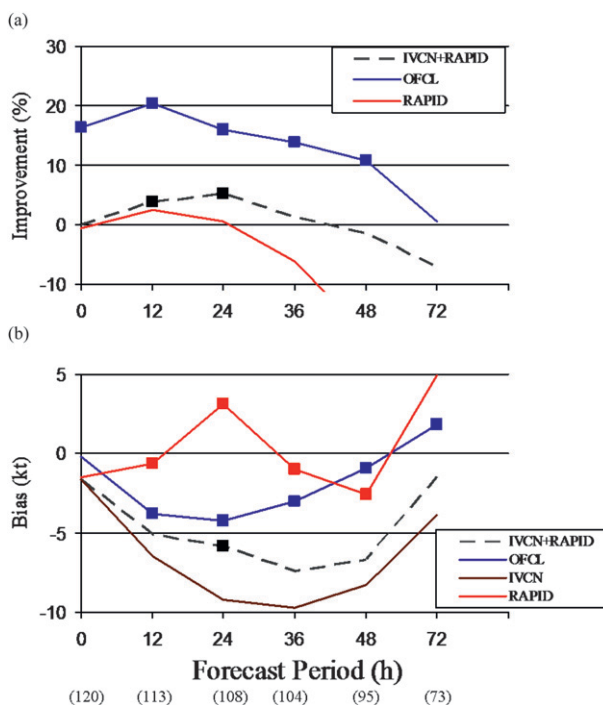

FIG. 3. (a) Intensity forecast improvement relative to IVCN (%) and (b) bias for forecasts in which the RI index reached the 40% probability threshold. Objective aids include RAPID, IVCN+RAPID, and OFCL. The 40% probability threshold defines small subsets of the combined 2006–08 Atlantic and eastern North Pacific seasons. Significant improvement is shown with filled squares. Acronyms are defined in Table 1 and numbers of cases are shown in parentheses.

existing operational consensus aids, RAPID improves the mean forecast errors. Significant improvements are obtained starting at a probability threshold of 30%. RAPID also reduces the negative bias of the consensus aids for these cases. As an example, the authors selected a 40% threshold to produce RAPID and the modified consensus (IVCN+RAPID) on the entire 2006–08 dataset and found significant improvement in skill and reduction of bias. Results for the 2009 independent data confirm those results. Also, the official forecasts retain their superiority over the guidance, both at analysis and all forecast leads.

Based on these findings, the authors suggest a 40% threshold for producing RAPID and recommend including it in IVCN. Using a 40% threshold yields a RAPID that is available for approximately 8% of all verifying forecasts and produces approximately a 4% reduction in mean forecast errors when included in IVCN (IVCN+RAPID). It also reduces IVCN negative biases by 15%–20% out to 24 h. Including RAPID in the IVCN consensus improves the consensus out through 36 h, indicating that the development of an aid specifically designed for RI at 36 and 48 h is feasible.
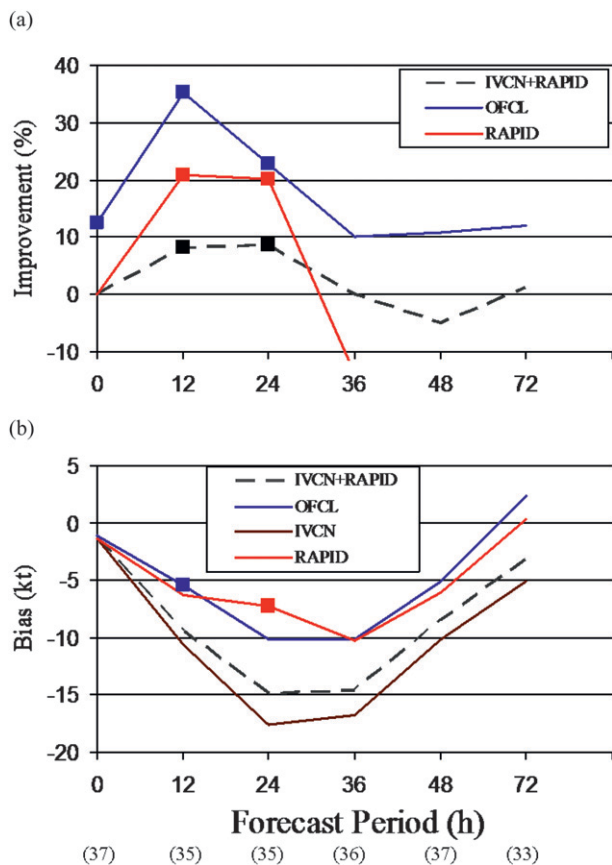
FIG. 4. (a) Intensity forecast improvement relative to IVCN (%) and (b) bias for forecasts in which the RI index reached the 40% probability threshold for the 2009 independent Atlantic and eastern North Pacific dataset. Objective aids include RAPID, IVCN+RAPID, and OFCL. Significant improvement is shown with filled squares. Acronyms are defined in Table 1 and numbers of cases are shown in parentheses.

As for the intensity consensus forecasts, it is suspected that improvements in the consensus members and the addition of other independent, skillful forecast aids would further improve their performance. Recent GFDL, GFDN, HWRF, SHIPS, and LGEM upgrades should all increase model skill. The Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS[1]; Reynolds et al. 2010) and the Coupled Hurricane Intensity Prediction System (CHIPS; Emanuel et al. 2004) have shown promise in Atlantic and eastern North Pacific intensity forecasts and could possibly be added to consensus aids. There is also guidance developed in the Hurricane Forecast Improvement Program (HFIP), the Joint

Hurricane Test Bed, and other programs that show promise for improving RI and intensity forecasts. Many of these developments naturally lend themselves to consensus and ensemble applications and should provide fruitful results in the future.

In a relevant development, the NHC in 2010 extended the lead times of tropical storm hurricane watches and warnings by 12 h to 48 and 36 h, respectively. Although the results of this manuscript suggest that the RI index can be used to improve forecasts through 36 h, it is hypothesized that the development of versions of the RI index specifically designed for use at the 12-, 36-, and 48-h lead times will provide further improvements in forecast skill. Thus, the development of probabilistic and deterministic forecast guidance (e.g., RAPID) at lead times of 12–48 h is thought to be an important topic for future research.

REFERENCES

Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The operational GFDL coupled hurricane–ocean prediction system and a summary of its performance. *Mon. Wea. Rev.,* **135,** 3965–3989.

Bernardet, L., and Coeditors, cited 2010: Hurricane Weather Research and Forecasting (HWRF) Model scientific documentation. [Available online at http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRF_final_2-2_cm.pdf.]

DeMaria, M., 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Mon. Wea. Rev.,* **137,** 68–82.

——, 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.

——, J. A. Knaff, and J. Kaplan, 2006: On the decay of tropical cyclone winds crossing narrow landmasses. *J. Appl. Meteor. Climatol.,* **45,** 491–499.

——, ——, and C. R. Sampson, 2007: Evaluation of long-term trends in tropical cyclone intensity forecasts. *Meteor. Atmos. Phys.,* **97,** 19–28.

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

Hogan, T. F., and T. E. Rosmond, 1991: The description of the Navy Operational Global Atmospheric Prediction System's spectral forecast model. *Mon. Wea. Rev.,* **119,** 1786–1815.

Kaplan, J., and M. DeMaria, 1995: A simple empirical model for predicting the decay of tropical cyclone winds after landfall. *J. Appl. Meteor.,* **34,** 2499–2512.

——, and ——, 2001: A note on the decay of tropical cyclone winds after landfall in the New England area. *J. Appl. Meteor.,* **40,** 280–286.

——, ——, and J. A. Knaff, 2010: A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Wea. Forecasting,* **25,** 220–241.

Knaff, J. A., M. DeMaria, C. R. Sampson, and J. M. Gross, 2003: Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting,* **18,** 80–92.

——, C. Guard, J. Kossin, T. Marchok, B. Sampson, T. Smith, and N. Surgi, 2006: Operational guidance and skill in forecasting structure change. *Proc. Sixth Int. Workshop on Tropical Cyclones,* San Jose, Costa Rica. WMO, 1.5. [Available online at http://severe.worldweather.org/iwtc/.]

Rennick, M. A., 1999: Performance of the navy's tropical cyclone prediction model in the western North Pacific basin during 1996. *Wea. Forecasting,* **14,** 3–14.

Reynolds, C. A., J. D. Doyle, R. M. Hodur, and H. Jin, 2010: Naval Research Laboratory multiscale targeting guidance for T-PARC and TCS-08. *Wea. Forecasting,* **25,** 526–544.

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. S. Goerss, and H. C. Weber, 2006: Operational performance of a new barotropic model (WBAR) in the western North Pacific basin. *Wea. Forecasting,* **21,** 656–662.

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