

Cluster Analysis of Philippine Tropical Cyclone Climatology: Applications to Forecasting

Corporal-Lodangco IL* and Leslie LM

School of Meteorology, University of Oklahoma, Norman, Oklahoma, USA

Abstract

This study aims at providing an increased understanding of tropical cyclone (TC) activity in the Philippines, to assist in reducing the fatalities and economic costs of TC impacts. A cluster analysis, using K-means, is applied to Philippine region TCs for the period 1950-2011. The clustering is carried out for TC genesis and decay locations, and TC tracks. Silhouette coefficient values and key meteorological and oceanic variables determine the optimal cluster numbers. It is found that, for the Philippine region, there are 4 genesis location, 5 decay location and 6 track clusters. The classification of TC genesis locations captures the longitudinal separation of cyclogenesis regions. The formation area east of the Philippines (west of 140°E) is the most active region, with 398 genesis points. The main TC dissipation area is Southeast Asia, with 352 decay points. Clustering the TC tracks identifies various track types by separating them into discrete patterns. Several distinct types of straight moving and recurring trajectories emerge. Short, straight west northwestward tracks directed towards Indochina have the highest trajectory frequency, with 248 TC tracks. The spatial and temporal behavior of Philippine TCs is determined from the clusters of genesis locations, decay locations, and tracks, for specific months. Because the TC genesis locations define the subsequent TC paths and landfall locations, they consequently also provide valuable TC forecasting guidance. Moreover, the monthly distribution of genesis and decay locations, and tracks, enables the variability of seasonal cycles between the clusters to be calculated.

Keywords: Tropical cyclones; Cluster analysis; K-means; Silhouette coefficients; Philippines

Introduction

The Philippine domain exhibits a high level of tropical cyclone (TC) activity that makes the country extremely vulnerable to hazards associated with TCs. TCs are multi-faceted natural hazards that are dependent upon oceanic and atmospheric conditions contributing to their formation and development, movement, and eventual decay. Seasonal variations in the large-scale atmospheric circulation also cause variations in TC characteristics. Studying the formation and movement of Philippine TCs provides valuable information necessary to understand their behavior more fully. Towards this end, cluster analysis using a K-means method is used to investigate the distinct characteristics of TCs over the Philippine domain. Here, the aim of TC clustering is to partition a set of TC genesis and decay locations, and tracks, into homogenous groups, such that group patterns are similar, and assist in identifying different genesis and decay regions, and track types.

Several studies use clustering to classify TC tracks over various TC basins. Blender et al. [1] used a K-means cluster analysis [2] for extra tropical cyclone tracks in the North Atlantic. Elsner [3] also used TC track data for the Western North Pacific (WNP) and showed that K-means clustering can be applied to TCs, using positions of maximum and final hurricane intensities. The same method was applied to North Atlantic TCs [4] and the value of k , defining *a priori* number of clusters, was set at three. McCloskey et al. [5] used K-means clustering to group TC tracks into coherent clusters and analyzed in terms of various climate modes including the North Atlantic Oscillation (NAO), the Atlantic Meridional Mode (AMM), the El Niño Southern Oscillation (ENSO) and the Madden Julian Oscillation (MJO). However, Gong et al. [6] recognized that K-means clusters are affected by the initial seed locations. The K-means method also cannot directly accommodate tracks of different lengths. This limitation is readily overcome. For example, [7] employed a probabilistic clustering technique, based on

polynomial regression models. The regression Finite-Mixture Model (FMM) is described by Gaffney [8] and applied to Atlantic extratropical cyclones to fit the “shape” of the trajectories. Track patterns classified using this model exhibit TC characteristics and physical relationships with their large-scale environments. The same model has been used for the WNP [9,10], the eastern North Pacific [11], Fiji [12,13], the Southern Hemisphere [14] and the North Atlantic [15]. Zhang et al. [16] applied an FMM-based clustering algorithm to TC tracks over China and identified three clusters. A probabilistic clustering method based on regression mixture models has grouped TCs over southwestern Indian Ocean, resulting in track shape groups [17]. Unlike K-means, the regression mixture model allows TC track clustering of varying shapes and lengths. Nakamura et al. [18] solved the problem of K-means by using the first and second moments of TC tracks to approximate the shape and lengths of tracks. The mass moments were then applied to K-means that demonstrated reliable clustering results for North Atlantic TC tracks, and produced an optimum of six clusters. Colbert [19] focus on TCs forming in the main North Atlantic development region (MDR). Their classification method for track types is based on defined threat regions that yield similar results to ref. [15,18]. The fuzzy c-means is another cluster method applied to TC tracks of Korean peninsula land falling TCs [20]. Kim [21] used the same method for the WNP and identified seven clusters as the optimal number. This

***Corresponding author:** Corporal-Lodangco IL, School of Meteorology, University of Oklahoma, Norman, Oklahoma, USA, Tel: 405-620-5893, Fax: 405-325-7689; E-mail: ireneacl@ou.edu

Received December 23, 2015; **Accepted** January 23, 2016; **Published** January 30, 2016

Citation: Corporal-Lodangco IL, Leslie LM (2016) Cluster Analysis of Philippine Tropical Cyclone Climatology: Applications to Forecasting. J Climatol Weather Forecasting 4: 152. doi:10.4172/2332-2594.1000152

Copyright: © 2016 Corporal-Lodangco IL, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

method is known to produce more realistic classification results. Harr [22] used fuzzy cluster analysis and empirical orthogonal functions to identify recurrent large-scale circulation patterns associated with TC characteristics. Tracks were classified into four types, straight moving, south recurving, north recurving, and those in the South China Sea, with tracks not fitting into these four types being discarded.

K-means clustering is well suited to this study for a number of reasons. It is not sensitive to outliers and is commonly used and computationally efficient. Furthermore, the necessary pre-specification of cluster number is possible because of the well-known characteristics of the behavior of Philippine region TCs. Most previous studies have focused on TCs over the western and eastern North Pacific and North Atlantic basins, but this study is the first attempt to classify the attributes of Philippine TCs. Here, the K-means method analyzes not only TC tracks but also genesis and decay locations. The major difference in track clustering is that the track is represented by single point, based on the average of the positions of genesis (initial position), maximum intensity position and decay (final position), allowing K-means clustering. The present study is significant because it provides useful information and patterns of the Philippine TCs. The specific goals of the study are: (1) to examine the spatial and temporal behavior of TCs by looking at monthly patterns of genesis and decay locations and tracks; and (2) to identify the seasonal cycle of each cluster.

Study Area

The study region covers latitudes 5°N-25°N and longitudes 115°E-135°E, shown in Figure 1 [23] as the black inset and referred to here as the Philippine domain. The irregular box (red broken line) shows the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) area of responsibility for TCs. PAGASA monitors and forecasts TCs that affect the Philippines.

Materials and Methods

Data and definitions

All of the 1,161 TCs that formed in, or entered, the Philippine

domain in the period 1950-2011 were included in this study. Cluster analysis was applied to the genesis and decay locations and the tracks. The location, maximum winds and central pressure are available every six hours from the Joint Typhoon Warning Center (JTWC) best-track data. The TC latitudes and longitudes must lie inside the domain during at least part of their lifetimes. All TCs, regardless of their intensity classification (TD – tropical depression; TS – tropical storm; and TY – typhoon) are included in the cluster analysis, providing the maximum possible sample size.

To interpret the clustering results, composite analyses of sea surface temperature (SST) were performed using the monthly NOAA Extended Reconstructed SST version 3 (ERSST V3) product [24]. Vorticity, sea level pressure (SLP) and relative humidity (RH) composites were produced using the National Center for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) reanalysis fields [25]. In clustering the genesis locations, two variables are used, latitude and longitude. Genesis location is the first reported location of the TC. Similarly, the decay location is the latitude and longitude of the last reported position of the TC. Each TC track is the mean of the locations of TC genesis, maximum intensity, and decay. Thus, for the purposes of this study a single point can be used to represent a TC track.

K-means cluster algorithm

Cluster analysis is a valuable method for identifying homogenous groups of objects. Objects in a specific cluster share many characteristics, but are dissimilar to objects not belonging to that cluster. Other techniques are available for clustering, but herein, the K-means [2] cluster algorithm is used to obtain and describe classification of TC genesis locations, tracks, and decay locations. The K-means cluster algorithm follows a partitioning procedure. This widely applied algorithm is not based on distance measures from one observation to another observation, but it uses the within-cluster variation as a measure to form homogenous clusters. Following [26], the sum, S , of the squared distances within the clusters is defined as:

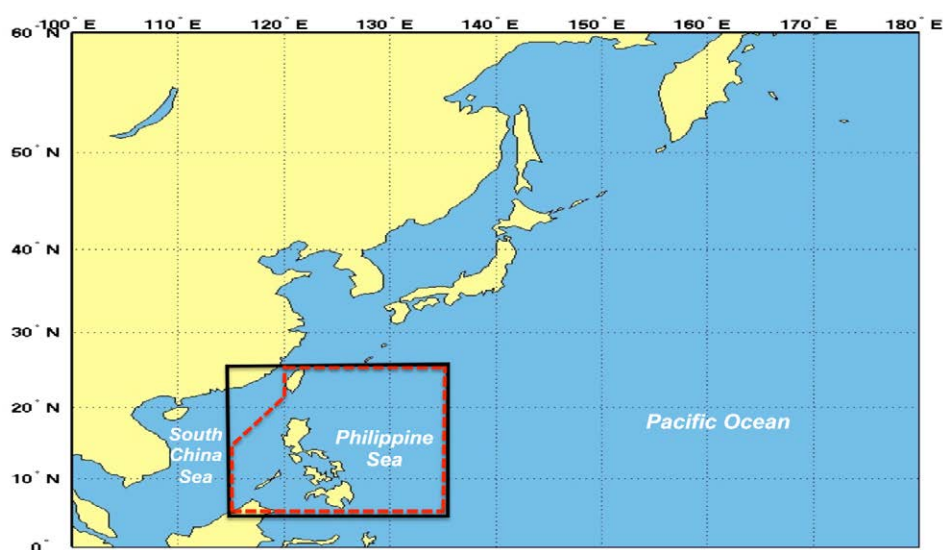


Figure 1: The study region covers latitudes 5°N-25°N and longitudes 115°E-135°E, shown as the black inset and referred to here as the Philippine domain. The irregular box (red broken line) shows the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) area of responsibility for TCs. PAGASA monitors and forecasts TCs that affect the Philippines. The figure is adapted from [23].

$$S = \sum_{j=1}^K \sum_{n \in S_j} |x_n - \mu_j|^2$$

where $|x_n - \mu_j|^2$ is a distance measure between a data point, x_n , and the cluster center, μ_j , indicating the distances of the n data points from their respective cluster centers.

Following [26], the clustering process starts by assigning objects to a number of clusters, k , a pre-specified parameter, also corresponds to the number of centroids. Objects are successively reassigned to other clusters to minimize within-cluster variations, which is the distance from observations 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. The process is repeated until there is no change in cluster membership. With K-means, cluster affiliation can change in the course of the clustering process.

K-means has many advantages over other methods. The software is readily available in various tool kits. In this study, the software framework employed was Matlab as it contains the built-in K-means and silhouette code. The computational space requirements for K-means are modest because only the data points and centroids are stored [27]. In particular, the time requirements are linear in the number of data points. K-means is less affected by outliers and, by making multiple passes through the data, the final solution optimizes within-clusters homogeneity and between-clusters heterogeneity. Furthermore, it can be applied to very large data set, as the procedure is not computationally demanding [28,29].

Determining the Optimum Number of Clusters

Silhouette coefficients

Two of the most difficult tasks in cluster analysis are deciding on the appropriate number of clusters, and deciding how to tell a poor cluster from a good one. The K-means cluster morphology is dependent upon the choice of cluster number that must be specified in advance before carrying out the cluster algorithm process. In this study, the cluster number is determined objectively by finding the maximum mean silhouette coefficient value, defined below in the next paragraph, with no negative members and the total sum of point-to-centroid distances and, in addition, it is determined subjectively, by examining the meteorological and oceanic patterns that best match each cluster. The value of k was sequentially set to 3, 4, 5, 6, and 7 to optimize the number of clusters to achieve a cluster structure that agrees with known characteristics of Philippine TC formation, tracks and decay regions. The final selected cluster number, k , is then based on the relationship between the clusters and the large-scale environmental conditions, especially in genesis clustering. This step requires detailed knowledge of the region.

Silhouettes, as defined by Kaufman [30], are a significant aid to the interpretation and validation of clusters of data. The silhouette value is a measure of the cohesiveness of each cluster and how well the clusters are separated. It is also a measure of the clustering solution's overall goodness-of-fit. The silhouette coefficient values range from -1 to +1 and can be used to compare the clustering solutions quantitatively. It is based on the average distance between the objects. A silhouette $s(i)$ is defined in ref. [29] as,

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

As noted in ref. [26], silhouette values <0.20, between 0.20 and

0.50, and >0.50 suggest poor, fair and good solutions, respectively. The sum and mean silhouette values reflect the overall cohesiveness of all TCs. (Table 1) shows the values of the sum and average of positive silhouettes, and the total sum of point-to-centroid distances. The graph of silhouette coefficient values for genesis and decay locations, and tracks are in Figure 2.

Note that application of K-means clustering to TC genesis and decay locations, and tracks, produced a small number of negative Silhouette coefficient values. To improve cluster cohesiveness, several rounds of clustering were performed. For each clustering process, the TCs that produced negative Silhouette coefficient values (typically several to about 10 TCs) were removed from the cluster, until those negative Silhouette values were eradicated. This process reduced the total number of TCs by <5%, to 1,109, 1,128, and 1,108 for genesis, decay, and track datasets, respectively. Some TCs were removed because they were located over land, which is not possible, so the Silhouette also can act as a TC quality control check.

Meteorological and oceanic variables

There is no objectively incorrect cluster method. However, as noted above, in grouping the TC genesis locations, decay locations and tracks, it is important to understand the large-scale environmental conditions that affect them. Perrone [31] linked various thermodynamic and dynamic variables to TC genesis. To choose the appropriate number of clusters, the results are interpreted and justified from meteorological and oceanic perspectives. The NCEP/NCAR reanalysis of the SST climatology is given in Figure 3a. Warm SSTs are necessary for TC initiation and development. TC formation requires a sea surface temperature of at least 26.5°C [32-34]. TC movement and tracks are largely influenced by large-scale circulation variables such as vorticity [35,36], SLP [37,39], and RH [40]. The composites for vorticity, SLP, and RH are in Figures 3b-3d. TC tracks are also influenced by the steering flow [41-44]. Interactions between steering flow and TC dynamics were investigated by creating and examining composites (not shown).

Results and Discussion

This section describes the clusters produced after applying the K-means to Philippine TCs and discusses the seasonality, and monthly spatial and temporal distributions, of each cluster. For the present study, the TC counts for genesis, decay and track points are computed in 5° latitude by 5° longitude grid box.

Optimal Cluster Number

The most appropriate number of clusters is determined using the largest silhouette coefficients and the smallest total sum of point-to-cluster distances as an objective measures, whereas the meteorological and oceanic variables serve as subjective measures. Based on the above-mentioned objective and subjective measures, the optimum number of clusters for genesis locations was $k=4$, with the sum of silhouette values equal to 676 and the mean silhouette value of 0.61. The composites of the relevant large-scale fields also support the selection of four clusters. The meteorological and oceanic features of TC regions are different, so

	Genesis, $k=4$	Decay, $k=5$	Tracks, $k=6$
Sum of Positive Silhouette Value	676	681	626
Mean of Positive Silhouette Value	0.61	0.61	0.56
Total Sum of Point-to-Centroid Distances	40078	62150	25438

Table 1: Silhouette coefficients and total sum of point-to-centroid distances.

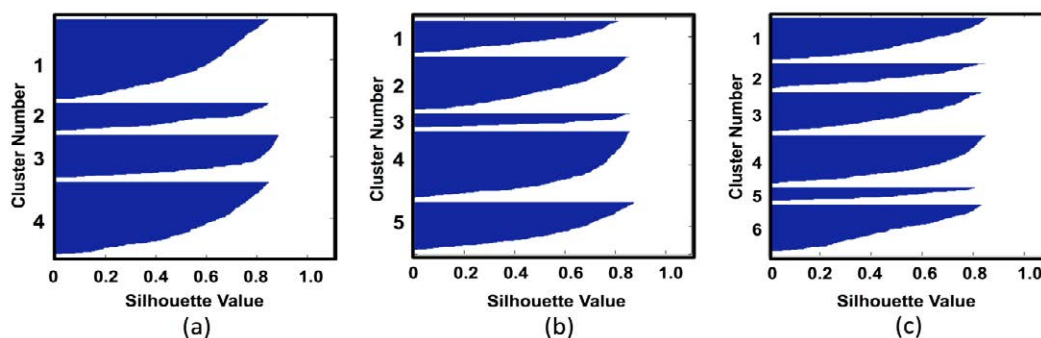


Figure 2: Silhouette graphs for (a) genesis, (b) track, and (c) decay clusters.

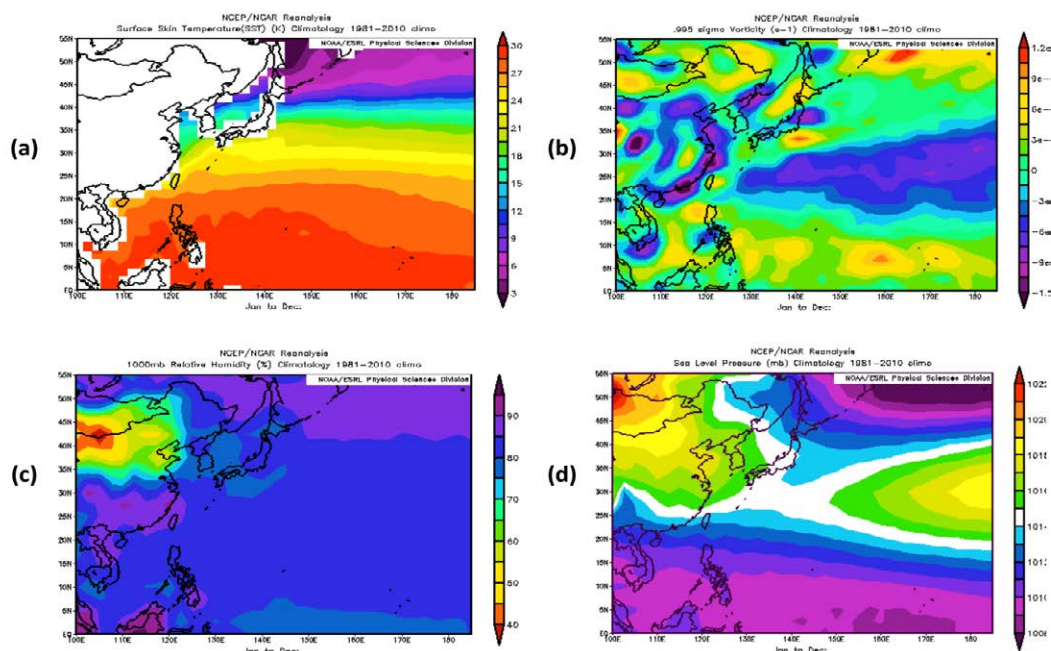


Figure 3: Key variables conducive to TC development that were used to determine the most appropriate number of clusters for (a) SST, (b) vorticity, (c) RH, and (d) SLP.

choosing for too few clusters would mean generalizing the TC region of genesis and therefore losing the distinct characteristics of other genesis regions. The optimum number of clusters for the decay regions was found to be five. For $k=5$ decay clusters, the sum of silhouette values is 6801, with 0.61 as the mean silhouette value. Track clustering suggested that six was the best choice for the number of clusters. The silhouette coefficient values for $k=6$ track clusters gives 626 for the sum of silhouette values and 0.56 for the mean silhouette value. All silhouette values were physically realistic. Figure 4 is the optimal number of clusters for genesis locations, decay locations and tracks. The clusters differ in the physical properties of their TCs, as shown by their geographical positions.

Clustering of Genesis Locations

Locations of genesis of TCs in the Philippine domain occur in a broad region west of the date line, within latitudes 2.5°N to 27.5°N and longitudes 107°E to 179.5°E . Figure 4a shows TC genesis locations in 4 clusters, color-coded by cluster number, and the black asterisks

are cluster centroids. Figure 5 illustrates the cumulative density of TC genesis points in the grid boxes that shows the preferred area of TC formation. The cluster analysis of TC genesis locations captures the longitudinal separation of TC formation regions. Clusters are discussed starting from west of the dateline to the South China Sea. The formation region near the Marshall Islands is cluster 2, the smallest cluster of 136 TCs, or $\sim 12\%$ of all TCs. The genesis locations in cluster 2 have the widest spread compared with other clusters and are farther east and closer to the Equator. TCs that developed near the Northern Marianas Islands and over the central part of Micronesia belong to cluster 4, with 363 TCs ($\sim 33\%$). The largest cluster is the formation region over the Philippine Sea (cluster 1) and comprises 398 TCs ($\sim 36\%$). Cluster 3 is cyclogenesis over the South China Sea, accounting for 212 TCs ($\sim 19\%$).

The genesis locations provide an understanding of possible TC track type, as genesis location greatly influences the track and decay point of the TC. The corresponding TC tracks of each of the four genesis clusters are in Figure 6. The tracks of TCs in clusters 2 and 4 (east of 135°E) have recurving and long straight tracks. The straight-

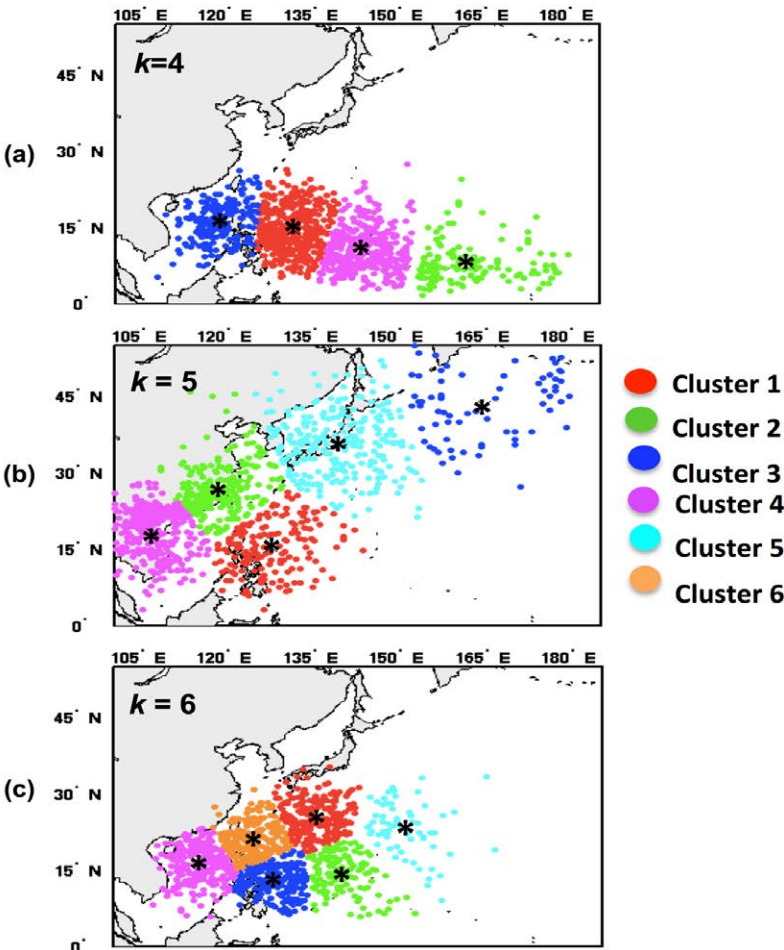


Figure 4: K-means clusters of (a) genesis locations, (b) decay locations, and (c) tracks. The black asterisks are the centroids of each cluster. The number of clusters, k , is shown in the top left of each panel.

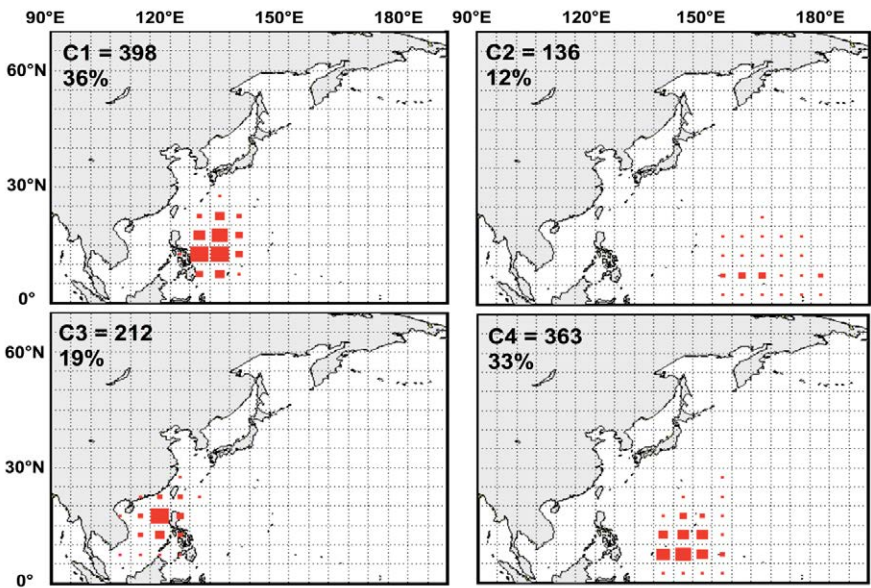


Figure 5: The cumulative density of genesis points in each 5° latitude by 5° longitude grid box and the percentage of TC genesis in each cluster.

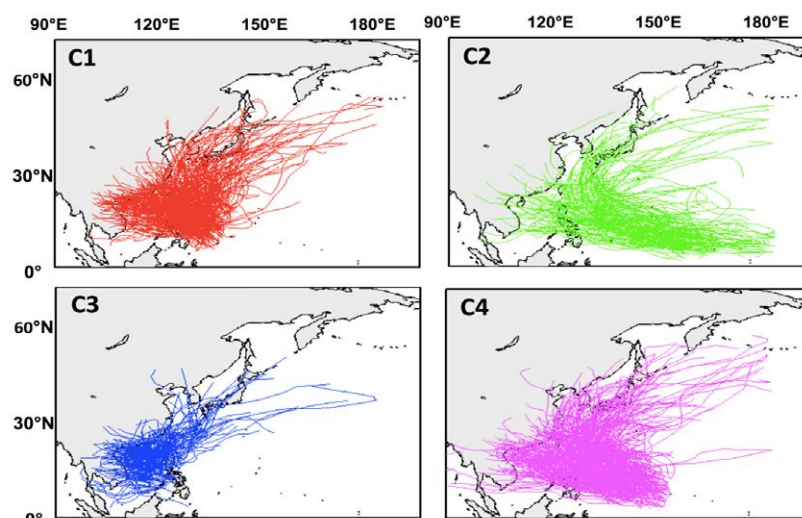


Figure 6: The TC tracks corresponding to genesis points in each of the four genesis clusters.

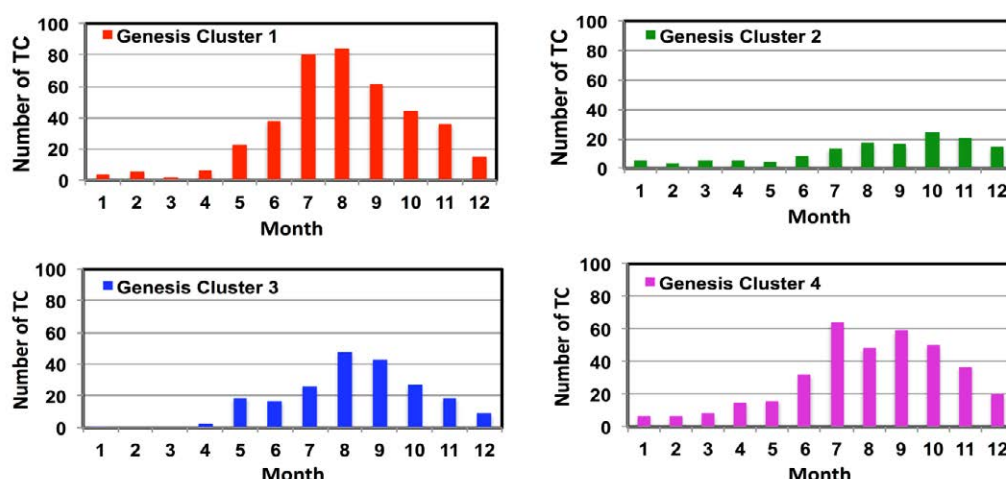


Figure 7: The number of TCs by calendar month for each genesis cluster.

moving TCs move towards Southeast Asia and the recurving TCs veer northeastward and decay over the open sea. Cyclogenesis in cluster 1 and 3 mainly consists of short straight tracks heading toward Southeast Asia, and slightly recurving tracks dissipating over the sea, whereas long straight tracks generally indicate northward motion.

Genesis Seasonality

Figure 7 shows the number of TCs by calendar month for each cluster. The seasonal evolution of TC genesis varies from cluster to cluster. The TC genesis in cluster 1 increases starting from April, peaks in August then decreases significantly in September and continues to diminish until December. Cluster 2 has a smaller seasonal cycle compared with other clusters. It has a lower frequency as only 136 TCs are classified in this cluster and October is the peak month of TC genesis. Cluster 3 is similar to cluster 1, but has fewer TCs in April and none in January, February and March. The TC formation in cluster 3 peaks in August and September is the next most active month. Cluster 4 shows a smooth evolution of TC activity from January to June and

has a bimodal distribution with one maximum in July and another in September. The largest group in genesis clustering is cluster 1, which accounts for TC formation over the Philippine Sea. Environmental conditions in this cluster are the most conducive for TC development. This is attributed to warmer SSTs and weaker wind shears that enhance TC formation. The differences in seasonality between clusters are attributed to changes in large-scale environmental factors associated with TC formation [14].

The seasonal genesis cluster cycle is summarized in box plots (Figure 8) showing the outlier months (red crosses). The box plots also demonstrate season length, mean, median, and upper and lower quartiles. Clusters 1 and 3 have 9-month seasons from April to December, but TCs in January to March for cluster 1 and January for cluster 3 are considered outliers. Cluster 2 has a year-round season while cluster 4 has 10-month season and TCs in January and February are outliers. The all TC boxplot shows January and February as statistical outliers, and the 10-month season from March to December shows August is the peak of TC development in the Philippine domain.

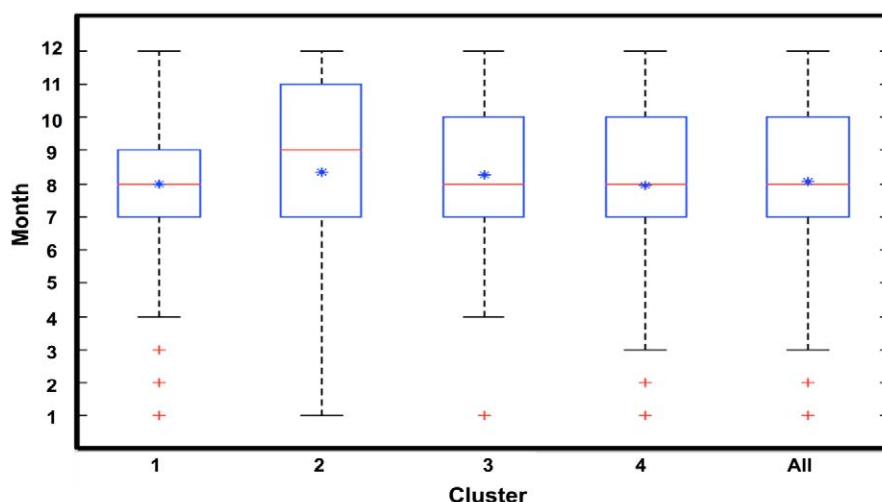


Figure 8: Boxplots of genesis clusters and all TCs showing the 25th and 75th percentiles (upper and lower limits of the box), the mean (asterisk), the median (red line) and outliers (red crosses) of the temporal distribution of TC genesis.

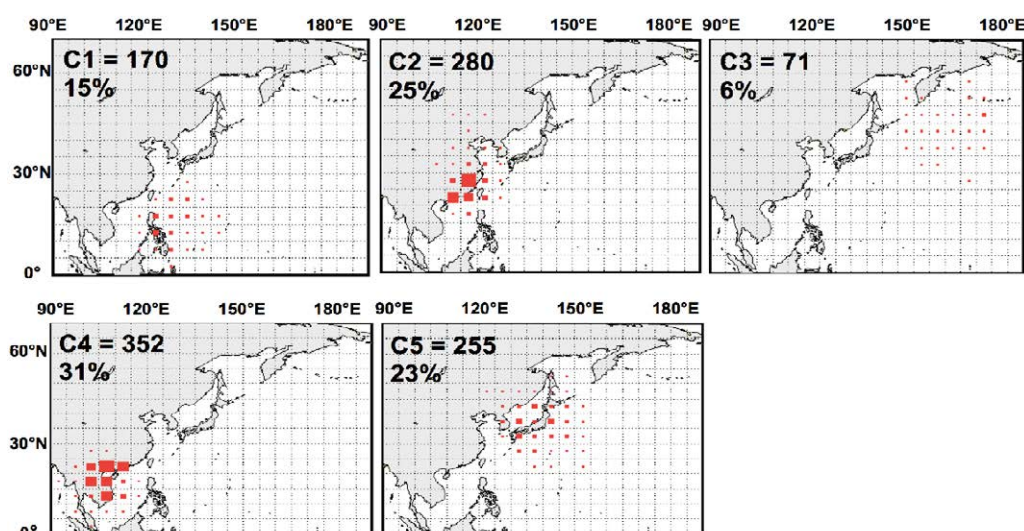


Figure 9: The cumulative density of decay points in each 5° latitude by 5° longitude grid box and the percentage of TC decays in each cluster.

Clustering of Decay Locations

TC landfall is contingent upon TC track. The main impact of a TC usually is at and after landfall [45]. The track of the TC is critical in determining the eventual cost. In the preceding section, TC decay locations are clustered and the TC tracks from decay clusters are plotted. This demonstrates the potential of predicting possible decay points as a function of genesis location. Figure 4b shows decay points of all TCs in the Philippine domain, color coded to signify the 5 decay clusters and the distinct features of each cluster. The number of TCs decaying in each grid box for each cluster is shown in Figure 9. The cumulative density of the distribution indicates the most preferred area of TC decay.

Decay clusters reflect the classification according to the threat region. Decay cluster 1 accounts for 170 TCs (~15%) that includes TCs that threaten the Philippines. Decay cluster 4 has the most cases, corresponding to land falling TCs over the South China and mainland

Southeast Asia or Indochina, comprising 352 TCs (~31%). Decay clusters 2 are the TC decay locations over Eastern China with 280 TCs (~25%). TCs that dissipate over Taiwan and Japan are in cluster 5 comprising 255 TCs (~23%). TCs decaying in higher latitudes of North Pacific, south of the Bering Sea, comprise cluster 3 with only 71 TCs (~6%), representing the smallest cluster that occurs least frequently.

Tracks of TCs from decay clusters are in Figure 10, exhibiting the region of genesis and track type for each decay cluster. From the decay clustering, straight tracks are readily separated from recurving tracks. Of 1,128 Decay points, ~29% of TCs have recurving tracks, including clusters 3 and 5, thereby posing a threat to Taiwan and Japan, and include TCs that decayed south of the Bering Sea. Straight-moving tracks, account for ~71% of the entire TC count. Decay clusters 1, 2, and 4 have straight tracks, making landfall over Philippines, Eastern China, and Indochina, respectively. TCs that strike in Taiwan and Japan and east of Japan have recurving tracks.

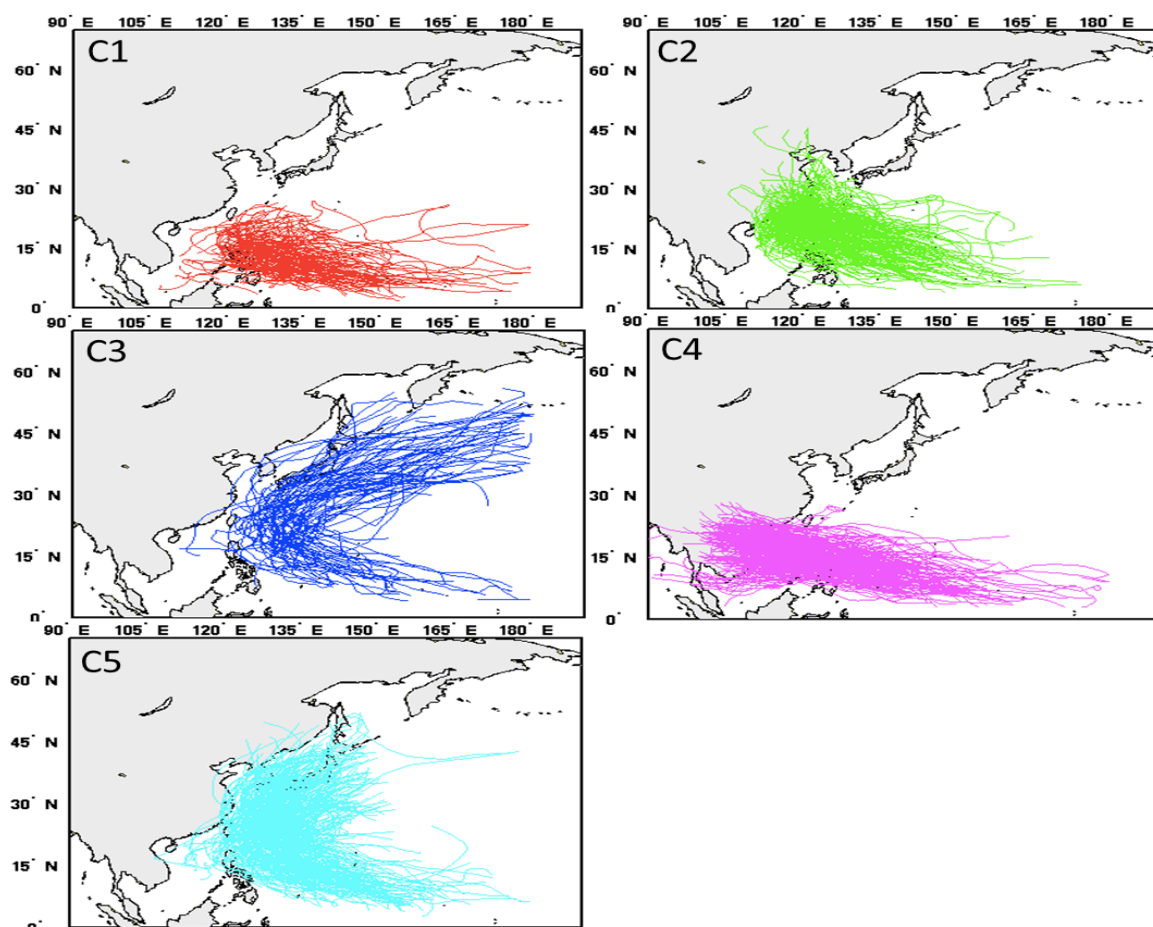


Figure 10: The TC tracks corresponding to decay points in each of the five decay clusters.

Decay Seasonality

The monthly frequency of each decay cluster is shown in Figure 11. Clusters 1 and 3 have flatter seasonal cycles, but cluster 1 has the broadest TC seasonal distribution, stretching from January to December. The smallest TC decay region, defined by cluster 3, is active only from April through December with September as the peak month. There are no TC decays in cluster 2 in January and February and almost none in March and December. A significant increase occurs in the July TC decay numbers, then decreases continuously starting in August and persisting until December. Cluster 4 has no TCs in February but an almost uniform January, March, and April TC decay numbers then continue to increase up to October, which is also the peak of TC season; followed by a notable dip in November and December. Cluster 5 has a steady monthly evolution of TC decay numbers. No TC decay occurs in January and February. From May, TC decay numbers increase, peaking in August then decreasing in September through November, with almost none in December.

A succinct description of the seasonal cycle of each decay cluster is shown in Figure 12. Cluster 1 has a year-round season while clusters 2 and 3 have nine-month seasons and TCs in March of cluster 2 are treated as outliers. Cluster 4 has the shortest season with only eight months length and TC decays in January, March and April considered as outliers. Cluster 5 and the cluster that represents all the TCs have 10-month seasons.

Clustering of Tracks

To isolate predictable aspects of the movement or landfall of TCs affecting the Philippines and in mitigating the damage, it is necessary to understand the characteristics of various TC tracks and the large-scale environmental factors that affect them. The landfall risk of a TC depends on its trajectory. The TC trajectory varies strongly with the season [46,47], as well as on interannual [41] and interdecadal time scales [48]. Cluster analysis has long been used to discover configurations hidden in historical TC tracks. The quantitative characteristics of clusters in TC tracks can provide valuable assistance in TC track and landfall prediction. Previous researchers noted that an effective way to explain the characteristics of various TC tracks is to classify TC trajectories into definite numbers of patterns [3-4,9-11,18,22,47,49-51].

The geographical position of the individual point that represents each TC track is presented in Figure 4c, and the colors indicate the cluster number. The density of TC passages per grid box is given in Figure 13. The density and size of the red square marker indicate number of TC passages per grid box. Also in Figure 13, in parenthesis, is the percentage of TCs for each cluster relative to all TCs. Clustering of TC tracks produced 6 clusters but the two main trajectory types identified by the cluster analysis correspond to straight-moving and recurring tracks. The geographical positions of individual TC tracks, separated by cluster, are in Figure 14. The 6 track clusters correspond to track locations that are more detailed, highlighting differences between

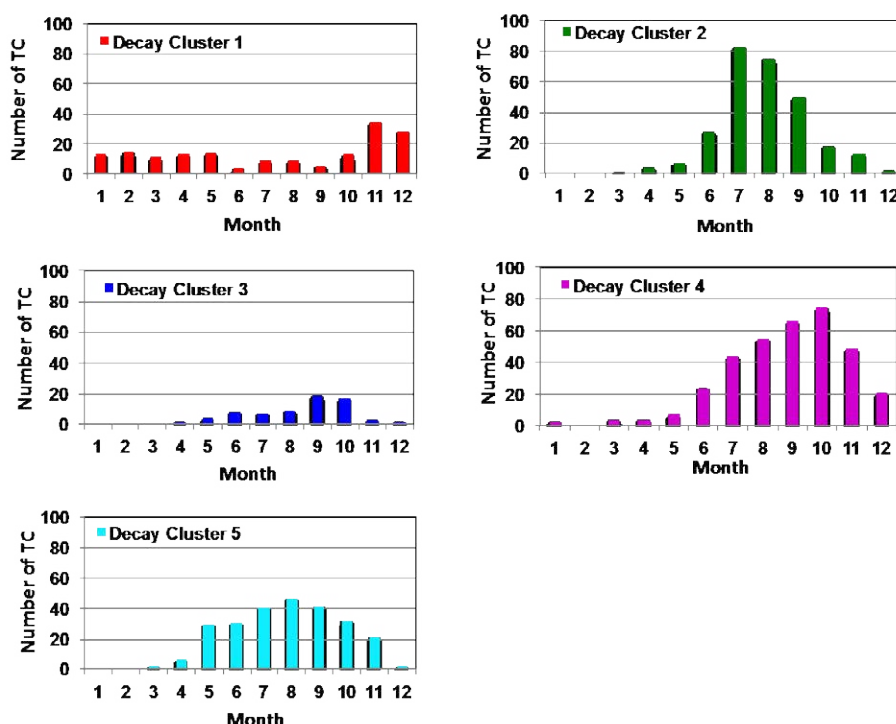


Figure 11: The number of TCs by calendar month for each decay cluster.

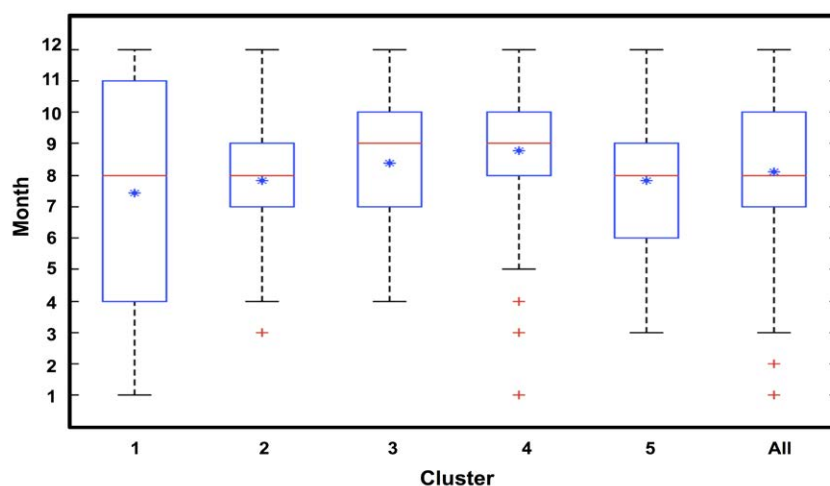


Figure 12: Boxplots of decay clusters and all TCs showing the 25th and 75th percentiles (upper and lower limits of the box), the mean (asterisk), the median (red line) and outliers (red crosses) of the temporal distribution of TC decay.

these two main types. There are 3 clusters with patterns of recurring tracks (clusters 1, 2, and 5) and also 3 clusters with straight tracks (clusters 3, 4, and 6). The clusters corresponding to straight tracks share great similarity exhibiting limited geographical cluster spread, whereas, the recurring types are more diffuse, occupying a much larger area of the WNP.

Cluster 1 consists of 219 TCs (~20% of the total), the second largest cluster, with short recurring tracks heading to Japan and some hitting Eastern China, Taiwan, and Korea. TCs in cluster 2 have long straight tracks that move northwestward before recurving northeastward toward south of Japan. Cluster 2 consists of 128 TCs (~12%). The

majority of the TCs in this cluster recurved northeastward while a few crossed the Philippines and headed straight to the east coast of China. Cluster 3 has 202 TCs (~18%) that represent west northwestward, long, straight tracks across the Philippines and the South China Sea heading to inland regions of the southeastern coast of mainland China and mainland Southeast Asia (Indochina). Cluster 4 is the most frequent trajectory type, with 248 TCs (~22%), and the track type resembles that of cluster 3, also heading west northwestward, but with shorter straight tracks and making landfall over regions similar to those of cluster 3. TCs in cluster 5 remain mostly offshore and dissipate over the open sea east of Japan, posing no threat to land. The TC tracks have a similarity

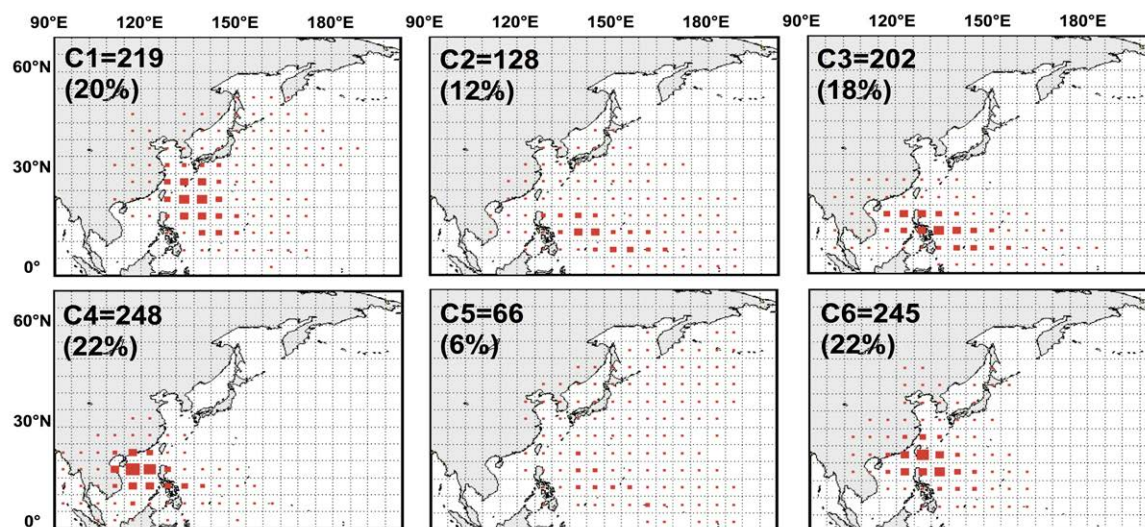


Figure 13: The cumulative density of TCs passing in each 5° latitude by 5° longitude grid box and the percentage of TC tracks in each cluster.

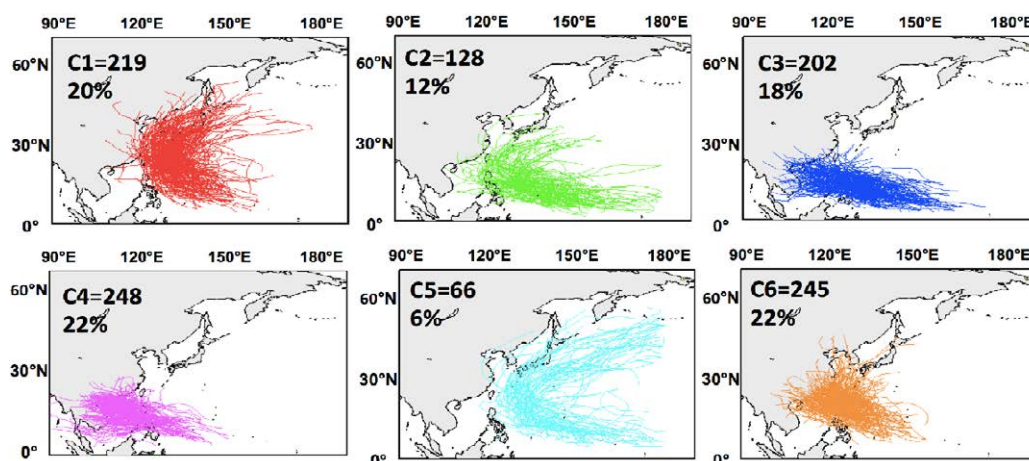


Figure 14: The TC Tracks in each of the six track clusters.

with cluster 1, characterized by longer recurving or parabolic shapes. The smallest cluster, cluster 5, has only 66 TCs (~6%). Cluster 6 is one of the two most frequent tracks with 245 TCs (~22%), comparable to clusters 3 and 4, characterized by straight-moving tracks with northwestward direction passing through the northern Philippines then striking Taiwan and the east coast of mainland China. The three dominant clusters thus are 1, 4, and 6, each cluster accounting for at least 20% of the tracks. Clusters 2 and 3 have lower percentages (~12% and ~18%, respectively), whereas cluster 5 is relatively rare, with ~6% of the TC count.

TC Track Seasonality

Monthly TC distributions for track clusters are summarized in Figure 15, showing the seasonality of TC activity for each track type. Track clusters 1, 4, and 6 have similar shapes and appearances. They have narrower seasonal distributions, and commence in April and end in December, except for cluster 1 that ends in November. Peak months are August, September, and July for clusters 1, 4, and 6, respectively. Track clusters 2 and 3 have smaller amplitude but broader seasonal

cycles with TC occurrences year round and both clusters reach their highest TC frequency in November with sub-peaks in April and July, respectively. Track cluster 5 is flatter and narrower, starting in April and running through December, with a peak in October. Figure 16 provides box plots of track clusters. Clusters 4 and 6 have 9-month seasons but TCs in January, from cluster 4, are treated as outliers. Clusters 2 and 3 have year-round seasons. Cluster 5 has the shortest season of 8 months and TCs in April are regarded as outliers. The box plot for all TCs has a 10-month season, and thus January-February TCs are flagged as outliers.

Monthly Analysis of TC Activity

The motivation for the monthly analysis is to show the important temporal and spatial behavior of TC activity. The plots of TC genesis and decay locations and tracks provide the distinct characteristics of each cluster with time. The graphs show the TC frequency of each cluster by month. The TCs are color coded to signify the cluster number to which it belongs. January, February, and March (JFM) are the quietest months and represents the “calmest” phase of TC activity

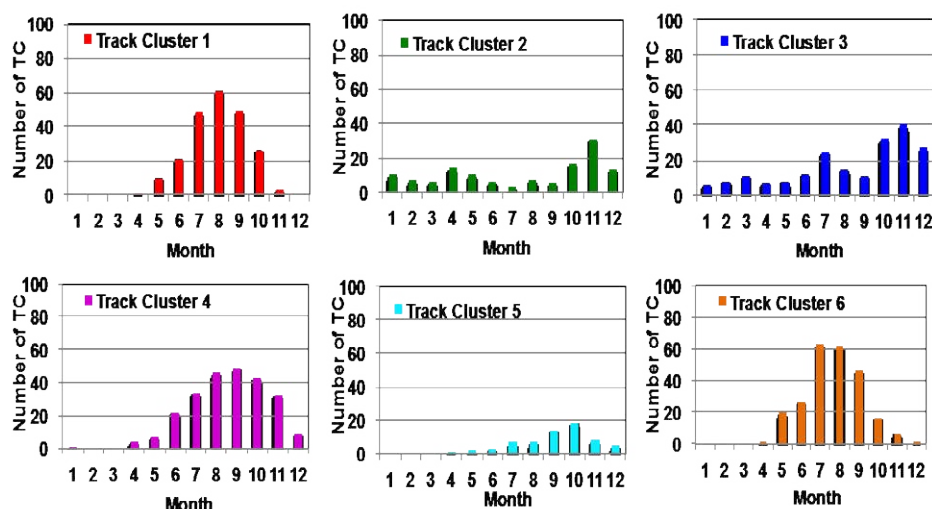


Figure 15: The number of TCs by calendar month for each track cluster.

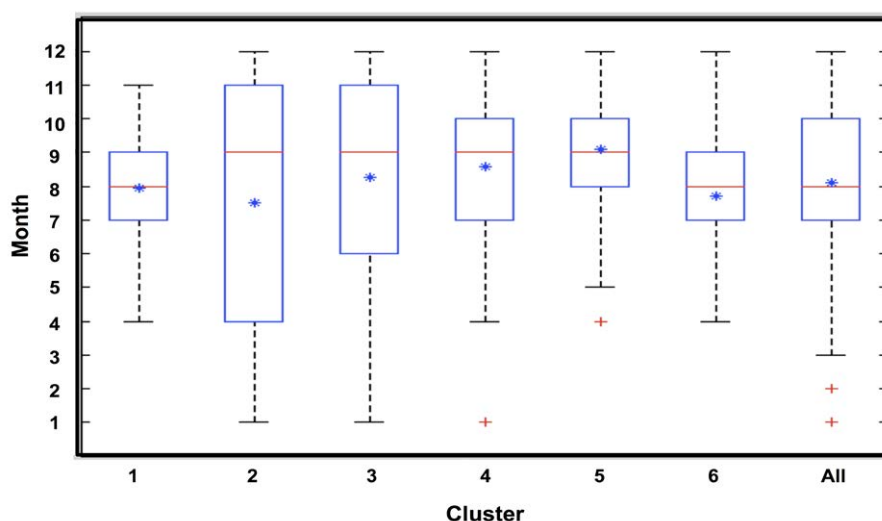


Figure 16: Boxplots of track clusters and all TCs showing the 25th and 75th percentiles (upper and lower limits of the box), the mean (asterisk), the median (red line) and outliers (red crosses) of the temporal distribution of TC tracks.

and TC formations occur east of the Philippines (Figure 17), over the Philippine Sea (genesis cluster 1), Northern Marianas Islands and central Micronesia (genesis cluster 4), and the Marshall Islands (genesis cluster 2). They are confined to lower latitudes. No TCs formed over the South China Sea (genesis cluster 3). In January, TCs make landfall over the Philippines including the Philippine Sea and South China Sea (decay clusters 1 and 4). TC landfalls in February are solely from cluster 1. TCs in March decay in regions associated with clusters 1, 2, 4, and 5, posing a threat to the Philippines, and Indochina. The most active decay region in JFM is cluster 1, which includes the Philippines and the Philippine Sea. The TC tracks occupy the lower latitudes, up to 23°N, and follow straight path (track clusters 3 and 4). However, some TCs veer northeastward (track cluster 2).

The second quarter, April-June (AMJ) is marked by an increase in genesis numbers over that of JFM (Figure 18). The birthplaces extend farther north to 22°N, about 6° latitude above the JFM birthplaces. All

genesis clusters are represented in this quarter including the South China Sea (genesis cluster 3). More than half (~52%) of the TCs in April originate from genesis cluster 4, which are those that form near Northern Marianas Islands and central Micronesia. Very few TCs belong to genesis cluster 3, at just 10% of the total TC genesis number. Notably, during the 62-year period, there is only one TC genesis in April over the South China Sea. Straight tracks are the major trajectory type, accounting for >50% of TCs for that month, and decay cluster 1 accounting for ~45% of April landfalling TCs over the Philippines and the Philippine Sea. In May and June, TC formation increases over the South China Sea and more TCs develop in higher latitudes, especially in June. TC development is also common in genesis clusters 1 and 4, with TC formation over the Philippine Sea, Northern Marianas Islands and central Micronesia.

Figure 19 shows the July-September (JAS) TC activity. The frequency and areal extent of TC genesis are largest during JAS,

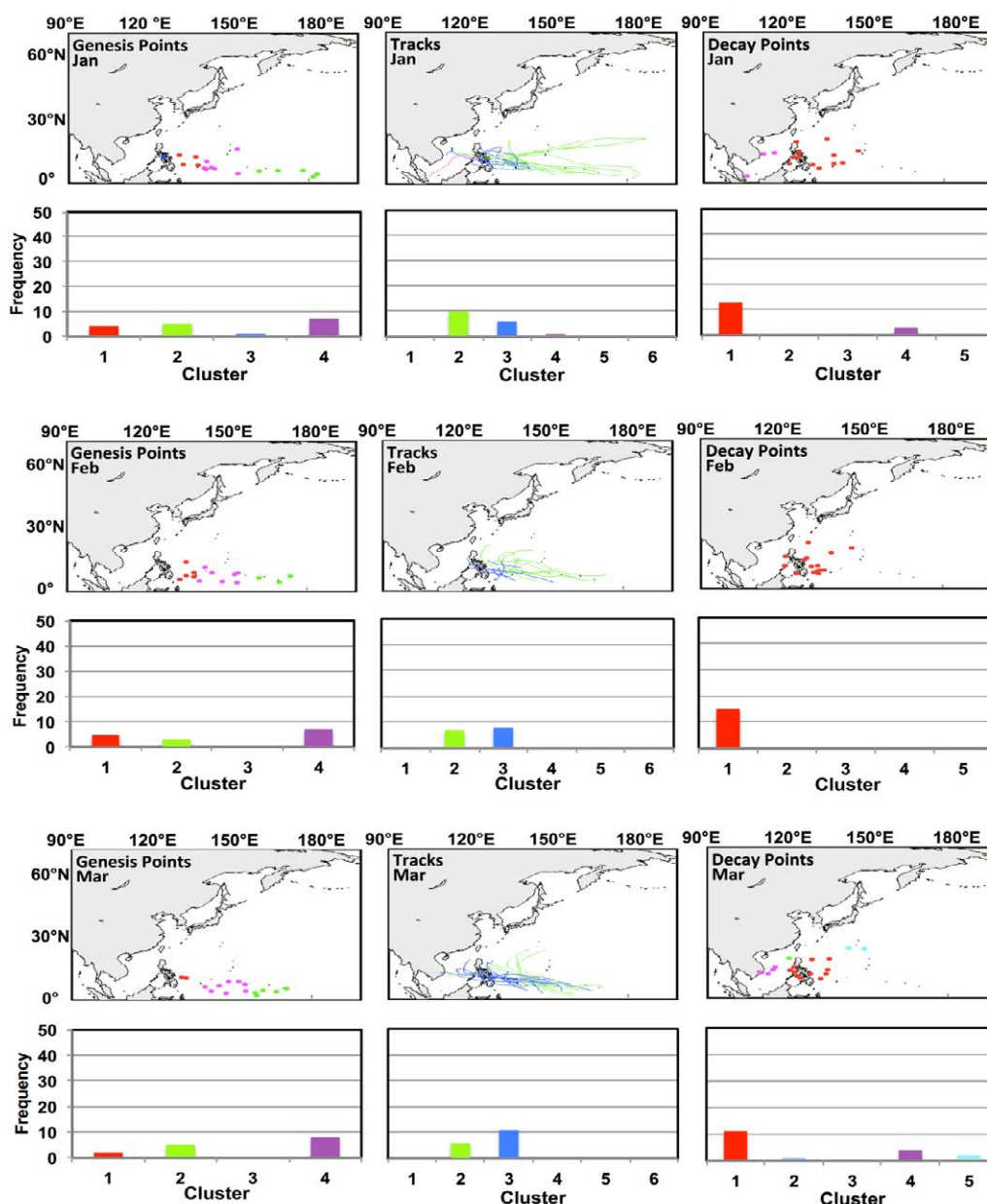


Figure 17: The monthly variations of TC activity over the Philippine domain during the months January-March.

so TC tracks are densest compared with other quarters. In July, the most frequent genesis clusters are 1 (Philippine Sea) and 4 (northern Marianas Island and central Micronesia) with ~79% of total TC genesis. Typically, they decay over eastern China (decay cluster 2, ~45%) while some TCs decay over the South China Sea, including mainland Southeast Asia, Taiwan and Japan. Short straight tracks toward eastern China are typical of July and short recurving tracks heading toward Japan are the next most common. TC genesis reaches its northernmost position in August but southernmost cyclogenesis also slightly shifts northward. Genesis cluster 1 (Philippine Sea) has the highest TC formation in August, with ~43% of the August TCs. Genesis clusters 3 (South China Sea) and 4 (Northern Marianas and central Micronesia) as the next major regions of formation, both with ~24% of August TC genesis. Most straight-moving TCs during this

month decay over mainland Indochina and over Eastern China, while most of the recurving TCs dissipate over Taiwan and Japan.

In September, TC genesis shifts slightly southward and decreases. Genesis clusters 1 and 4 have distinctive maxima during this month. TC tracks are mainly short, straight tracks toward Eastern China including the southern coast of Eastern China, and short recurving tracks head towards Taiwan and Japan. In contrast, decay cluster 1 has the lowest TC dissipation; they are TCs decaying over or near the Philippines. In September and October, recurving tracks extend farther northeast.

Genesis and decay locations and TC tracks are in Figure 20, during the October - December (OND) quarter. OND has less TC genesis than JAS and the latitudinal extent of genesis and tracks is not as large and dense. However, OND has the greatest landfall probabilities and more

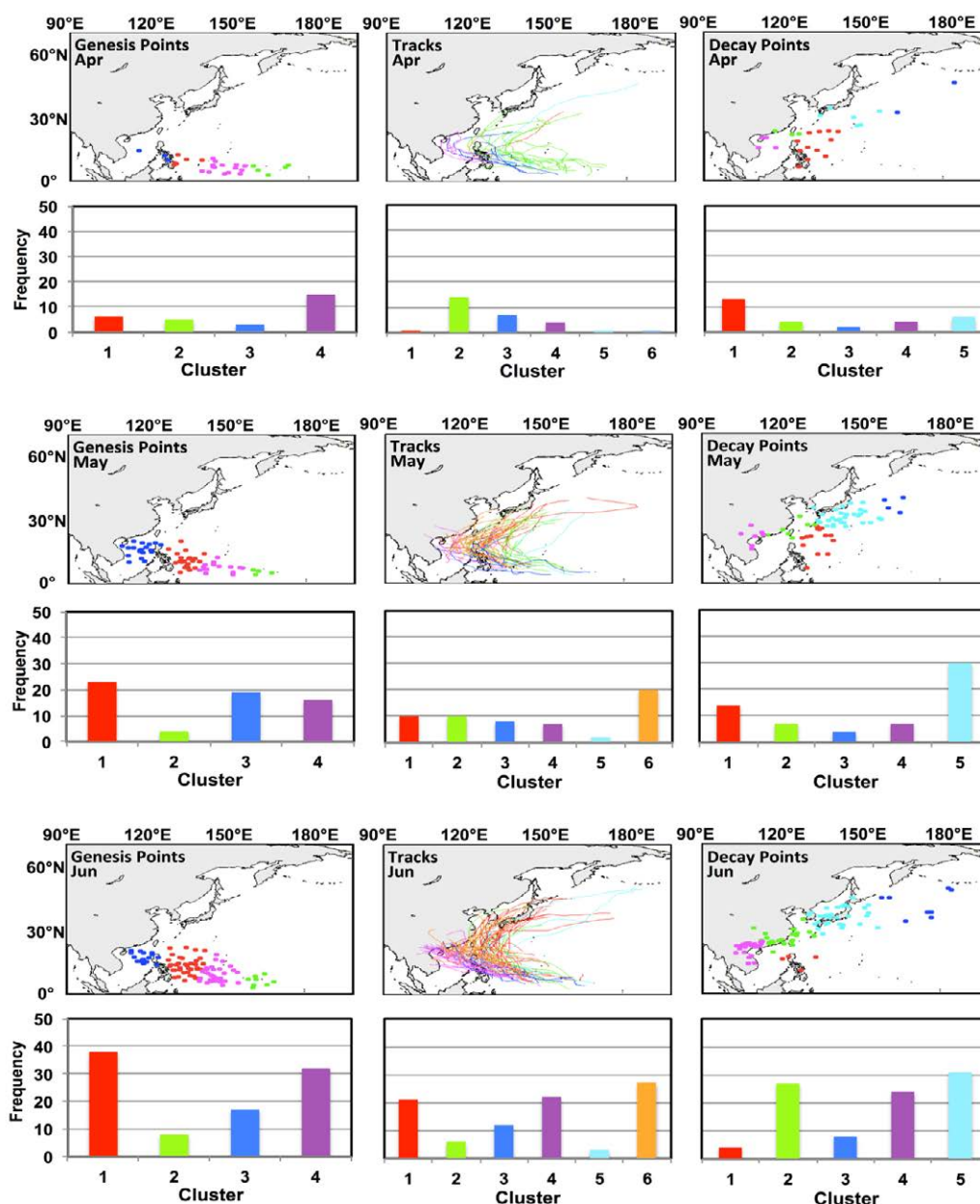


Figure 18: The monthly variations of TC activity over the Philippine domain during the months April-June.

straight-moving TCs. Genesis points dominate clusters 1 and 4, due to TCs that form over the Philippine Sea and near the Northern Marianas Islands, including central Micronesia. TC formation decreases in December. TC tracks in OND occur mainly from clusters 3 and 4, and have straight trajectories crossing the Philippines and South China Sea, before reaching Southeast Asia and Eastern China. In OND, TC genesis is closer to the dateline and straight-moving TCs occur in these months. TC decays are mostly from decay clusters 1 and 4.

Summary and Conclusions

This study includes all TCs in the Philippine domain, regardless of intensity classification, maximizing sample size. It is an extension of existing studies of TCs with storm intensity [3] and TCs occurring

in the peak season, June-November [47]. Harr [47] excluded tracks considered to be unusual. Here, TCs are eliminated objectively using negative silhouette values to improve cluster cohesiveness.

The mean genesis location of the TCs in the WNP has a well-defined annual cycle [52,53], with the average latitude northernmost in August and closest to the equator in February. This cycle is consistently present in monthly cluster analyses. Formation clusters of Philippine TCs occupy the western Pacific warm pool and lower latitudes, because of low vertical wind shear. TC genesis in fall and winter is more equatorward, whereas summer TC genesis points are northernmost. The cluster with highest TC genesis occurs in the Philippine Sea, just east of the Philippines, because the climatological SSTs in the Philippine Sea are above 26°C year round. Harr [47] suggest that the

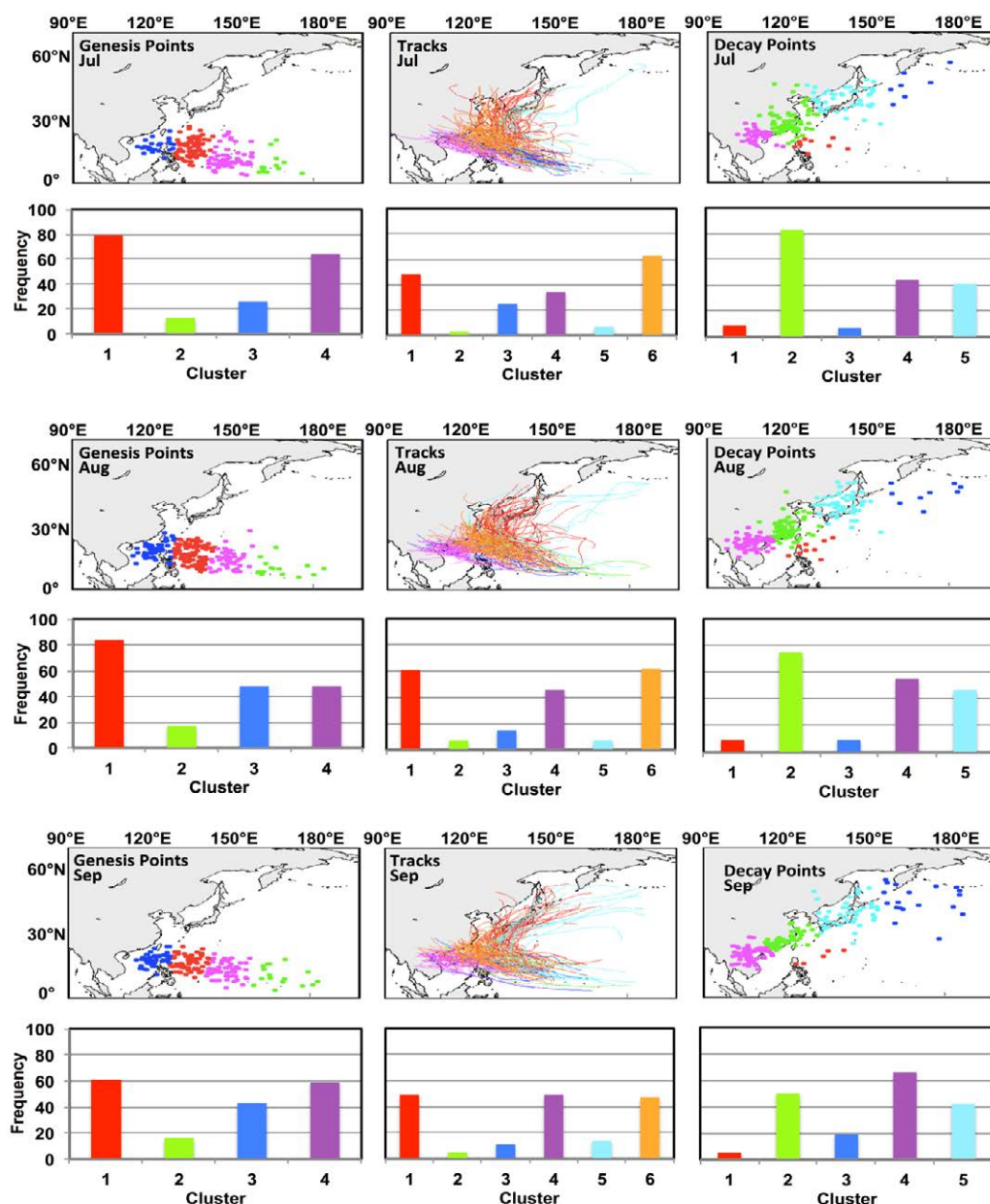


Figure 19: The monthly variations of TC activity over the Philippine domain during the months July-September.

combination of the broad belt of anomalous equatorial westerlies between 75°E-145°E and anomalous easterlies between 20°N-30°N and 120°E-150°E form a large region of cyclonic horizontal shear; possibly associated with an enhanced monsoon trough, a favorable environment for TC genesis [54]. Anomalous easterlies between 20°N-30°N are produced by an enhanced gradient between the active monsoon trough to the south, and the strong subtropical ridge to the north. Anomalous anticyclonic circulation over the East China Sea and southern Japan implies a strengthened subtropical ridge; these large-scale circulations provide an environment favorable for TC genesis in the monsoon trough. It is likely that the Philippines will experience land falling TCs; TCs cross the Philippines before striking Eastern China, although some pass north of the Philippines. TCs in genesis clusters close to the dateline are most likely to follow recurving tracks depending on

the prevailing wind affecting the Philippines. During the Northeast monsoon peak, TCs have straight trajectories and strike or cross the Philippines but, during the Southwest monsoon peak, they generally recurve northeastward. TCs originating over the Philippine Sea usually decay over mainland Indochina and Eastern China. Genesis clusters (2 and 5) with TCs developing closer to the dateline and Equator have the longest tracks, the longest mean durations compared to other clusters. Some TCs pass southern Japan while others track east of Japan. Other genesis clusters are associated with straight-moving TCs. TC genesis locations contribute to the intraseasonal variability of TC tracks but large- and synoptic-scale circulations also influence tracks.

The two main trajectory types identified in this study, namely recurving and straight-moving, are consistent with two principal track types identified previously [9,22,50,55]. The 6 track clusters, that present

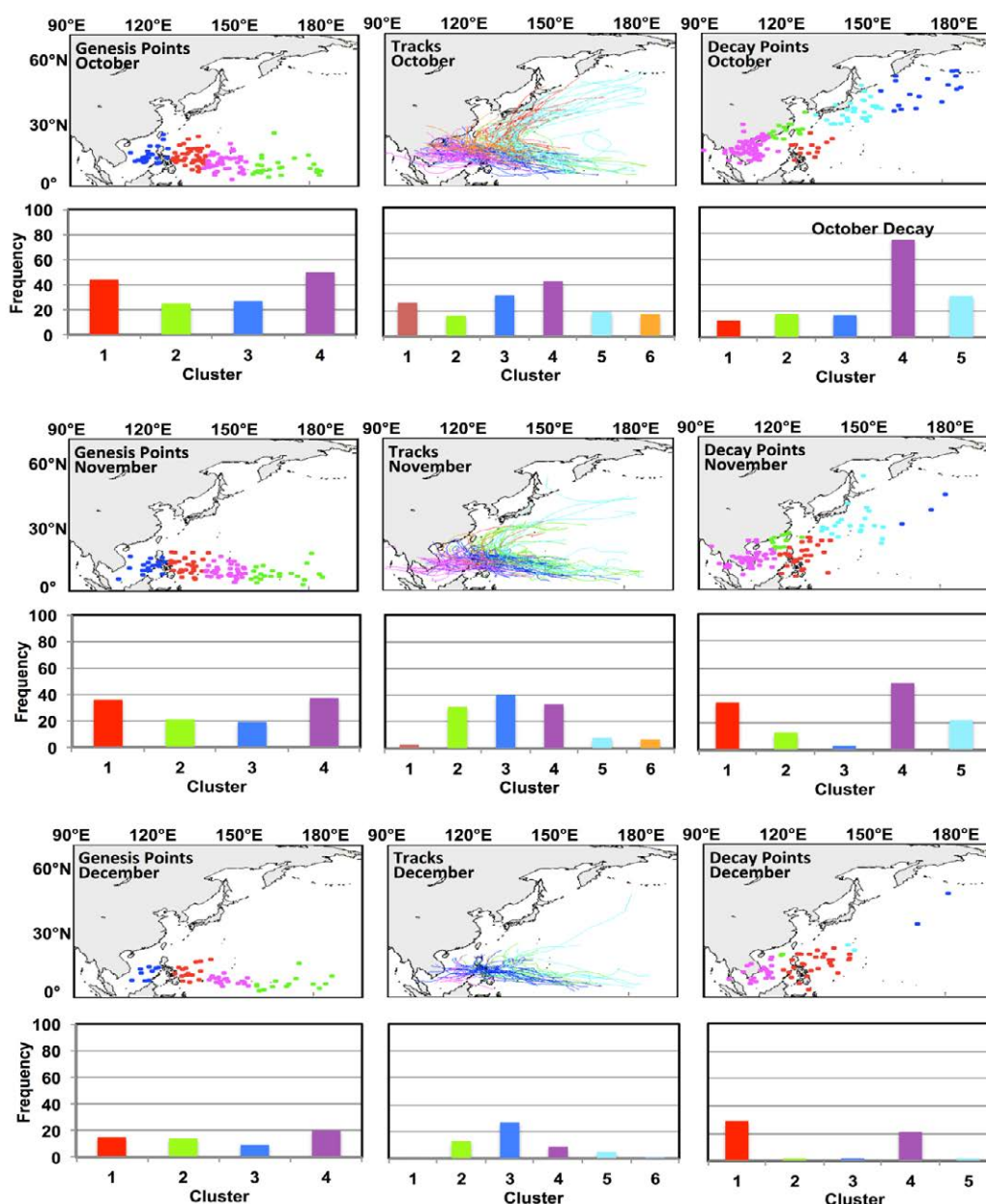


Figure 20: The monthly variations of TC activity over the Philippine domain during the months October-December.

broader trajectory types, are very similar with cluster results of Camargo [9] except for the cluster that represents the TCs outside the Philippine domain. Although the two studies used two different cluster methods, the results are similar; implying that the K-means employed in this study provides accurate clustering. The clustering of Philippine TC tracks reveals ~62% of TCs follow a straight moving trajectory type corresponding to TCs that belong in track clusters 3, 4, and 6. Track clusters 1, 2, and 5 have recurving trajectory types. The clustering also suggests that TCs originating east of ~150°E have a higher probability (~67%) of recurving, depending on time of year and prevailing wind, whereas the probability of straight-moving tracks remains higher for TCs forming west of 140°E, at ~76%.

The monthly distribution of genesis, decay and track clusters demonstrates that seasonality and general characteristics of TC activity

differ from cluster to cluster. This variation in TC characteristics is attributable to the large-scale circulations. Philippine TC genesis positions shift north from June-August, but regress southward in September. These seasonal variations reflect the positions of the monsoon trough and the North Pacific subtropical high [52]. Monthly cluster analyses identify genesis, decay and tracks that dominate in particular months, providing a useful forecast tool. Patterns in genesis locations can be used in track and landfall forecasts, once the genesis location of a TC is known. Notably, analysis of the seasonal variations of the genesis, decay and tracks produces spatial/temporal characteristics that provide valuable predictive guidance.

Acknowledgement

The authors express their gratitude to the Cooperative Institute for Mesoscale

Meteorological Studies (CIMMS), University of Oklahoma, for supporting the first author. We also thank Dr. Michael B. Richman for his assistance in interpreting the silhouette coefficient analysis.

References

- Blender R, Fraedrich K, Lunkeit F (1997) Identification of cyclone track regimes in the North Atlantic. *Quarterly Journal of the Royal Meteorological Society* 123: 727-741.
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1: 281-297.
- Elsner JB, Liu KB (2003) Examining the ENSO-typhoon hypothesis. *Climate Research* 25: 43-54.
- Elsner JB (2003) Tracking hurricanes. *Bulletin of the American Meteorological Society* 84: 353-356.
- McCloskey TA, Bianchette T, Liu KB (2013) Track patterns of land falling and coastal tropical cyclones in the Atlantic Basin, their relationship with the North Atlantic Oscillation (NAO), and the potential effect of global warming. *American Journal of Climate Change* 2: 12-22.
- Gong X, Richman MB (1995) On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *Journal of Climate and Applied Meteorology* 8: 897-931.
- Gaffney SJ (2004) Probabilistic curve-aligned clustering and prediction with regression mixture models. Ph.D. thesis, University of California, Irvine, CA 281.
- Gaffney SJ, Robertson AW, Smyth A, Camargo SJ, Ghil M (2007) Probabilistic clustering of extratropical cyclones using regression mixture model. *Climate Dynamics* 29: 423-440.
- Camargo SJ, Robertson AW, Gaffney SJ, Smyth P, Ghil M (2007) Cluster analysis of typhoon tracks. Part I: General properties. *Journal of Climate* 20: 3635-3653.
- Camargo SJ, Robertson AW, Gaffney SJ, Smyth P, Ghil M (2007) Cluster analysis of typhoon tracks. Part II. Large-scale circulation and ENSO. *Journal of Climate* 20: 3654-3676.
- Camargo SJ, Robertson AW, Barnston AG, Ghil M (2008) Clustering of eastern North Pacific tropical cyclone tracks: ENSO and MJO effects. *Geochimistry, Geophysics, Geosystems* 9: 1-4.
- Chand SS, Walsh KJ (2009) Tropical cyclone activity in the Fiji region: spatial pattern and relationship to large-scale circulation. *Journal of Climate* 22: 3877-3893.
- Chand SS, Walsh KJ (2010) The influence of the Madden-Julian oscillation on tropical cyclone activity in the Fiji region. *Journal of Climate* 23: 868-886.
- Ramsay HA, Camargo SJ, Kim D (2011) Cluster analysis of tropical cyclone tracks in the southern hemisphere. *Climate Dynamics* 39: 897-917.
- Kossin JP, Camargo SJ, Sitkowski M (2010) Climate modulation of North Atlantic hurricane tracks. *Journal of Climate* 28: 3057-3076.
- Zhang W, Leung Y, Wang Y (2012) Cluster analysis of post-landfall tracks of land falling tropical cyclones over China. *Climate Dynamics* 40: 1237-1255.
- Ash KD, Matyas CJ (2012) The influences of ENSO and the subtropical Indian Ocean Dipole on tropical cyclone trajectories in the southwestern Indian Ocean. *International Journal of Climate* 32: 41-56.
- Nakamura J, Lall U, Kushnir Y, Camargo SJ (2009) Classifying North Atlantic tropical cyclone tracks by their mass moments. *Journal of Climate* 22: 5481-5494.
- Colbert AJ, Soden BJ (2012) Climatological variations in North Atlantic tropical cyclone tracks. *Journal of Climate* 25: 657-673.
- Choi KS, Kim BJ, Byun HR (2008) Relationship between Korean Peninsula land falling tropical cyclones and interannual climate variabilities. *Journal of Korean Earth Science Society* 29: 375-385.
- Kim HS, Kim JH, Ho CH, Chu PS (2011) Pattern classification of typhoon tracks using the fuzzy c-means clustering method. *Journal of Climate* 24: 488-508.
- Harr PA, Elsberry RL (1995) Large-scale circulation variability over the tropical western North Pacific. Part I: Spatial patterns and tropical cyclone characteristics. *Monthly Weather Review* 123: 1225-1246.
- Corporal-Lodangco I, Leslie L, Lamb P (2015) Impacts of ENSO on Philippine Tropical Cyclone Activity. *J Climate*
- Smith TM, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). *Journal of Climate* 21: 2283-2296.
- Kalnay E (1996) The NCEP/NCAR 40-year re-analysis project. *Bulletin of the American Meteorological Society* 77: 437-471.
- Mooi E, Sarstedt M (2011) A Concise Guide to Market Research: The Process, Data, and Methods using IBM SPSS statistics 11-23.
- Tan PN, Steinbach M, Kumar V (2006) *Introduction to Data Mining*. Pearson, Boston.
- Aldenderfer MS, Blashfield RK (1984) *Cluster analysis*. Sage Publications, Newbury Park, CA.
- Hair JF, Tatham R, Black, WC (1992) *Multivariate Data Analysis with Readings*, Macmillan, New York.
- Kaufman L, Rousseeuw PJ (2005) *Finding groups in data: An Introduction to Cluster Analysis*.
- Perrone TJ, Lowe PR (1986) A statistically derived prediction procedure for tropical storm formation. *Monthly Weather Review* 114: 165-177.
- Gray WM (1993) Predicting Atlantic basin seasonal tropical cyclone activity by 1 August. *Weather and Forecasting* 8: 73-86.
- Wendland WM (1977) Tropical storm frequencies related to sea surface temperatures. *Journal of Applied Meteorology* 16: 477-481.
- Emanuel KA (1999) Thermodynamic control of hurricane intensity. *Nature* 401: 665-669.
- Wu L, Wang B (2004) Assessing impacts of global warming on tropical cyclone tracks. *Journal of Climate* 17: 1686-1698.
- Holland GJ, Wang Y (1995) Baroclinic dynamics of simulated tropical cyclone recurvature. *Journal of Atmospheric Sciences* 52: 410-426.
- Wu CC, Emanuel KA (1993) Interaction of a baroclinic vortex with background shear: application to hurricane movement. *Journal of Atmospheric Sciences* 50: 62-76.
- Wang Y, Holland GJ (1996) Tropical cyclone motion and evolution in vertical shear. *Journal of Atmospheric Sciences* 53: 3313-3332.
- Corbosiero KL, Molinari J (2003) The relationship between storm motion, vertical wind shear, and convective asymmetries in tropical cyclones. *Journal of Atmospheric Sciences* 60: 366-376.
- Carlson TN (1971) An apparent relationship between the sea-surface temperature of the tropical Atlantic and the development of African disturbances into tropical storms. *Monthly Weather Review* 99: 309-310.
- Chan JCL (1985) Tropical cyclone activity in the Northwest Pacific in relation to the El Niño/Southern Oscillation phenomenon. *Monthly Weather Review* 113: 599-606.
- Chan JCL, Li KK (2005) Ensemble forecasting of tropical cyclone motion using a barotropic model Part III: Combining perturbations of the environment and the vortex. *Meteorology and Atmospheric Physics* 90: 109-126.
- Holland GJ (1983) Tropical cyclone motion: Environmental interaction plus a beta effect. *Journal of Atmospheric Sciences* 40: 328-342.
- Holland GJ (1993) Tropical cyclone motion. *Global guide to tropical cyclone forecasting*, G. J. Holland, Ed., World Meteorological Organization.
- Powell M, Soukup G, Cocke S, Gulati S, Morisseau-Leroy N et.al, (2005) State of Florida hurricane loss projection model: Atmospheric science component. *Journal of Wind Engineering and Industrial Aerodynamics* 93: 651-674.
- Gray WM (1979) Hurricanes: Their formation, structure and likely role in the tropical circulation. *Meteorology over the Tropical Oceans*, Royal Meteorological Society 155-218.
- Harr PA, Elsberry RL (1991) Tropical cyclone track characteristics as a function of large-scale circulation anomalies. *Monthly Weather Review* 119: 1448-1468.
- Ho CH, Baik JJ, Kim JH, Gong DY (2004) Interdecadal changes in summertime typhoon tracks. *Journal of Climate* 17: 1767-1776.

49. Hodanish S, Gray WM (1993) An observational analysis of tropical cyclone recurvature. *Monthly Weather Review* 121: 2665-2689.
50. Lander MA (1996) Specific tropical cyclone tracks and unusual tropical cyclone motions associated with a reverse-oriented monsoon trough in the western North Pacific. *Weather and Forecasting* 11: 170-186.
51. Hall TM, Jewson S (2007) Statistical modeling of North Atlantic tropical cyclone tracks. *Tellus* 59: 486-498.
52. Chia HH, Ropelewski CF (2002) The interannual variability in the genesis location of tropical cyclones in the northwest Pacific. *Journal of Climate* 15: 2934-2944.
53. Camargo SJ, Sobel AH (2005) Western North Pacific tropical cyclone intensity and ENSO. *Journal of Climate* 18: 2996-3006.
54. Love G (1982) The role of the general circulation in Western Pacific tropical cyclone genesis. *Colorado State University Atmospheric Science Paper* 340.
55. Sandgathe SA (1987) Opportunities for tropical cyclone motion research in the northwest Pacific region. *Naval Postgraduate School, Monterey, CA* pp: 36.