Investigation of Rapid Intensification Conditions of Tropical Cyclones with Data Mining Techniques

Ruixin Yang and Jiang Tang
Department of Geography and Geoinformation Science
George Mason University
Fairfax, VA 22030-4444
ryang@gmu.edu

Abstract

Rapidly intensifying (RI) tropical cyclones (TC) are the major error sources in TC intensity forecasting. In order to improve the estimates of RI probability, association rules as a data mining technique are used to facilitate the process of looking for candidate sets of conditions which have strong interactions with rapidly intensifying TCs. Compared to the relation analysis method, the technique of association rules can not only simplify the exploration of associations among multiple conditions but also provide an as complete as possible picture of relations among those conditions. The association rule data mining was used to investigate the RI conditions based on the database for SHIPS (Statistical Hurricane Intensity Prediction Scheme), an operational statistical-dynamical hurricane intensity forecasting model. The mining results identified a reduced predictor set with fewer factors but improved RI probability estimates compared to the results based on relation analysis. That is, the RI probability with three conditions satisfied: low vertical shear, high humidity, and TC being in intensification phase is higher than that with five satisfied conditions including the above three plus high sea surface temperature and a intensity far away from the maximum potential intensity. Moreover, for a given number of constraints affecting the RI process, the data mining technique can identify the combination of the factors which give the largest RI probabilities. One such combination (high latitude, low longitude, the TC being in an intensification phase, an initial intensity far away from the maximum potential intensity, high steering layer value, and low relative eddy flux convergence) gives such a high RI probability that the combination can be considered as a "sufficient" condition for RI, which almost guarantees an RI will take place. In this paper, above results are described and an outline is sketched on how to use data mining techniques to improve the TC intensity forecasting skills.

Key words: tropical cyclones, rapid intensification, data mining

1. Introduction

Tropical Cyclone (TC) is one of most costly natural disasters when the TC intensity is high. Accurate prediction of TC behavior including tracking and intensity is necessary in order to reduce the potential damages. Although the TC track forecast is of relatively high skill, the intensity forecasting is still a challenge (DeMaria et al. 2007; Franklin 2008; Rappaport et al. 2009).

The difficulty of the TC intensity forecasting reflects the fact that there are many factors controlling TC intensity changes (Wang & Wu 2004; DeMaria et al. 2007). Broad studies are carried out on the favorable factors for TC intensification, which include warm ocean eddies (Shay et al. 2000; Hong et al. 2000; Wu et al. 2007), the contraction of an outer eyewall (Willoughby et al. 1982; Willoughby and Black 1996; Lee and Bell 2007; Kossin & Sitkowski 2009; Kuo et al. 2009), an environment with low vertical shear (Gray 1968; Merrill 1988; DeMaria and Kaplan 1994; DeMaria 1996; Frank and Ritchie 1999, 2001; Zeng et al. 2007, 2008), interactions between the upper-level trough and a TC (Molinari and Vollaro 1989, 1990; DeMaria et al. 1993; Bosart et al. 2000), dissipative heat (Jin et al. 2007) and even cloud microphysics (Wang 2002) and isotopic concentrations (Gedzelman et al. 2003).

Most of the previous studies were largely focusing on only one of three categories of factors, ocean characters, inner-core processes, and environmental interactions, and it is well known that intensity changes depend on a combination of those factors (Zhu et al. 2004). Holliday and Thompsom (1979) examined the rapidly intensifying northwest Pacific typhoons and observed that a sufficiently deep layer of warm water, the development at night time, and a smaller size of eye were necessary for those RI typhoons. DeMaria and Kaplan (1994) studied Atlantic TCs and observed that the TCs with a smaller size, with a greater potential to reach their maximum potential intensity, with a faster intensification history, and in an environment with low vertical shear and weak upper-level forcing could have the largest 48-hour intensification rates. In a study of the rapid intensification of Hurricane Opal (1985), Bosart et al. (2000) concluded that its RI was resulted from a combination of several factors: enhanced divergence, low vertical shear and the enhanced heat and moisture from a warm Gulf eddy. Kaplan and DeMaria (2003) examined the large-scale characteristics of rapidly intensifying Atlantic tropical cyclones from 1989-2000 using the NHC HURDAT file and the SHIPS database. Their results confirmed the aforementioned studies. Furthermore a scheme to estimate the probability of RI was developed in their study by combining the thresholds of the five persistence and synoptic predictors: the persistence of intensity change, the vertical shear, the sea surface temperature, the potential to reach the maximum potential intensity, and the moisture content in the lower atmosphere.

The goal of this study is to introduce the technique of association rules from the data mining research area as an "unsupervised," "automatic" data exploration method to discover "multiple-to-one" associations among a large number of environmental characteristics that are responsible for tropical cyclones which will be rapidly intensifying. The results from this technique can then be used to formulate the hypotheses regarding the underlying physical mechanisms, which can be used as the guidance in the traditional statistical analysis, and even to improve RI forecasting.

To satisfy the goal, the SHIPS 2003 dataset is used here to examine the ability of the association rule algorithm to discover the associated environmental conditions in rapid tropical cyclone development. The dataset construction, the RI definition, RI thresholds and RI probability definitions are followed those in Kaplan and DeMaria (2003, hereafter KD03) as much as possible. The purpose is to create a similar context as KD03 so that the results from the association rules can be easily mapped to and verified by the traditional statistical terms and methods.

2. Data and Methods

Data-SHIPS Dataset

The datasets for this study are the NHC HURDAT file (Jarvinen et al. 1984) and the SHIPS 1982-2003 database (DeMaria and Kaplan 1994, 1999; DeMaria et al. 2005). A detailed description of the variables in the HURDAT and SHIPS datasets can be found in KD03. The HURDAT file consists of 6-hr estimates of position and maximum sustained surface wind speed for all named Atlantic TCs from 1851 to the present. The SHIPS database contains synoptic information for every 12 hr for all Atlantic TCs from 1982 to the present. In this study, the time period is limited to 1982-2003 due to the initial data availability.

The two data sets are merged based on a methodology which is identical to that described in KD03 except for the consideration of the systems remaining over both water and tropical regions during the period from t – 12hr to t + 24hr, and for the non-developing tropical depressions. In total, there are 34 values (variables) for each case, and each variable was evaluated at the beginning (t=0h) of each 24-h period. The 0000 and 1200 UTC synoptic predictor values in the SHIPS database were averaged to estimate the magnitude of the corresponding values at 0600 and 1800 UTC. KD03 uses only SHIPS data from 1989 to 2000. Since we have more data from 1982 to 2003, we divided the data into 3 subsets: 1989-2000, 1982-1988, and 2001-2003. In other words, we used the first subset as a training set to repeat the results of KD03, and used the last two subsets as testing sets for the conclusion.

Furthermore, we divided the data into rapidly intensifying cases (RI) and non-rapidly intensifying cases (non-RI). In the merged data set, there are 7497 records with 265 RI cases. However, to mine/analyze the data further, records with missing values were removed from the data set, which resulted in a total of 5505 valid records and the same 265 RI cases. For the 1989 to 2000 period, the reconstructed dataset contains a total of 3306 cases, which were from 135 distinct Atlantic TCs (0 tropical depressions, 54 tropical storms, and 81 hurricanes). Abided by the RI definition proposed in KD03, that is, at least 30 knots of intensity increase in the next 24 hours, the 3306 cases are comprised of 169 RI cases and 3137 non-RI cases, from 53 distinct TCs. A similar process is applied to datasets from 1982-1988 and from 2001-2003. The numbers of cases after such divisions are listed in Table 1.

Table 1. Total case numbers, RI case numbers, and RI probabilities (RIP) based on the sample means of RI cases for various time periods.

Time Coverage	Total	RI	RIP
1982-1988	1170	60	5.1%
1989-2000	3306	169	5.1%
2001-2003	1029	36	3.5%
1982-2003	5505	265	4.8%

Method-Association Rule Algorithm

To discover "multiple-to-one" associations among a large number of factors favoring rapidly intensifying TCs, we will "mine" the above cleaned data by using an association rule algorithm. Association rule induction (Agrawal et al. 1993) is a powerful method for market basket analysis, which aims at finding regularities in the shopping behavior of customers. An association rule is a rule like "Z←X, Y." The items X and Y are called antecedents in the rule and Z is the consequent. This rule expresses an association between items X, Y, and Z. It states that if a customer is picked randomly and the customer selected items X and Y, it is likely that the customer also selected item Z.

Usually, three parameters, support, confidence, and lift, are reported for mined association rules. The support estimates the probability P({X,Y,Z}), and the confidence estimates the probability P(Z|{X, Y}). An association rule "Z←X, Y" is strong if it has a large support and a high confidence. The third parameter, the lift (Silverstein et al. 1998) is introduced as the ratio between the actual probability of the item set containing both antecedent and consequent divided by the product of the individual probabilities of the antecedent set and the consequent. That is, lift = P({X,Y,Z})/[P({X,Y})*P(Z)]. The lift measures the dependency among the antecedents and the consequent of a rule. Lift values above one indicate a positive dependence, while those below one indicate a negative dependence.

The version of the association rule algorithm used in this study is implemented by Borgelt (2009). The support value in this implementation is defined as P({X,Y}) instead of P({X,Y,Z}). The practical description of the rule measurements (support, confidence, and lift) for this application is postponed to a later section with a real rule for RI. As in most data mining applications, data preprocessing is necessary before the association rule data mining algorithm can be applied.

Data Discretization for Mining

For the mining task, the original 34 attributes are reduced to 11 independent predictors as listed in Table 2. These predictors are chosen because KD03 found that the mean initial conditions of these predictors for RI cases and non-RI cases are statistically different at least at the 95% significance level based on an unequal-variance, two-sided *t* test. To mine those attributes with the association rule algorithm, we should convert continuous values of the attributes into disjoint conditions. Here, we divide the values into "High" or "Low" conditions based on the threshold values provided in KD03, which was derived from the mean values of the RI samples, as listed in Table 2. For example, "SHRD=L" means that the 850-200hPa vertical shear is less than 4.9 m/s. After the discretization processing, the antecedent set is of 22 (11X2) entries.

Table 2. The 11 statistically significant predictors (from KD03, Table 4). The predictors DVMX, SHR, and SLYR in KD03 are renamed as IV12, SHRD, and PSLV here.

Name	Description	Threshold
IV12	Intensity change during the previous 12 hours.	4.6 m/s
SHRD	850-200 hPa vertical shear.	4.9 m/s
SST	Sea surface temperature.	28.4 ^{o}C
POT	Maximum potential intensity (MPI) – initial intensity	47.6 m/s
RHLO	850-700 hPa relative humidity.	69.7 %
LAT	Latitude	19.7 ^{o}N
LON	Longitude	63.2 ^{o}W
USTM	Zonal (u) component of storm motion.	-3.1 m/s
U200	200 hPa zonal (u) component of wind	-0.6 m/s
REFC	200 hPa relative eddy angular momentum flux convergence	0.9 m/s/day
PSLV	Pressure of the center of mass of layer for which the environmental winds best match the current storm motion.	583.4 hPa

3. Results

Compared to statistical analysis, the advantage of the association rule data mining is that the technique can explore exhaustively associations among multiple conditions because it examines all possible combinations of frequent condition sets automatically. As a successful scientific data mining example, a rule "*RI←SHRD=L, PD12=H, RHLO=H (supp=1.3%, conf=47.6%, lift=931.5%)*" for the 1989-2000 period was mined out (Yang et al. 2007). In simple language, the rule tells us that there are 1.3% cases satisfying the three conditions, low vertical shear of horizontal wind (SHRD=L [<4.9m/s]), high humidity in the 850-700hPa level (RHLO=H [≥69.7%]), and the TC being in intensification phase (IV12=H [≥4.6m/s]). Among those cases for the 1989-2000 period, 47.6% of them underwent RI, and the ratio of this RI probability to the sample mean probability (5.1%) is 9.3, or 930% as given by the lift value (the inconsistency between the ratio and the given lift value is due to the round-off errors in the mining algorithm). The most noticeable feature about this rule is that the RIP mined out here with three conditions is higher than that found in KD03, which include the three conditions above and two additional conditions, high sea surface temperature (SST=H), and an intensity far away from the maximum potential intensity (POT=H). Following the same procedure and on the same data set here, the RIP with the five conditions is only 43.5%.

One step further, one can use the association rule data mining technique to search for the "optimal" conditions which result in the highest RIP for a given number of conditions among the selected set (Yang et al. 2008). Figure 1 shows the changes in highest RI probabilities with the number of the thresholds in the 22 predictor pool for different time periods. All curves demonstrate the same trend. That is, the highest RIPs increase with the numbers of predictors initially, reach the peak values when the number of predictors approaches five to seven (N=5-7), and then decrease with further increases of the numbers of predictors. These results demonstrate that the multiple factors together are responsible for the RI process of TCs. However, the number of factors will saturate at certain numbers. After that, the impacts of individual factors may cancel each other out, or may be replaced by other factors.

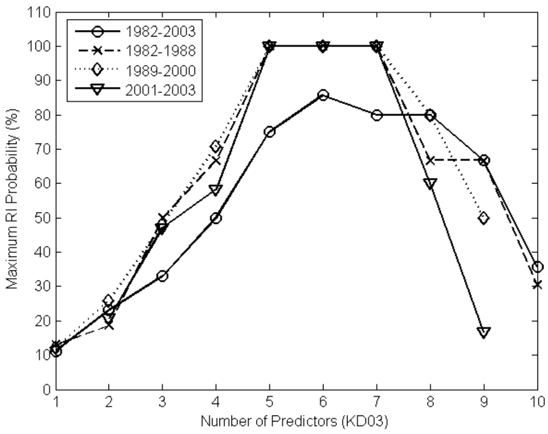

Figure 1. The highest RI probabilities for different number of thresholds and multiple time periods.

The most striking result in Figure 1 is that the data mining algorithms identify certain cases with a 100% RI probability. These "perfect" results take place only for the three sub-periods but not for the whole period. This plausible result comes from the fact that the detailed "optimal" conditions for each subset period are different from each other, and as a result, the conditions and the RI probability for the whole time period are also different from those in individual sub-periods. For example, Table 3 lists the detailed conditions for the N=6 cases for different time periods. By carefully checking, one can see that no two groups give the same conditions although the RIP (confidence) is 100% for all three short periods.

Table 3. Optimal conditions when N=6 for different time periods.The columns represented by "s," "c," and "i" give the values of support, confidence, and lift in percentages, respectively.

Periods	Detailed Conditions for N=6	s	c	i
1982-1988	IV12=L,LAT=L,ZONX=L,POT=L,SHRD=L,RHLO=H	0.2	100	1950
1989 -2000	POT=L,SHRD=L,IV12=H,ZONX=H,RHLO=H,PSLV=H	0.2	100	1956.2
2001 -2003	IV12=L,LAT=L,POT=L,U200=L,SST=H,REFC=H	0.2	100	2858.3
1982 -2003	LON=L,REFC=L,IV12=H,LAT=H,POT=H,PSLV=H	0.1	85.7	1780.6

The sub-period "optimal" rules trigger one natural question: how well are those rules for other sub-periods. If a rule is useful, that rule should be mined as a valid one in other time periods, too. Table 4 lists the support, confidence and lift values of every optimal condition combination in other sub periods. These "optimal" rules defined for one sub-period do appear as good rules (i.e., lift > 1) in most of other sub periods, although with lower support, confidence, and lift values. In some cases, the exact rules with all conditions satisfied were not found, and rules will fewer conditions satisfied were found instead.

One prominent result is that the optimal rules mined out for the three sub-periods work well for the whole time period. Although the confidence levels are lower than that for the optimal rule for the whole period, the support level is the same, and the lift values are much higher than one. This finding suggests that those rules are useful to estimate RIP values for TC cases because mining data for sub-periods may capture the RI conditions not common for a long time period but important in certain conditions and time periods. Those rules will be helpful if the mining results are used to help RI forecasting.

Table 4. Support, confidence, and lift values in percentage (from left to right in each entry) of the optimal rules in other sub periods.

	82-88 data	89-00 data	01-03 data	82-03 data
82-88 rule	0.2,100.0,1950.0	X (only 5 cond.)	0.1,25,714.6	0.1,42.9,890.3
89-00 rule	0.1,33.3,650.0	0.2,100.0,1956.2	X (only 5 cond.)	0.1,57.1,1187.1
01-03 rule	0.1,100.0,1950.0	X (only 4 cond.)	0.2,100.0,2858.3	0.1,25.0,519.3
82-03 rule	0.1,100.0,1950.0	0.2,83.3,1630.2	X (only 5 cond.)	0.1,85.7,1780.6

One potential challenge for using the association rule mining results in RI studies and forecasting is the low support values, that is, small percentage of cases for each mined rule. Actually, due to the large number of predictors, the chance for any condition combination to appear is low, and the mined out rules give a relatively high confidence on the RIP values. Take the example of the whole time period, 1982-2003, the highest RIP for the period is 85.7% when six conditions are working together. Detailed investigation showed that there are seven cases satisfying the six conditions in the whole data set, and six of them underwent RI (6/7=85.7%). The seven cases are traced back to the original data and it is found that the seven cases are actually from four TCs (Yang et al. 2008). The fact that relatively large numbers of divers TCs underwent RI when the above given conditions were satisfied and the extreme high RIP lead us to believe that the results from data mining are significant. Moreover, based on the individual tracing result, it is found that the only case in which the TC did not undergo a RI process is the case for Hurricane Karl at 0h of September 24, 1998. However, the intensity of Hurricane Karl increased from 35 knots to 50 knots in 24 hours and then to 75 knots in another 24 hours. Karl continued the intensification process after that until 0h of September 27 at 90 knots. Therefore, it is quite reasonable to say that the six conditions together are favorable to the RI process in almost all cases. The mining results identify at least one "sufficient" condition for RI (Yang et al. 2008). Therefore, the results will not only shed light for understanding the RI processes and but also help to guide future RI forecasting.

4. Concluding Remarks

In this study the association rule mining technique has been successfully applied to rapid intensification of tropical cyclones, one major challenge for operational intensity forecasting (Rappaport et al. 2009; Kaplan et al. 2010). The long term goal of this work is to make the mining results useful for forecasting of TC intensity changes in general and

RI in particular. However, several issues in the data pre-processing and post-processing processes should be investigated first since there is no clear guidance on what a scientist should be looking for before mining.

The first challenge appears at how to choose the split value to separate a parameter into a "High" or "Low" value range during data pre-processing because the RIP values do depend on the threshold values of POT (Yang et al. 2008). Since the sample sizes of RI cases and non-RI cases are highly skewed, sophisticated methods such as the decision trees or Bayesian theory on two-class classification should be considered.

Another challenge is to identify the "interesting" mined rules. Although the fact that association rule mining exhausts all of the combinations of the parameters is the major advantage to traditional one-to-one statistical comparison method, it also presents a stack of hay for a needle finder which limits its own usage, especially when the number of parameters are large as in the SHIPS database. Several rule pruning techniques must be considered in finding the most interesting rules. In our study the measurements of support, confidence and lift are used first as a standard practice to find truly correlated rules. In order to have a manageable number of truly correlated rules, the levels of support, confidence and lift are adjusted. Moreover, our study benefits from concise rule searching which further reduces the overlap among the truly correlated rules. Additionally, through hyper-edge search the hidden relationships among multiple parameters can be found even without any background knowledge of geosciences

Although SHIPS is the most skillful among the NHC operational intensity forecasting models and it has been improved in the past, the linear feature limits further significant improvements (DeMaria 2009). Moreover the unanticipated RI process in TC development, which is one major challenge for intensity forecasting (KD03; Rappaport et al. 2009; Kaplan et al. 2010), is difficult to be correctly forecasted based on a global linear regression model. As such, an outline is given below for leveraging data mining techniques to improve statistical TC intensity forecasting in future studies.

One possible solution is to use several piecewise regression models (Breiman et al. 1993). A set of association rules which have the complete coverage of the rapidly intensifying cases should be found first. The qualified rules must have a lift greater than 1 to guarantee the correlation is true. The qualified rules must be concise rules to avoid redundancy. The rule selection can start at the concise rules with the most support then go to the concise rules with smaller support. The confidence for the selected rules may vary. In order to avoid the problem of over-fitting, a qualified rule is better to cover cases from different TCs. The goal is to cover most of the rapidly intensifying cases, if not all, with as few rules as possible. The rules with large support may indicate the major processes controlling rapid intensification of TCs. The rules with small support reflect a particular process controlling several TCs. From the set of selected association rules, we can gain more understanding of the processes behind rapid intensifications. Later on a set of regression models can be established to estimate the actual amount of intensity change based on the probabilities of rapid intensification.

The above piece-wise model needs a relatively accurate estimate of probabilities of rapid intensification. An RI Index (RII) should be established to assess this probability for each forecast. Kaplan et al. (2010) developed an enhanced version of RII based on the work of KD03. The results from this work are most suitable to develop a similar RI index with a totally different methodology.

Another possible future work is to consider other data mining techniques such as the log-linear model proposed by Wu et al. (2003) and the chi-square test proposed by DuMouchel and Pregibon (2001) to explore infrequent but surprising association rules.

References

Agrawal, R., T. Imielinski, and A. Swami, 1993: Mining association rules between sets of items in large databases, *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, Washington D.C., May 1993, 207-216.

Borgelt, C., 2009: Apriori - Association Rule Induction/Frequent Item Set Mining, http://www.borgelt.net/apriori.html, last access on May 11, 2009.

Bosart, L.F., C.S. Velden, W.E. Bracken, J. Molinari, and P.G. Black, 2000: Environmental influences on the rapid intensification of Hurricane Opal (1995) over the Gulf of Mexico. *Mon. Wea. Rev.*, **128**, 322–352.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1993: *Classification and regression trees*. Chapman & Hall, New York, USA.

DeMaria, M. and J. Kaplan, 1999: An Updated Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic and Eastern North Pacific Basins Mark, *Wea. Forecasting,* **14**, 326–337.

DeMaria, M., 1996: The effect of vertical shear on tropical cyclone intensity change, *J. Atmos. Sci.*, **53**, 2076-2087.

DeMaria, M., 2009: A Simplified Dynamical System for Tropical Cyclone Intensity Prediction. *Mon. Wea. Rev.*, **137**, 68–82.

DeMaria, M., and J. Kaplan, 1994: A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin, *Wea. Forecasting*, **9**, 209-220.

DeMaria, M., J.A. Knaff, and C.R. Sampson, 2007: Evaluation of Long-Term Trends in Operational Tropical Cyclone Intensity Forecasts. *Meteor. and Atmos. Phys.*, **59**, 19-28.

DeMaria, M., J.-J Baik, and J. Kaplan, 1993: Upper-level eddy angular momentum flux and tropical cyclone intensity change, *J. Atmos. Sci.*, **50**, 1133-1147.

DeMaria, M., M. Mainelli, L. K. Shay, J. A. Knaff, and J. Kaplan, 2005: Further Improvements to the Statistical Hurricane Intensity Prediction Scheme (SHIPS), *Wea. Forecasting,* **20**, 531–543.

DuMouchel, W., and D. Pregibon, 2001: Empirical bayes screening for multi-item associations, in *Proceedings of the 7th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, August 26-29 2001, 67-76.

Frank, W. M. and E. A. Ritchie, 1999: Effects of environmental flow upon tropical cyclone structure. Mon. Wea. Rev., *127, 2044–2061.*

Frank, W. M., and E. A. Ritchie, 2001: Effects of vertical wind shear on hurricane intensity and structure. *Mon. Wea. Rev.*, **129**, 2249-2269.

Franklin, J. L., 2008: 2007 National Hurricane Center forecast verification report. NOAA/NWS/NCEP/Tropical Prediction Center, 68 pp. online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2007.pdf, last accessed on October 21, 2008.

Gedzelman, S., J. Lawrence, J. Gamache, M. Black, E. Hindman, R. Black, J. Dunion, H. Willoughby, and X. Zhang, 2003: Probing Hurricanes with Stable Isotopes of Rain and Water Vapor, *Mon. Wea. Rev.,* **131**, 1112–1127.

Gray, W. M., 1968: Global view of the origin of tropical disturbances and storms, *Mon. Wea. Rev.*, **96**, 669–700.

Holliday, C. R., and A. H. Thompson (1979), Climatological characteristics of rapidly intensifying typhoons. *Mon. Wea. Rev*., 107, 1022–1034.

Hong, X., S. W. Chang, S. Raman, L. K. Shay, and R. Hodur, 2000: The interaction between Hurricane Opal (1995) and a warm core ring in the Gulf of Mexico. *Mon. Wea. Rev*., **128**, 1347–1365.

Jarvinen, B.R., C.J. Neumann, and M.A.S. Davis, 1984: A tropical cyclone data tape for the North Atlantic basin, 1886-1983: Contents, limitations, and uses, *NOAA Technical Memorandum* NWS NHC 22.

Jin, Y., W.T. Thompson, S. Wang, and C.S. Liou, 2007: A Numerical Study of the Effect of Dissipative Heating on Tropical Cyclone Intensity. *Wea. Forecasting*, **22**, 950–966.

Kaplan J., M. DeMaria, J.A. Knaff, 2010: A Revised Tropical Cyclone Rapid Intensification Index for the Atlantic and Eastern North Pacific Basins. *Wea. Forecasting*, **25**:1, 220-241.

Kaplan, J. and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin, *Wea. Forecasting*, **18**, 1093-1108.

Kossin, James P., Matthew Sitkowski, 2009: An Objective Model for Identifying Secondary Eyewall Formation in Hurricanes, *Mon. Wea. Rev.*, **137**, 876-892

Kuo, Hung-Chi, Chih-Pei Chang, Yi-Ting Yang, Hau-Jang Jiang. (2009) Western North Pacific Typhoons with Concentric Eyewalls. *Monthly Weather Review* **137**:11, 3758-3770.

Lee, W.-C., and M. M. Bell, 2007: Rapid intensification, eyewall contraction, and breakdown of Hurricane Charley (2004) near landfall, *Geophys. Res. Lett*., **34**, L02802, doi:10.1029/2006GL027889.

Merrill, R.T., 1988: Environmental Influences on Hurricane Intensification. *J. Atmos. Sci.*, **45**, 1678–1687.

Molinari, J., and D. Vollaro, 1989: External influences on hurricane intensity. Part I: Outflow-layer eddy angular momentum fluxes. *J. Atmos. Sci*., **46**, 1093–1105.

Molinari, J., and D. Vollaro, 1990: External influences on hurricane intensity. Part II: Vertical structure and response of the hurricane vortex. *J. Atmos. Sci*., **47**, 1902–1918.

Rappaport, E.N., J.L. Franklin, L.A. Avila, S.R. Baig, J.L. Beven, E.S. Blake, C.A. Burr, J.G. Jiing, C.A. Juckins, R.D. Knabb, C.W. Landsea, M. Mainelli, M. Mayfield, C.J. McAdie, R.J. Pasch, C. Sisko, S.R. Stewart, and A.N. Tribble, 2009: Advances and Challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395–419.

Shay, L. K., G. J. Goni, and P. G. Black, 2000: Effects of a warm oceanic feature on Hurricane Opal. *Mon. Wea. Rev*., **128**, 1366–1383.

Silverstein, C., S. Brin, and R. Motwani, 1998: Beyond market baskets: generalizing association rules to dependence rules. *Data Mining and Knowledge Discovery*, **2**, 39-68.

Wang, Y., 2002: An Explicit Simulation of Tropical Cyclones with a Triply Nested Movable Mesh Primitive Equation Model: TCM3. Part II: Model Refinements and Sensitivity to Cloud Microphysics Parameterization *Mon. Wea. Rev.*, **130**, 3022–3036.

Wang, Y., and C.C. Wu, 2004: Current understanding of tropical cyclone structure and intensity changes – a review, *Meteorology and Atmospheric Physics*, **87**, 257–278.

Willoughby, H. E., J. A. Clos, and M. G. Shoreibah, 1982: Concentric eyewalls, secondary wind maxima, and the evolution of the hurricane vortex, *J. Atmos. Sci.*, **39**, 395–411.

Willoughby, H.E. and P.G. Black, 1996: Hurricane Andrew in Florida: Dynamics of a Disaster, *Bull. Am. Meteorol. Soc.,* **77**, 543–549.

Wu, C.C., C.Y. Lee, and I.I. Lin, 2007: The Effect of the Ocean Eddy on Tropical Cyclone Intensity. *J. Atmos. Sci.*, **64**, 3562–3578.

Wu, X., D. Barbara, and Y. Ye, 2003: Screening and interpreting multi-item associations based on log-linear modeling, in *Proceedings of the 9th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, Washington, DC, USA, August 24-27 2003, pp 276-285.

Zeng, Z., L. Chen, and Y. Wang, 2008: An Observational Study of Environmental Dynamical Control of Tropical Cyclone Intensity in the Atlantic. *Mon. Wea. Rev.*, **136**, 3307–3322.

Zeng, Z., Y. Wang, and C.C. Wu, 2007: Environmental Dynamical Control of Tropical Cyclone Intensity—An Observational Study. *Mon. Wea. Rev.*, **135**, 38–59.

Zhu, T., D.L. Zhang, and F. Weng, 2004: Numerical Simulation of Hurricane Bonnie (1998). Part I: Eyewall Evolution and Intensity Changes, *Mon. Wea. Rev.,* **132**, 225–241.