Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 140 (2018) 359–366

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference with Theme: Cyber Physical Systems and Deep Learning, CAS 2018, 5 November – 7 November 2018, Chicago, Illinois, USA

Multidimensional Kernel Principal Component Analysis of False Alarms of Rapidly Intensifying Atlantic Tropical Cyclones

Andrew Mercer[a,b*], Alexandria Grimes[a,b], Kimberly Wood[a,b]

[a]*Department of Geosciences, Mississippi State University, Mississippi State, MS 39762-5448*
[b]*Northern Gulf Institute, Mississippi State University, Mississippi State, MS 39762*

Abstract

Previous work investigating rapid intensification (RI) predictability within Atlantic tropical cyclones (TCs) has revealed numerous challenges in identifying parameters useful in predicting RI. In particular, false alarm RI forecasts are notably high, driving down forecast skill despite higher RI detection rates. RI forecast improvement will be achieved by identifying spatial patterns distinct to RI false alarm cases that help distinguish them from correctly forecast RI events. A previously developed machine learning ensemble was used to hindcast on a database of Atlantic TCs spanning 1999 – 2016 to identify false alarm RI forecasts, as well as correct RI forecasts. Each set of cases (correct and false alarm) was used to develop kernel principal component (KPC) based composites of important meteorological features within RI TCs to facilitate meteorological comparisons against correctly forecast events. Optimality was determined using a simple metric based on a cluster analysis silhouette and compared against a baseline of traditional k-means cluster analysis, hierarchical analysis, and rotated PC analysis. Once optimal cluster configurations were identified for each meteorological feature, events within each cluster were averaged to yield composites, which in turn were used to facilitate meteorological comparisons between false alarm and correctly forecast RI events.

Keywords: Kernel principal component analysis, cluster analysis, composite analysis, tropical cyclones, rapid intensification

1. Introduction

Classification problems within atmospheric sciences continue to weigh the benefits of adequately detecting occurrences of important meteorological events (e.g. tropical cyclone rapid intensification – RI) at the expense of increased false alarm ratios [1]. Recent advances in RI forecast algorithms (primarily the Statistical Hurricane

10.1016/j.procs.2018.10.303

Intensity Prediction System Rapid Intensification Index – SHIPS-RII [2]) for Atlantic RI have maintained steadily high probabilities of detection, where probability of detection (POD) is defined as the percentage of RI occurrences identified by the forecast algorithm. In particular, the SHIPS-RII had a 97.4% detection rate for RI events during the 2017 Atlantic hurricane season (18 tropical cyclones (TCs), 6 of which were major hurricanes). However, this result came at the expense of a high false alarm ratio (FAR; 85.4%), defined as the percentage of RI forecasts in which no RI occurred. This elevated FAR is primarily responsible for decreased forecast skill within Atlantic TC forecasts (SHIPS-RII produced a Heidke Skill Score – HSS - of 0.097, where 1 is a perfect forecast and 0 has no skill). These performance issues exacerbate public communication challenges, as public inactivity during natural disaster situations increases with increasing FAR [3].

Recent developments in RI prediction [4-6] have produced a machine-learning RI classification ensemble to issue operational RI forecasts, currently under testing by the National Hurricane Center (NHC). Each ensemble member issues a single RI/non-RI class forecast and contributes to a global RI probability for the full ensemble. A preliminary version of this ensemble [6] yielded a POD of 31.6% for the 2017 Atlantic hurricane season, though it did reduce the FAR to 74.4%, which yielded an increased HSS of 0.193.

While the higher HSS was encouraging, additional feature selection and ensemble development was implemented to further improve performance. In this update, we considered TC events spanning 1999-2016 (a total of 5409 TC timesteps) and limited predictors to the SHIPS-RII and NHC HURDAT2 database [7]. The optimal feature set was determined using a leave-one-year out cross-validation approach with logistic regression and resulted in 7 features (24-hour peak wind gust change, 6-hour peak wind gust change, an ocean heat content predictor, a cloud-top brightness temperature predictor, average 200 mb zonal wind speed, 200 mb divergence, and a shear predictor). Once the optimal feature set was identified, numerous support vector machine (SVM), random forest (RF), and backpropagation neural network (NNET) configurations were cross-validated to identify ensemble membership (through maximizing HSS). As each ensemble member (Table 1) yielded an individual RI forecast, a global ensemble RI probability was obtained by training NNET, SVM, and RF configurations on the ensemble output. The optimal global probability configuration that resulted was a 2nd degree polynomial SVM with a cost of 1 and a γ of 0.05. This improved ensemble still produced output with FAR values exceeding 0.5, showing the need for further investigation into the false alarm problem.

The primary objective of this research is to identify meteorological characteristics associated with false alarm RI events and compare them with RI events correctly forecast by the previously described machine-learning ensemble. Composite fields will be created by testing several clustering algorithms to identify optimal groupings of correctly forecast RI events (hereafter referred to as hits) and false alarm events. Meteorological composites, based on averaging the resulting clusters, will be analyzed for differences to identify unique patterns associated with hits and false alarms. These revelations will inform future machine-learning ensemble development to potentially reduce false alarms while maintaining high detection rates.

2. Data and Methodology

During the training of the machine learning ensemble (described previously), each TC timestep was used for independent testing. Each timestep where half the ensemble members issued a false alarm forecast (RI predicted when no RI was seen in the observation wind field) was retained as a false alarm (368 total), while each instance where at least half of the ensemble members issued a correct RI forecast was added to a database of hits (110 total). These cases were used to formulate composites, as described below.

2.1 Composite Datasets

To quantify the TC environment, the meteorological composites required a three-dimensional continuous dataset. Additionally, a dataset that emulates an operational forecasting environment was ideal, as this type of data would include forecast errors and biases inherent in real-time TC forecasts. Such datasets are only available from archived historical numerical weather prediction databases, and for this project the Global Forecast System Final analysis (FNL – [8]) fields were utilized.

Table 1. Machine-learning ensemble members (24 total). SVMs provide the associated kernel (RBF for radial basis function and Poly for a polynomial kernel with its associated degree) and cost. RFs with a number of trees and features used to formulate each tree provided. NNETs refers to a single layer backpropagation neural network with a number of hidden nodes and decay rate provided.

SVMs	Kernel	Cost	**RFs**	Trees	Features	**NNETS**	Hidden Nodes	Decay
SVM1	RBF	1000	**RF1**	5000	3	**NNET1**	2	0.0025
SVM2	RBF	1000	**RF2**	15000	2	**NNET2**	3	0.0025
SVM3	RBF	1000	**RF3**	15000	3	**NNET3**	2	0.005
SVM4	Poly-2	1000	**RF4**	15000	4	**NNET4**	3	0.005
SVM5	Poly-3	1000	**RF5**	15000	5	**NNET5**	2	0.0075
SVM6	Poly-3	10	**RF6**	25000	4	**NNET6**	4	0.0075
SVM7	Poly-4	100				**NNET7**	2	0.01
SVM8	Poly-4	10				**NNET8**	2	0.0125
						NNET9	3	0.0125
						NNET10	3	0.015

These fields are updated and corrected versions of traditional Global Forecast system (GFS) analysis fields, which are global model-derived analysis fields generated every 6 hours for operational forecast purposes. The FNL data, which span the 1999-2016 timeline of the TC databases described above, are provided on a 1° x 1° latitude-longitude grid with 26 vertical levels spanning 1000 mb to 10 mb. All base-state meteorological fields (e.g. temperature, pressure, wind component magnitudes, humidity) are provided at all levels, as well as some derived fields (e.g. precipitable water, vorticity).

A variety of meteorological research studies have shown the importance of upper level wind fields (primarily divergence at 200 mb), vertical shear, and low and mid-level thermodynamics in diagnosing the probability of RI [2,9-11]. To quantify these TC characteristics, 850 mb and 500 mb equivalent potential temperature θ_e [12], 850-200 mb wind shear (where shear magnitude is determined as the vector difference in the winds at both levels), and 200 mb wind field divergence were computed from the FNL fields. θ_e is widely considered to be a good estimate of the thermodynamic potential energy of the atmosphere [12] as it incorporates heat added by phase change of all water vapor into the column (thereby incorporating atmospheric humidity) and an adiabatic compression of the atmospheric parcel to a reference pressure level (1000 mb).

Additionally, composite analyses benefit from Lagrangian reference frames (those which follow the TC) to ensure minimal damping of important features due to the compositing methodology. To create storm-centric domains, the storm center was determined by first identifying the National Hurricane Center's storm center estimate from their Hurricane Databases (HURDAT2) dataset [7], and then searching the mean sea level pressure fields (using a 300 km search radius) within the associated FNL dataset (at the same observation time) for a local minimum in pressure, thereby removing any model initialization errors in initial storm position. Once centers were established, 19° x 19° latitude longitude storm-centric grids (361 gridpoints, roughly 1000 km x 1000 km domains) were retained for all tested meteorological variables for use in the composite analysis.

2.2 Composite Methodology

Meteorological research has employed a variety of composite analysis techniques, from simple averaging [13] to hierarchical and k-means clustering [14] and more sophisticated methods such as k-means clustering on rotated principal components (RPCA) [15]. Recent work [16] has shown benefits from kernel principal component analysis (KPCA) combined with k-means cluster analysis owing to its relaxation of linearity assumptions and its configurability. This study assessed the benefits of KPCA over traditional methods by considering four clustering methods (hierarchical clustering using Wards method, k-means clustering, RPCA with k-means clustering on the loadings, and KPCA with k-means clustering on the loadings).

Prior to separating the data into hit and false alarm subsets, the full FNL database (i.e. all 5409 Atlantic TC timesteps) was standardized by computing standard anomalies by gridpoint (which revealed spatial locations with abnormally high or low magnitudes for each variable). Hit and false alarm databases were retained from these standardized FNL fields.

As 8 total datasets were considered independently, cluster analyses were completed on each standardized database to identify meteorological characteristics unique to each event type (hit or false alarm). All previously described clustering methods were tested, including hierarchical clustering (using Ward's method and a Euclidean distance matrix) and k-means cluster analysis. Additionally, both RPCA and KPCA were implemented, where the optimal number of retained PCs was determined using North's test [17]. KPCA also required the selection of an optimal kernel similarity matrix, of which two were tested: a radial basis function (RBF – equation 1) and a polynomial kernel (equation 2).

$$k(x,y) = \exp\left[-\frac{\left||x-y|\right|}{2\sigma^2}\right] \tag{1}$$

$$k(x,y) = (xy+1)^d \tag{2}$$

Here, x and y are mapped vectors to compute a kernel matrix. RBF kernels were tuned by altering σ (Gaussian spread) values (ranging from 25 to 500 at intervals of 25 for a total of 20 possible configurations), and polynomial kernels were altered by modifying the polynomial's degree d (ranging from 2 to 5 for a total of 4 configurations). The polynomial KPCA cluster analysis consistently provided poorly dispersed clusters (e.g. the 200 mb divergence 4th degree polynomial false alarm grouping yielded groups of 30, 5, and 333 events), which is too similar to a simple mean composite of all cases and was not ideal. As a result, we only used RBF kernels in the KPCA.

Given the large number of clustering methods tested, a metric used to determine the optimal grouping was required. The silhouette coefficient (SC – [18]) provides a unique measure of cluster analysis success by simultaneously assessing cluster quality and revealing misclustered events. Values of 1 for the SC suggest maximum separation between clusters and minimal intracluster distance (a high-quality clustering), while negative SCs imply cohesion among the cluster is larger than separation between clusters, meaning the event is likely misclustered. As a global cluster analysis performance metric was needed to optimize the clustering, the average SC for all events in a given cluster analysis was scaled by the percentage of correctly clustered events. This metric would be notably small for analyses in which a large percentage of events were misclustered, which would typically be associated with a reduced average SC. Resulting graphs of the scaled SC by clusters retained (e.g. Fig. 1) revealed the optimal configuration for each meteorological variable and each database (false alarms and hits – Table 2). The results were consistent with previous work [16], showing KPCA to be the preferred clustering method for all 8 considered databases.

3. Composite Results

The selected analysis methods produced numerous clusters (and thus composites) for each variable and each event type (either false alarm or hit – Table 2). For the sake of brevity, the hit database composite with the highest membership (Table 2) is shown (Figs. 2-5, left panel), as well as the false alarm composite with the highest positive correlation to the hit composite (thereby the most spatially consistent with the hit field – right panels of Figs. 2-5).

The composite fields revealed important differences in the lower-level (850 mb) θ_e fields, including a notable θ_e gradient that spans west – east in the false alarm composites that was largely non-existent in the hit composite fields. In fact, all false alarm clusters but 2 and 3 contained this notable θ_e gradient, and clusters 2 and 3 showed θ_e values up to 1 standard deviation below average (thereby containing less thermodynamic potential energy). Interestingly, several of the 850 mb θ_e hit composites did not show elevated θ_e values but instead revealed patterns as much as 1 standard deviation below the mean (not shown). These patterns, while surprising, were much less common (only 28 of the 110 hit events) than the elevated θ_e pattern shown in the most common hit pattern (left panel of Fig. 2). These results strongly suggest most false alarm events are characterized by a strong horizontal 850 mb θ_e gradient and/or below average θ_e values.

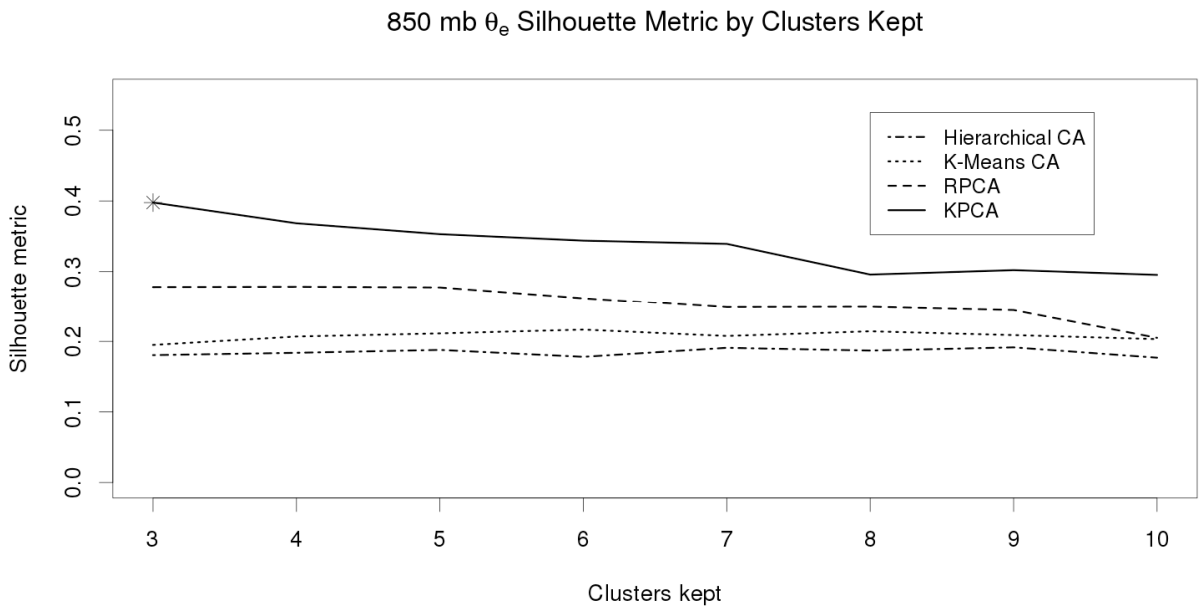

Fig. 1. 850 mb θ_e silhouette coefficient metrics by clusters kept for the false alarm databases. The star shows the optimal configuration. Similar figures were generated to determine the optimal cluster configurations for each dataset.

Table 2. Optimal cluster analysis configuration for each dataset. All databases yielded optimal silhouette metric results using KPCA with varying RBF σ values (given below) with associated SC metrics and cluster sizes. The list of numbers under cluster sizes shows the number of clusters in each composite set. Note that the cluster order is arbitrary. A column identifying the variance explained by the KPCA and the kept KPCs is provided as well.

False Alarms					
Variable	σ	KPCs Kept	Variance Explained	Silhouette Metric	Cluster Sizes
850 mb θ_e	25	6	0.768	0.398	57, 22, 34, 35, 61, 31, 31, 34, 23, 40
500 mb θ_e	75	4	0.822	0.424	166, 59, 143
850-200 mb shear	75	3	0.626	0.417	65, 241, 44, 45
200 mb divergence	500	8	0.526	0.245	29, 51, 35, 75, 46, 54, 43, 35
Hits					
Variable	σ	KPCs Kept	Variance Explained	Silhouette Metric	Cluster Sizes
850 mb θ_e	25	6	0.834	0.489	3, 15, 60, 10, 22
500 mb θ_e	200	3	0.857	0.550	27, 68, 15
850-200 mb shear	75	3	0.772	0.604	14, 81, 15
200 mb divergence	175	6	0.501	0.306	5, 25, 63, 6, 11

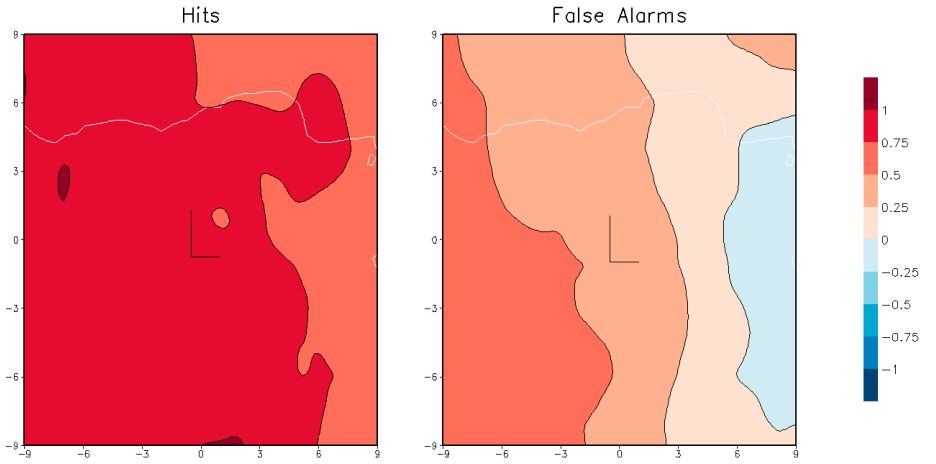

Fig. 2. 850 mb θ_e composites for the most common (N=60) hit pattern (left) and the most highly correlated (r = 0.48) false alarm composite pattern (right). Contours are standard anomalies.

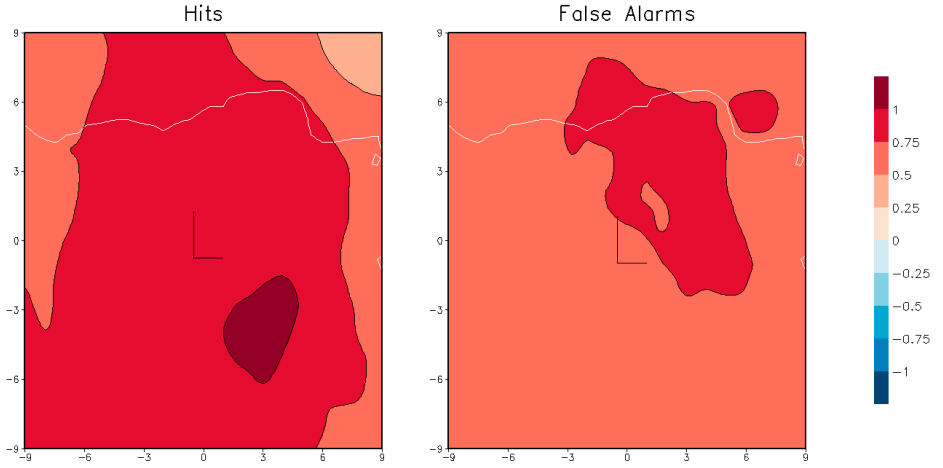

Fig. 3. 500 mb θ_e composites for the most common (N=68) hit pattern (left) and the most highly correlated (r = 0.22) false alarm composite pattern (right). Contours are standard anomalies.

The θ_e results at 500 mb were less revealing (Fig. 3). The most common hit pattern (N=68 – left panel of Fig. 3) showed elevated θ_e patterns throughout the domain, while the other two hit clusters showed similar θ_e patterns that were below average (not shown). This result suggested no prevalent spatial pattern in θ_e associated with hit events but instead revealed domain-wide elevated or reduced θ_e values. The false alarm composite patterns showed similar results to the hits, with one dominant pattern that shows elevated (albeit slightly less elevated) θ_e values (Fig. 3, right panel), and the remaining composites also showing cooler than average θ_e values and no discernable spatial patterns within the fields. These limited differences suggest using mid-level thermodynamics as predictors in RI forecasting schemes could increase false alarms, thus removing these fields may improve the false alarm problem.

The 850-200 mb shear patterns (Fig. 4) were similar, both showing below-average vertical shear values in the most common hit pattern and most highly correlated false alarm pattern. This is not surprising as low-shear environments are known to support RI [2]. However, one hit composite (cluster 1, N=14) showed a strong east to west shear gradient, with elevated shear on the eastern side of the TC domain. A similar false alarm composite (cluster 1, N=65) showed the same gradient, but flipped, suggesting shear on the eastern side of the domain is more conducive to an

accurate machine learning RI forecast and on the west side can lead to a false alarm forecast. Physical reasons for this discrepancy will be considered in future work.

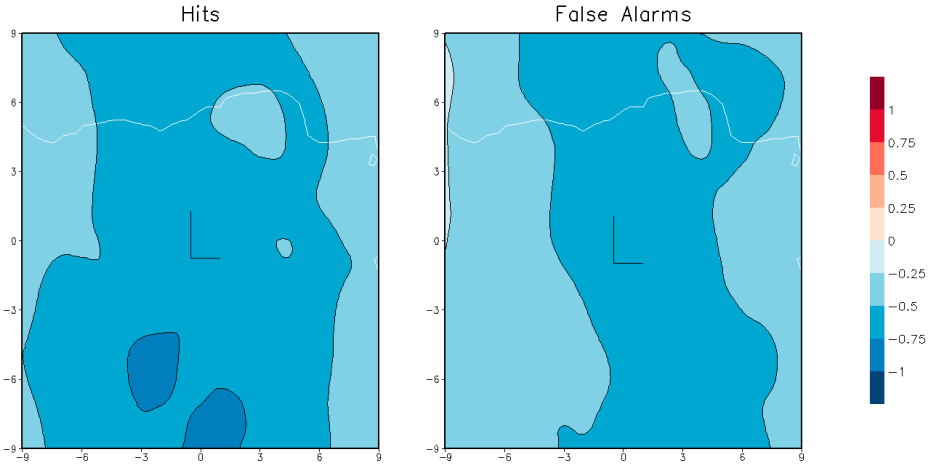

Fig. 4. 850-200 mb vertical shear composites for the most common (N=81) hit pattern (left) and the most highly correlated (r = 0.44) false alarm composite pattern (right). Contours are standard anomalies.

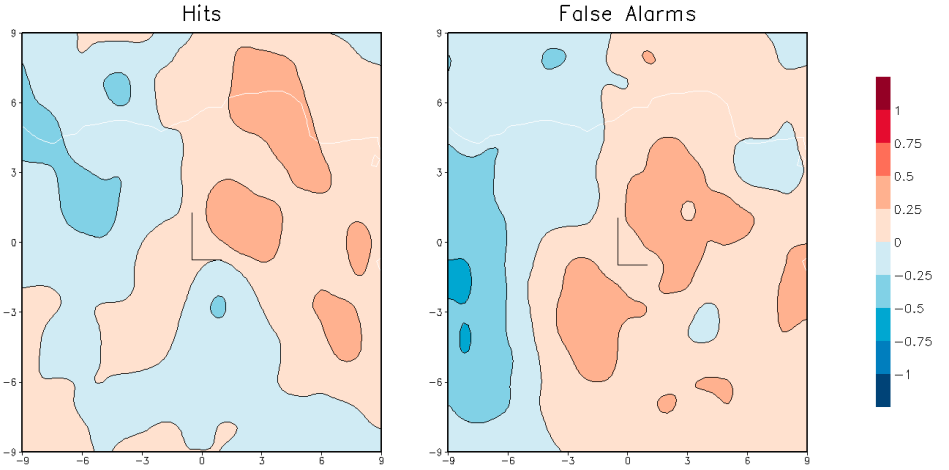

Fig. 5. 200 mb divergence composites for the most common (N=63) hit pattern (left) and the most highly correlated (r = 0.49) false alarm composite pattern (right). Contours are standard anomalies.

The 200 mb divergence patterns (Fig. 5) revealed very little in terms of spatial pattern differences, as both hit composites and false alarm composites were quite noisy (which is expected in a divergence field). However, the fifth hit divergence composite field (N=11) showed a coupling of above and below average divergence, with a similar pattern showing up in the third false alarm composite (N=35), but with higher magnitude negative divergence regions. The pattern's orientation suggests RI is being forecast when upper level stretching deformation is observed, but the orientation of that deformation pattern may help distinguish hits from false alarms in cases with high magnitudes in the upper levels. As this result was relatively uncommon, increased understanding of kinematic patterns may help reduce false alarms in some instances. However, improved representations of low-level thermodynamics are likely key for reducing the false alarm problem within Atlantic RI forecasts.

4. Summary and Conclusions

False alarms within Atlantic TC RI forecasts continue to be a major challenge and contribute to a significant reduction of forecast skill. Improvements to RI forecasts require better understanding of what types of variables and

patterns are associated with false alarm RI forecasts. The objectives of this study were to identify spatial patterns and variables that help distinguish false alarm RI predictions from correct RI forecasts to reduce this false alarm problem.

False alarm and hit events were determined using a machine learning ensemble currently in testing for operational forecasts of RI in Atlantic TCs. In total, 368 false alarm events and 110 hit events were identified. Four meteorological characteristics (850 mb and 500 mb θ_e, 850-200 mb shear, and 200 mb divergence) were retained, and each variable and database was clustered using a variety of clustering methods (8 total datasets resulted). By using the SC metric to identify the optimal cluster analysis method, we found that KPCA continually outperformed all other considered clustering techniques from previous work (Table 2). Once the best clustering was found, we averaged all members in each cluster to create composite fields. The resulting patterns (Figs. 2-5) revealed that low-level thermodynamic fields showed the greatest distinctions between hits and false alarms, notably the existence of a horizontal θ_e gradient in false alarms absent from hit events. Other notable kinematic results revealed major distinctions in some of the smaller clusters, suggesting limited false alarm reduction is possible by considering kinematics, but most effort should focus on low-level thermodynamics.

Future work will analyze additional FNL fields to identify other notable pattern differences between hits and false alarms. We will then incorporate important findings into the existing machine learning ensemble for RI forecasts. Ultimately, improved RI forecast skill by reduction of false alarms is the primary goal of this research, and this study is a first step towards that outcome.

Acknowledgements: This research was funded by NOAA-NA17OAR4590140.

References

[1] Wilks, D. (2011) *Statistical Methods in the Atmospheric Sciences*. Academic Press, Burlington, MA, 704 pp.

[2] Kaplan, J., M. DeMaria, and J. Knaff. (2010) "A revised tropical cyclone rapid intensification index for the Atlantic and Eastern North Pacific basins." *Weather and* Forecasting, **25**: 220-241.

[3] Trainor, J. (2015) "Tornadoes, social science, and the false alarm effect." *Weather, Climate, and Society*, **10**(3): 333-352.

[4] Grimes, A., and A. Mercer. (2015) "Synoptic-scale precursors to tropical cyclone rapid intensification in the Atlantic Basin. *Advances in Meteorology*, 17pp.

[5] Mercer, A., and A. Grimes. (2015) "Diagnosing tropical cyclone rapid intensification using kernel methods and reanalysis datasets." *Procedia Computer Science*, **61**: 422-427.

[6] Mercer, A., and A. Grimes (2017) "Atlantic tropical cyclone rapid intensification probabilistic forecasts from an ensemble of machine learning methods." *Procedia Computer Science*, **114**, 333-340.

[7] Landsea, C., and J. Franklin. (2013) "Atlantic hurricane database uncertainty and presentation of a enw database format. *Monthly Weather Review*, **141**: 3576-3592.

[8] National Center for Atmospheric Research Data Support Center (2018) "NCEP FNL Operational Global Tropospheric Analyses, continuing from July 1999." Available at http://rda.ucar.edu/datasets/ds083.2.

[9] Nolan, D., Y. Moon, and D. Stern. (2007) "Tropical cyclone intensification from asymmetric convection: Energetics and efficiency." *Journal of Atmospheric Science*, **64:** 3377-3405.

[10] Molinari, J., and D. Vollaro. (2010) "Rapid intensification of a sheared tropical storm." *Monthly Weather Review*, **138**: 3869-3885.

[11] Rogers, R., P. Reasor, and J. Zhang. (2015) "Multiscale structure and evolution of Hurricane Earl (2010) during rapid intensification." **143**: 536-562.

[12] Bolton, D. (1980) "The computation of equivalent potential temperature." *Monthly Weather Review*, **108**: 1046-1053.

[13] Bauer, M., and A. Del GEnio. (2006) "Composite analysis of winter cyclones in a GCM: Influence on climatological humidity. *Journal of Climate*, **19**, 1652-1672.

[14] Camargo, S., and A. Robertson (2007) "Cluster analysis of typhoon tracks. Part I: General properties." *Journal of Climate*, **20**, 3635-3653.

[15] Mercer, A., C. Shafer, C. Doswell, M. Richman, and L. Leslie. (2012): Synoptic composites of tornadic and nontornadic outbreaks. *Monthly Weather Review*, **140**: 2590-2608.

[16] Richman, M., and I. Adrianto. (2010) "Classification and regionalization through kernel principal component analysis." *Physics and Chemistry of the Earth*, **35**: 316-328.

[17] North, G., T. Bell, R. Galahan, and F. Moeng. (1982) "Sampling errors in the estimation of empirical orthogonal functions." *Monthly Weather Review*, **17**: 699-706.

[18] Rosseeuw, P. (1987) "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." *Journal of Computational and Applied Mathematics*, **20**: 53-65.