P.9 THE DEVELOPMENT OF A STORM TYPE CLIMATOLOGY USING AN AUTOMATED STORM CLASSIFICATION SYSTEM

Angelyn Kolodziej*, Vailliappa Lakshmanan, and Travis Smith
CIMMS/Univ. of Oklahoma and NOAA/NSSL, Norman, OK
e-mail: angelyn@ou.edu

1. INTRODUCTION

Climatological storm information is important to National Weather Service (NWS) forecast offices where warning success is highly dependent on the predominant regional storm type (Guillot et al. 2007). Based on anticipated hazards, it is natural to categorize storms into types (supercells, squall lines, etc.) by their distinct characteristics. However, hand-classification of storm types can be quite tedious. For this reason, investigators have focused on techniques to automatically identify storm type from weather imagery to better predict potentially hazardous events (Morel et al. 2002; Rigo and Llasat 2004; Short et al. 2004; Baldwin et al. 2005; Guillot et al. 2007; Gagne et al. 2009).

This study describes the development of an algorithm for an automated multiscale storm classification system. The technique allows radar reflectivity to be partitioned meaningfully storm-by-storm, instead of pixel-by-pixel.

2.0 METHODOLOGY

The Warning Decision Support System-Integrated Information (WDSS-II; Lakshmanan et al. 2007) tool allows the capability to process, visualize, cluster, and classify data. In the first intricate step, archived WSR-88D radar data was manually categorized into storm types using WDSS-II (see Table 1). Storms were manually classified every hour for each case by “drawing” polygons around specific storms and dividing them into broad categories, “supercells,” “ordinary cells,” “short-lived convective cells,” and “convective lines.” These storms were considered the ideal storm types.

Supercells were identified by intense radar reflectivities associated with a mesocyclone, as indicated by mid-level azimuthal shear (Doswell and Burgess 1993; Smith and Elmore 2004). Ordinary cells tended to be intense, long-lasting storms with no rotation (Browning 1977). Short-lived convective cells included quick, pulse-like storms with moderately high reflectivity and no rotation. Convective lines were identified first by their linear appearance, then by the presence of an outflow boundary as indicated in the low-level azimuthal shear value (Doswell 2001).

A WDSS-II algorithm identified all storm clusters (greater than 30 dBZ). Storms were clustered into two scales using the WDSS-II algorithm, which employed a k-means clustering and watershed segmentation approach (Lakshmanan et al. 2003; Wilks 2006). Supercells, ordinary cells, and short-lived convective cells were segmented at a scale of 200-km^2, while convective lines were identified on a scale of 2000-km^2. Mean and maximum values of each gridded field were extracted from each cluster and are referred to as “attributes.” Example attributes are as follows: composite reflectivity, Maximum Estimated Hail Size (MESH), azimuthal shear, surface temperatures, and Most Unstable CAPE. A full dataset was collected that included thousands of clusters with radar and environmental attributes, as well as the subjectively determined storm type for all dates listed in Table 1.

The manual classification resulted in 259 short-lived convective cell clusters. Supercells were the smallest group, with 153 clusters. Since ordinary cells were comprised of non-rotating storms that could be subdivided into smaller groups (i.e. clustered storms, multicells, etc), they accounted for more than half of the clusters (565 clusters). On the 2000-km^2 scale, 89 convective lines were classified.

Five cases were used as a training dataset for the creation of an automated storm classifier system; the other five cases generated testing data.

Table 1 - Selected analysis dates and regions, with start time and duration. Asterisks (*) identify training dataset.

Date of case	Region (States affected)	Start Time/Duration
May 23, 2008*	Central Plains (CO, KS, NE, OK, TX, WY)	00 UTC / 15 hours
June 5, 2008*	Central Plains/Upper Midwest (CO, IA, IL, KS, MN, MO, NE, OK, SD, TX)	17 UTC / 7 hours
July 10, 2008	Upper Midwest/Upper Great Plains (IA, IL, MN, ND, SC, SD, WI)	18 UTC / 13 hours
December 9, 2008	Southeast (AL, LA, MS)	15 UTC / 15 hours
December 27, 2008	South Central Plains and Southeast (AR, IL, IN, KY, MO, OK, TN, TX)	06 UTC / 13 hours
February 10, 2009*	Southern Plains (AR, KS, LA, MO, OK, TX)	19 UTC / 12 hours
March 27, 2009	Southeast and Gulf Coast (AL, LA, MS, OK, TX)	18 UTC / 11 hours
April 19, 2009*	Southeast and Gulf Coast (AL, GA, MS)	18 UTC / 10 hours
May 8, 2009*	South Central Plains and Southeast (AR, IA, KS, KY, MO, SC, TN, TX, VA)	12 UTC / 22 hours
July 20, 2009	Central Plains (CO, KS, ND, NE, OK, TX)	20 UTC / 11 hours

The next process employed the Waikato Environment for Knowledge Analysis (WEKA) system to create a decision tree model to predict storm types using the hand-classification dataset. The decision trees were compared to the True Skill Statistic (TSS). The TSS was used as it reduces the penalty for incorrectly classified low-probability events and more strongly penalizes incorrectly classified high-probability events (Wilks 2006).

For the 200-km^2 scale and its three categories of storms (supercell, ordinary cell, and short-lived convective cell), the TSS was calculated by use of a 3x3 contingency table. For the larger 2000-km^2 scale indicating convective lines, the TSS was calculated using an ordinary 2x2 table.

Using a method described in Kolodziej et al. (2011), decision trees with a high and consistent TSS were chosen for both scales (see Figures 1 and 2).

3.0 RESULTS

The decision trees were automated and carried out over a seasonally diverse national dataset to obtain a preliminary climatology of storm types. The algorithm ran on every radar scan for fifteen days a month within an eleven-month period in 2009.

3.1 Regional storm type characteristics

The United States was divided into "regions" for analysis: the Northwest, Southwest, Northern Plains, Southern Plains, Midwest, Southeast, and Northeast. Regional storm-type frequency was investigated (see Figure 3). Storm clusters were evaluated within each region. Short-lived convective cells were the most prevalent storm type within each region. Percentages of storm types in Figure 3 refer to the ratio of regional storms as compared to the national average. The Southern Plains was the leading producer of supercells and convective lines. The next highest supercell-producing regions were the Northern Plains and Southeast, with 17% and 12.2%, respectively. The Southeast provided the highest fraction of ordinary cells and short-lived convective cells as compared to the national average. The Southeast also produced the second-highest fraction of convective lines.

Vertically Integrated Liquid (VIL), mid-level shear, low-level shear, composite reflectivity, Probability of Severe Hail (POSH), and Maximum Estimated Hail Size (MEHS or MESH) were also extracted from storms and compared within each region.

Regional storms varied in individual characteristics, specifically VIL, MESH, POSH, reflectivity, low- and mid-level shear, as follows:

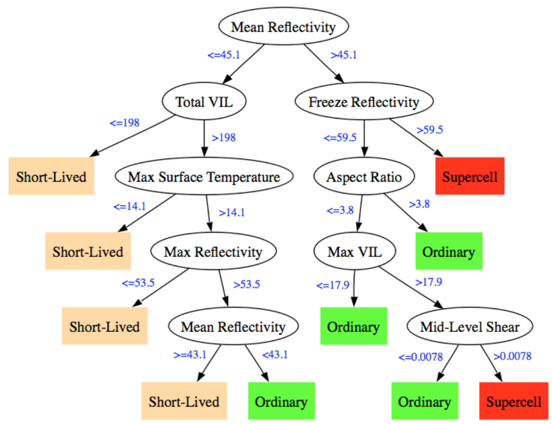

Figure 1 - Final decision tree for the 200-km scale. Reflectivity units are in dBZ, VIL in kg/m^2, temperature in Celsius, and mid-level shear in s^-1.

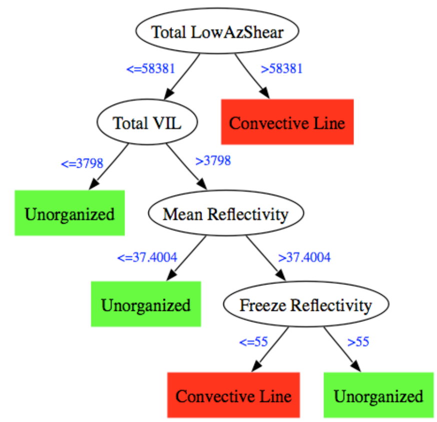

Figure 2 - Final decision tree for the 2000-km scale. Reflectivity units are in dBZ, VIL in kg/m^2, temperature in Celsius, and mid-level shear in s^-1

- Storms within the Southern Plains had consistently higher values for each characteristic than the national average.
- The Northeast tended to have lower storm characteristic values than the national average.
- Southeast supercells were of a different nature from those of the Southern Plains, with lower VIL, reflectivity, MESH, and POSH than the national average, but with high values of low- and mid-level shear.
- Southern U.S. VIL values tended to be higher than the northern United States.
- The western U.S. had higher values of MESH and POSH than the eastern U.S.
- Although variations in characteristics occurred in the Northwest, Southwest, and Midwest, the variations were not consistent among storm types.

3.2 Seasonal storm type characteristics

Each season was divided into seasonal subsets, where December and February are considered winter, March, April, and May are considered spring, June, July and August are considered summer, and September, October and November are autumn months. To account for the missing January data, a monthly average was determined for each season (see Table 2).

Distinct seasonal variations were identified. A primary peak for all storm types occurred in summer. A secondary peak occurred in spring for all storm types except ordinary cells, with the secondary peak for ordinary cells apparent in autumn. Although winter produced fewer smaller-scale storms, it had similar numbers of monthly convective lines when compared to the other seasons.

Table 2 - A monthly average of storm types for each season. Numbers represent "clusters" rather than individual storms

	Supercells	Ordinary Cells	Short-lived convective cells	Convective lines
Spring	1,334	25,247	100,646	909
Summer	2,900	64,017	118,993	960
Autumn	847	47,792	116,317	803
Winter	105	12,134	74,013	746

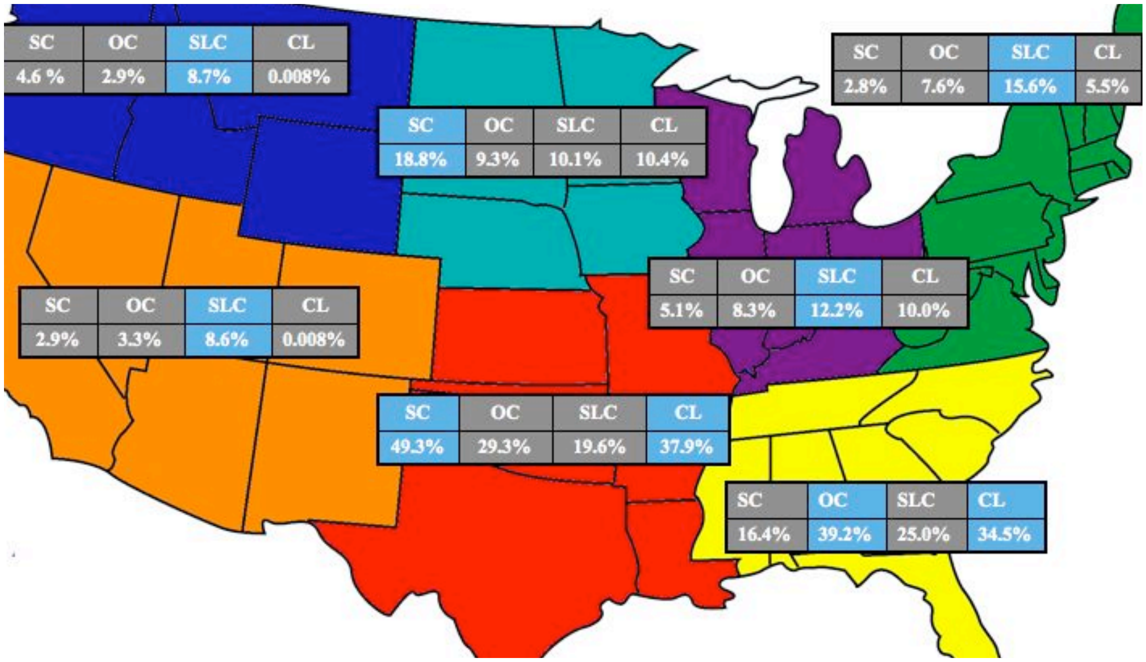

Figure 3 - The fraction of storm types that occurred within each region with the highest regional percentage(s) highlighted; SC stands for supercell, OC stands for ordinary cell, SLC stands for short-lived convective cell, and CL stands for convective line.

4.0 CONCLUSIONS

Storms were hand-classified into four storm categories for ten cases. WSR-88D radar reflectivity was partitioned into storm clusters, including short-lived convective cells, supercells, ordinary cells, and convective lines. A dataset of radar parameters, environmental information, and ideal storm types was developed from extracted attributes of each cluster. Decision tree models were generated from this dataset and used to create an automated storm type classification system.

The storm type classification system was implemented to generate a preliminary climatology of storm types. The system ran on every radar scan within an eleven-month period in 2009. The algorithm clustered each storm, extracted its attributes, and assigned to it a storm type. The dataset was used to determine statistics for the frequencies of storm types within these seven regions of the U.S.

4.1 Future Work

The individual storm characteristics identify distinctions between similar storm types, as well as the impact of locations upon storm type. Such climatological information is of considerable importance to National Weather Service forecast offices, where warning success tends to be highly dependent on the predominant regional storm type (Guillot et al. 2007). These developments and advances improve technology to generate more accurate and detailed nowcasts of storm type and storm severity.

5.0 REFERENCES

Baldwin, M., J. Kain, and S. Lakshmirvarahan, 2005: Development of an automated classification procedure for rainfall systems. *Mon. Wea. Rev.,* **133**, 844-862.

Browning, K. A., 1977: The structure and mechanism of hailstorms. *Meteor. Monogr.,* No. 16, Amer. Meteor. Soc., 1-39.

Doswell, C. A. III, 2001: Severe convective storms – an overview. *Severe Convective Storms. Meteor. Monogr.,* No. 50, Amer. Meteor. Soc., 1-26.

------- and D.W. Burgess, 1993: Tornadoes and tornadic storms: A review of conceptual models. *The Tornado: Its Structure, Dynamics,* *Prediction, and Hazards*, Amer. Geophys. Union, *Geophys. Monogr.*, **79**, 161-172.

Gagne, D. J., II, A. McGovern, and J. Brotzge, 2009: Classification of convective areas using decision trees. *J. Atmos. Ocean. Tech.*, **26**, 1341-1353.

Guillot, E. M., T. M. Smith, V. Lakshmanan, K. L. Elmore, and D. W. Burgess, 2007: Tornado and severe thunderstorm warning forecast skill and its relationship to storm type. Preprints, *24th Conf. on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*, New Orleans, LA, Amer. Meteor. Soc., 4A.3.

Kolodziej, A. G., V. Lakshmanan, T. Smith, M. Richman, 2011: An automated technique to categorize storm type. *J. Atmos. Tech.*, Submitted.

Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *J. Atmos. Res.*, **67**, 367-380.

-------, T. Smith, G. J. Stumpf, and K. Hondl, 2007: The warning decision support system - integrated information. *Wea. Forecasting*, **22**, 596-612.

Morel, C., and S. Senesi, 2002: A climatology of mesoscale convective systems over Europe using satellite infrared imagery. I: Methodology. *Quart. J. Roy. Meteor. Soc.*, **128**, 1953-1992.

Rigo, T., and M. C. Llasat, 2004: A methodology for the classification of convective structure using meteorological radar: Application to heavy rainfall events on the Mediterranean coast of the Iberian Peninsula. *Nat. Hazards Earth Syst. Sci.*, **4**, 59-68.

Short, D. A., J. E. Sardonia, W. C. Lambert, and M. M. Wheeler, 2004: Nowcasting thunderstorm anvil clouds over Kennedy Space Center and Cape Canaveral Air Force Station; *Wea. Forecasting*, **19**, 706-713.

Smith, T. M. and K. L. Elmore, 2004: The use of radial velocity derivatives to diagnose rotation and divergence. Preprints*, 11th Conf. on Aviation, Range, and Aerospace*, Hyannis, MA, Amer. Meteor. Soc., CD-ROM, P5.6.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Elsevier, Inc. 627 pp.