Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 114 (2017) 333–340

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems, CAS October 30 – November 1, 2017, Chicago, Illinois, USA

Atlantic Tropical Cyclone Rapid Intensification Probabilistic Forecasts from an Ensemble of Machine Learning Methods

Andrew Mercer[a]* and Alexandria Grimes[a]

[a]*Department of Geosciences, Mississippi State University, 108 Hilbun Hall, Mississippi State, MS 39762-5448*

Abstract

Atlantic tropical cyclone (TC) rapid intensification (RI) continues to be a major forecasting challenge, with forecast skill scores only about 15% better than climatology. To date, RI forecasts have been completed using linear discriminant analysis (LDA) on predictors optimized for RI forecasts, and no study has directly addressed machine learning's (hereafter AI) capability in forecasting RI. As such, the objective of this study is to quantify the RI predictability using proxy forecast model data and an ensemble of AI methods to generate probabilistic RI forecasts. Atlantic RI events from 1985 to 2011 were retained for all valid times (over water) for each TC, and these cases were used to train an AI ensemble optimized (through three steps) for RI prediction. First, backwards elimination feature selection was used on a blend of the proxy forecast data, predictors from the currently utilized LDA model, and observed TC track information (such as intensity and position) to optimize the predictor suite. Second, numerous configurations of three AI methods (support vector machines [SVMs], artificial neural networks [ANNs], and random forests [RFs]) were tested using bootstrap-based cross-validation to ascertain the best configurations of each AI method. Finally, the best AI configurations were used to generate probabilistic output for RI, weighted by each ensemble member's individual cross-validation performance. Resulting probabilistic forecasts were in line with the current LDA method, though the upper skill limit of the ensemble exceeded 30% improvement over climatology, which far exceeds the current LDA scheme.

Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.

Keywords: Artificial neural networks; support vector machines; random forests; ensemble; probabilistic forecasts; tropical cyclone rapid intensification

* Corresponding author. Tel.: +1-662-325-3915; fax: +1-662-325-9423.
E-mail address: a.mercer@msstate.edu

1877-0509
Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.
10.1016/j.procs.2017.09.036

1. Introduction

Tropical cyclones (hereafter TCs) are important meteorological phenomena owing to their potential for major impacts along coastal areas. Despite their importance, the inherently complex thermodynamic and kinematic processes that drive TC intensification and weakening are poorly forecast by current dynamic weather models. To address this issue, statistical methods (such as the Statistical Hurricane Intensification Predictive Scheme – SHIPS[1]) have been implemented in operational TC intensification forecasts with modest success. Further confounding the forecasting problem is the nature of TCs which undergo rapid intensification (hereafter RI). RI is considerably more challenging to predict, as current operational predictive schemes[3] demonstrate skill scores relative to climatological RI forecasts of only 0.2[3]. This is a critical issue, as most hurricane-strength TCs undergo RI at some point in their life cycle, and all major (category 3 or stronger) storms undergo RI[4]. To allow for better preparation against these events, improved RI prediction is needed.

RI prediction within TCs is hampered by several key issues. First, an agreed upon definition of RI is lacking, as the National Hurricane Center defines RI as a 30-kt increase in sustained peak wind speed in 24 hours, while the National Weather Service defines RI as 42-mb or greater central pressure falls within a 24 hour period. The latent heat processes driving TC intensification are poorly represented in current dynamic weather models owing to their complexity and that thermodynamic processes are more commonly modeled using statistical methods than formal analytic equations. Further, many of the processes related to RI are inherently nonlinear, and current forecast implementations have only focused on linear statistical methods[3,5].

Despite these challenges, Atlantic Basin RI prediction has evolved from a more simplified probabilistic approach[2] which employed exceedance probabilities on five fundamental intensification predictors (many of which were included in the original SHIPS implementation) to more advanced linear discriminant analysis (hereafter LDA) methods[4], (the current operational SHIPS Rapid Intensification Index [SHIPS-RII]). Within this evolution, Kaplan and DeMaria specifically noted the need for implementation of machine learning methods to improve upon their exceedance probability work[1]. A subsequent study by Rozoff et al.[5] was the first effort to address this lack of machine learning effort, but their study utilized Bayesian inference and logistic regression.

There is a notable dearth of machine learning work with RI prediction. Grimes and Mercer[7] attempt to address this issue by revisiting the feature selection problem. Current operational models[3] use field-averaged characteristics of TCs, including relative dry air abundance, ocean heat content, previous 12-hour intensity changes, and satellite imagery. However, since the LDA methods require individual predictors as input, spatial information is averaged out to yield single values for the full TC domain. Grimes and Mercer[7] addressed this by using individual gridpoints within TC-centric spatial domains of Global Forecast System (hereafter GFS) reforecast fields[7], a suitable proxy for operational weather forecast data. Their study found that the predictors with the greatest discrimination power included equivalent potential temperature (which incorporates a measure of total heat within the TC) and several kinematic fields, including vertical wind shear and upper-level divergence. These fields were used in an initial predictive study of machine learning RI forecasting[8] which found modest predictability improvements using observational datasets. Mercer and Grimes[9] advanced the work further, assessing the importance of dynamic weather model resolution on RI and non-RI forecasts. They found that coarser resolution had better predictive skill in a support vector machine (hereafter SVM)[10] as considerable noise was introduced into the model at higher spatial resolutions. They also found that, for their subset of 10 RI events and 10 non-RI events, skill scores for 24-hour RI forecasts exceeded 0.2 consistently, which is a modest improvement over current forecast implementations. However, the limited dataset size likely influenced these results, thus the need for further analysis. Regardless, there is clear potential for further enhancement of RI prediction using machine learning methods.

To address remaining shortcomings, the objective of this current study was to assess RI predictability using a proxy forecast database on the full dataset of Atlantic Basin TC events from 1985-2009. In particular, the primary objective was to quantify improvements in current RI forecast methods by utilizing an ensemble of machine learning methods, which include SVMs, multi-layer perceptrons (MPs)[11], and random forests (RFs)[12]. These objectives were accomplished through two primary research phases. First, robust feature selection identifying those predictors within the forecast database that were most distinct between RI and non-RI environments was completed. Second, model tuning of each of the machine learning methods was done to obtain a 41-member machine learning ensemble from which an RI forecast probability RI could be obtained.

2. Data

For suitable training of machine learning methods, a thorough and robust database of TCs is required. In this study, the HURDAT2[13] database, a dataset maintained by the National Hurricane Center of all Atlantic Basin TCs and their characteristics, was employed. This database allowed for the determination of RI and non-RI timesteps, based on 24-hour wind speed changes. While other studies[1,2,4] have considered multiple RI definitions, given the initial scope of this project, only the 30-kt definition of RI was utilized. Each TC's temporal evolution included multiple timesteps, and each timestep was classified as an RI or non-RI timestep based on the previously mentioned definition. There were two important criteria for establishing these timesteps. First, the TC must have existed for at least 5 HURDAT2 analysis periods (30 hours) so that a 24-hour wind speed change could be measured. Second, the TC's final timestep was no later than at least 24 hours prior to TC landfall, since RI is not predicted once TCs make landfall. These criteria resulted in a database of 949 timesteps, of which 7.4% (70 timesteps) met the criteria of RI.

In addition to the required TC information from HURDAT2, meteorological features were needed to serve as RI predictors for the machine learning methods. Hamill et al.[6] developed the GFS-reforecast (hereafter GFSR) dataset, which provides actual GFS forecasts at 1° global spatial resolution for all days from 1985 to present using a hindcast approach. The GFSR contain numerous meteorological fields, including temperature, geopotential height, u and v wind components, mean sea level pressure, sea surface temperature, and specific humidity. Additionally, several fields, including wind shear, static stability, and equivalent potential temperature, were derived from these base-state fields (italicized in Table 1). The base-state data are provided on 8 vertical levels at 1° global resolution every 24 hours at 0000 UTC. Each reforecast time has a 168-hour forecast (at 6 hour intervals) available for use (though only the analysis time was used for this study). This dataset has several advantages for this type of work. First, the global nature of the dataset ensures data are uniformly distributed over data sparse regions such as the Atlantic Ocean where RI processes are likely to dominate. Second, it serves as a proxy for a true GFS forecast dataset, meaning that any algorithms developed herein will easily transition into a forecasting mode by simply switching the input dataset.

Table 1. GFSR extracted fields (165 gridpoints for each field, 59 total fields). Derived fields are italicized.

Variable name	Vertical levels (mb)
Geopotential height (m)	1000, 925, 850, 700, 500, 300, 200, 100
Temperature (K)	1000, 925, 850, 700, 500, 300, 200, 100
Zonal (u) wind speed (m s^{-1})	1000, 925, 850, 700, 500, 300, 200, 100
Meridional (v) wind speed (m s^{-1})	1000, 925, 850, 700, 500, 300, 200, 100
Specific Humidity (kg kg^{-1})	1000, 925, 850, 700, 500, 300
Mean Sea Level Pressure (Pa)	Surface
Sea Surface Temperature (K)	Surface
Latent Heat Flux (K m s^{-1})	Surface
Sensible Heat Flux (K m s^{-1})	Surface
Convective Available Potential Energy (J kg^{-1})	Surface
Convective Inhibition (J kg^{-1})	Surface
Pressure Vertical Velocity (Pa s^{-1})	850
Static Stability (m^4 s^2 kg^{-2})	925, 850, 700, 500
Equivalent Potential Temperature (K)	1000, 850, 700, 500, 300
Divergence (s^{-1})	200
Vorticity (s^{-1})	700, 500, 200
Vertical Shear (m/s)	850-200 mb layer

In this study, it was expected the GFSR analysis (0-hour) fields would be the most robust and minimize forecast model error (ideally maximizing classification skill), and as such TC centric domains were retained for each timestep. TC centers were established by utilizing the timestep's center point from the HURDAT2 database and finding the local minimum of mean sea level pressure nearest that center point (ensuring the GFSR fields were centered over the GFS initialization of the TC, not the HURDAT2 position, which were often similar but not co-located). Domains of 11° x 15° longitude/latitude were extracted for each timestep, resulting in 165 gridpoints over 59 different data layers (incorporating all three-dimensional fields and surface fields from the GFSR – Table 1). In total, 9735 features were retained from the GFSR for each timestep in the study. In addition to retaining GFSR fields for each timestep, seven SHIPS-RII predictors available for the full study period were also retained as additional features, which have been shown in other studies[2,3,5] to be skillful in discriminating RI and non-RI environments. However, data availability of the SHIPS-RII predictors limited the study size to only 658 timesteps, of which 52 (7.9%) were RI timesteps. Four predictors based on the TC's currently observed characteristics were included as well (TC latitude and longitude, 24-hour wind speed change, and movement speed). Combining the SHIPS-RII features and the GFSR points, a total of 9746 features were available for discrimination on the 658 cases.

3. Methodology

The proper development of the machine learning ensemble required both a robust feature selection procedure and significant tuning of the models. These procedures were completed in a series of three steps, outlined below.

3.1 GFSR Feature Selection by Layer

Given the considerable number of features and possible feature combinations (over 47 million), an initial feature selection procedure was implemented that considered each of the 59 layers individually. Layer averages of RI and non-RI environments were computed, and permutation tests on the 165 gridpoints in each layer average were formulated to identify those layers which possessed the greatest discrimination potential. Layers were retained if their permutation test *p*-values were smaller than 0.01 (significant at the 99% confidence level), which was selected to ensure maximum separation between RI and non-RI groups. This procedure resulted in eleven layers which were significantly different enough to be used in additional feature selection, including 1000 mb and 925 mb specific humidity, sea surface temperature, 100 mb temperature, 1000 mb, 850 mb, 700 mb, and 500 mb equivalent potential temperature, and *u*-component wind at 200 mb, 300 mb, and 500 mb. These eleven layers had an associated 1815 individual gridpoints which were further reduced in a second feature selection method to ensure maximum separability between the RI and non-RI classes.

3.2 Pointwise GFSR Feature Selection

To filter down the reduced feature set to a more reasonable size, a second series of permutation tests was conducted on the GFSR individual gridpoints for those layers identified in section 3.1. However, owing to the highly unbalanced nature of the RI and non-RI data, permutation tests were conducted using a bootstrapping methodology in which a random subset of non-RI event data was retained that was equal in size to the RI event data size (70 cases). The procedure was then repeated 500 times to retain 500 *p*-values for each tested feature. The percentages of those *p*-values which were smaller than 0.01 (significant at the 99% confidence level) were retained for all features, and those percentages were used as a secondary cutoff for feature selection. Features were only retained if 99% of their associated bootstrapped permutation tests met the significance criterion, which resulted in only 6 GFSR points being kept. These 6 points were *u*-wind components at the 200 and 300 mb levels (5 points were at the 200 mb level), which were highly unique between RI and non-RI events. Interestingly, further analysis of the *u*-wind component at 200 mb suggests a stronger anticyclonic circulation at 200 mb in the RI events and possibly increased outflow, which means these events may be associated with TCs with increased high magnitude wind speeds aloft (and thus minimal vertical wind shear, which is a known factor for TC intensification)[14]. Maps illustrating those differences at 200 mb are provided in Fig. 1 below.

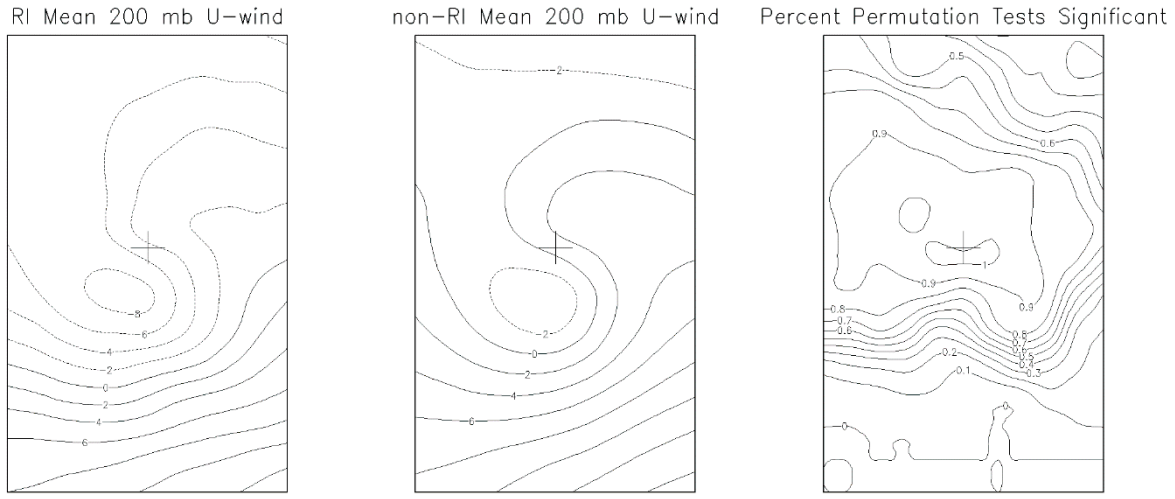

Fig. 1. Mean *u*-wind component for all RI timesteps (left panel), all non-RI timesteps (middle panel), and the percentages of permutation tests that were significantly different ($p < 0.01$) between the two. The cross in the center indicates the TC center for each panel.

3.3 Creation of Machine Learning Ensemble

The work in sections 3.1 and 3.2 resulted in 6 GFSR predictors that were combined with the 7 retained SHIPS-RII predictors and the 4 other HURDAT2-derived predictors to result in a total of 17 features used in the creation of a machine learning ensemble. These 17 features were used as input into three machine learning methods (MPs, RFs, and SVMs), from which an ensemble product was created. Each machine learning method requires optimization of its tuning parameters[10,11,12], which was done using a bootstrapping-based cross-validation routine, in which 20% of the timesteps were randomly withheld as independent testing and the remaining 80% were used to train the method. Optimization was based on maximizing the Heidke skill score (HSS)[15], which measures the quality of an RI forecast relative to random chance (which has been adjusted to reflect the unbalanced nature of the data) and is a common measure of performance in meteorology. Higher HSS values support better classifications, while negative HSS values suggest random chance is a better model. Each machine configuration was cross-validated 300 times in a pairwise manner (ensuring all training and testing sets were consistent for all configurations). The likelihood of overfitting all methods was reduced by the inclusion of this 300-iteration cross-validation methodology.

For the machine learning optimization, the tested tuning parameters included:

- For SVMs – cost functions of 1,10,100, and 1000; radial basis functions with $\gamma = 0.01$, 0.1, and 0.05 as well as polynomial kernel functions with degrees 2 to 5; a total of 28 permutations
- For MPs – stopping criteria of 30000, 50000, and 100000 epochs; hidden node counts of 10 to 13; hidden layer counts of 1 to 4; a total of 48 permutations
- For RFs – number of trees to grow of 100, 200, 300, 400, and 500; cutoff criteria of 0.1, 0.2, 0.3, 0.4, and 0.5; and predictors used at each leaf split between 4 and 8; a total of 125 permutations

Note that the cutoff criteria for RFs is the method by which the unbalanced RI/non-RI datasets was addressed. Similarly, weights were used with the SVMs and MPs based on the relative frequency of RIs and non-RIs for each cross-validation iteration. Despite these issues and the availability of many more possible configurations, a wide spread of model performance was observed from these combinations, giving high confidence in the HSS results used to include model configurations in the ensemble.

The final step was to establish the best performing machine learning configurations for inclusion in the ensemble. For each of the 300 bootstrap iterations, the highest HSS configuration for each method (SVMs, RFs, or MPs) was noted, and those configurations which performed best in 10 or more of the 300 bootstrap iterations were

retained. This resulted in 5 SVM members, 18 RF members, and 18 MP members. These members, as well as their associated HSS statistics, are provided in Tables 2-4. Once the 18-member ensemble was established, it was cross-validated as a full ensemble on the same 300 events to evaluate probabilistic output. However, since the HSS statistic is based on discrete RI/non-RI forecasts, a more appropriate statistic, the Brier skill score (BSS)[15], was used to measure probabilistic performance. Probabilistic output was formulated in two ways. First, the sum of RI forecasts was divided by the total number of forecasts to yield RI forecast frequency, which is a reasonable first-guess probability estimate. Second, each ensemble member was weighted by its previously determined HSS, such that those members whose performance was better were weighted more heavily in computing the probability. These methods yielded comparable results, as shown below.

4. Results

The initial results in Tables 2-4 show interesting performance characteristics within the different ensemble members. With the exception of two ensemble members (SVM1 and RF9), HSS values were consistently above 0.2, which is still quite poor and comparable to what is observed in current operational forecasts[2]. The best performing ensemble member was SVM4, which is consistent with previous work[9] which suggests the use of the radial basis kernel function with relatively small costs yields optimal SVM classification performance. Interestingly, the MP members all performed very similarly, though their performance was not consistent, as each member's predictions were correlated between 0.51 and 0.6 and only 5% of all MP-predicted RIs were agreed upon by all 18 MP ensemble members. Interestingly, there was a notable, consistent improvement in performance when looking at RFs with 200 trees grown, and the RF ensemble members showed little sensitivity to the cutoff selection, suggesting the unbalanced nature of the data was not necessarily degrading RF performance.

While individual ensemble member performance was a useful initial measure of model skill, the ensemble performance was of primary interest to the objectives of this study. Ensemble performance was characterized using both contingency statistics and BSS values, comparing the BSS value against the SHIPS-RII baseline of 0.13[2]. To assess performance, a 1000 iteration bootstrap cross validation was completed, where 80% of cases were withheld to train each ensemble member and the remaining 20% were tested upon. This approach yielded 1000 bootstrap replicates of BSS and other contingency statistics to evaluate overall ensemble performance. This approach was repeated for both the initial ensemble where all members were weighted equally and the weighted ensemble probabilities (based on individual member performance characterized in Tables 2 and 3). Additionally, RI/non-RI ensemble decisions formulated by popular votes (in the case of a tie, an RI was selected), and these decisions were used to formulate ensemble contingency statistics. Results are presented in Fig. 2.

Fig. 2. Ensemble performance bootstrap replicates for BSS values for the weighted ensemble (left panel), the unweighted ensemble (center panel), and the bias-corrected HSS (right panel). The red line shows the operational baseline BSS value for the SHIPS-RII of 0.13.

Table 2. SVM ensemble member configurations and performance statistics. Larger HSS values mean better performance.

Member	Kernel	Cost	γ-value	HSS
SVM1	Poly-2	1	0.05	0.183
SVM2	RBF	1	0.05	0.270
SVM3	RBF	10	0.05	0.306
SVM4	RBF	1	0.1	0.319
SVM5	RBF	10	0.1	0.277

Table 3. Same as Table 2, but for RFs (left side) and MPs (right side)

Member	Trees	Predictors	Cutoff	HSS	Member	Layers	Nodes	Epochs	HSS
RF1	100	4	0.2	0.258	MP1	4	10	100000	0.308
RF2	100	4	0.3	0.248	MP2	2	11	100000	0.309
RF3	100	5	0.2	0.263	MP3	2	12	100000	0.310
RF4	100	5	0.3	0.263	MP4	1	8	100000	0.308
RF5	100	6	0.2	0.265	MP5	4	8	100000	0.310
RF6	100	6	0.3	0.264	MP6	4	10	30000	0.309
RF7	100	6	0.4	0.265	MP7	1	11	30000	0.310
RF8	100	6	0.5	0.214	MP8	1	12	30000	0.309
RF9	100	6	0.6	0.131	MP9	3	8	30000	0.309
RF10	100	7	0.3	0.267	MP10	4	8	30000	0.306
RF11	100	7	0.4	0.274	MP11	2	9	30000	0.314
RF12	100	5	0.2	0.265	MP12	4	9	30000	0.309
RF13	200	5	0.4	0.259	MP13	1	10	50000	0.310
RF14	200	6	0.2	0.267	MP14	3	11	50000	0.307
RF15	200	6	0.3	0.264	MP15	4	12	50000	0.306
RF16	200	7	0.2	0.267	MP16	1	8	50000	0.313
RF17	200	7	0.3	0.267	MP17	1	9	50000	0.309
RF18	200	7	0.4	0.277	MP18	2	9	50000	0.313

The ensemble performance was consistent with what is currently observed in operational forecasts, which was a bit unexpected given the anticipated improvements offered from the machine learning methods. BSS values consistently fell around the 0.2 operational threshold, with a median bootstrap BSS for both the weighted and unweighted ensembles falling slightly below that threshold (0.103 for weighted, 0.098 for unweighted). The weighting scheme did offer a slight amount of improvement, which suggests further experimentation with the weighting scheme could result in better overall predictions. Additionally, the upper confidence interval for both the weighted (0.383) and unweighted (0.380) ensembles offer significant improvements over operations. Evidently, more work is needed to optimize the ensemble and increase the median BSS values towards this upper limit.

The contingency statistics revealed some of the important issues that arose from using the ensemble. First, the bias statistics had 95% confidence intervals falling over 0.214 and 1.5, with a median of 0.62. This value clearly suggests the voting scheme is significantly underforecasting RI events, which is an expected issue owing to the unbalanced nature of the data. Bias correction was attempted by modifying the voting threshold to suggest an RI occurred with only 35% of the ensemble voting RI, and the remaining statistics were updated to the bias-corrected ensemble (including the HSS panel in Fig. 2). After bias correction, it was clear the primary issues resided around a high false alarm ratio (95% confidence intervals of 0.3 to 0.9), which makes sense given the voting threshold selected. Similarly, probability of detection values were appreciably lower than the false alarm ratio, with values falling between 0.111 and 0.714. Finally, the critical success index values fell between 0.059 and 0.429. These contingency statistics

results suggest only modest improvements relative to climatology and performance values that are in line with what is currently observed in operational RI forecasts. Overall, the resulting performance statistics were in line with current operational RI forecasts, but do not yet offer the significant bump in performance that was expected at the outset of this study.

5. Conclusions

The goals of this project were to ascertain predictability of RI environments using an ensemble of machine learning methods. Overall, results were mixed, as performance remained comparable to current operational standards. However, the upper limits of performance far exceeded that which is currently seen operationally, suggesting that additional work may result in improved predictability of RI. It is expected that additional timesteps added to the ensemble will result in additional learning and improved performance.

Future work will revisit the feature selection methodology, parsing through each of the 9745 GFSR gridpoints individually to assess their relative value (as was done with the limited 1815 set described previously). Data reduction methods such as rotated principal component analysis, which was considered in this study but not explored fully, will be added to attempt to deal with the volume of predictors. Genetic algorithms will also be implemented to assist and possibly improve upon the feature selection procedures outlined in this study. Finally, the framework will be transitioned into an operational product that could be used by the National Hurricane Center to complement the current SHIPS-RII model used operationally.

6. Acknowledgements

We would like to thank the director of the Northern Gulf Institute, Robert Moorhead, for financial support in the submission of this manuscript. This work was funded from NOAA-OAR grant number #NA11OAR4320199.

7. References

1 DeMaria M, Kaplan J. A statistical hurricane intensity prediction scheme for the Atlantic and Pacific basins. *Wea. Forecasting* 1994;**9**:209-220.

2 Kaplan J, Rozoff, C, DeMaria M. Sampson C., Kossin J, Velden C, Cione J, Dunion J, Knaff J, Zhang J, Dostalek J, Hawkins J, Lee T, Solbrig J. Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical methods. *Wea. Forecasting* 2015;**30**:1374-1396.

3 Kaplan J, DeMaria M. Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic Basin. *Wea. Forecasting* 2003;**25**:1093-1108.

4 Grimes A, Mercer A. Synoptic-scale precursors to tropical cyclone rapid intensification in the Atlantic Basin. *Adv. Meteor.* 2015;**1**:17 pp.

5 Rozoff C, Velden C, Kaplan J, Kossin J, Wimmers A. Improvements to the probabilistic prediction of tropical cyclone rapid intensification resulting from the inclusion of passive microwave observations. *Wea. Forecasting* 2015;**30**:1016-1038.

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

7 Grimes A, Mercer A. Diagnosing tropical cyclone rapid intensification through rotated principal component analysis of synoptic-scale diagnostic fields. In: Lupo, A, editor. *Recent Developments in Tropical Cyclone Dynamics, Prediction, and Detection*. Intech publishing; 2016. p. 26-49.

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

9 Mercer A, Grimes A. Importance of model resolution on discriminating rapidly and non-rapidly intensifying Atlantic basin tropical cyclones. *Procedia Comp. Sci.* 2015;**95**:223-228.

10 Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*. Cambridge University Press; 2000.

11 Haykin S. *Neural networks and learning machines.* Pearson Prentice Hall, 1999.

12 James G. *An introduction to statistical learning*. New York, Springer, 2013.

13 Landsea C, Franklin J. Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.* 2013;**141**:3576-3592.

14 Frank W, Ritchie E. Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes. *Mon. Wea. Rev.* 2001;**129**:2249-2269.

15 Wilks D. *Statistical Methods in the Atmospheric Sciences, 3rd ed.* Academic Press, 2011.