Mining NEXRAD Radar Data: An Investigative Study

Xiang Li, Rahul Ramachandran, John Rushing and Sara Graves
Information and System Center, University of Alabama in Huntsville

Kevin Kelleher
National Severe Storms Laboratory, NOAA

S. Lakshmivarahan
School of Computer Science, University of Oklahoma

Douglas Kennedy
Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma

Jason Levit
Center for Analysis and Prediction of Storms, University of Oklahoma

A collaborative team of meteorologists and data mining experts conducted a case study to detect and classify mesocyclone signatures in WSR-88D radar data using mining techniques. Radar data for May 6th, 1994 and May 11th, 1992 from Norman and Tulsa, Oklahoma was used in this case study. Two Mesocyclone Detection Algorithms (MDA) were used in this study. One (NSSL MDA) is a computationally optimized version of the National Severe Storm Laboratory (NSSL) MDA. The second (UAH MDA) was created based upon the original NSSL algorithm, but with some image processing techniques. The primary difference between the two algorithms is in the technique used to segment the two dimensional (2D) mesocyclone signatures. The UAH MDA uses a region growing technique; hence the shape of the feature is no longer a restriction. True or false labels were assigned to the mesocyclone features generated from the MDA by comparing with a truth set derived by an expert using the NSSL algorithm. This labeled feature data set was then used in a series of analysis experiments. The objective of these experiments was to: Evaluate the performance of different classifiers in their ability to distinguish between a true mesocyclone signature versus an artifact or noise; Optimize the classification results by selecting key parameters from the feature data set using different feature reduction techniques; Explore the patterns in the feature data set using various clustering algorithms; and determine the suitability of the algorithm as a data mining tool.

1 Introduction

Mesocyclones are rotating updraft/downdraft structures inside severe thunderstorms. Detection of mesocyclones is important for severe weather forecast because over 90% of mesocyclones are accompanied by severe weather such as tornadoes or large hail [1]. Mesocyclone signatures can be identified from the Weather Surveillance Radar –1988 Doppler (WSR-88D, [2]) and appear as couplets of incoming and outgoing radial velocities. WSR-88D, the next generation weather radar deployed in the United States a decade ago, has proved to be extremely valuable for severe weather forecasts and warning. Experienced radar analysts can identify mesocyclone signatures in WSR-88D data. However, this manual identification process is tedious and time-consuming, and can be overwhelming during a severe weather outbreak. Consequently, several mesocyclone detection algorithms (MDA) have been developed ([3], [4], [5]) for automated mesocyclone signature detection. The National Severe Storm Laboratory (NSSL) MDA [3] is one such algorithm and is an enhancement to the Build 9 WSR-88D Mesocyclone Algorithm (B9MA) [4]. Compared to B9MA, NSSL MDA identifies a broader spectrum of mesocyclones and has an improved probability of mesocyclone feature detection. The latest version of the NSSL MDA also includes a neural network classifier to filter out false mesocyclone signatures [6].

This paper describes an investigative case study that was conducted to detect and classify mesocyclone signatures from WSR-88D data using data mining techniques. This research consisted of two components.

The first component was to develop a mining algorithm to detect mesocyclone signatures from WSR-88D radar data. In addition to the accuracy, the major requirement for the algorithm was computational speed. The algorithm should be fast enough to efficiently and expeditiously process large volumes of archived WSR-88D data. The second component of this research was to identify "true" mesocyclone signatures using the detected algorithm output. In order to meet this objective, we 1) evaluated the performance of different classifiers in their ability to distinguish between a true mesocyclone signature versus an artifact or noise, 2) optimized the classification results by selecting relevant features from the mesocyclone feature dataset using different feature reduction techniques, and 3) explored the patterns in the feature data set using various clustering algorithms.

Section 2 describes the data used in this case study. Section 3 briefly describes the mesocyclone detection algorithm developed in this study as well as the classifiers, feature selection and clustering algorithms that were used. Results and analyses are described in section 4. Finally, section 5 summarizes this investigative study.

2 Data Description

Dealiased WSR-88D Level II data for the following dates were used in this study; 1) May 11, 1992 over Norman, Oklahoma (N35°13′48″, W97°16′80″) with radar station name of KOUN; 2) May 6, 1994 over Tulsa, Oklahoma (N36°10′30″, W95°33′53″) with radar station name of KINX. Tornado outbreaks took place at these places during these selected time periods. The WSR-88D data contains three fields of measurement: reflectivity, velocity, and spectrum width. The velocity field was used for mesocyclone detection. The reflectivity field was also used as an additional parameter for the detected mesocyclone signatures. For the May 11, 1992 case (also referred to as KOUN case), there were a total of 50 WSR-88D volume scans available, covering the time period from 17:40 UT to 22:30 UT. For the May 6 1994 case (also referred to as KINX case), there were a total of 117 volume scans available, covering the time period from 20:30 UT to 08:40 UT the next day. For the two cases, validation data sets were available from NSSL derived from the original NSSL MDA. These data sets were used in the classification process as the "truth" data.

3 Methodology

3.1 Mesocyclone Detection Algorithm

Mesocyclones contain a velocity signature known as a Rankine Vortex [7], representing incoming and outgoing radial velocity couplets in the radar velocity field. For a segment of consecutive gate measurements with the same range and increasing scan angle, this signature is known to have continuously increasing velocity values followed by gates with similar values, but opposite signs, that decrease in magnitude with increasing scan angle. Identifying shear segments based on this signature is the most intuitive approach. This algorithm has three steps: 1) identifying 1-dimensional (1D) shear segments, 2) identifying two-dimensional (2D) shear regions in a scan sweep, and 3) collocating shear regions across elevation scans into 3-dimensional (3D) mesocyclone signatures. The criteria used to determine the strength of the shear segments were similar to the ones used in NSSL MDA. The major difference between the MDA developed in this study (hereafter referred to as UAH MDA) and the NSSL MDA is in the 2D shear region identification. The NSSL MDA builds 2D shear regions by examining all 1D shear candidates, selecting the ones that satisfy location, shape and size constraints and then grouping them into 2D shear region. The UAH MDA builds 2D shear regions using a region-growing technique [8], i.e., grouping those 1D shear segments adjacent to each other into a 2D region. There is no shape restriction for the identified 2D shear regions.

A feature vector is calculated for each 2D shear region and its features include central azimuth angle, central range, central height, diameter, shear, maximum Gate-to-Gate Velocity Difference (GTGVD), rotational velocity difference (Vrot), and mesocyclone strength index, as defined in [3]. For a 3D

mesocyclone signature, a number of features such as base and top heights, core base and top heights are calculated in addition to the mean values of the corresponding 2D features. Thus, each 3D mesocyclone signature contains twenty features: 1) base height, 2) depth, 3) core base height, 4) core depth, 5) base diameter, 6) base shear, 7) base Vrot, 8) base GTGVD, 9) mean diameter, 10) maximum diameter, 11) mean shear, 12) maximum shear, 13) maximum GTGVD, 14) height of maximum GTGVD, 15) maximum Vrot, 16) mean strength index, 17) maximum strength index, 18) mean reflectivity, 19) maximum reflectivity, and 20) rotational. The rotational feature indicates whether or not a mesocyclone contains velocity values with opposite signs.

For classification purposes, each identified 3D mesocyclone signature was labeled as either "true" if it matched with one of the NSSL true signatures or "false" if it did not match with any NSSL-validated signature. A 3D signature from UAH MDA was matched with a 3D signature from NSSL MDA if the spatial locations of the two signatures were within 8 km and the depth of 3D signatures overlapped with each other. The 3D signature data set was then randomly split into equal halves to be used as the training and testing data in the classification process.

3.2 Classification, feature selection and clustering

A classifier is necessary to weed out the false mesocyclone signatures identified by the detection algorithm. These false signatures may be due to noise or other artifact. In this study, two classifiers are examined, naïve Bayes classifier and Support Vector Machine (SVM) classifier [9].

Bayes classifier is a well-known supervised classifier. It classifies samples based on the Bayes rule that an unknown sample is assigned to a known class with maximum posteriori probabilities. The Bayesian classifier is optimal with respect to minimizing the classification error probability. Support Vector Machine (SVM) is a new classification technique based on statistical learning theory. SVM classifiers can produce a non-linear decision boundary that separates two classes with the largest margin. Detailed description of the two classifiers can be found in Hastie, et al. [9].

Feature selection is an important step in the classification process. Computational complexity and classifier generalization are two of the major reasons that make feature selection important. In this study, three feature selection methods were applied to the two classifiers to optimize classifier performances. The three methods were: 1) forward feature selection, 2) backward feature elimination, and 3) feature selection using the Genetic Algorithm (GA). The features that are most frequently selected are considered crucial in classifying the mesocyclone signatures.

Two clustering algorithms were also applied on the MDA output data set to examine the clustering tendency of the data. The clustering algorithms used were: 1) Maximin-distance algorithm [10], and 2) K-means algorithm [10]. The maximin-distance algorithm seeks clusters that are farthest apart. The algorithm is iterative and terminates when the distance of any feature to the cluster center is within a pre-defined threshold. The K-means algorithm identifies the cluster centers so that the sum of distance from a feature sample to its cluster center for all samples is minimized. For the K-means algorithm, the number of clusters has to be specified; whereas the maximin-distance algorithm determines the number of clusters based on the data.

All of the classification, feature selection and clustering algorithms used in this study are implemented as the component modules in the Algorithm Development and Mining System (ADaM) toolkit [11] developed by the Information Technology and System Center (ITSC), UAH. The ADaM toolkit consists of a variety of data mining and image processing modules in addition to data preprocessing modules, and has been extensively used in the Earth and Space sciences research and applications [12], [13], [14].

4 Results and analyses

4.1 UAH MDA performances

Using the two cases, several aspects of NSSL MDA and UAH MDA outputs were compared. Comparisons between the two MDAs included the number of mesocyclone signatures identified, the spatial locations of the signatures, and feature distributions of the identified signatures. For the KOUN case, NSSL MDA identified 996 signatures, whereas the UAH MDA identified a total of 1097 signatures. For the KINX case, NSSL MDA identified 3862 signatures, while the UAH MDA identified 1876 signatures. With only two case studies it is difficult to draw general conclusions, but the UAH MDA identified significantly less mesocyclone signatures than the NSSL MDA for the KINX case. One reason for this is that UAH MDA does not identify "long-range 2D features", that is, the 2D features detected at lowest elevation scan and located greater than 175 km from the radar, as NSSL MDA does [3]. The difference in 2D shear region detection between the two algorithms may also cause some differences in detected mesocyclones. Figs. 1a and 1b show the location distributions of the identified signatures from both of MDAs for KOUN and KINX, respectively. It can be seen that both algorithms have similar spatial coverage for the detected mesocyclone signatures. Figs. 2 and 3 show the distributions of the mesocyclone features from the two MDAs. The distributions of features from the two algorithms are similar except for the mesocyclone depth. The correlations of feature distributions are higher than 0.80 for base height, mean diameter and mean shear, whereas the correlation for depth distribution is about 0.55. The relative low correlation for depth is mainly due to the distribution discrepancy at depth values smaller than 1.0. The histograms of depth feature peak at about 0.75 km for the NSSL features, which is mainly attributed by the "long-range 2D features", while it is not true for the UAH depth features.

The computational speed performance for the NSSL and UAH MDA's were also examined. Note that neither MDA algorithm detected any mesocyclone signatures for clear air conditions. Using a 400 MHz Pentium 2 PC with 256 MB memory, for a volume scan the UAH MDA mean processing times for the KOUN case and the KINX case were 1.02 sec/volume and 0.90 sec/volume, respectively. In contrast, for the NSSL MDA running on the same system, the mean processing times were 1.98 seconds and 1.86 seconds, respectively. Consequently, the UHA MDA ran roughly twice as fast as the NSSL MDA on a Pentium 2.

4.2 Mesocyclone signature classifications

Four metrics were used by NSSL to evaluate the performance of a classifier on mesocyclones in [3]. The four metrics are: 1) Probability of detection (POD), 2) False alarm rate (FAR), 3) Critical success index (CSI), and 4) Heidke skill score (HSS). Due to the heavy skewness of detected mesocyclone signatures towards the "false" class, NSSL determined the CSI is the proper measure in evaluating classifier performance [3].

Tables 1 and 2 contain the CSI results from Bayes and SVM classifiers, for the KOUN and KINX cases, respectively, for the three feature selection methods. In these tables, the bitstring is a mask indicating the elimination (1) or selection (0) of a feature. The first 20 bits in the bitstring correspond to the 20 features calculated by the UAH MDA. The last bit in the bitstring represents the class index, and it is always selected (0). From Table 1, it can be seen that six combinations of classifiers with feature selection methods produced similar CSI values. The GA feature selection method gives a slightly better CSI result. This is also true for KINX case, as shown in Table 2. Also, the CSI value for KINX case is about 10% better than that for KOUN case.

The CSI values ranged from 0.30 to 0.40, similar to the values in [3]. However, signatures used to train and test the neural network classifier in [3] were based on the actual tornado observations. Using NSSL mesocyclone signatures with 8 features, the best CSI values for KOUN and KINX cases were 0.56 and 0.61, respectively. One of the possible causes for the lower CSI values from the UAH MDA could be the propagation of error in the truth labeling process for UAH data sets. The UAH MDA mesocyclone

signature is labeled as either a true or false signature based on whether or not it matches with one of the NSSL signatures.

From Tables 1 and 2, it is seen that although the CSI values for different feature selection methods are very close, the features optimized for the classifiers varied from method to method. Also, the total number of features selected varied for different classifiers, ranging from 10 to 13. The occurrence of a feature in the feature selection process was calculated and shown in Table 3. The maximum occurrence that a feature could be selected is 12. Base height, base Vrot and maximum strength index features had the highest scores. The depth and height of maximum GTGVD also scored high. Maximum Vrot also scored high but is highly correlated with Base Vrot. Thus, the most important features to classify the true mesocyclones are the vertical description of the signatures (base height, depth and height of maximum GTGVD) and the strength of the rotation (strength index and maximum Vrot). The secondary features include the maximum diameter and maximum reflectivity features. Surprisingly, reflectivity did not have a big impact on classification accuracy.

4.3 Mesocyclone clustering

The Maximin-distance clustering algorithm was applied to the entire mesocyclone data sets for both the KOUN case and KINX case. Fifteen out of the twenty parameters that were used are: 1) base height, 2) depth, 3) base diameter, 4) base shear, 5) base Vrot, 6) base GTGVD, 7) mean diameter, 8) maximum diameter, 9) mean shear, 10) maximum shear, 11) maximum GTGVD, 12) height of maximum GTGVD, 13) maximum Vrot, 14) mean strength, and 15) mean reflectivity. The core depth, core base height, maximum strength index and maximum reflectivity parameters are highly correlated with depth, base height, mean strength index, and mean reflectivity parameters, respectively, and they were not used in clustering process to reduce information redundancy. The rotational parameter was not a key parameter for mesocyclones as shown in previous section, and was not included either. The results are shown in Tables 4 and 5 for KOUN case and KINX case, respectively.

For the KOUN case, the maximin algorithm generates 23 clusters, of which 11 clusters have less than 1% of the total samples and are not placed into the table. Of the 12 clusters in table 4, clusters 1, 14, 18, 19 probably represent the true mesocyclones based on the large storm depth, low base height, large rotational velocity, high shear and high shear strength index. Also, the reflectivity values are high for these clusters. Clusters 10 and 13 most probably consist of false mesocyclones signatures due to the shallow depth of the storm, and small values of reflectivity. Similarly, clusters 5, 9, 12 and 17 probably contain false mesocyclones signatures due to small strength index, small rotational velocity and small reflectivity values. Clusters 15 and 21 are ambiguous having large GTGVD, rotational velocity and large reflectivity values, but also have high storm bases.

For the KINX case, the maximin algorithm generates 10 clusters, out of which 4 clusters have less than 1% of total samples and are not placed into the table. Clusters 7, 8, 9 and 10 most probably represent true mesocyclones due to high strength index, large rotational velocity, deep storm depth, high GTGVD and high reflectivity except for cluster 10. Clusters 1 and 5 more likely represent the false mesocyclones signatures due to weak strength index, and small rotational velocity.

The values in the last 4 columns in tables 4 and 5 are the number of samples, number of samples labeled as true (#1) and false (#0), and cluster index. For the clusters containing false signatures that we interpreted based on the feature characteristics, most of the samples are labeled as false samples except clusters 9 and 17 in the KOUN case and cluster 1 in the KINX case. In these three clusters, 15% to 25% of the samples are labeled as true samples. On the other hand, for mesocyclone clusters containing true signatures, a large percentage of the samples is labeled as false signatures. The mesocyclone samples from NSSL MDA were also examined using the maximin clustering algorithm. For the NSSL samples, 7 features were available, including base height, depth, mean diameter and shear, maximum diameter and shear, and maximum rotational velocity. Similar results were observed for both the KINX and KOUN cases. The results from the maximin cluster algorithm suggest that the clustering tendency for true mesocyclones feature set is not consistent with expert labeling. This explains the low CSI scores for the classifiers. It may also suggest

that additional features may be required to adequately characterize a mesocyclone so as to correctly classify the true mesocyclones.

The K-means clustering algorithm was also applied on the mesocyclone samples. Table 6 has the results from the KINX case with k = 10. In table 6, clusters 5, 6 and 9 are interpreted as true mesocyclone clusters due to their high strength index, high GTGVD, rotational velocity, low base height and deep storm depth. The rest of clusters are interpreted as false clusters due to one or many of the following reasons:, low strength index, low rotational velocity, shallow depth and high storm base. The results from the K-means are similar to the ones from maximin clustering.

5 Summary

In this study, a mesocyclone detection algorithm (UAH MDA) for WSR-88D radar data was developed. The performance of this MDA was compared with the NSSL MDA for two cases. The spatial coverage and feature distribution of identified mesocyclones from the two MDAs for the two cases were very similar.

The identified mesocyclone signatures for the two cases were then applied to two classifiers using 3 different feature selection methods to reach the best classification performance. The best CSIs for the two cases are 0.30 and 0.40, respectively. The classifier optimization using feature selection found that the most important features to classify the true mesocyclones are the vertical description of the signatures and the strength of the rotation. Two clustering algorithms were also applied to the identified mesocyclone samples. The clustering results show no clear separation between true and false mesocyclone signatures. This ambiguity could be attributed to several different factors, such as incorrect labeling and the need for additional features.

The UAH MDA running on a Pentium 2 PC requires about 1 second to process a full volume scan of radar data and is very fast compared to the six minutes required to create a volume scan. In addition, the UAH MDA algorithm is about twice as fast as a variant of the full NSSL MDA (which was computationally optimized for this study). Both the NSSL and UAH algorithms appear to identify similar mesocyclone features and feature locations. Although additional analysis is recommended to insure the UAH MDA is correctly identifying important mesocyclone signatures, it appears from this first study that the UAH MDA could be considered for use as a tool to mine the mesocyclones from large volumes of archived radar data. Clearly the National Climatic Data Center could benefit by such a tool. The UAH Mesocyclone Detection Algorithm is available as a stand-alone application and can be easily used by individual researchers.

6 References

[1] Burgess, D.W., 1976: Single Doppler radar vortex recognition. Part I: Mesocyclone signatures. Preprints, *17th conf. on Radar Meteorology*, Seattle, WA, Amer. Meteor. Soc., 97-103

[2] Crum, T.D., R.L. Alberty, and D.W. Burgess, 1993: Recording, Archiving, and Using WSR-88D Data, *Bull. Amer. Meteor. Soc.*, Vol. 74, No. 4, 645-653

[3] Stumpf, G.J., A. Witt, E. D. Mitchell, P. L. Spencer, J.T. Johnson, M.D. Eilts, K.W. Thomas, and D.W. Burgess, 1998: The National Severe Storms Laboratory Mesocyclone Detection Algorithm for the WSR-88D, *Weather and Forecasting*, Vol.13, 304-326

[4] Zrnic, D.S., D.W. Burgess, and L.D. Hennington, 1985: Automatic detection of mesocyclonic shear with Doppler radar, *J. Atmos. Oceaniic Technol*., Vol. 2, 425-438

[5] Desrochers, P.R., and R.J. Donaldson Jr., 1992: Automatic tornado prediction with an improved mesocyclone-detection algorithm, *Weather Forecasting*, Vol. 7, 373-388

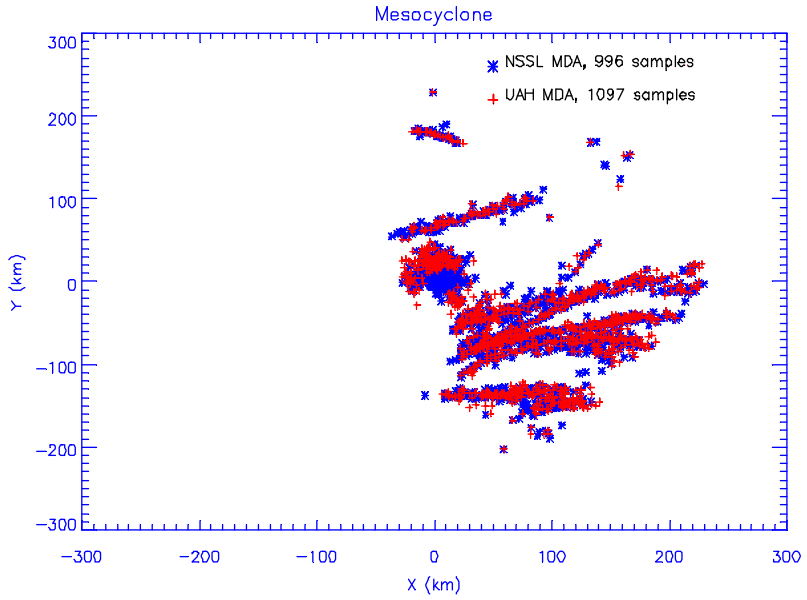
[6] Marzban, C. and G.J. Stumpf, 1996: A Neural Network for Tornado Prediction Based on Doppler Radar-Derived Attributes, *J. Applied Meteor*., Vol. 35, 617-626

[7] Donaldson, R.J., 1970: Vortex signature recognition by a Doppler radar, *J. Appli., Meteor*., Vol. 9, 661-670

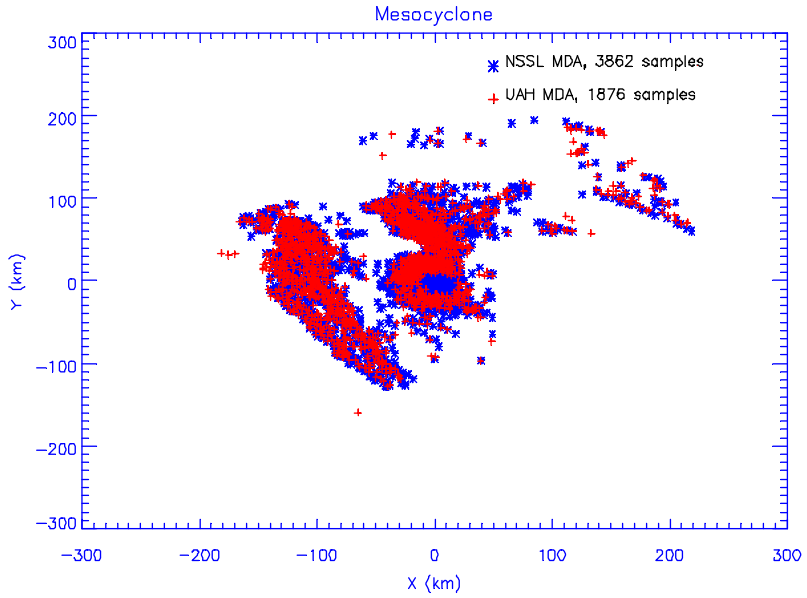
[8] Gonzalez, R.C, and R.E. Woods, 1992: Digital Image Processing, *Addison-Wesley Publishing Company Inc*. 716pp
[9] Hastie, T., R. Tibshirani, and J. Friedman, 2001: The Elements of Statistical Learning, - Data Mining, Inference, and Prediction, *Springer-Verlag publisher*, 533pp
[10] Tou, J.T. and R.C. Gonzalez, 1974: Pattern Recognition Principles, *Addison-Wesley Publishing Company*, 377pp
[11] Graves, S.J. and R. Ramachandran, 1999: ADaM: Algorithm Development and Mining System, *Earth Observing System Investigators Working Group (IWG) meeting*, Vail, CO, June 15-17
[12] Nair, U. J., J. Rushing, R. Ramachandran, K.-S. Kuo, S. J. Graves, and R. Welch, 1999: Detection of Cumulus Cloud Fields in Satellite Imagery, *The International Symposium on Optical Science, Engineering and Instrumentation, Denver*, July 1999
[13] Ramachandran R., H. Conover, S. J. Graves, and K. Keiser, 2000: Algorithm Development and Mining (ADaM) System for Earth Science Applications, *Second Conference on Artificial Intelligence, 80th AMS Annual Meeting*, Long Beach, CA, January 2000
[14] He Y., R. Ramachandran, X. Li, J. Rushing, H. Conover, S. J. Graves, W. Lyatsky, A. Tan, and G. Germany, 2003: Framework for Mining and Analysis of Space Science Data, *SIAM International Conference on Data Mining*, San Francisco, CA, May 1 - 3, 2003

7 Acknowledgement

This research was partially funded by NOAA through support from Stephen Del Greco of the National Climatic Data Center. The authors appreciate the assistance of Arthur Witt of the National Severe Storms Laboratory for his help in obtaining data and his assistance in interpreting the data. The Radar Software Library (RSL) (http://trmm-fc.gsfc.nasa.gov/trmm_gv/software/rsl/index.html) developed by NASA TRMM Satellite Validation Office was used to read WSR-88D data.



(a)



(b)

Figure 1. Spatial location distributions of the identified mesocyclone signatures detected from NSSL MDA and UAH MDA, respectively. (a) KOUN case, (b) KINX case

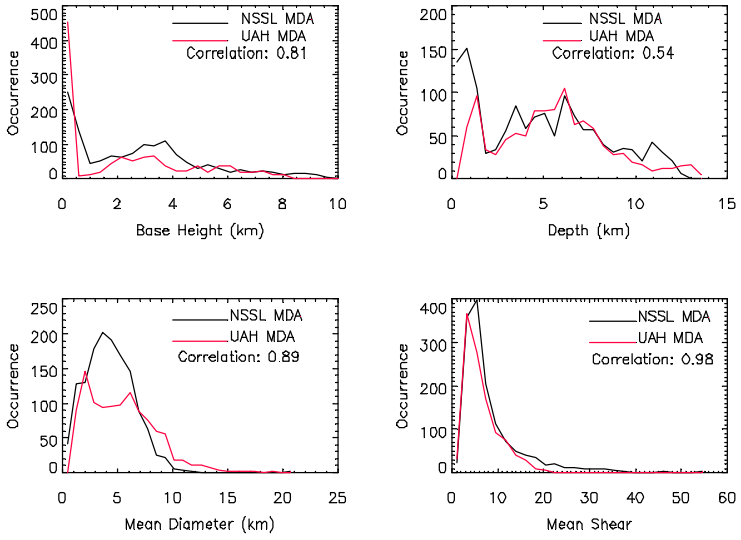


Figure 2. Comparison of feature distributions of the mesocyclone signatures detected from NSSL MDA and UAH MDA for KOUN case

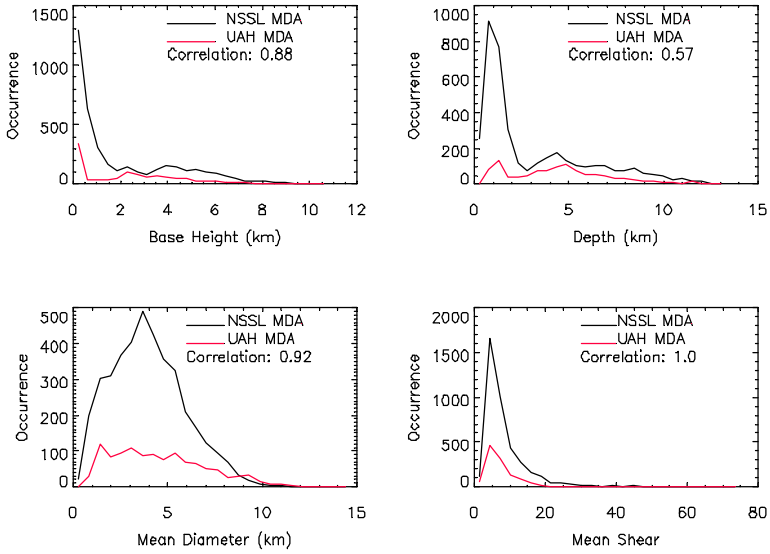


Figure 3. Comparison of feature distributions of the mesocyclone signatures detected from NSSL MDA and UAH MDA for KINX case

Table 1. The optimal CSI results of Bayes and SVM classifiers using three feature selection methods for the KOUN case

Feature Reduction	Classifier	BitString	CSI
Backward Elimination	Bayes	011110110000101000000	0.298578
Backward Elimination	SVM	000100000001000100010	0.307393
Forward Selection	Bayes	100001000000111000000	0.306604
Forward Selection	SVM	100000000000100000010	0.301527
Genetic Algorithm	Bayes	011001001000100101100	0.309278
Genetic Algorithm	SVM	010100111011100110010	0.313008

Table 2. The optimal CSI results of Bayes and SVM classifiers using three feature selection methods for the KINX case

Feature Reduction	Classifier	BitString	CSI
Backward Elimination	Bayes	001001101011101111110	0.400000
Backward Elimination	SVM	000111011111110111010	0.392283
Forward Selection	Bayes	000001010011100101110	0.410072
Forward Selection	SVM	000111011101110100010	0.397394
Genetic Algorithm	Bayes	00001101011100101110	0.423488
Genetic Algorithm	SVM	000011011111110100110	0.403226

Table 3. Frequency that features of mesocyclone signature are selected in feature selection process for the KINX and KOUN cases. There are a total of 12 feature selection occurrences for the two cases.

Feature Name	Occurrence	Percentage (%)
Base Height	10	83.3
Base Vrot	10	83.3
Maximum Strength Index	10	83.3
Depth	9	75
Height of Maximum GTGVD	9	75
Maximum Vrot	9	75
Core Base Height	8	66.7
Core Depth	8	66.7
Maximum Diameter	8	66.7
Base Diameter	7	58.3
Mean Diameter	7	58.3
Mean Shear	7	58.3
Base GTGVD	6	50
Maximum Reflectivity	6	50
Maximum Shear	5	41.7
Mean Reflectivity	5	41.7
Rotation	4	33.3
Base Shear	3	25
Mean Strength Index	3	25
Maximum GTGVD	1	8.3

Table 4 Results from maximin-distance clustering algorithm for the KOUN case

BsHt	depth	BsDm	BsShr	BsVrt	bsGTG	mnDm	mxDm	mnShr	mxShr	mxGTG	HtGTG	mxVrt	mnStr	mnRef	num	#1	#0	cls
0.6	6.9	1.9	16.8	14.5	26.7	4.1	6.9	10	20	29.4	1.5	17.1	3.8	36.6	16	4	12	1
5.7	5.2	5.5	3.6	8.8	13.2	6.2	7.5	3.6	4.4	16.2	8.8	11.5	2.4	28.8	59	1	58	5
0.9	6.6	7.9	2.8	10.1	14.2	8	10	3.1	4	16.9	4.2	12.5	2.5	39.4	194	28	166	9
0.5	1.3	2.2	10.5	9.8	12.1	2.1	2.6	10.5	13	13.9	1.2	11	2.4	4.5	188	2	186	10
5.3	6.3	13.4	1.8	11.4	13.3	12.5	15	2.2	2.8	15.7	8.8	14.7	2.5	35.4	17	1	16	12
0.1	1.5	2.8	11.9	13.6	17.4	2.3	3.1	14.9	20.9	25	1	17.7	4.2	5.4	25	2	23	13
0.6	8.6	2.5	18.7	21.9	41.8	4.7	7.3	10.9	19.6	44	2.4	23.4	6.3	45.3	30	13	17	14
5.6	5.6	5.6	6.3	14.9	21.2	6.3	7.8	5.7	7.3	23.7	7.9	16.5	4.4	40.9	64	14	50	15
2.5	5.1	3.2	6.7	9.5	14.7	3.7	4.9	6.8	9	17.9	5.2	12	2.6	37.4	222	41	181	17
0.4	11.2	9.1	2.6	11.1	15.6	8.5	11.4	4.1	6.2	25.9	8.6	20.2	4.5	41.9	15	0	15	18
0.8	7.6	5.1	7.1	15.5	25.1	6	8.2	5.7	8.2	27.5	2.7	18	4.1	42.2	117	41	76	19
5.1	6	8.1	3.5	13.2	16.4	8.1	10.1	3.9	5	19.9	8.7	16.4	3	39	117	24	93	21

Table 5 Results from maximin-distance clustering algorithm for the KINX case

bsHt	depth	BsDm	BsShr	BsVrt	bsGTG	mnDm	mxDm	mnShr	mxShr	mxGTG	HtGTG	mxVrt	mnStr	mnRef	num	#1	#0	cls
3.3	4.8	5.4	4.9	11	14.9	5.5	6.9	5	6.4	17.7	5.8	13.3	2.5	37.3	1037	201	836	1
0.6	1.2	2.1	11	9.9	12.7	2.1	2.7	10.8	13.5	14.5	1.2	11	2.4	7.7	571	5	566	5
3.4	6.1	2.3	21.4	21.9	41.6	3.9	6.5	13	22.5	42.7	4.6	22.6	5.2	40.1	20	9	11	7
1	7.9	6.6	6.3	16.8	24	7.1	9.6	5.7	8.2	28.7	4.2	21.1	3.9	41.5	135	80	55	8
1.3	6.5	3.8	7.2	12.5	17.9	3.1	4.5	11.6	20.7	32.6	4.9	18.5	4.4	43.6	25	3	22	9
0.7	3.2	3.3	12.4	17.7	26	3.5	4.4	10.3	13.2	27.4	1.7	18.6	4.6	22.5	47	7	40	10

Table 6 Results from k-mean clustering algorithm for the KINX case, k = 10

bsHt	depth	bsDm	bsShr	bsVrt	bsGTG	mnDm	mxDm	mnShr	mxShr	mxGTG	HtGTG	mxVrt	mnStr	mnRef	num	#1	#0	cls
5.7	4.7	5.6	6	15.4	19.5	5.6	6.7	5.9	7.2	21.7	7.8	17.2	3.7	40.9	126	34	92	1
4.1	4.8	7.9	3.3	12.1	14.3	7.5	9.3	3.9	5	17.7	7	15.1	2.2	38.9	187	43	144	2
3	3.8	3.5	6.1	9.4	13	3.9	5.1	6	7.7	15.5	5.1	11.6	2.3	34.9	361	41	320	3
0.4	1.1	1.6	16.8	12.3	16.4	1.8	2.2	14.6	18.6	17.8	0.7	13.2	3.1	5.9	184	1	183	4
2.4	6.1	4.6	10.3	19.7	31.4	5.1	6.9	9.4	14	38.1	4.6	23.5	5.5	39.7	79	28	51	5
1.1	6.7	5.2	6	13.5	19.3	5.5	7.5	6.1	8.9	24.8	4	17.5	3.5	39.4	235	106	129	6
1.1	7	9.6	2.4	10.8	14	9	11.2	2.7	3.5	16.7	4.8	12.7	2.1	37.5	126	27	99	7
0.6	1.1	2.4	8.7	9.1	11.2	2.3	2.8	9.2	11.4	13	1.2	10.2	2.2	6.5	369	1	368	8
1.4	4.3	2.1	28.3	26.2	45.6	3.1	4.7	19.6	30.7	47.4	2.3	27.1	6.9	27.9	29	7	22	9
3.5	4.7	4.5	4.9	9.3	14.4	4.7	5.7	4.8	5.9	16.1	5.8	10.6	2.2	35.5	180	23	157	10