Clustering analysis of multi-channel microwave satellite imagery to classify tropical cyclone intensity

Dhrumil Shukla[1*], Natalia N. Solorzano[1,2], Barnabas Bede[1], Jeremy N. Thomas[1,2,3], and Connor Bracy[2,3]

1. DigiPen Institute of Technology, Redmond, WA, USA
2. NorthWest Research Associates, Redmond, WA, USA
3. University of Washington, Seattle, WA, USA

Introduction

This work uses machine learning (ML) methods applied to 11 channels of passive microwave imagery to identify distinct convection patterns commonly observed in tropical cyclones (TCs) in different storm intensity ranges for all basins 2006-2019.

Numerous studies have analyzed TCs to understand how they form, what factors affect them and what can be done to reduce their damages. Infrared (IR) and visible imagery are obtained from geostationary satellites, which can provide continuous monitoring of TCs as they evolve. Geostationary satellite datasets also present some drawbacks. IR and visible channels are not ideal for detecting hydrometeors, and relevant features of the TCs such as the eye can be obscured by high cirrus clouds.

Low earth orbit (LEO) satellites supply microwave imagery. Images of a certain location are typically available twice a day. This sparsity of observations constitutes the drawback of such dataset. However, microwave radiation can pass through clouds detecting middle and lower level features beneath the high cirrus clouds. This allows finer details of the convective structure of TCs to be observed.

Based on patterns observed on satellite imagery, intensity estimation techniques have been developed. The most widely used method to determine the intensity from satellite data is the Dvorak Technique (Dvorak, 1984), which makes use of visible and IR imagery and relies on a "decision tree" methodology based on human interpretation. An updated intensity estimation technique which uses passive microwave imagery from LEOs is the enhanced and automated version of the Advanced Dvorak Technique (Olander and Velden, 2019). This new version motivated our work.

Recently, ML techniques have been utilized to understand satellite imagery and address the difficult problem of estimating TC intensity (Nelson, 2019). We use clustering, a type of unsupervised ML, to discover similarity between datapoints based on various criteria. Clustering techniques can be classified into centroid-based methods (k-means, c-means), distribution-based models (expectation maximization), fuzzy k-means (Xu, 2016), density-based methods such as DBSCAN (Ester, 1996), and hierarchical clustering (Jain, 2010). In the present work, we cluster TCs using various clustering algorithms and different criteria for cluster similarity.

[*] Corresponding author address: Dhrumil Shukla, DigiPen Institute of Technology, Redmond, WA 98052; e-mail: dhrumil.shukla@digipen.edu

Our work applies similar methods and reproduces results of Nelson (2019) who used the 91GHz channel on Special Sensor Microwave Imagery/Sensor Imagery (SSMIS) provided by Defense Meteorological Satellite Program (DMSP) satellites from February 2006 to 2017. In Nelson (2019), k-means and hierarchical clustering methods are implemented to identity patterns observed while using the Advanced Dvorak Technique to infer storm intensity. We expand the analysis of Nelson (2019) to 11 SSMIS microwave channels and use additional clustering techniques.

Data and Methodology

We use the SSMIS channels on the DMSP satellites (F16 - F19) for the years 2006 - 2019 for all TCs in the North Atlantic, Central Pacific, East Pacific, Indian Ocean, Southern Hemisphere, and West Pacific basins. The data are Level 1C- R (Common Calibrated Brightness Temperatures Collocated) data available from NASA (https://pmm.nasa.gov/data-access/downloads/gpm).
Channel details for SSMIS are shown in Figure 1. Each dataset consists of latitude, longitude, and the SSMIS brightness temperatures (Tbs) channels in Figure 1. For each storm, best track location, winds, and pressure are provided by Naval Research Laboratory or the SHIPS model. An example of a real-time tool using SSMIS imagery along with lightning is WWLLN-TC described in Solorzano et al. (2018).

Center Frequencies (GHz)	19.35	22.235	37.0	91.655	150.0	183.31+/-1	183.31+/-3	183.31+/-7
Polarization	V/H	V	V/H	V/H	H	H	H	H
Bandwidth (MHz)	350	410	160	1410	1640	510	1020	1530
Sensitivity (K)	0.35	0.45	0.22	0.19	0.53	0.38	0.39	0.56
Instantaneous Field of View (km x km)	73x47	73x47	41x31	14x13	14x13	14x13	14x13	14x13
Sampling Interval (km x km)	45x74	45x74	28x45	13x16	13x16	13x16	13x16	13x16

Figure 1 - SSMIS Characteristics

The following pre-processing is done prior to clustering:

- Data must be over the ocean, since there are significant changes in the TCs structural patterns when it approaches over land.
- Files with missing or incorrect data were removed.
- Data must be between the start and end dates provided in the best track files.
- Images are cropped to 800 x 800 km squares about the storm center.

A total of 11 images (one for each channel) are created if they follow these prerequisites. Figures 2 and 3 show how an image looks before and after applying pre-processing.

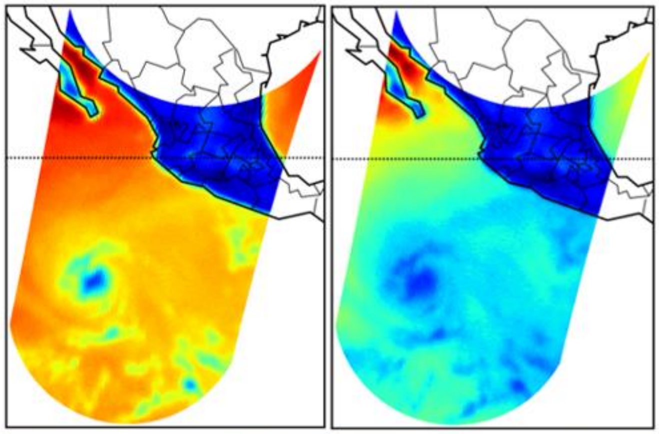


Figure 2 - 2014 EPAC Amanda Storm of 19H and 22V GHz frequencies plotted without applying pre-requisites

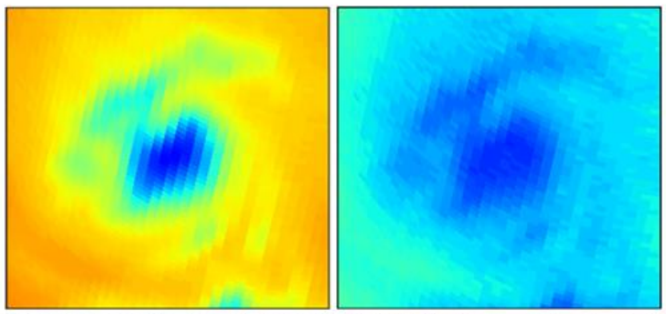


Figure 3 - 2014 EPAC Amanda Storm of 19H and 22V GHz frequencies plotted after applying pre-requisites

Three clustering techniques have been explored: fuzzy k-means, k-means, and hierarchical clustering. For this preliminary study, only hierarchical clustering is implemented, which allows the change of additional parameters: linkage method and distance metric[†]. Figure 4 is a hierarchical clustering dendrogram, which shows how seven clusters are merged with one another and a single cluster is created. Clusters are merged based on the linkage method and distance metric we define.

[†] https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html

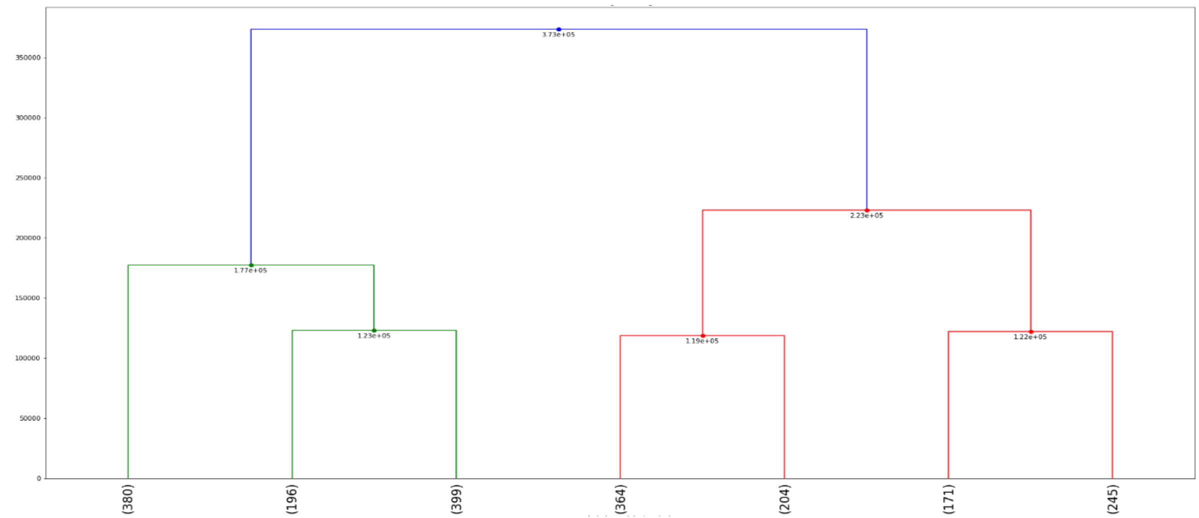


Figure 4 - Dendrogram for 19H frequency and Cluster Size 7

After clustering, we define a selection criterion to identify significant clusters and patterns. The method is:

- Only images whose silhouette value (Rousseeuw, 1987) is positive are valid.
- Images having the same pattern must be from at least three different storms.
- Patterns were identified by visually inspecting each cluster.

Results

In these preliminary results, clustering is performed on the all TCs in the North Atlantic basin for 19H, 91H, 150H and 183$\pm$1H GHz channels from 2006 to 2019 (Figures 5-8). The method used is hierarchical clustering, with linkage method set as Ward, distance method equal to Euclidean and cluster size set to seven. To analyze common pattern across different channels, the image created should be present in all the channels; i.e. an image created in 19H GHz must be present in other channels. After this process, significant patterns are identified as shown in Figures 5-8 along with storm intensity using the Saffir-Simpson and Dvorak scales.

We note that in clusters formed for each channel:

- Color is given priority by the algorithm. i.e. an image having blue color, and another with green-yellow color will be in different clusters even though they have a similar pattern.
- Intensity bins from T-5.0 to T-6.0 have similar patterns and are in the same clusters for 19H, 91H and 150H GHz.
- Patterns for 91H GHz are similar to the one achieved by Nelson, (2019).

Figure 5 - Significant patterns identified by hierarchical cluster analysis from T-3.0 to T-6.0, when clustering by all pixels for Frequency channel 19H for all storms from Atlantic basin from 2006 to 2019.

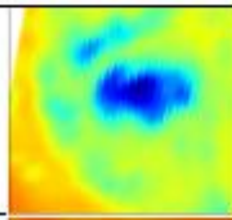
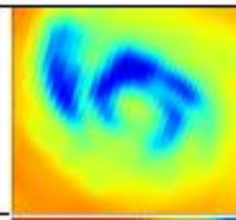
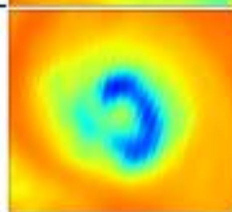
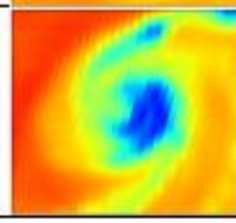
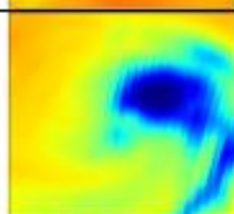
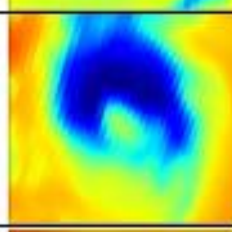
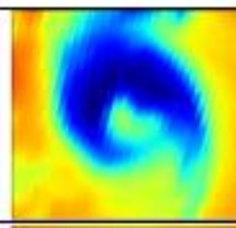
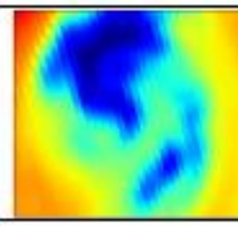
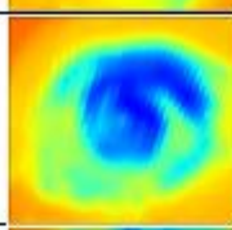
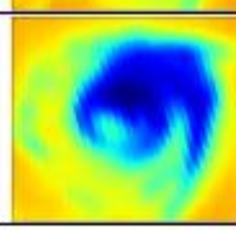
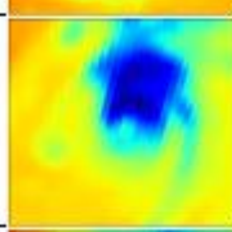
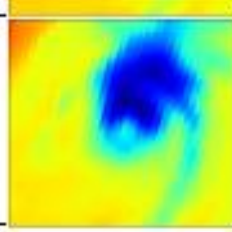
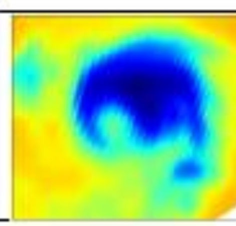
Saffir Simpson Scale	Intensity Number	Images (For frequency 19H)		
Tropical Storm	T - 3.0			
	T – 3.5			
Cat 1	T – 4.0			
	T – 4.5			
Cat 2	T – 5.0			
Cat 3	T – 5.5			
Cat 4	T – 6.0			

Figure 6 - Significant patterns identified by hierarchical cluster analysis from T-3.0 to T-6.0, when clustering by all pixels for Frequency channel 91H for all storms from Atlantic basin from 2006 to 2019.

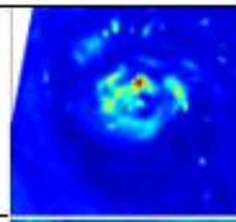
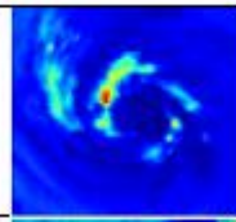
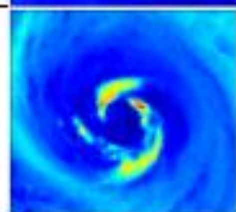
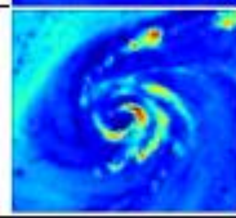
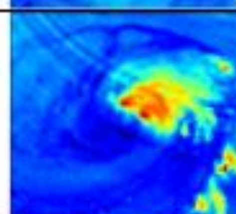
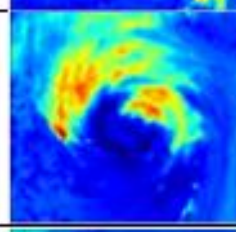
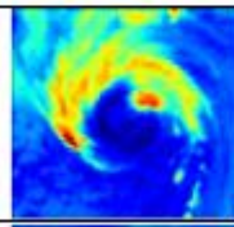
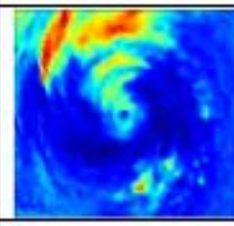
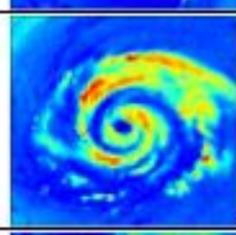
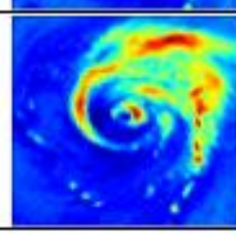
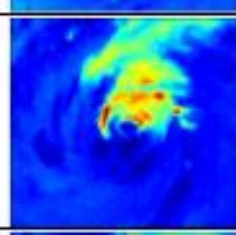
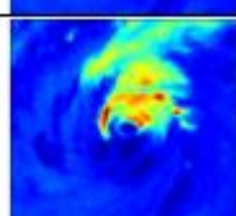
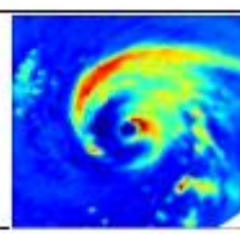
Saffir Simpson Scale	Intensity Number	Images (For frequency 91H)		
Tropical Storm	T - 3.0			
	T – 3.5			
Cat 1	T – 4.0			
	T – 4.5			
Cat 2	T – 5.0			
Cat 3	T – 5.5			
Cat 4	T – 6.0			

Figure 7 - Significant patterns identified by hierarchical cluster analysis from T-3.0 to T-6.0, when clustering by all pixels for Frequency channel 150H for all storms from Atlantic basin from 2006 to 2019.

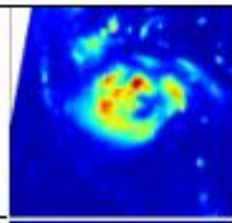
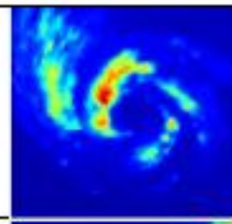
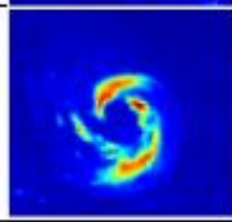
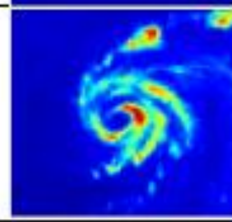
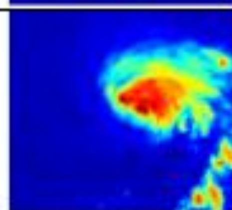
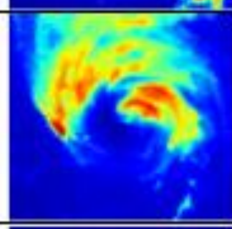
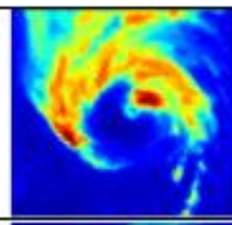
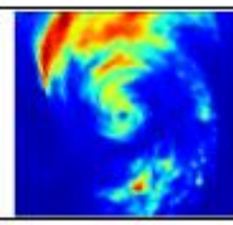
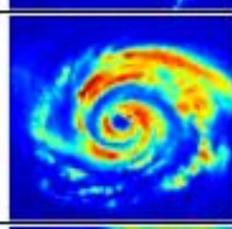
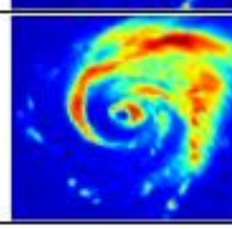
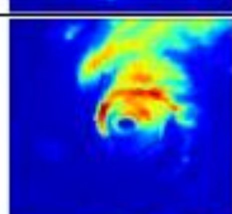
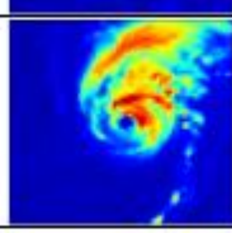
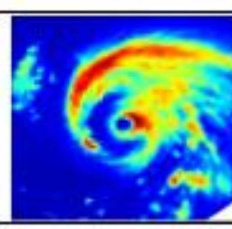
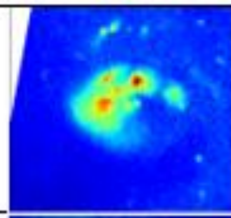
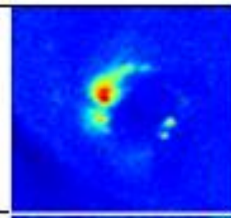
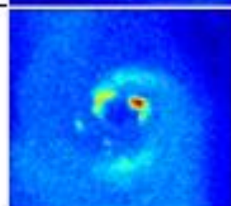
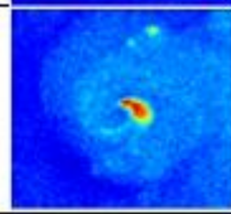
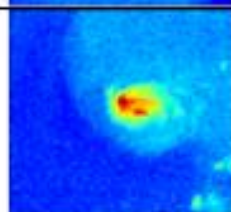
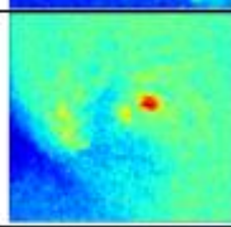
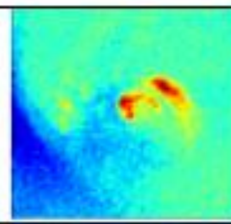
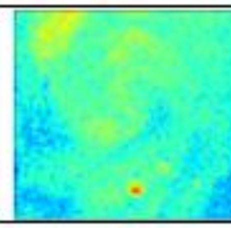
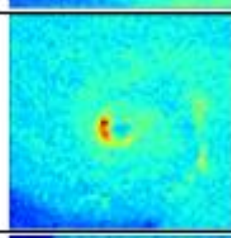
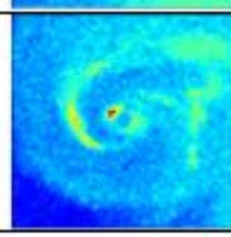
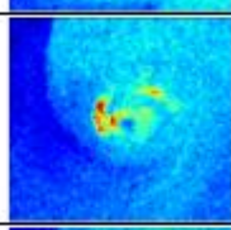
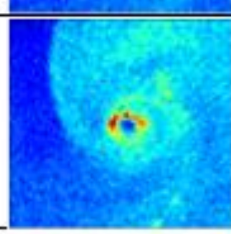
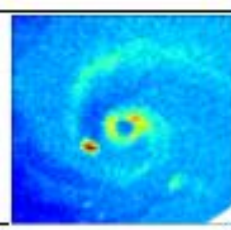
Saffir Simpson Scale	Intensity Number	Images (For frequency 150H)		
Tropical Storm	T - 3.0			
	T – 3.5			
Cat 1	T – 4.0			
	T – 4.5			
Cat 2	T – 5.0			
Cat 3	T – 5.5			
Cat 4	T – 6.0			

Figure 8 - Significant patterns identified by hierarchical cluster analysis from T-3.0 to T-6.0, when clustering by all pixels for Frequency channel 183_1H for all storms from Atlantic basin from 2006 to 2019.

Saffir Simpson Scale	Intensity Number	Images (For frequency 183_1H)		
Tropical Storm	T - 3.0			
	T – 3.5			
Cat 1	T – 4.0			
	T – 4.5			
Cat 2	T – 5.0			
Cat 3	T – 5.5			
Cat 4	T – 6.0			

Conclusions

In this preliminary analysis of TCs from 2006-2019 in the North Atlantic basin, clustering of higher frequency channels (150H and 183$\pm$1H GHz) better depict intense convection structures when compared to lower frequency channels (19H GHz). Addition of density-based clustering (Ester, 1996) might be helpful in exploring the patterns that hierarchical clustering is not able to achieve. By clustering all the SSMIS channels using various techniques, our work identifies additional precipitation (e.g., 19 GHz) and convection (91- 183 GHz) features that match Dvorak patterns, which may lead to better storm intensity classification.

References

DeMaria, R. T., 2016: Automated tropical cyclone eye detection using discriminant analysis. M.S. thesis, Dept. of Computer Science, Colorado State University, 63 pp., https://dspace.library.colostate.edu/bitstream/handle/10217/170410/DeMaria_colostate_ 0053N_13387.pdf.

Dvorak, V. F., 1984: Tropical cyclone intensity analysis using satellite data. NOAA Tech. Rep. NESDIS 11, 45 pp., http:// severe.worldweather.wmo.int/TCFW/RAI_Training/Dvorak_ 1984.pdf.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X.,1996: A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of Second International Conference on Knowledge Discovery and Data Mining, E. Simoudis, J. Han, and U. Fayyad, Eds. AAAI Press, Portland, Oregon, 226–231.

Jain, A.K. Data clustering: 50 years beyond K-means. Pattern Recognition Lett. (2009), doi:10.1016/j.patrec.2009.09.011

Knaff, J. A., and R. T. DeMaria, 2017: Forecasting tropical cyclone eye formation and dissipation in infrared imagery. Wea. Forecasting, 32, 2103–2116, https://doi.org/10.1175/WAF-D-17-0037.1.

Nelson, A. M., 2019: Characterization of Tropical Cyclone Intensity Using Microwave Imagery. Theses and Dissertations. 2358. https://scholar.afit.edu/etd/2358

Olander, T. L., and C. S. Velden, 2019: The Advanced Dvorak Technique (ADT) for estimating tropical cyclone intensity: Update and new capabilities. Wea. Forecasting, 34, 905–922, https://doi.org/10.1175/ WAF-D-19-0007.1.

Rousseeuw, P. J., 1987: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20:53-65.

Solorzano, N. N., J. N. Thomas, and C. Bracy, 2018: Monitoring tropical cyclones with lightning and satellite data, Eos, 99, https://doi.org/10.1029/2018EO092439 .

Xu, J., Han, J., Xiong, K., and Nie, K. 2016: Robust and Sparse Fuzzy K-means Clustering. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16). AAAI Press, 2224–2230.