16C.2 EVALUATION OF MACHINE-LEARNING BASED RAPID INTENSIFICATION FORECAST PERFORMANCE DURING THE 2017 ATLANTIC HURRICANE SEASON

Alexandria Grimes, Andrew E. Mercer *, and Kimberly M. Wood
Mississippi State University, Mississippi State, MS

1. INTRODUCTION

Forecasting rapid intensification (RI) within Atlantic tropical cyclones (TCs) remains a significant challenge in operational meteorology. Kaplan et al. (2010) provided the Statistical Hurricane Intensity Prediction Scheme- Rapid Intensification Index (SHIPS-RII) as the first true RI classification model using linear discriminant analysis (Wilks 2011). They found Brier skill scores (BSS) of 0.1 to 0.13 (dependent on RI threshold). Rozoff and Kossin (2011) expanded the SHIPS-RII framework with Bayesian and logistic regression models, boosting BSS up to 0.22 for logistic regression and 0.15 for Bayesian modeling. Kaplan et al. (2015) detailed the implementation of the Rozoff and Kossin (2011) methodology, noting that consensus models with all three techniques provide BSS values at or below 0.2 consistently. Overall, the blended performance was significantly hampered by large false alarm ratios exceeding 0.6, likely owing to the application of linear techniques to an inherently nonlinear forecast problem. Additionally, SHIPS-RII is trained on average characteristics of predictors within a specific radius of the cyclone center, rather than incorporating spatial information around the TC.

Recent efforts have addressed these challenges by employing artificial intelligence (AI) based prediction models. Grimes and Mercer (2015) utilized spatial fields of base-state and derived meteorological variables from Modern Era Retrospective Analysis for Research and Applications (MERRA – Rienecker et al. 2011) to isolate distinct RI classification predictors through non-parametric hypothesis testing approaches. Resulting predictors, employed in a support vector machine (SVM) designed to predict the first instance of RI, yielded skill scores consistent with Kaplan et al. (2015). Grimes and Mercer (2016) expanded this work, applying rotated principal component analysis (RPCA – Richman 1986, Wilks 2011) over time (T-mode) and space (S-mode) to gridded fields from the Global Forecast System Reforecast datasets (GFSR – Hamill et al. 2013) to isolate fields that are most distinct between RI and non-RI events. Subsequent work utilized a SVM ensemble with reanalysis datasets to forecast the first instance of RI within Atlantic TCs (Mercer and Grimes 2015). They showed the 20th Century Reanalysis (Compo et al. 2011) provided the best forecast skill with BSS near 0.3, a 35% improvement over current operational values noted in Kaplan et al. (2015).

These studies suggested potential for major improvements in TC RI forecasts through the inclusion of machine learning in the forecasting process but were limited by a lack of direct forecast application. Additionally, other machine learning methods (such as random forests – RFs and artificial neural networks – ANNs) have not been tested for RI prediction. This study emulates operational forecast data with GFSR fields and assesses operational performance of machine learning methods through a developed ensemble of SVMs, RFs, and ANNs.

2. METHODOLOGY

Data from 1985–2009 were collected from the National Hurricane Center and SHIPS for every TC in the Atlantic of at least tropical or subtropical depression strength. TC times classified as extratropical were removed from the dataset. All GFSR gridpoints within 19° of the TC center were included as well.

Development of the AI ensemble requires three steps: feature selection, AI ensemble optimization, and the blending of AI ensemble members into a single probabilistic forecast.

2.1 Feature Selection

For the AI to distinguish between RI/non-RI events, it requires examples of input data that are distinct from one another. In this study, these features are selected via permutation testing following a two-step procedure. First, a permutation test is completed on the mean RI and non-RI GFSR layer by taking the gridpoints of each mean field as the basis for the test. Layers

* *Corresponding author address:* Andrew E. Mercer, 108 Hilbun Hall, P. O. Box 5448, Mississippi State, MS 39762-5448. E-mail: a.mercer@msstate.edu

resulting in p-values less than 0.01 are retained for a second permutation test, in which all RI/non-RI events are tested on a gridpoint basis. This method yields points significantly different between RI/non-RI events at $p \leq 0.01$, which are the most distinct spatial features in the database and thus useful in forecasting events undergoing RI.

2.2 AI Ensemble

In addition to feature selection, an optimal AI ensemble is necessary to best classify RI/non-RI events. By tuning numerous parameters of each of the three methods considered (SVMs, RFs, and ANNs), the optimal routine can be identified. Tunable parameters include node and layer selection (for ANNs), tree and branch selection (for RFs), and kernel and cost selection (for SVMs). Each configuration is tested by conducting 500 bootstrap-resampled cross-validation trials, where 85% of the events were withheld for training and 15% withheld for independent testing. Each AI configuration is assessed using Heidke Skill Score (HSS) values, retaining configurations which perform best in at least 10 of 500 trials (41 members).

Once the 41-member ensemble is established, a net probability of RI can be computed. To weight the relative contribution of each individual member to the total probabilistic forecast, cross-validation (1000 bootstraps) performance (based on Heidke skill score) is used. Resulting HSS values for the AI ensemble are on the order of 0.3. Average BSS is 0.12, but some bootstrap members yield performance as high as 0.38, which far exceeds performance in Kaplan et al. (2015).

3. 2017 ATLANTIC HURRICANE SEASON

The AI ensemble was run for each GFSR forecast for the 2017 Atlantic hurricane season to evaluate how well the AI ensemble would translate into a quasi-operational environment. BSS values were calculated globally for each TC event and for each timestep of a given event. Overall, global AI ensemble performance showed positive BSS values for 10 of the 17 (59%) named storms for the 2017 season, with perfect performance for 3 of those storms (Fig. 1). However, the AI ensemble performed the worst on Tropical Storm Don, where one false alarm for an individual timestep greatly influenced the ensemble performance for this event. The AI ensemble had BSS near 0.5 for 4 of the 17 (24%) named storms including Tropical Storm Emily, Hurricane Gert, Hurricane Maria, and Hurricane Ophelia.

Upon evaluation of each forecast timestep, BSS results were influenced by at least one outlying poor forecast in terms of either a miss or a false alarm for the majority of the 2017 events. Despite this, overall AI ensemble BSS performance was skilled, with percentage of BSS values greater than 0.89[1] at each timestep shown in Table 1.

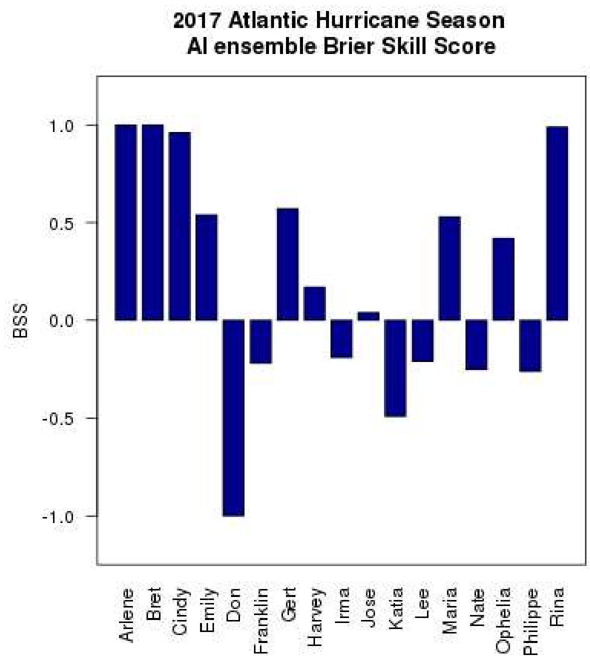

FIG.1. Global AI ensemble performance for the 2017 Atlantic hurricane season. BSS values reflect performance of the 41-member AI ensemble over the lifecycle of each event.

4. CONCLUSIONS

Results for the AI ensemble present a strong complement to the SHIPS-RII. Comparisons for the 2017 Atlantic hurricane season revealed the AI ensemble outperformed SHIPS-RII on individual forecasts for 13 of the 17 named storms (76%). Future work will extend the current dataset through 2016 to update the baseline forecast model. Additionally, a test statistic will be included to evaluate real-time performance of individual AI ensemble members, allowing for further optimization of the AI ensemble.

[1]One false alarm from the 41 ensemble members results in a BSS of 0.89.

TABLE 1. Performance evaluation of the AI ensemble for each timestep for each event. Percentage of BSS greater than 0.89, percent correct, false alarm rate, and probability of detection for each event provided.

TC Event	Percent of BSS values > 0.89	PC	FAR	POD
Arlene	100	1	0	--
Bret	100	1	0	--
Cindy	100	1	0	--
Don	67	0.83	0.17	--
Emily	67	1	0	--
Franklin	63	0.81	0.07	0
Gert	87	1	0	--
Harvey	78	0.89	0.03	0.4
Irma	72	0.78	0.07	0
Jose	88	0.92	0.02	0.2
Katia	38	0.63	0.17	0
Lee	75	0.86	0.05	0
Maria	82	0.92	0.05	0.71
Nate	27	0.47	0.42	0
Ophelia	96	1	0	--
Philippe	75	1	0	--
Rina	100	1	0	--

5. ACKNOWLEDGEMENTS

This work was funded through the NOAA Joint Hurricane Testbed Grant #NA17OAR459014.

6. REFERENCES

Compo, G., and Coauthors, 2011: The twentieth century reanalysis project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1-28.

Grimes, A., and A. Mercer, 2015: Synoptic-scale precursors to tropical cyclone rapid intensification in the Atlantic Basin. *Adv. Meteor.*, **2015**, 17 pp.

Grimes, A., and A. Mercer, 2016: Diagnosing tropical cyclone rapid intensification through rotated principal component analysis of synoptic-scale diagnostic fields. *Chapter 2: Recent Developments in Tropical Cyclone Dynamics, Prediction, and Detection*, Intech publishing, 26-49. *Invited book chapter*

Hamill, T., G. Bates, J. Whitaker, D. Murray, M. Fiorino, T. Galarneau, Y. Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553-1565.

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

Kaplan, J., and Coauthors, 2015: Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical methods. *Wea. Forecasting*, **30**, 1374-1396.

Mercer, A., and A. Grimes, 2015: Diagnosing tropical cyclone rapid intensification using kernel methods and reanalysis datasets. *Procedia Comp. Sci.,* **61**, 422-427.

Richman, M. B., 1986: Rotation of principal components. *J. Climate,* **6,** 293-335.

Rienecker, M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate,* **24**, 3624-3643.

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

Wilks, D., 2011: Statistical Methods in the Atmospheric Sciences. Academic Press, Burlington, MA, 704 pp.