Joint Hurricane Testbed Year 1 Final Report
September 1, 2011 to September 1, 2012

Improvements to the SHIPS Rapid Intensification Index

Principle Investigator: John Kaplan
Hurricane Research Division
NOAA/AOML

Co-PIs: Christopher M. Rozoff (CIMSS), Charles R. Sampson (NRL), James P. Kossin (NOAA/NCDC), Christopher S. Velden (CIMSS)

Collaborator: Mark DeMaria (NOAA/NESDIS)

Computer scientist support: Paul A. Leighton NOAA/HRD

NHC Points of Contact: Stacy R. Stewart, Eric S. Blake, and Christopher W. Landsea

1. Project overview

The goal of this Joint Hurricane Testbed (JHT) funded project is to improve the forecasting utility of the current operational SHIPS rapid intensification index (RII) (Kaplan et al. 2010) by implementation of the following model enhancements. First, ensemble-based versions of the RII that employ both the current SHIPS discriminant RII as well as new Bayesian and logistic versions (Rozoff and Kossin 2011) are to be derived for the added forecast lead-times of 12-h, 36-h and 48-h as well as the current operational lead-time of 24 h. Secondly, microwave imagery-based versions of the RII are to be derived at each of the aforementioned lead times. Lastly, revised versions of the recently developed deterministic rapid intensity aid (Sampson et al. 2011) are to be developed utilizing the newly developed ensemble-based RII models. A brief description of our year-1 JHT project accomplishments related to the aforementioned tasks as well as an evaluation of the current SHIPS RII is provided below.

2. Year-1 project accomplishments

a. Evaluation of SHIPS RII

Figure 1 shows a verification of the 2008-2011 operational 24-h lead-time SHIPS RII forecasts that were evaluated based upon a Brier Skill Score (BSS) utilizing the methodology described in Kaplan et al. (2010). It can be seen that the RII forecasts that were made for the above time period were generally skillful when verified against both climatology and a forecast of no rapid intensification (RI) although the skill was rather limited for the Atlantic basin.

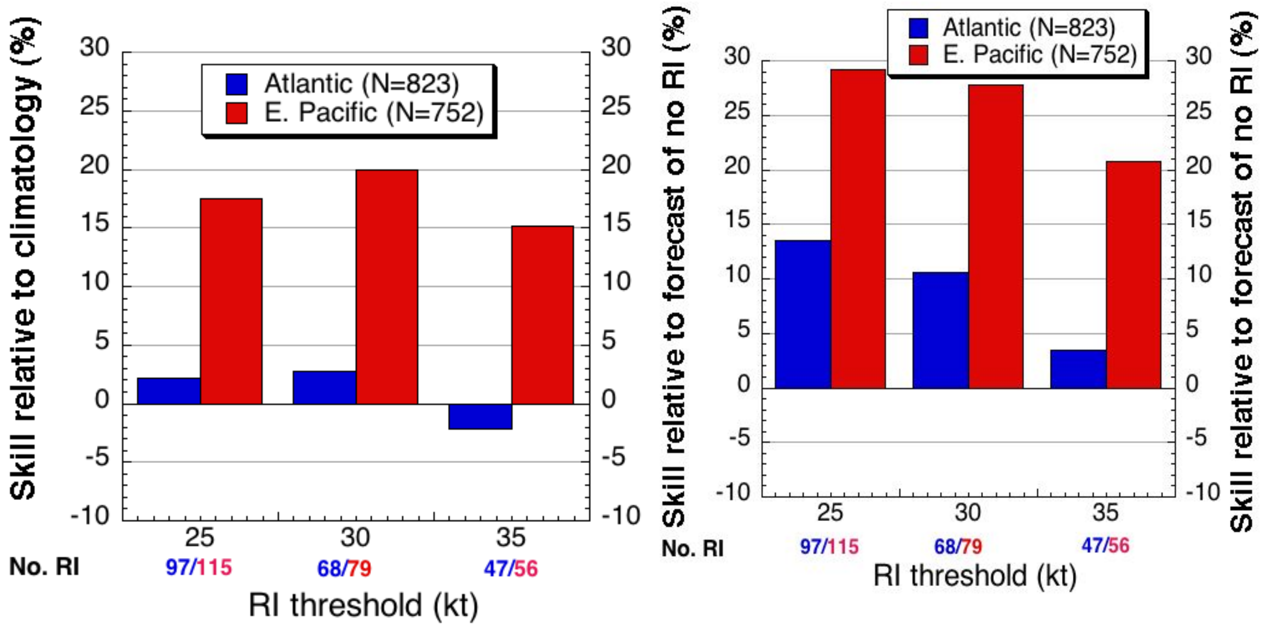

Fig. 1. Skill of the operational 2008-2011 Atlantic and eastern North Pacific SHIPS RII forecasts relative to climatology (left panel) and a forecast of no RI (right panel). The total number of 24-h forecasts is provided in the legend while the number of RI cases is depicted along the x-axis.

Since, as noted above, the skill of the current operational RII is still somewhat limited, particularly for the Atlantic basin, an experimental version of the SHIPS RII was developed as part of a recently completed JHT project (Kaplan et al. 2011). This new experimental RII included two additional RI predictors that had not been previously employed in the old operational SHIPS RII as well as two replacement predictors; thus, it included a total of ten RI predictors. Specifically, predictors computed utilizing total precipitable water and the second principle component of GOES infra-red imagery were employed in place of predictors that had been previously derived based upon the 850-700 mb relative humidity and the percentage of the area within the storm's inner-core region with GOES infra-red pixels < -30 °. The inner-core dry air flux and the maximum sustained wind at time t=0 h were the two additional predictors that were used in the new experimental version of the RII but not in the old operational version. This new experimental version of the RII was officially accepted for operational implementation by the NHC in May of 2012.

Figure 2 shows a comparison of the skill of the new 2012 operational ten-predictor SHIPS RII and the old eight-predictor 2011 version for the independent 2008-2011 sample. It should be noted that the 2012 operational SHIPS-RII included all of the previously described model improvements as well as a slightly modified method for scaling the individual RI predictor magnitudes that was implemented to produce smoother changes in the RI probabilities between 6-hourly forecast times. To obtain the independent results depicted in Fig. 2, both the new 2012 ten-predictor and old 2011 eight-predictor versions of the RII were first re-derived for each of the individual years that comprise the four-year sample (i.e., 2008, 2009, 2010, and 2011) after excluding all cases from each of those years. Then, utilizing the operational input data that were archived for each season, both versions of the SHIPS-RII were re-run for all forecast times for each of the four independent years by employing the version of each model that had been previously derived by excluding the cases from that specific year. It can be seen that the new experimental version of the RII was generally superior to the current operational version in each basin with absolute skill improvements of up to about 7% (4%) observed in the Atlantic (eastern North Pacific) basins, respectively.

Figure 3 shows examples of the performance of the 2012 operational version of the RII relative to the 2011 version for eastern North Pacific Hurricane Adrian (2011) and Atlantic basin Hurricane Ophelia (2011). Although it can be seen that the 2012 version of the RII produced higher and less variable probability of RI estimates for Adrian, both versions produced similar RI probabilities for Ophelia. Nevertheless, the overall results depicted in Figs. 2 and Figs. 3 indicate that, on average, the new 2012 operational version of the RII is superior to the old 2011 version and thus will be employed to derive the new ensemble versions of the RII described below.

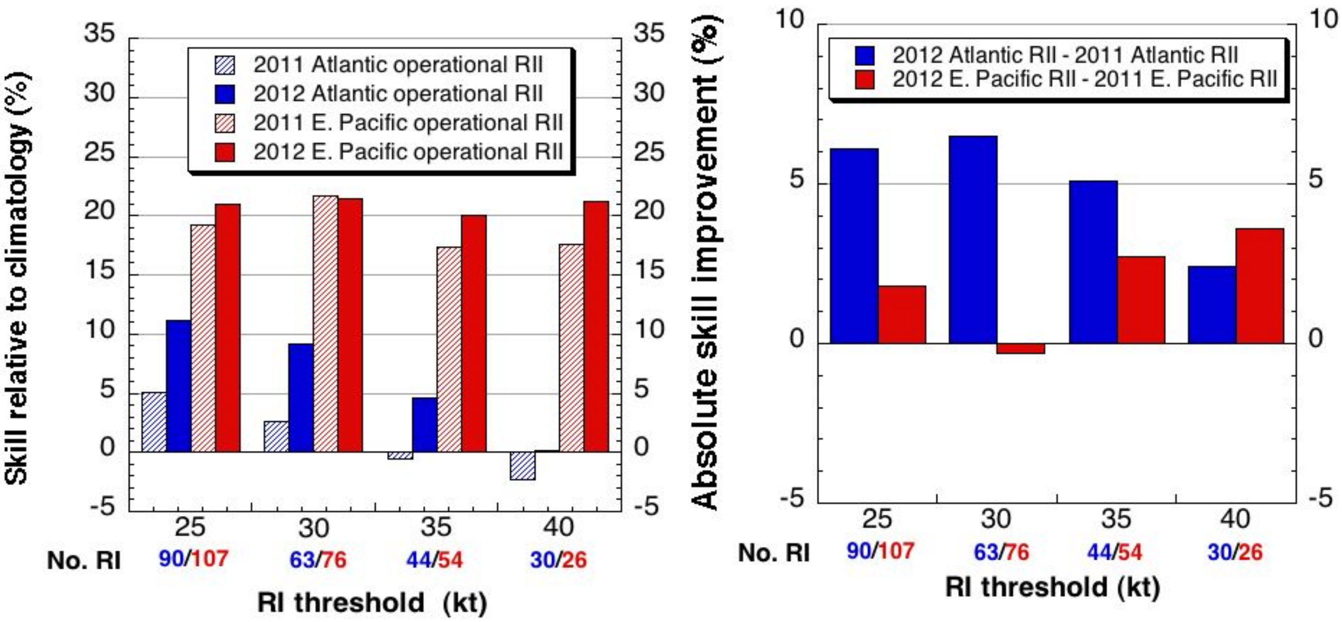

Fig. 2. The skill of the old 2011 operational (solid) and new 2012 operational (hatched) versions of the RII for the 2008-2011 Atlantic (blue) and eastern North Pacific (red) independent re-run forecasts (left panel) and the percentage improvements in skill of the 2012 version of the RII over the 2011 version (right panel). The total number of 24-h forecasts was 823 for the Atlantic and 752 for the eastern North Pacific basin.

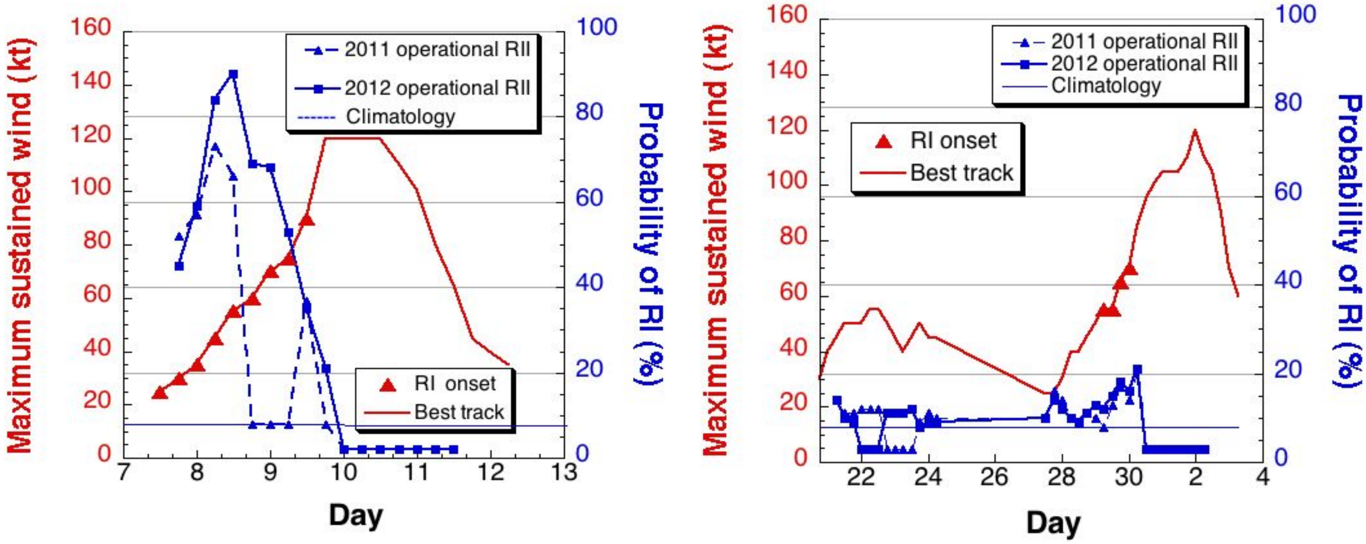

Fig. 3. Examples of the differences between the predicted 30-kt RII probabilities for the 2012 (solid blue line) and 2011 (dashed blue line) versions of the RII for eastern North Pacific basin Hurricane Adrian (left panel) and Atlantic basin Hurricane Ophelia (right panel) from the 2011 Hurricane Season. The NHC best track maximum sustained wind estimates (sold red line) and time of onset of each 24-h period of RI (red triangles) as well as the climatological probability of RI (horizontal blue lines) for each basin are also depicted.

b. Development of an ensemble-based RII

Rozoff and Kossin (2011) showed that the skill of the RII could be improved by averaging the probabilities computed from two new versions of the RII based on the Bayesian and logistic regression (Logistic) methods and those obtained from the current SHIPS discriminant version of the RII. As an example, Fig. 4 shows the skill of the three RII models (i.e., the Discriminant, Logistic, and Bayesian models) and their ensemble average for the 30-kt RI threshold in both Atlantic and eastern North Pacific basin for the 24-h lead-time. It is worth noting that each model utilizes the 1995-2011 developmental data. Skill was obtained here by using leave-one-year-out cross validation. It can be seen that the skill of the ensemble version of the RII slightly exceeds that of the other three versions in each basin. Although the increase in skill of the ensemble-based RII over that of the other RI models is relatively small, results obtained for the other RI thresholds and lead-times showed that the ensemble-based version of the RII is nearly always more skillful than any of the individual RII models.

Figure 5 shows the skill of both the discriminant and ensemble-based versions of the RII for the Atlantic and eastern North Pacific basins for the 12-h, 24-h, 36-h and 48-h lead times. The RI threshold representing approximately the 95% percentile of over-water intensity change was chosen for each lead-time and each ocean basin. It can be seen that the skill of the ensemble RII exceeds that of the discriminant RII by up to 4% in the Atlantic basin and 9% in the eastern North Pacific basins, respectively. Interestingly, a general increase in skill as a function of forecast lead-time is observed (particularly in the Atlantic basin) even though the magnitude of the intensity change that corresponds to the longer lead times also increases. This is likely due in part to the fact that, at the longer lead-times, forecasting the timing of the onset of RI is not as crucial since RI may actually only last for a period of perhaps 18 h and thus need only take place within that 48-h time window.

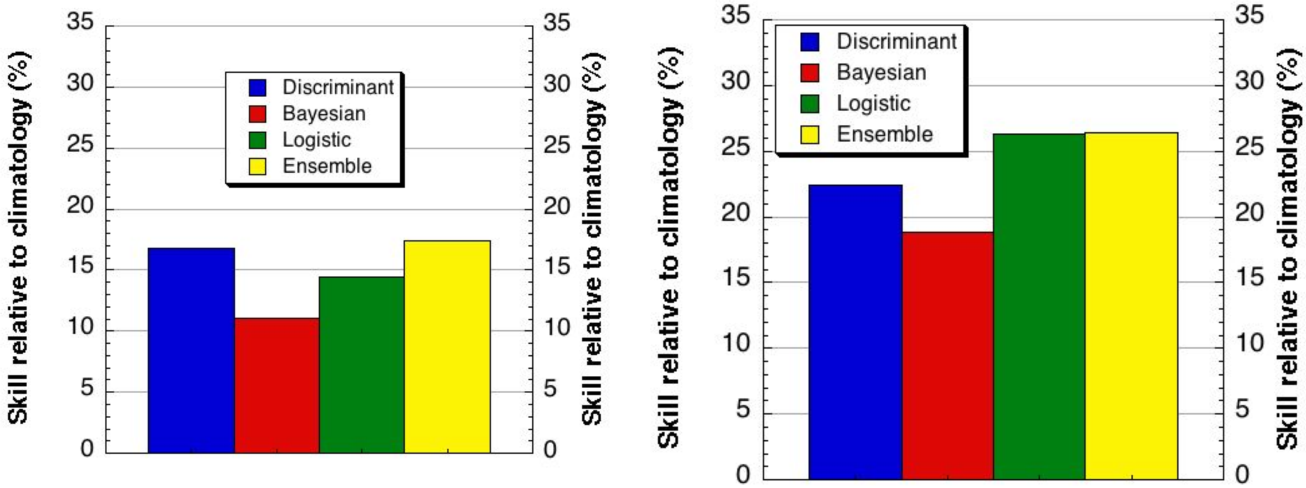

Fig. 4. The skill of the Discriminant, Bayesian, Logistic, and Ensemble-based versions of the RII models for the Atlantic (left panel) and eastern North Pacific (right panel) basins for the 30-kt RI threshold at the 24-h lead-time for the 1995-2011 cross-validated sample.

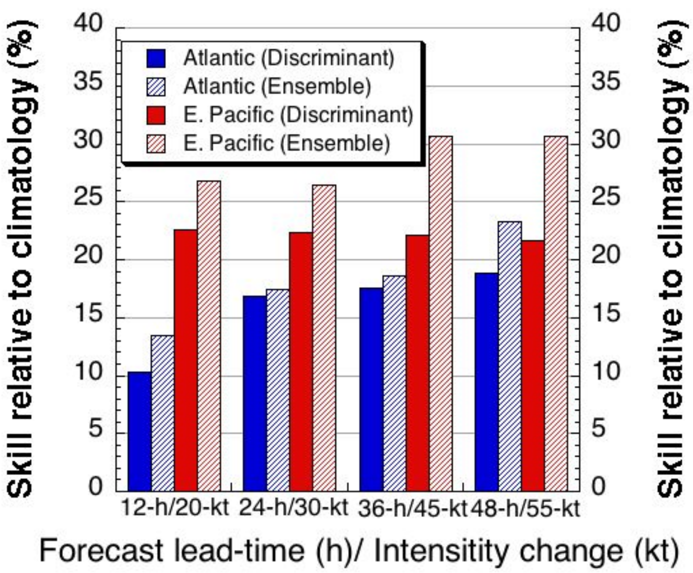

Fig. 5. Skill of the Atlantic (blue) and eastern North Pacific (red) cross-validated Ensemble RII forecasts as a function of forecast lead time for the 1995-2011 cross-validated sample.

Figure 6 shows an example of the cross-validated Ensemble-based RI guidance performance for Hurricane Wilma (2005). It can be seen that the RI probabilities for the 48-h lead-time (yellow line) that correspond to the likelihood of an intensity change of $\geq$ 55 kt in the next 48 h are generally higher than those observed for any of the shorter lead times and exceed 70% for the time period late on the 17th when Wilma was only a weak 45-kt tropical storm.

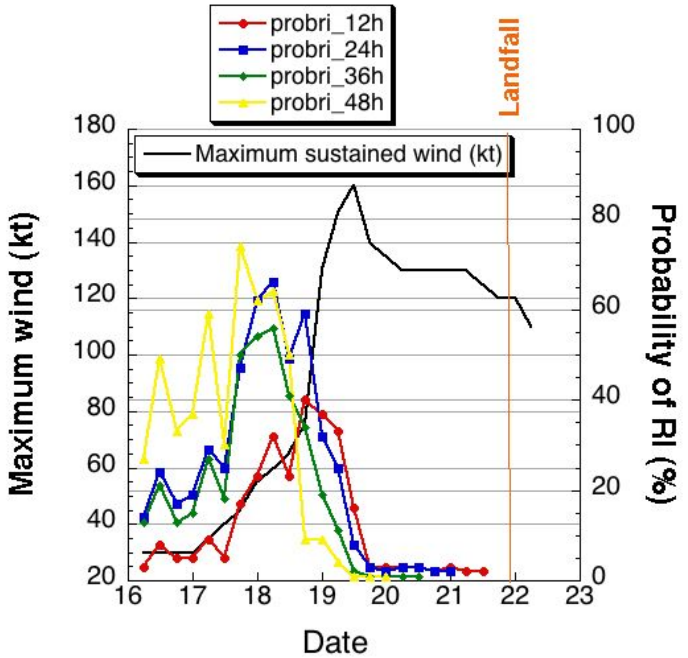

Fig. 6. The probability of RI of the 12-h (red), 24-h (blue), 36-h (green) and 48-h (yellow) lead-time Hurricane Wilma (2005) Ensemble RII forecasts. The solid black line depicts the National Hurricane Center (NHC) best track maximum-sustained wind estimates. The time of Wilma's landfall along the Yucatan Peninsula is also shown.

In preparation for real-time testing of the newly developed multiple lead-time (i.e., 12-h, 24-h, 36-h, and 48-h) version of the RII described above, code was written to compute each of the three different RII models (Discriminant, Bayesian, and logistic) as well as their ensemble at each forecast lead-time. This code was implemented and has been running in real-time at the Cooperative Institute for Research in the Atmosphere (CIRA) located on the Colorado Statue University (CSU) campus since the beginning of August with the output (Table 1) being made available to forecasters at the NHC via an ftp site. Soon after the code was installed, one minor bug that resulted in missing probabilities occasionally being coded as something other than missing was detected and subsequently corrected and the revised version of the code has been running successfully ever since in both the Atlantic and eastern North Pacific basins in real-time.

Table 1. Example of output for the multiple lead-time version of the RII that is currently running in real time at CIRA on the CSU campus for eastern North Pacific basin storm number seven on 9 August 2012 at 0600 UTC. The output shows the RI probabilities for the 20-kt/12-h, 25-kt/24-h, 30-kt/24-h, 35-kt/24-h, 40-kt/24-h, 45-kt/36-h and 55-kt/48h RI thresholds and lead times for the SHIPS-RII (i.e., Discriminant), Logistic, Bayesian, and Ensemble versions of the RII.

```
+++++++ SECTION 3, RII WITH MULTIPLE TIMES ++++++++
                   AND CONSENSUS FOR JHT

** 2012 E. Pacific EXPERIMENTAL  RI INDEX EP072012 EP07       08/09/12  06 UTC **

Prob RI for 20kt/ 12hr RI threshold=  12% is   2.0 times sample mean ( 6.2%)
Prob RI for 25kt/ 24hr RI threshold=  18% is   1.5 times sample mean (12.5%)
Prob RI for 30kt/ 24hr RI threshold=  15% is   1.7 times sample mean ( 8.3%)
Prob RI for 35kt/ 24hr RI threshold=  11% is   1.9 times sample mean ( 5.7%)
Prob RI for 40kt/ 24hr RI threshold=   7% is   1.8 times sample mean ( 4.0%)
Prob RI for 45kt/ 36hr RI threshold=   2% is   0.3 times sample mean ( 5.9%)
Prob RI for 55kt/ 48hr RI threshold=   2% is   0.3 times sample mean ( 5.5%)

Matrix of RI probabilities
---------------------------------------------------------------------
 RI (kt / h)   | 20/12 | 25/24 | 30/24 | 35/24 | 40/24 | 45/36 | 55/48
---------------------------------------------------------------------
   SHIPS-RII:     12.1%    18.5%    14.5%    10.5%     7.4%     1.6%     1.5%
    Logistic:      1.5%     2.7%     1.3%     0.7%     0.3%     0.2%     0.0%
    Bayesian:      0.5%     0.2%     0.1%     0.0%     0.0%     0.1%     0.0%
   Consensus:      4.7%     7.1%     5.3%     3.7%     2.6%     0.7%     0.5%
```

c. RI models incorporating microwave imagery

We seek to improve the statistical prediction of RI using predictors derived from passive microwave imagery (MI). In contrast to infrared and visible satellite imagery, MI can depict the distribution of precipitation underneath thick, overlying clouds. To this end, we have created physically based statistical features exploiting information about the distribution and intensity of precipitation, including warm rain and ice hydrometeors in the inner core of developing and mature TCs. We have used these predictors to enhance our probabilistic models for RI.

In particular, the Bayesian and logistic regression RI models described in Rozoff and Kossin (2011) have been updated to include microwave imagery (MI)-based predictors and we plan to eventually update the SHIPS-RII. The model includes a climatology of MI brightness temperatures (T_b) from the Special Sensor Microwave/Imager (SSM/I), Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), the Advanced Microwave Scanning Radiometer-EOS (AMSR-E), and WindSat calibrated via histogram matching to the frequencies of 19.4, 37.0, and 85.5 GHz. This dataset covers the Atlantic and eastern North Pacific basins over the years of 1998–2008. Multiple MI channels are chosen because of each channel's unique perspective on storm structure. Low-frequency channels (i.e., 19.4 and 37.0 GHz) measure the emission of microwave radiation from liquid hydrometeors, and, as such, can often depict the low-level precipitation structure. Higher frequency channels (37.0 and 85.5 GHz) capture ice scattering from precipitation at upper levels of the troposphere.

Storm-centered predictors are chosen such that they are statistically independent and maximize the BSS in independent leave-one-year-out cross validation of the various RI models. Also, in order for an MI-based predictor to be considered, the difference in the composite means of the RI and no-RI samples must be statistically significant at the 95% level according to a two-sided student-*t* test. To properly compare the BSS of the MI-enhanced RI models with the MI-free versions, each model is trained and evaluated on only the forecast times in which all of the SHIPS-based and optimal MI-based predictors are available. Despite the uneven temporal coverage of satellite passes over a storm, the models are developed for forecasting at the synoptic times of 00, 06, 12, and 18 UTC. To deal with this, the MI-based schemes are only developed for forecasts in which the MI is less than 6-h old. Depending on the quality-control criteria used in defining microwave predictors, this means microwave data are available about 40-60% of the time.

As an example of the improvements that can be achieved by including simple MI-based predictors in the logistic regression RI model, we focus our attention on RI prediction for an RI threshold of 35-kt per 24 h. The higher RI threshold typically yields a lower BSS than those of the lower RI thresholds, giving it the most potential for improvement (Kaplan et al. 2010; Rozoff and Kossin 2011). Table 2 shows the optimal, statistically independent MI-based predictors for the logistic scheme in the North Atlantic basin. Overall, the composite means of the RI and non-RI samples show that precipitation is more intense and centralized in TCs that are currently or about to undergo RI.

Table 3 shows the BSS for the Atlantic basin logistic regression-based RI model with and without MI-based predictors for the 35-kt RI threshold. Whether we consider TCs with current $v_{max} \geq 25$ kt or only those with $v_{max} \geq 45$ kt, the BSS is improved by the inclusion of MI-based predictors. One notable improvement is that the MI-enhanced RI model can more often produce higher probabilities of RI and with a higher degree of accuracy. Figure 7 shows that the BSS can be improved at other RI thresholds as well, although the greatest relative improvement in forecast skill is achieved for higher RI thresholds.

Table 2. Microwave-determined features applied to the logistic regression model in the North Atlantic basin. Also shown are the relative values of the predictor means for the RI sample relative to the non-RI sample.

Feature Description	RI avg.
19.4-GHz avg. $T_{b,v}$ (r = 100–300 km)	higher
19.4-GHz min eye $T_{b,v}$	higher
19.4-GHz avg. ring $T_{b,h}$	higher
37.0-GHz radius of max $T_{b,h}$	lower
37.0-GHz avg. $T_{b,v}$ (r = 0–100 km)	higher

Table 3. The Brier skill score for TCs tested in the Atlantic basin for an RI threshold of 35-kt per 24 h for the logistic regression model. Results are shown for the model incorporating TCs with $v_{max} \geq 25$ kt (N = 1360) and TCs with $v_{max} \geq 45$ kt (N = 1013).

Model	Brier Skill Score
$v_{max} \geq 25$ kt without MI	7.5%
$v_{max} \geq 25$ kt with MI	19.6%
$v_{max} \geq 45$ kt without MI	11.9%
$v_{max} \geq 45$ kt with MI	24.7%

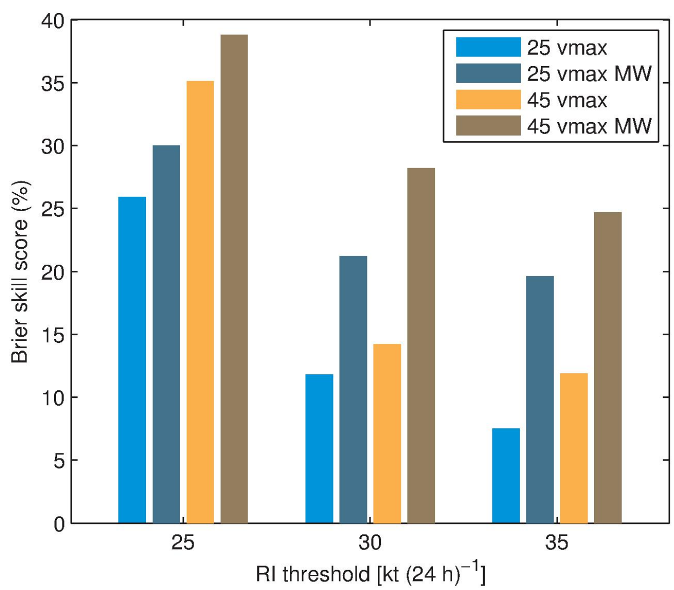

Figure 7. Brier skill scores for Atlantic basin TCs with RI thresholds of 25, 30, and 35-kt per 24 h for the RI model excluding MI predictors (light blue and orange for TCs with $v_{\max} \geq 25$ kt and $v_{\max} \geq 45$ kt, respectively) and including MI predictors (dark blue and brown for TCs with $v_{\max} \geq$ 25 kt and $v_{\max} \geq 45$ kt, respectively).

The optimal predictors for the eastern North Pacific basin differ from the Atlantic for the 35-kt per 24 h RI threshold (Table 4). In this case, predictors from each of the three tested channels are found to improve the forecast skill of the logistic regression model. Similar to before, more intense convection is favorable for RI. Moreover, in the RI sample, the convection tends to be more widespread. At first glance, it is counterintuitive that the RI sample shows more asymmetry at both 19.4 and 85.5 GHz. Studying individual cases suggest this may be due to strong convective bursts dominating the signal, but further study of this impact is necessary. Table 5 indicates that while the MI-based predictors improve the BSS for the 35-kt RI threshold in the eastern North Pacific basin the relative improvement is far less than in the Atlantic. This lack of relative improvement may be, in part, due to the already elevated BSS for the RI model that does not use MI-based predictors. In other words, MI-based predictors may be most beneficial for less predictable ocean basins such as the Atlantic. Nevertheless, Fig. 8 indicates that inclusion of the MI-based predictors also improves the BSS for all three of the RI thresholds

Table 4. Same as Table 3, but for the eastern North Pacific basin.

Feature Description	RI avg.
19.4-GHz std. dev. $T_{b,h}$ (r = 0–100 km)	Higher
19.4-GHz % area $T_{b,v}$>245K (r = 50-200 km)	Higher
37.0-GHz avg. $T_{b,h}$ (r = 100–300 km)	Higher
85.5-GHz std. dev. $T_{b,v}$ (r = 100–300 km)	Higher

Table 5. The Brier skill score for TCs tested in the eastern North Pacific basin for an RI threshold of 35-kt per 24 h for the logistic regression model. Results are shown for the model incorporating TCs with $v_{max} \geq$ 25 kt (N = 1470) and TCs with $v_{max} \geq$ 45 kt (N = 939).

Model	Brier Skill Score
$v_{max} \geq$ 25 kt without MI	24.7%
$v_{max} \geq$ 25 kt with MI	28.2%
$v_{max} \geq$ 45 kt without MI	24.2%
$v_{max} \geq$ 45 kt with MI	27.5%

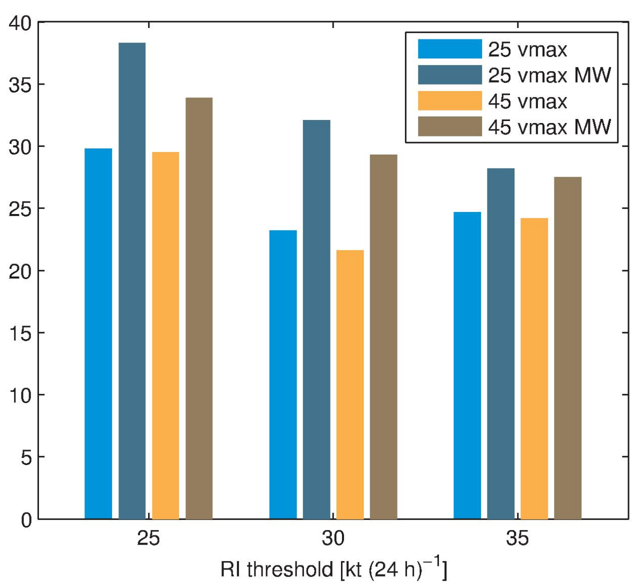

Fig. 8. Same as Fig. 7, but for the eastern North Pacific basin.

A number of improvements are underway to improve our MI-based RI models for the 2013 experimental testing. We are augmenting our developmental dataset with Special Sensor Microwave Imager/Sounder (SSMI/S) MI (at 19.4, 37.0, and 91.7 GHz) and the Advanced Microwave Sounding Unit-B (AMSU-B) MI (at 89.9 GHz). In addition, more years of data (2009-2011) are being added to enhance our developmental dataset.

Currently, we are finishing the implementation of real-time MI-based RI models for both the Atlantic and eastern North Pacific basins. We are currently ingesting real-time data from NESDIS for SSM/I, SSMI/S, AMSU-B, and TRMM-TMI and creating predictors for forecasts at synoptic times. The automated coupling between real-time predictor creation and the SHIPS model is nearly complete. Because the real-time scheme is being completed during the middle of the 2012 hurricane season, only TCs in September through December will be tested in real time. However, we will retroactively test the 2012 Atlantic and eastern North Pacific basin seasons at CIMSS in the end of the 2012 season so that seasonal statistics can be obtained. Plans are on track to have an experimental set of microwave-based models ready for real-time testing during the 2013 hurricane season.

d. Deterministic RI Aid

A deterministic rapid intensification aid (RAPID; Sampson et al. 2011) has been using the 2009 version of the Rapid Intensification Index (RII; Kaplan et al. 2010). The 2009 version of RII produced probabilities of three intensification rates (25, 30, 35 knots) through a 24-h period. The threshold used to trigger whether the deterministic RI aids were run is 40%, although this could be adjusted upward to improve performance. In the first year of our JHT project we developed a deterministic intensity forecast aid that runs as part of the objective aid suite at NCEP within the SHIPS model. We also added a deterministic RI aid for the 40-kt RII guidance and deployed it in the 2012 version of SHIPS. The resultant output from all this deterministic RI guidance is then transmitted back to NHC and ingested into the ATCF in real-time. The RAPID guidance ran successfully for the 2011 and 2012 seasons to date at NRL. We recently recomputed an intensity consensus (see Sampson et al. 2008) using RAPID and the members of IVCN (DSHP, LGEM, GHMI, GFNI and HWFI) in order to evaluate its impact on the consensus. As seen in Figure 9, we found improved forecast errors and reduced biases. The improved forecast errors for IVRI are about 4-5% and are moderately significant. The IVRI bias improvements are approximately 2.5 and 2.8 kt at 12 and 24 h, respectively.

We continue to run and evaluate the RAPID aid in real-time at NRL this season. In addition, we also recently processed the latest version of the RII guidance that includes additional forecast times (36 and 48 h) and RII ensemble probabilities. This was done utilizing the cross-validated forecasts from the 2008-2010 seasons so that we could compare with IVCN, and the results look promising (Fig. 10). The RAPID aid (RX25) outperforms both consensus aids in both mean absolute error and bias. The reduction of mean forecast errors garnered by including the RAPID aids in the consensus (IVRI vs. IVCN) is approximately 10% (highly significant) and the biases improvements are approximately 6-12 kt. We do have reservations about the new RAPID aids depicted in Fig. 10 performing this well when it is tested on truly independent data; however, we do expect them to perform at least as well or better than the RAPID guidance that is currently being run as part of the SHIPS guidance suite. The new 36- and 48-h RAPID aids also appear to improve performance of the IVRI consensus over IVCN, though the number of cases for those forecast times is quite limited.

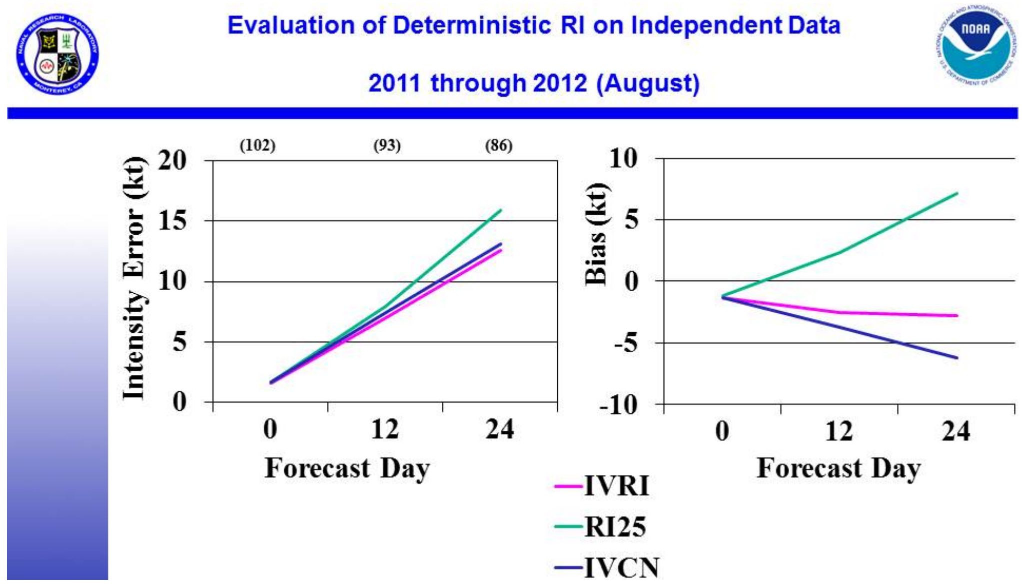

Fig. 9. Results from 2011 and 2012 Atlantic and eastern North Pacific seasons for an intensity consensus (IVRI) that includes the deterministic RI aid that was run post-season shows lower mean errors and less bias than IVCN (the operational NHC consensus). RI25 is the deterministic RI guidance for a 25-kt RI event.

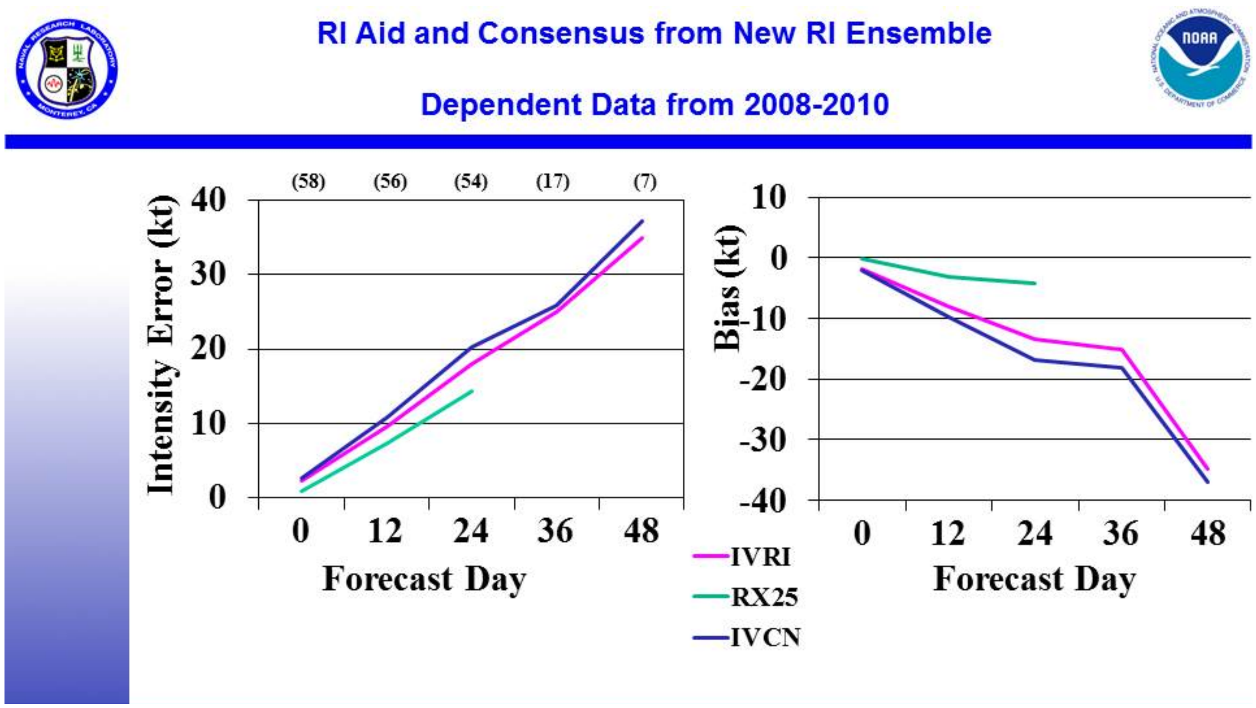

Fig. 10. Results from 2008-2010 Atlantic and eastern North Pacific seasons for an intensity consensus (IVRI) that includes the deterministic RI aid shows lower mean errors and less bias than IVCN (the operational NHC consensus). RX25 is the deterministic RI guidance for a 25-kt RI event using the ensemble RI probabilities and the most recent version of the RII algorithm. Note that these results were obtained using the cross-validated 2008-2010 developmental dataset.

3. References

Kaplan, J, M. DeMaria 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.

______, J. Cione, M. DeMaria, J. Dostalek, J. Dunion, J. Knaff, J. Zhang, T. Lee, J. Hawkins, J. Solbrig, E. Kalina, and P. Leighton, 2011: Improvement in the rapid intensity index by incorporation of inner-core information. JHT final report. Available from http://www.nhc.noaa.gov/jht/09-11reports/final_Kaplan_JHT11.pdf.

Rozoff, C. M., and J. P. Kossin, 2011: New probabilistic forecast schemes for the prediction of tropical cyclone rapid intensification. *Wea. Forecasting,* **26,** 677-689.

Sampson, C. R., J. L. Franklin, J. A. Knaff and M. DeMaria, 2008: Experiments with a Simple Tropical Cyclone Intensity Consensus. *Wea. Forecasting***, 23**, 304-312.

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