Complex Adaptive Systems, Publication 4
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2014-Philadelphia, PA

Cluster Analysis of North Atlantic Tropical Cyclones

Irenea L. Corporal-Lodangco[a,]*, Michael B. Richman[a,b], Lance M. Leslie[a,b],
Peter J. Lamb[a†]

[a]*Cooperative Institute for Mesoscale Meteorological Studies*
[b]*University of Oklahoma*

Abstract

Tropical cyclones (TCs) in the North Atlantic (NA) basin pose an annual risk to coastal regions, with hurricane Katrina (2005) the costliest TC in US history. This study employs K-means cluster analysis (CA) to detect the distinctive, important NA TC paths and lifecycles. Unlike previous TC cluster analyses, which examined TC tracks, the present work documents TC genesis and decay locations. Application of the silhouette coefficient provided an objective method to determine the optimal number of clusters (7 for genesis locations, 6 for preferred tracks, and 5 for decay locations). Additionally, silhouette coefficients provided the information necessary to remove storms that did not fit specific clusters, improving cluster cohesiveness.

For TC genesis, K-means CA captured the separation between tropical and higher-latitude TCs. Clustering of genesis points identifies formative areas. The western NA cluster is the most active. TCs have distinct decay locations, notably in the western NA, Gulf of Mexico and western Caribbean Sea. Clustering TC tracks reveals that TCs moving to higher latitudes recurve generally, whereas Caribbean and Gulf coast TCs have straight-line tracks. Temporally, early season TC clusters form in the western Caribbean Sea, Gulf of Mexico, and western NA. Mid-season TC clusters shift eastward, extending from the tropical NA to Africa. Late season TC clusters recur in the Caribbean Sea, Gulf of Mexico and western NA.

* Corresponding author. Tel.: +1-405-620-5893; fax: +1-405-325-3098.
  E-mail address: ireneacl@ou.edu

† Died May 28, 2014

1. Introduction

Examining the formation, movement, and decay of NA TCs provide valuable information necessary to understand the behavior and risk associated with these storms. Previous studies have investigated NA TC tracks using CA [1]-[5]; however, no analysis has documented regions of TC decay. Moreover, little attention has also been paid to cluster the genesis locations. Given the importance of both genesis and decay locations, this study is an important addition to the knowledge of NA TC behavior. Existing works on CA of NA TCs are all focused on classifying the trajectories. To synthesize knowledge from a large dataset of TCs, this study employs a K-means cluster analysis to partition a set of TCs for genesis and decay locations, and tracks, into homogenous or coherent groups.

2. Data and methodology

2.1. Data and definitions

The hurricane data 2nd generation (HURDAT2) from the National Oceanic and Atmospheric Administration (NOAA) National Hurricane Center (NHC) are used in this study. Information on the location, maximum winds and central pressure is obtained every six hours. This study examines 940 TCs that formed over the NA basin for the period 1950-2012. CA is applied to these data for locations of TC genesis and decay, and on the TC tracks.

To gain a coherent physical insight for each cluster, those cases that fall within each cluster are subject to composite analyses of sea surface temperature (SST) using the climatology of NOAA Extended Reconstructed SST Version 3 (ERSST V3) product [6]. Vorticity, sea level pressure (SLP) and relative humidity (RH) composites are constructed using the National Center for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) reanalysis fields [7].

In clustering genesis locations, two variables are used: latitude and longitude. The genesis location is defined as the first reported location of the TC. Similarly, in classifying the decay locations, the latitude and longitude of the decay point are employed. Decay location is defined as the last reported location of the TC. Each TC track is represented by the average of three locations: genesis, maximum intensity, and decay.

2.2. K-means cluster algorithm

A number of clustering techniques are available. K-means [8] cluster algorithm is used herein for classification of TC genesis locations, tracks, and decay locations. This algorithm is not based on distance measures from one observation to another observation, but uses the within-cluster variation as measure to form homogenous clusters. Specifically, the procedure aims at segmenting the data in such a manner that the within-cluster variation is minimized. The clustering process starts by randomly assigning objects to a number of clusters k, where k is a user-specified parameter. The importance of the assignment is illustrated in [9]. The objects are then successively reassigned to other clusters to minimize the within-cluster variations, which is the distance from each observation to the center of associated cluster. If the reallocation of an object to another cluster decreases the within-cluster variation, this object is reassigned to that cluster.

2.3. Silhouette coefficients

A necessary question is how to determine the optimal number of clusters, k. [10] define a set of silhouette values that provide a graphical aid to the interpretation and validation of clusters of data. The silhouette coefficient is a measure of the cohesiveness of each cluster and how well the clusters are separated. It is also a useful tool to assess the overall goodness-of-fit for a given clustering solution as the values can be used to compare the clustering solutions quantitatively. The silhouette coefficient (s) is based on the average distance between the objects and can vary between -1 to +1, and is defined as,

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

Specifically, a silhouette value of less than 0.20 indicates a poor solution quality; a value between 0.20 and 0.50

indicates a fair solution, whereas values of more than 0.50 indicate a good solution [11].

The initial application of K-means CA to TC genesis locations, tracks and decay locations produced a small number of negative silhouette coefficient values, indicating that some cases were fit poorly to a cluster. To improve the cluster cohesiveness, several rounds of clustering were performed. For each clustering process, the TCs that produced negative silhouette coefficients are removed from the dataset, until those negative silhouette coefficients are eradicated. The process reduced the number of TCs to 916, 919, and 915 for genesis, track, and decay data, respectively. Another useful summary statistic is the average silhouette value across all objects. This summarizes how well the current configuration fits the data. An easy way to select the appropriate number of clusters is to choose that number of clusters which maximizes the average silhouette. The sum of silhouette values and average silhouette value are presented in Table 1. The plots of positive silhouette coefficients after applying the K-means CA to TC genesis locations, tracks, and decay locations are presented in Fig. 1.

	Genesis, *k*=7	Tracks, *k*=6	Decay, *k*=5
Sum of positive silhouette values	595.9	545.6	532.2
Average of positive silhouette values	0.651	0.596	0.579

Table 1. Sum and average of positive silhouette coefficients for genesis, tracks, and decay
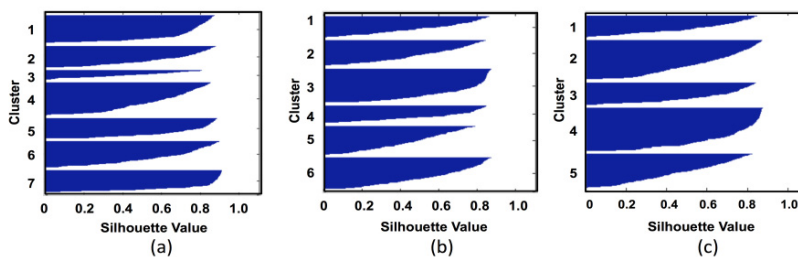

Fig.1. Silhouette plots for (a) genesis, (b) track, and (c) decay clusters.

2.4. Oceanic and meteorological variables

In classifying the TC genesis locations, tracks, and decay locations, it is important to understand the large-scale environmental conditions that affect them. TC formation and development requires a sea surface temperature of at least 26.5ºC [12]. In the summer months, the SSTs in the Caribbean and portions of the NA Ocean can reach 29ºC, making them prime locations for TC inception. TC movement and tracks are largely influenced by large-scale circulation parameters such as vorticity [13], SLP [14] and RH [15]. TC tracks are also influenced by the steering flow [16]. The reanalysis and composites of SST, vorticity, SLP, and RH are constructed (not shown) and subjectively examined to gain a physical understanding of what the CA represent.

3. Results and discussion

3.1. Genesis locations

Using the aforementioned methods, the 916 TC genesis locations were found to map to 7 clusters. The most appropriate number of clusters is determined using silhouette coefficients as an objective measures, whereas the oceanic and meteorological variables are examined to present the physical mechanisms behind the clustering. The composites of large-scale environmental fields also support the choice of 7. The meteorological features of one TC genesis region are different from the others so settling for a fewer number of clusters would mean generalizing the TC region of genesis and losing the distinct characteristics of other genesis regions. For example, the SST and RH in clusters 1 and 5 are almost the same but they differ in SLP and vorticity.

Fig. 2a shows the genesis locations of TCs, color-coded by cluster number. The genesis locations for all 7 clusters are distributed within the confines of the 8°N to 45°N latitudes. The cumulative density of TC genesis per

5° latitude x 5° longitude grid box for each cluster is given in Figs. 2b-h that show the preferred area of formation. The size of the red square marker indicates the number of TC passages per grid box. The genesis regions are the Gulf of Mexico (cluster 1) that accounts for 158 TCs (17%); east of Caribbean Sea (cluster 2) comprising 124 TCs (14%); the TC formation in central NA accounting for only 53 TCs (6%); the smallest cluster that represents cluster 3, western NA (cluster 4) with 190 TCs (21%); western Caribbean Sea (cluster 5) consists of 116 TCs (13%); central tropical NA (cluster 6) with 150 TCs (16%); and the eastern tropical NA (cluster 7) accounting for 125 TCs (14%).

The official TC season over the NA basin begins on June 1 and ends on November 30. During these months, 96% of all TCs form. June and July are the early season months and November is considered as a late-season month [17]. Figure 3 shows the number of TCs per calendar month for each cluster. All clusters, other than cluster 5, indicate September to be the peak of TC genesis followed by August. However, for cluster 3, October is the second most active month. Clusters 6 and 7 have almost the same seasonality with, no TC genesis from January to May. Cluster 6 also has no TC genesis in November. Nearly none in November and none at all in December for cluster 7. Despite the official TC season in the NA, certain clusters had "out of season storms". Clusters 2 and 4 also have similarity, with the TC season starting in April and ending in December. The TC occurrence in cluster 4 continues to increase as the season progresses, peaking in September, then decreasing significantly in October and November, with almost none in December. Cluster 3 has lower frequency as only 53 TCs are classified. TC activity commences in June and ends in December. Clusters 1 and 5, both on the western periphery of the domain, exhibit a lower probability in July. TC activity begins in May and concludes in November for cluster 1, whereas cluster 5 activity commences in April, and concludes in December, though an outlier was present in February (Fig. 3).

The spatial cumulative density distribution (Figs. 2b-h), allows deduction of distinct TC genesis regions. It is of note that fewer TCs formed poleward of 30°N due to cooler SSTs and stronger wind shears inhibiting TC formation [18]. The separate clustering of Gulf Coast and East Coast TCs is explained by the large-scale changes in SLP patterns over the NA [17]. [17] also suggest an inverse relationship between TC activity at low and high latitudes, this support the latitudinal separation of TC genesis locations.
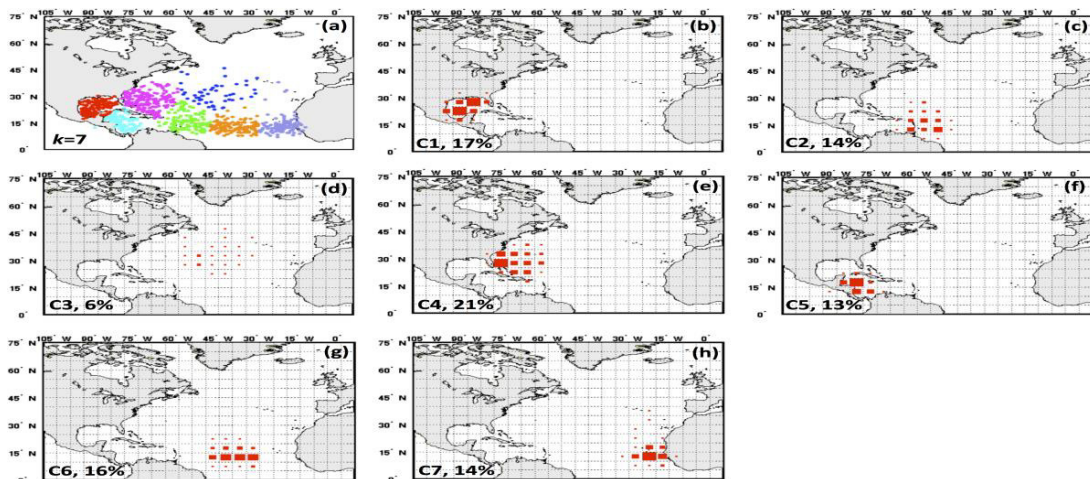
Fig. 2. (a) K-means genesis clusters; (b-h) cumulative density distribution of genesis locations per 5°latitude x 5°longitude grid box.
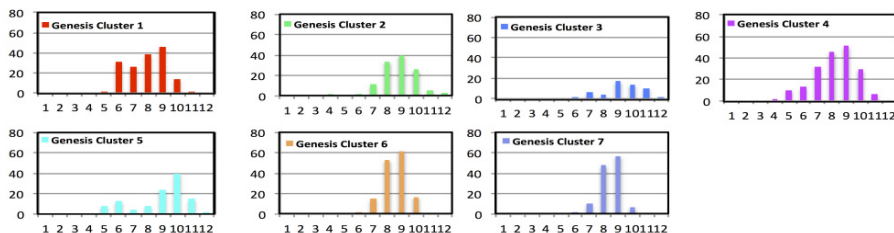
Fig. 3. Numbers of TCs per calendar month for each genesis cluster.

3.2. Tracks

Classifying TC tracks identifies the characteristics of various TC trajectories by separating them into distinct numbers of patterns. An individual TC track is represented by the average position of the TC at genesis, maximum intensity, and decay. The silhouette coefficients for tracks indicate a good solution (Fig. 1b). The actual geographical positions of the TCs average positions are presented in Fig. 4a, with color-coded cluster number. Figures 4b-g are the 6 clusters of NA TC tracks. Each cluster has specific, distinct geographical paths.

The cumulative density of TC passages per 5°latitude x 5°longitude grid box for each cluster is shown in Fig. 5. Cluster 2 is consists of 149 TCs (16%), where track passages are heavily concentrated in the domain of 15°N to 40°N and 65°W to 35°W (Fig. 5b). TCs in this cluster generally follow a recurving track pattern and tend to remain offshore (Fig. 4c). Recurving TCs in clusters 5 and 6 with 169 (18%) and 181 (20%) TCs, respectively, tend to threaten the North American East Coast and some recurve into the open ocean; whereas, straight moving TCs threaten the western Caribbean Sea and the Gulf Coast. TCs in these clusters also include those that veered northwestward, posing threat to the Caribbean Islands (Figs. 4f,g). Clusters 1 (with 121 TCs, 13%) and 4 (with 101 TCs, 11%) have resemblance in their general track shape (Figs. 4b,e). TCs in these clusters that develop and dissipate in the tropical NA have a straight moving track pattern. TCs that progress northeastward have a recurving track type. TCs in the compact cluster 4 also exist mainly over the ocean. The 198 TCs (22%) in cluster 3 are concentrated in the western Caribbean and Gulf of Mexico (Fig. 4d). The randomness of the tracks is different from the other clusters. Track clusters 3, 5, and 6 contain the highest frequencies of TCs, as together they comprise 60% of the TCs in the basin (Figs. 5c,e, f). Clusters 1 and 4 that contain the fewest cases also possess similar track shape. All low-latitudes TCs are prevented from recurving because of the southward jet stream shift with the Bermuda high [19], increasing the likelihood for some TCs in cluster 5 to make Gulf Coast landfall (Figs. 4f,g).

The monthly TC frequency for each cluster is summarized in Fig. 6. Clusters 1, 2, 3, 4, and 6 have a pronounced peak in TC activity in September, followed by August, whereas, cluster 5 reaches its highest frequency in August. Cluster 4 has a narrower seasonal distribution when compared with other clusters.
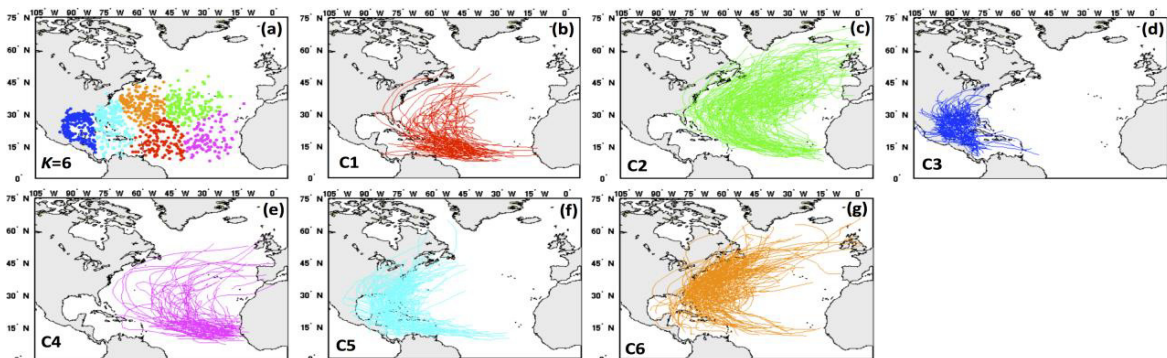

Fig. 4. (a) TC points that represent TC tracks; (b-g) Track clusters after applying K-means clustering to points that represent the NA TC tracks.
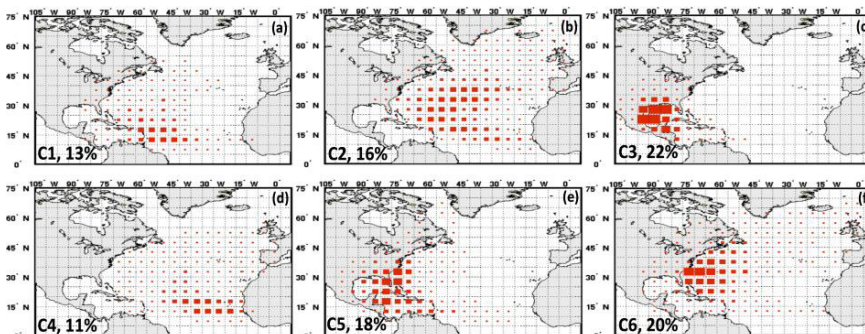

Fig. 5. Cumulative density of TC passages per 5°latitude x 5°longitude for each track cluster.
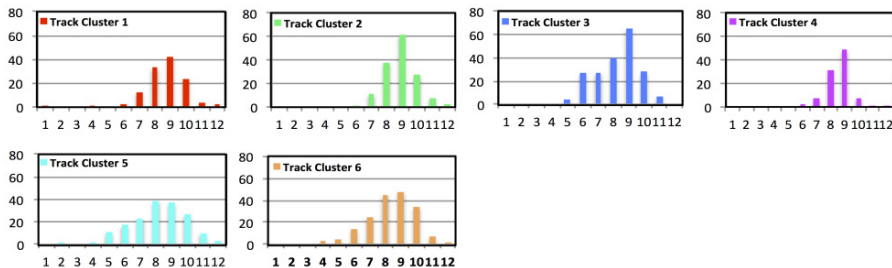
Fig. 6. Numbers of TCs per calendar month for each track cluster.

[1] identified two recurring track clusters, representing TCs that threaten the NA north of 35°N, and one straight moving track cluster that represents the TCs that threaten the NA south of 35°N. [2] also identified six NA track clusters but used a different method to derive the TC tracks. [3] objectively separated the NA TC tracks into four groups. The overall zonal and meridional separation of their tracks also occurs in the present study, but the present study produced a greater of clusters (six rather than four).

3.3. Decay locations

The decay locations of 915 TCs over NA during the 63-year period are grouped into 5 clusters by the silhouette coefficient of the K-means algorithm. Each cluster has its distinct features as seen in Fig. 7a, color-coded to indicate the cluster number. Figures 7b-f show the spatial distribution of decay locations by cluster, per 5° latitude by 5° longitude grid box. The cumulative density distribution suggests the most preferred area of TC decay. Cluster 1 corresponds to TCs that decay over the tropical NA region that accounts for 121 TCs (13%). The biggest cluster, 4, accounting for 247 TCs (27%), represents the TCs that decay over the Gulf of Mexico and western Caribbean including those that decay inland, over Southern US and Central America. Cluster 5 corresponds to TC decay in the lower latitudes of western NA (south of North American Coast) with 193 TCs (21%). TCs that dissipate in the higher latitudes of western and central NA define cluster 2, with 225 TCs (25%), whereas cluster 3 signifies TC decay in the higher latitudes of central and eastern NA with 129 TCs (14%). It is noteworthy that TCs that formed over the tropical eastern NA follow a track that moves westward, then northwestward and decays over the coastline at higher latitudes, a feature noted by [19]. Figure 8 shows the monthly frequency of each decay cluster. September is the peak of TC decay and followed by August as the second active month. Most clusters, except cluster 3 have quasi-Gaussian seasonal distributions.

3.4 Monthly TC activity

January through May is considered as outside the TC season in NA basin. The number of TC during June is more than twice that of May. TCs in June are characterized by clusters 1 and 5, those that form in western Caribbean and Gulf of Mexico and from cluster 4, the cyclogenesis in the western NA. Fewer TCs form in eastern and central tropical NA. This is observed in July where TC genesis is most common in clusters 1, 4, and 6. But the area of genesis moves eastward as more TCs develop in the eastern Caribbean and in the central and eastern tropical NA. This observation implies that during the early season of TC activity, the majority of TC genesis is in Gulf of Mexico including the western Caribbean and in the western NA.

TC decay points are predominantly seen in western NA and in the Gulf Coast including the western Caribbean Sea. The most prevailing track pattern from June through July is classified as straight moving TCs that head off to western Caribbean and Gulf of Mexico. TCs that reach the lower latitudes of western NA with northeastward movement are also prevalent.

The number of TCs in all clusters increases in the mid-season months. In August, the points of genesis are noticeably more numerous than in July. TC formations are concentrated in the tropical NA and in western NA, including the western Caribbean and Gulf of Mexico. The spatial distribution of genesis locations in September resembles that of August but there is an increase of TC formation in the higher latitudes of central NA. September is the peak of the TC season. TC frequency in each cluster is at its highest during this month. In October, cyclogenesis

over the NA basin has decreased, especially in the eastern and central tropical NA.

In August and September, decay points extend over a broad area of NA with some as far north as 65ºN. TC dissipation along the Gulf Coast, East Coast, and the Caribbean is also common. During the month of October TC decays follow the pattern of August and September, but with fewer TCs. Recurving TCs tend to threaten the North American Coast, whereas those that do not recurve threaten the Gulf Coast. TCs that originate over the western NA move toward the northeast. All decay and track clusters are well represented in mid-season months.

TC genesis during the late part of TC season has decreased significantly. In November, most of the TCs dissipate in the open waters of NA and some decays in the Gulf of Mexico and western Caribbean. Most TCs in December dissipate in the central NA including eastern Caribbean. TC tracks in the late part of the season are mostly moving northeast and remain offshore.
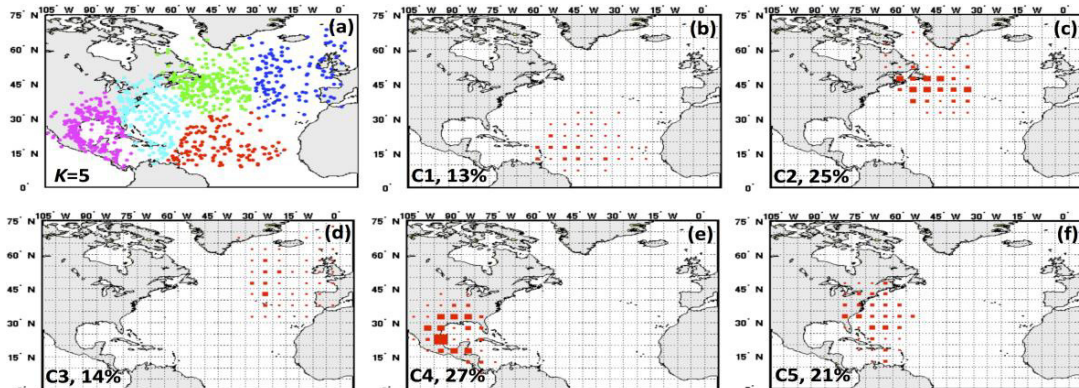

Fig. 7. (a) Decay clusters; (b-h) cumulative density distribution of decay locations per 5°latitude x 5°longitude grid box.
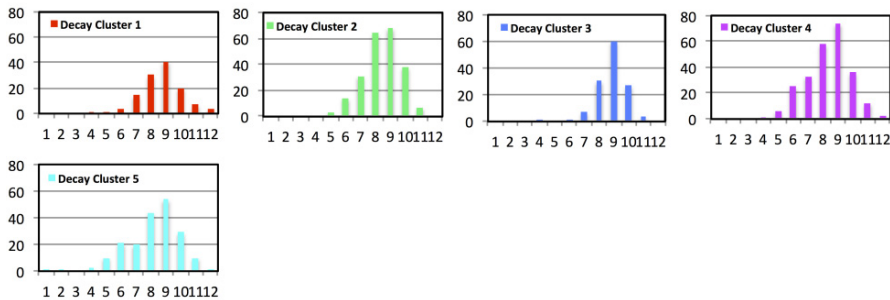

Fig. 8. Numbers of TCs per calendar month for each decay cluster.

4. Summary and conclusions

A K-means CA method is applied to 1950-2012 TC data to obtain 7, 6, and 5 clusters, for genesis points, tracks, and decay points, respectively. Clustering results are interpreted and justified from the statistical and meteorological perspectives. In clustering the genesis locations, the K-means method captured the latitudinal separation between tropical TCs and higher-latitudes TCs. Clustering also identified the longitudinal separation of TCs that form in Gulf of Mexico, Caribbean Sea, central and eastern tropical NA. The spatial cumulative density of genesis points provides an improved sense of areas that are potentially active during NA TC season. The cluster that represents the Gulf of Mexico and western Caribbean was found to have the highest number of TC formations.

TCs in the NA exhibit distinct decay locations. Of 915 decay locations used in the CA, 5 decay regions are identified. Prevalent areas of dissipation are the western NA, Gulf of Mexico and western Caribbean Sea. The clustering of genesis locations resulted in a higher number of clusters, as opposed to the number of clusters produced in clustering the decay locations. This is attributed to the tropics possessing more diverse characteristics, when compared to the higher latitudes where most TCs decay. Clustering TC tracks identifies the uniqueness of

various paths by separating them into distinct patterns. Each cluster has distinguishing features. TCs moving to the higher latitudes of NA and western NA are indicative of a general pattern of recurving storms. TCs that threaten the Caribbean and Gulf Coast are associated with straight-moving tracks.

Monthly frequencies of each cluster show differences in seasonality and general characteristics of TC activity in each cluster. Most TCs form between June and November, which defines the TC season over the NA with September and August as the most active months. This is the time of year when warm SSTs cover the largest expanse of the NA [17] and coincides with the vertical wind shear of horizontal winds in the tropical Atlantic reaching their minimum [18].

By examining the monthly TC activity, it is documented how early season TCs tend to develop in the Western Caribbean Sea and the Gulf of Mexico. By mid-season, the focus for development shifts eastward to include most of the tropical Atlantic and Caribbean Sea. TCs developing during this portion of the season often have an extended westward track. By the latter part of the season, TC development again shifts westward in the Caribbean Sea, Gulf of Mexico, and western Atlantic. According to [3], this is caused by the seasonal variations of the thermodynamics in the tropical Atlantic and also by the seasonality of easterly waves emanating from the west coast of Africa.

5. Acknowledgement

6. References

1. J. B. Elsner, Tracking hurricanes, Bull. Amer. Meteor. Soc. 84 (2003) 353-6.
2. J. Nakamura, U. Lall, Y. Kushir, and S. J. Camargo, Classifying North Atlantic tropical cyclone tracks by their mass moments, J. Clim. 22 (2009) 5481-94,.
3. J. P. Kossin, S. J. Camargo, and M. Sitkowski, Climate modulation of North Atlantic hurricane tracks, J. Clim. 28 (2010) 3057-76.
4. A. J. Colbert and B. J. Soden, Climatological variations in North Atlantic tropical cyclone tracks, J. Clim. 25 (2012) 657-73.
5. T. A. McCloskey, T. A. Bianchette, and K.B. Liu, Track patterns of landfalling and coastal tropical cyclones in the Atlantic Basin, their relationship with the North Atlantic Oscillation (NAO), and the potential effect of global warming, Amer. J. Clim. Change 2 (2013) 12-22.
6. T. M. Smith, R. W. Reynolds, T. C. Peterson, and J. Lawrimore, Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006), J. Clim. 21 (2008) 2283-96.
7. E. Kalnay and Co-authors, The NCEP/NCAR 40-year re-analysis project, Bull. Amer. Meteor. Soc. 77 (1996) 437-71.
8. J. MacQueen, Some methods for classification and analysis of multivariate observations, Proc. Fifth Berkeley Symp. on Mathematical Statistics and Probability, Berkeley, CA, University of California, (1967) 281–297.
9. X-F Gong and M.B.Richman, On the application of cluster analysis to growing season precipitation data in North America east of the Rockies, J. Clim. 8 (1996) 897-931.
10. L. Kaufman and P. J. Rousseeuw, Finding groups in data: An introduction to cluster analysis, (2005) ISBN: 978-0-471-73578-6.
11. E. Mooi and M. Sarstedt, A concise guide to market research: The process, data, and methods using IBM SPSS statistics, Springer, (2011).
12. W. M. Gray, Predicting Atlantic basin seasonal tropical cyclone activity by 1 August, Wea. Forecasting 8 (1993) 73-86.
13. L. Wu and B. Wang, Assessing impacts of global warming on tropical cyclone tracks, J. Clim. 17 (2004) 1686-98.
14. K. L. Corbostero and J. Molinari, The relationship between storm motion, vertical wind shear, and convective asymmetries in tropical cyclones, J. Atmos. Sci. 60 (2003) 366-76.
15. T. N. Carlson, An apparent relationship between the sea-surface temperature of the tropical Atlantic and the development of African disturbances into tropical storms, Mon. Wea. Rev. 99 (1971) 309-10.
16. G. J. Holland, Tropical cyclone motion. Global guide to tropical cyclone forecasting, G. J. Holland, Ed., World Meteorological Organization, (1993) WMO/TD-500.
17. J. B. Elsner and A. B. Kara, Hurricanes of the North Atlantic: Climate and society, Oxford University Press, (1999) 512 pp.
18. W. M. Gray, Hurricanes: Their formation, structure and likely role in the tropical circulation, Meteorology over the tropical oceans, D. B. Shaw, Ed. Royal Meteorology Society, (1979) 155-218.
19. J. B. Elsner, K. B. Liu, and B. Kocher, Spatial variations in major U.S. hurricane activity: statistics and physical mechanism, J. Clim. 13 (2000) 2293-2305.