Objective analysis of storm morphology in radar data

Magdelyn Zoerner,[1,2] Robert Jackson,[2] and Bhupendra Raut[2,3]
[1)]*Iowa State University, Department of Geological and Atmospheric Sciences*
[2)]*Argonne National Laboratory, Environmental Science Division*
[3)]*Northwestern-Argonne Institute of Science and Engineering*

(*Electronic mail: maggiezoerner@gmail.com)

(Dated: 05 August 2022)

Large convective systems can be associated with substantial precipitation, damaging winds, hail, lightning, and other forms of severe weather. Therefore, it is essential that these storms are forecasted accurately. The morphology of these storms can provide important information regarding the atmospheric dynamics that support a storm's development and impact on its surrounding environment. However, numerical weather prediction models often have difficulties forecasting the morphology of these storms. A large amount of model and observational data needs to be compared objectively in order to identify the sources of error, which brings about the need for an automated method of morphological analysis. Current morphological analyses are performed either by manual classification of storm shape, or by an automated computation of a basic measurement. However, these methods are often inconsistent or too simplistic to accurately analyze the morphology of convective systems. To determine whether the automated identification and classification of storm morphology is feasible, this study uses image processing and analysis to compute various shape parameters of radar reflectivity echoes. This study also uses machine learning techniques to cluster the echoes based on their morphology. The results are encouraging, as they indicate that each cluster contains a set of reflectivity echoes with their own distinct morphology, which indicate that this technique is feasible.

I. INTRODUCTION

The influence of atmospheric thermodynamic conditions on convective system structure and evolution is well documented. Synoptic scale features, such as fronts and upper-level disturbances, can determine the organization of convective systems, as investigated in earlier studies[1,2]. Additionally, the process of self-organization of convective storms in the absence of synoptic scale features has been investigated. The process of self-organization exhibits a dependence on vertical profiles of wind shear, temperature, and humidity[3,4], among other parameters. The structure and morphology of a convective system can also have an influence on the surrounding environment by affecting the momentum fluxes and thermodynamic behavior at the surface[5]. The morphology of a system indicates key atmospheric dynamics that contribute to its development as well as the impact that a system may have on its environment.

Significant advances have been achieved in our understanding of the atmospheric conditions that influence the development and evolution of convective systems. However, numerical models have shown deficiencies in representing the morphology of convective systems[6], which emphasizes the value of developing a method to identify and classify convective system morphology automatically.

Current methods of convective system morphological analysis commonly consist of manual classifications based on observations. Earlier studies implement the use of radar reflectivity echoes, radiosonde data, and satellite imagery to classify the convective morphology based on visual and spatial characteristics. For example, 150 cases of squall lines that produced severe weather in Oklahoma were sampled and classified into four categories based on their radar and satellite signatures[7]. Similarly, rainstorms in Oklahoma were analyzed based on their symmetry about an axis normal and parallel to the midpoint of the system, determining the position of the stratiform rain region relative to the convective precipitation[8]. And, mesoscale convective systems in the southern Great Plains were classified into different types in a manner similar to previous studies, but with internal structural characteristics in mind[9]. However, these methods of morphological analysis do not adequately represent the shape characteristics of convective systems.

Our approach entails computing multiple parameters that determine shape (e.g., elongation, skewness, compactness, etc.), as morphological parameters quantify the shape characteristics of binary objects in images, making it possible to objectively classify and analyze the shapes. We then supplied these parameters to an algorithm that clustered the radar images together based on their commonalities in morphology.

II. DATA AND METHODS

A. Radar data

We obtained Next Generation Weather Radar (NEXRAD) Level II reflectivity data for the summer (June through August) of 2020 at the Lincoln, IL radar site (KILX). The data was obtained via Amazon Web Services (AWS)[10]. NEXRAD data is available as part of a research agreement with NOAA and AWS that grants any user access to free and accessible environmental data.

B. Image processing

Image processing is a widely used technique to analyze shape characteristics. This technique is used across many fields, including biology[11] for the analysis of biological phe-

nomena (i.e., behavior within a cell). The radar scans were gridded using the Python ARM Radar Toolkit (Py-ART)[12] to map the radar data onto a Cartesian grid. As radar reflectivity values greater than 35 dBz typically indicate severe convective precipitation[13], areas containing values less than 35 dBz were removed. The radar images were converted to binary. Then, the images were further thresholded to remove smaller echoes that contained less than 175 pixels, as this 175-pixel limit adequately removed the smaller regions of reflectivity while keeping the main region of precipitation in the image.

After the pre-processing was completed, we performed image processing using the OpenCV library[14]. Using OpenCV contour analysis features, we extracted the largest exterior contour of the image. This technique allows us to outline the shape of the reflectivity echo, then perform computations based on its shape. The following parameters were found using this technique along with additional OpenCV modules:

1. Geometric moments

Geometric moments are an established technique in image analysis that provide information about the intensity of an image at a certain pixel[15]. OpenCV returns the spatial, central, and central normalized moments, as shown in (1).

$$m_{pq} = \sum_{x=0}^{M} \sum_{y=0}^{N} x^p y^q f(x,y) \tag{1}$$

Where M and N represent the size of the pixel space f(x,y) on the order of (p + q).

2. Moment invariants

Moment invariants are widely used in image processing as they aid in the identification of features in an image. Moment invariants are invariant to rotation, translation, and scale of an image[16]. OpenCV returns a set of moment invariants as defined by Hu[17], shown:

$H_1 = \mu_{20} + \mu_{02}$

$H_2 = (\mu_{20} - \mu_{02})^2 + 4(\mu_{11})^2$

$H_3 = (\mu_{30} - 3\mu_{12})^2 + (\mu_{03} - 3\mu_{21})^2$

$H_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{03} + \mu_{21})^2$

$H_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{03} + \mu_{21})^2)$

$H_6 = (\mu_{20} - \mu_{02})((\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03})$

$H_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) + (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{03} + \mu_{21})^2)$

3. Area

The number of pixels in the reflectivity echo.

4. Perimeter

The number of pixels in the boundary of the reflectivity echo.

5. Center of mass

The center of mass was calculated using geometric moments:

$$\bar{x} = \frac{M_{10}}{M_{00}}, \bar{y} = \frac{M_{01}}{M_{00}} \tag{2}$$

6. Skewness

The third order normalized geometric moment represents the skewness. An image that is positively skewed in the x-direction contains much of its mass in the positive x-domain.

7. Elongation

Elongation represents the height and width ratio of the bounding box of the reflectivity echo.

$$\text{elongation} = \frac{\text{bounding box width}}{\text{bounding box height}} \tag{3}$$

8. Equivalent diameter

Equivalent diameter is the diameter of a circle with the same diameter as the reflectivity echo.

$$\text{equivalent diameter} = \sqrt{\frac{(4 \times \text{area})}{\pi}} \tag{4}$$

9. Compactness

Compactness represents the relationship between the area of the reflectivity echo and the area of a circle with the same perimeter as the echo[18].

$$\text{compactness} = \frac{\text{perimeter}^2}{4\pi(\text{area})} \tag{5}$$

10. Roundness

Roundness represents the relationship between the area of the reflectivity echo and the area of a circle with the same convex perimeter as the echo[18].

$$\text{roundness} = \frac{4\pi(\text{area})}{\text{convex perimeter}^2} \tag{6}$$

11. Eccentricity

Eccentricity represents the ratio of the major axis length to the minor axis length of the reflectivity echo.

$$\text{eccentricity} = \frac{\text{major axis length}}{\text{minor axis length}} \tag{7}$$

12. Deviation from circle

The first normalized central moment represents the deviation from circular shape.

C. Statistical analysis: Correlation matrix

A correlation matrix displays the correlation coefficient between variables. Parameters with little to no correlation with other parameters likely tell more information about the shape. For example, if a parameter is closely related to another parameter, the two parameters likely tell similar information about the shape. However, a parameter with no relation to other parameters tells information about the shape that other parameters cannot.

D. Machine learning

Principal component analysis and K-Means clustering were performed using the scikit-learn[19] library.

1. Principal component analysis

Principal component analysis (PCA) reduces the dimensionality of the dataset. This technique allows for easier interpretation of the data while minimizing the amount of data that is lost[20]. We performed principal component analysis before clustering the dataset. In the next iterations of this study, we would like to cluster the dataset without principal component analysis and compare the results.

2. Clustering

Clustering is a machine learning technique used to group data points together based on a similar characteristic that aims to identify groups of homogeneity within a heterogeneous data set[21]. To choose the number of clusters, we used the elbow method (Fig. 1). This method is used to select the optimal number of clusters by plotting the explained variation versus the number of clusters, then selecting the "elbow" of the curve as the optimal number of clusters[22]. The four clusters found using the principal component analysis from the data set are shown in Figure 2.

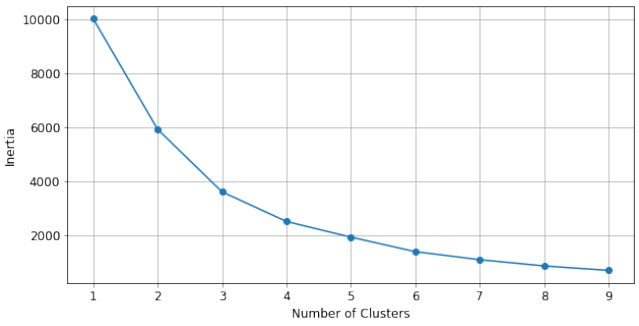

FIG. 1. Plot using the elbow method to determine the optimal number of clusters.

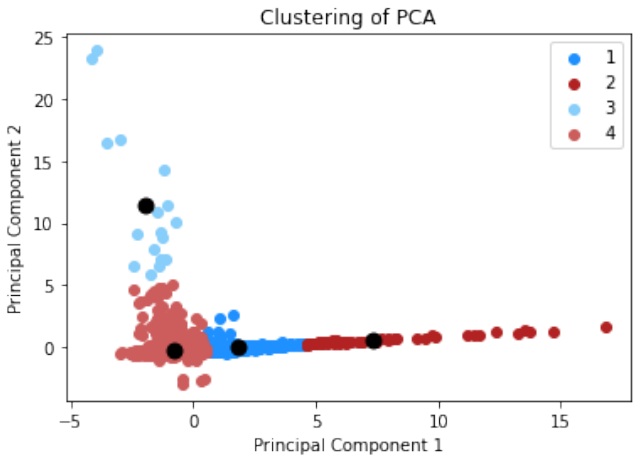

FIG. 2. Clustering of radar images after principal component analysis.

III. RESULTS AND DISCUSSION

The radar images were categorized into four clusters based on their commonalities in shape. Each of the four clusters exhibit their own set of distinct, unique morphological characteristics. Cluster 1 contained images with a more linear characteristic, with much of the reflectivity distributed along the y-axis. An example of a reflectivity echo found in Cluster 1 is shown in Figure 3. Cluster 2 contained similar linear characteristics, though much of the reflectivity was distributed along the x-axis, rather than the y-axis. An example of a radar image found in Cluster 2 is shown in Figure 4. The radar images in Cluster 3 did not show a distribution along any certain axis. These reflectivity echoes were more widespread with greater surface areas. An example of a radar image found in Cluster 3 is shown in Figure 5. Cluster 4 contained radar images with smaller, isolated groups of precipitation. These echoes were characterized by their relatively small surface areas and ovular shape. An example of a radar image found in Cluster 4 is shown in Figure 6.

The radar data classifications were concentrated in certain clusters. Cluster 4 contains 1295 radar images, which represents 77% of the data. Cluster 1 contains 317 images, repre-

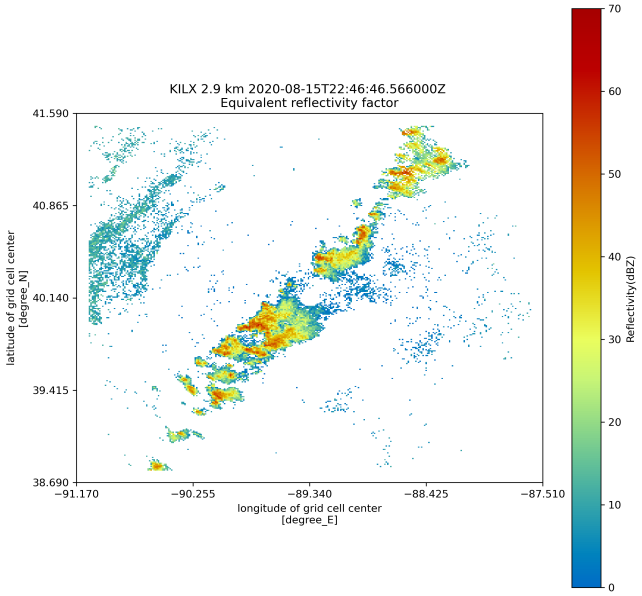

FIG. 3. Example of radar image in Cluster 1.

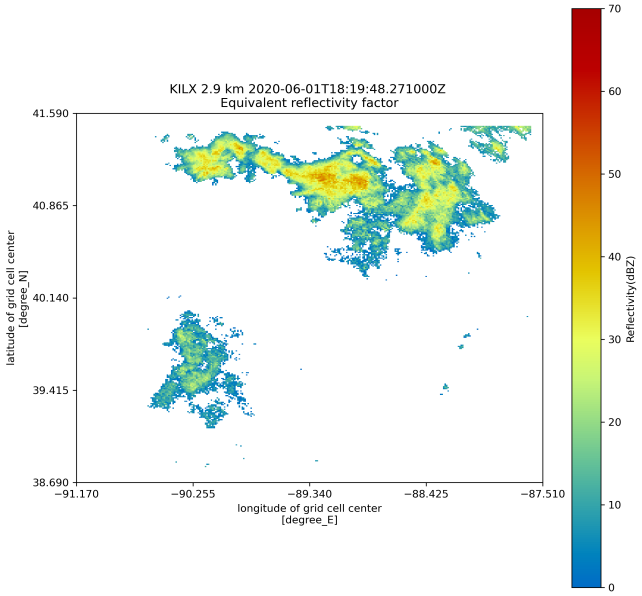

FIG. 4. Example of radar image in Cluster 2.

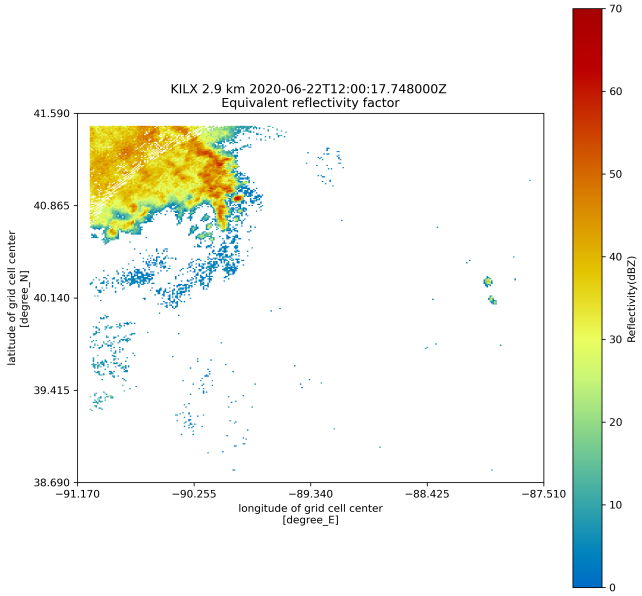

FIG. 5. Example of radar image in Cluster 3.

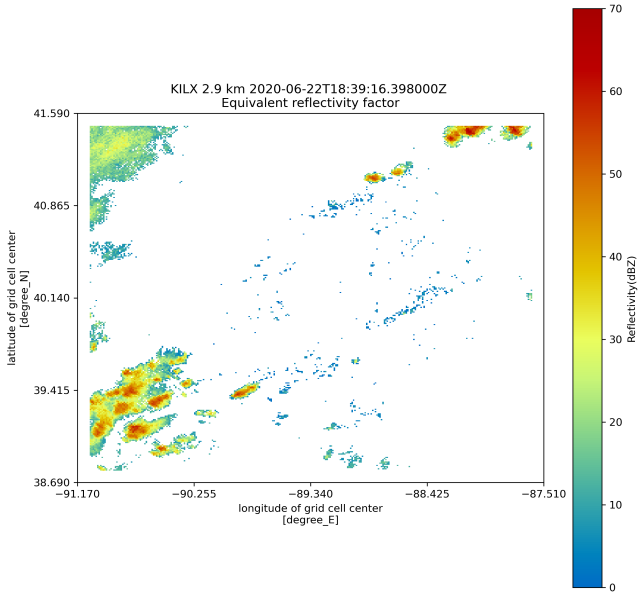

FIG. 6. Example of radar image in Cluster 4.

senting 19% of the data. Cluster 2 contains 62 radar images, which represents 4% of the data. Cluster 3 contains 17 radar images, representing 1% of the data (Fig. 7).

Several of the parameters are correlated with others (Fig. 8). Compactness and roundness are strongly related inversely (r = -0.78). Compactness is positively related to perimeter, area, and equivalent diameter, with (r = 0.63), (r = 0.82), (r = 0.65), respectively. Meanwhile, roundness is negatively related to perimeter, area, and equivalent diameter, with (r = -0.38), (r = -0.54), and (r = -0.45) respectively. Skewness in both the x- and y-directions are inversely related to deviation from circle with (r = -0.58) and r = (-0.58), respectively. Equivalent diameter is closely related to area and perimeter, with (r = 0.95) and (r = 0.94), respectively. While some correlations are strong, they are not perfectly positively nor negatively correlated. For example, some reflectivity echo shapes that are compact are not round, but this is not true for every compact case. This reinforces the idea that each reflectivity echo is unique, and can be defined by its own set of morphological characteristics.

In this study, we wanted to determine whether the automated process of identifying and categorizing convective systems based on their morphology would be feasible. Each of the four clusters found contains a set of radar images with

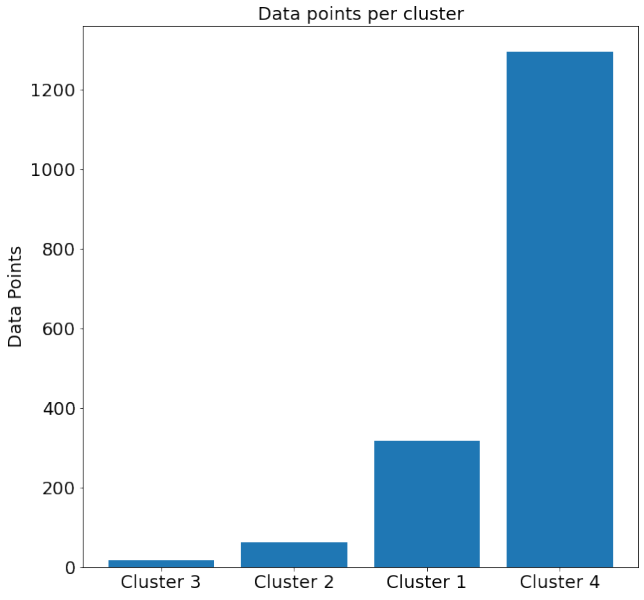

FIG. 7. Bar chart representing the number of data points in each of the four clusters.

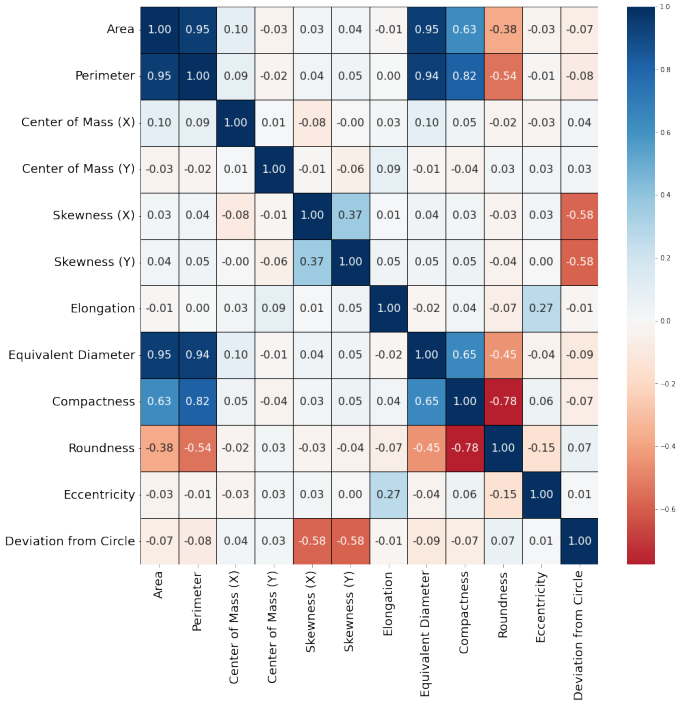

FIG. 8. Correlation matrix representing the degree of correlation between all parameters. 1 represents a perfectly positive correlation, 0 represents no correlation, -1 represents a perfectly negative correlation.

unique morphology. These results are encouraging, as they indicate that this process is feasible. However, this process requires refinement for future iterations. Special cases of convective systems, such as broken linear storms, require additional analysis that were not considered in this study. Future iterations would also require a more refined process of removing noise from the radar image to better represent the system. Additionally, this study only focuses on the convective region of the system, and does not consider stratiform rain, which can tell much about the development, evolution, and behavior of a convective system. Had time allowed, we would have liked to compute more morphological parameters to better characterize the shapes of the convective systems. Furthermore, more analysis can be performed to optimize the parameters that are more informative, and eliminate others.

IV. FUTURE WORK

With future iterations of this study, we hope to provide the United States Department of Energy Atmospheric Radiation Measurement (ARM) user facility with a database of these convective classifications. These classifications can be used to advance atmospheric research and improve numerical prediction models. We also hope to provide an open-source Python package that automatically performs morphological analysis for atmospheric data. As many of the parameters found are invariant to scale, rotation, and translation, this process can be applied to different types of imagery besides radar (i.e., satellite).

V. ACKNOWLEDGEMENTS

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships (SULI) program. Special thanks to my mentors, Robert Jackson and Bhupendra Raut for all their support.

VI. BIBLIOGRAPHY

[1]Carbone, Richard E. “A Severe Frontal Rainband. Part I. Stormwide Hydrodynamic Structure.” Journal of the Atmospheric Sciences, vol. 39, no. 2, 1 Feb. 1982, https://doi.org/10.1175/1520-0469(1982)039<0258:ASFRPI>2.0.CO;2.

[2]Hobbs, Peter V., et al. “The Mesoscale and Microscale Structure and Organization of Clouds and Precipitation in Midlatitude Cyclones. I: A Case Study of a Cold Front.” Journal of the Atmospheric Sciences, vol. 37, Mar. 1980, pp. 568–596., https://doi.org/10.1175/1520-0469(1980)037<0568:TMAMSA>2.0.CO;2.

[3]Moncrieff, Mitchell W. “The Dynamical Structure of Two-Dimensional Steady Convection in Constant Vertical Shear.” Quart. J. Roy. Meteor. Soc., vol. 104, 1978, pp. 543–567., https://doi.org/10.1002/qj.49710444102.

[4]Weisman, M. L., and J. B. Klemp. “The Dependence of Numerically Simulated Convective Storms on Vertical Wind Shear and Buoyancy.” Monthly Weather Review, vol. 110, 1982, pp. 504–520., https://doi.org/10.1175/1520-0493(1982)110<0504:TDONSC>2.0.CO;2.

[5]LeMone, Margaret A. “The Role of Environmental Shear and Thermodynamic Conditions in Determining the Structure and Evolution of Mesoscale Convective Systems during TOGA COARE.” Journal of the

Atmospheric Sciences, vol. 55, no. 23, 1 Dec. 1998, pp. 3493–3518., https://doi.org/10.1175/1520-0469(1998)055<3493:TROESA>2.0.CO;2.

[6]Moncrieff, Mitchell W, et al. “Simulation, Modeling, and Dynamically Based Parameterization of Organized Tropical Convection for Global Climate Models.” Journal of the Atmospheric Sciences, vol. 74, no. 5, 1 May 2017, pp. 1363–1380., https://doi.org/10.1175/JAS-D-16-0166.1.

[7]Bluestein, Howard B., and Michael H. Jain. “Formation of Mesoscale Lines of Precipitation: Severe Squall Lines in Oklahoma during the Spring.” Journal of Atmospheric Sciences, vol. 42, no. 16, Aug. 1985, https://doi.org/10.1175/1520-0469(1985)042<1711:FOMLOP>2.0.CO;2 .

[8]Houze, Robert A., et al. “Mesoscale Organization of Springtime Rainstorms in Oklahoma.” Monthly Weather Review, vol. 118, no. 3, 1 Mar. 1990, pp. 613–654., https://doi.org/10.1175/1520-0493(1990)118<0613:MOOSRI>2.0.CO;2.

[9]Blanchard, David O. “Mesoscale Convective Patterns of the Southern High Plains.” Bulletin of the American Meteorological Society, vol. 71, no. 7, 1990, pp. 994–1005., https://doi.org/10.1175/1520-0477(1990)071<0994:MCPOTS>2.0.CO;2.

[10]NEXRAD on AWS was accessed from https://registry.opendata.aws/noaa-nexrad.

[11]Uchida,Seiichi. "Image processing and recognition for biological images". Development, Growth, I& Differentiation, vol. 55, no. 4, 2013, pp. 523-549, https://doi.org/10.1111/dgd.12054.

[12]Helmus, J.J., and S.M. Collis. “The Python ARM Radar Toolkit (Py-ART), a Library for Working with Weather Radar Data in the Python Programming Language.” Journal of Open Research Software, vol. 4, no. 1, 2016, https://doi.org/10.5334/jors.119.

[13]Roberts, Rita D., and Steven Rutledge. “Nowcasting Storm Initiation and Growth Using GOES-8 and WSR-88D Data.” Weather and Forecasting, vol. 18, no. 4, 2002, pp. 562–584., https://doi.org/10.1175/1520-0434(2003)018<0562:NSIAGU>2.0.CO;2.

[14]Bradski, G. “The OpenCV Library.” Dr. Dobb’s Journal of Software Tools, 2000.

[15]Li, Erbo, et al. “Shape DNA: Basic Generating Functions for Geometric Moment Invariants.” ArXiv, 2017, https://doi.org/10.48550/arXiv.1703.02242.

[16]Nasrudin, Mohd Wafi, et al. “Moment Invariants Technique for Image Analysis and Its Applications: A Review.” Journal of Physics: Conference Series 1962 , 2021, https://doi.org/10.1088/1742-6596/1962/1/012028.

[17]Hu, Ming-Kuei. “Visual Pattern Recognition by Moment Invariants.” IRE Transactions on Information Theory, vol. 8, no. 2, 1962, https://doi.org/10.1109/TIT.1962.1057692.

[18]Wirth, Michael A. “Shape Analysis I& Measurement.” 2004, University of Guelph, Computing and Information Science Image Processing Group.

[19]Pedregosa , Fabian, et al. “Scikit-Learn: Machine Learning in Python.” Journal of Machine Learning Research, vol. 12, 2011, pp. 2825–2830.

[20]Jolliffe, Ian T., and Jorge Cadima. “Principal Component Analysis: A Review and Recent Developments.” Philosophical Transactions of the Royal Society A, vol. 374, 2016, https://doi.org/10.1098.rsta.2015.0202.

[21]Eick, C. F., et al. “Supervised Clustering - Algorithms and Benefits.” 16th IEEE International Conference on Tools with Artificial Intelligence, 2004, https://doi.org/10.1109/ICTAI.2004.111.

[22]Ketchen, D.J., and C.L. Shook. “The Application of Cluster Analysis in Strategic Management Research: Analysis and Critique.” The Journal of Strategy and Management, vol. 17, no. 6, 1996, pp. 441–458., https://doi.org/10.1002/(SICI)1097-0266(199606)17:6