Article ID: 1006-8775(2006) 01-0142-09

CLASSIFYING WESTERN NORTH PACIFIC TROPICAL CYCLONES BY PHYSICAL INDEX SYSTEM

YAN Dong-yi (颜东谊), XU Kui (徐 奎), MA Chao (马 超), MA Man-cang (马满仓)
(State Key Laboratory of Hydraulics Engineering Simulation and Safety, Tianjin University, Tianjin 300072 China)

Abstract: The classification of tropical cyclones (TCs) is significant to obtaining their temporal and spatial variation characteristics in the context of dramatic-changing global climate. A new TCs clustering method by using K-means clustering algorithm with nine physical indexes is proposed in the paper. Each TC is quantified into an 11-dimensional vector concerning trajectory attributes, time attributes and power attributes. Two recurving clusters (cluster A and E) and three straight-moving clusters (cluster B, C and D) are categorized from the TC best-track dataset of the western North Pacific (WNP) over the period of 1949-2013, and TCs' properties have been analyzed and compared in different aspects. The calculation results of coefficient variation (CV) and Nash-Sutcliffe efficiency (NSE) reveal a high level of intra-cluster cohesiveness and inter-cluster divergence, which means that the physical index system could serve as a feasible method of TCs classification. The clusters are then analyzed in terms of trajectory, lifespan, seasonality, trend, intensity and Power Dissipation Index (PDI). The five classified clusters show distinct features in TCs' temporal and spatial development discipline. Moreover, each cluster has its individual motion pattern, variation trend, influence region and impact degree.

Key words: tropical cyclone; physical index; K-means clustering; Nash-Sutcliffe efficiency; inter-cluster divergence; intra-cluster cohesiveness; power dissipation index
CLC number: P444 **Document code:** A
doi: 10.16555/j.1006-8775.2018.02.003

1 INTRODUCTION

Tropical cyclone (TC) is a kind of devastating meteorological disaster and induces countless casualties and economic losses. Western North Pacific (WNP), where TCs occur actively, suffers more than 30 TCs every year on average, accounting for nearly one third of the gross TCs on the globe (Wang et al.[1]). In China, practically half of the natural disaster losses are caused by TCs (Xiao and Xiao [2]), resulting in an annual economic loss of about 200 billion RMB (Zhang et al.[3]). In the context of dramatic-changing global climate initiated by human activities and natural factors (Dessai [4]), the influence of TCs becomes more and more serious (Wu et al.[5]). In the perspective of disaster system theory, TCs, as the hazard factor of TCs disasters, have much uncertainty and complexity. The analysis of the cybotaxis of hazard factor and its space, time and power principles would provide a scientific basis for the prevention and evaluation of TCs disasters.

Received 2017-08-14; **Revised** 2018-03-29; **Accepted** 2018-05-15
Foundation item: National Key Research and Development Program of China (2016YFC0401903); National Natural Science Foundation of China (51722906, 51679159, 51509179); Tianjin Research Program of Application Foundation and Advanced Technology (15JCYBTC21800)
Biography: YAN Dong-yi, primarily undertaking research on tropical cyclones.
Corresponding author: XU Kui, e-mail: jackykui@126.com

Clustering TCs into several patterns is significant to comprehending their characteristics of temporal and spatial variation. In previous studies, TCs were classified into different clusters according to large-scale synoptic pattern or subjective analysis. Lander [6] discriminated the TC trajectories into four categories (straight moving, recurving, north oriented and South China Sea) based on the large-scale pattern of the WNP monsoonal flow. Horvath et al.[7] classified the TCs of Apennine and Adriatic area into four main types by subjective analysis technique: Genoa, Adriatic, simultaneous Genoa and Adriatic, as well as non-Genoa and non-Adriatic. Hodanish and Gray [8] divided the WNP TC trajectories into four clusters (sharply recurving, slowly recurving, left turning, and non-recurving) by means of the large-scale synoptic pattern, which interacted with the TCs' environment prior to and during the recurvature.

In the case of TC trajectories, objective analysis methods based on mathematical theory and data-mining algorithm have become prevalent in recent decades. The finite mixture model was applied to the classification of TC trajectories according to the trajectory shape and location. Camargo et al. [9] identified three straight-moving trajectory clusters and four recurving clusters from the WNP TC trajectories from 1950 to 2002 with this method. The fuzzy c-means clustering method was suggested to be a useful approach to probabilistic-type clustering of multitudinous TC tracks

(Kim et al.[10]). In order to overcome the shortcomings of the finite mixture model that the model requires identical trajectories lengths, Kim et al.[10] interjected a TC track into 20 segments artificially, and divided the WNP TC trajectories into seven different clusters. The tree construction algorithm C4.5 has been employed to analyze TC tracks as well. Zhang et al.[11] focused on the rules of TCs recurvature, and the parameters affecting TCs recurvature were categorized into three groups: large-scale circulation, circulations surrounding TCs and straight-moving TCs, and they applied the method to the classification of TCs tracks of the WNP and the South China Sea (SCS).

K-means method, as a common data-mining method, has been proposed to explore the TCs in the WNP (Elsner and Liu[12]; Nakamure et al.[13]; Yu et al.[14]), Atlantic (Elsner [15]) and other basins (Lin et al.[16]). Elsner and Liu [12] pointed out that K-means method could be applied to the classification of TCs based on the location with maximum intensity and dissipation. However, this method was not applicable to the cases that TCs had different trajectories lengths (Camargo et al.[9]). In order to overcome this deficiency, Nakamure et al.[13] suggested that TC trajectories could be clustered by using K-means algorithm with mass moments of centroid and variance. This method was well applied to analyze the north Atlantic TCs. Furthermore, Yu et al.[14] modified Nakamure's method by adjusting the weight factors to emphasize the significance of TCs trajectory direction, length and shape in clustering process. The WNP TC trajectories were separated into seven groups by the method. Whereas, Lin et al.[16] recommended that TC trajectories could be decomposed into three coefficients of different orders: genesis location, trajectory direction and curvature. In addition, they clustered the TC trajectories over the Bay of Bengal into westward-moving type, northward-moving type and northwestward-moving type by K-means algorithm with those coefficients.

Commonly the TC trajectory determines the landing risk and the influence scope. Most scholars so far have paid much attention to it. Whereas, the TC time and power attributes are of great significance to the TC disasters and should not be neglected. TCs, with different occurrence time, intensity and influence time will vary in their impact degrees. Thus, aiming to analyze the TCs' cybotaxis and obtain the multiple characteristics of TC trajectory, time and power attributes, a new method using K-means clustering algorithm with physical indexes is proposed for TCs classification in this paper. A series of physical indexes, including genesis location, dissipation location, trajectory length, mean deflection angle, lifespan, seasonality, intensity and Power Dissipation Index (PDI) (Emanuel[17]), are suggested to describe TCs' space-time and power discipline. The TCs information is converted into 11-dimensional vectors applied to K-means clustering algorithm, whose contents emphasize not only trajectory attributes but also power attributes and time attributes. Moreover, in the purpose of evaluating the result of TC classification, the coefficient variation (CV) values and Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe [18]) are applied for contrastive analysis of degree of inter-cluster divergence and intra-cluster cohesiveness.

In what follows, the best-track dataset resource and the clustering methodology applied in the classification of the WNP TCs are interpreted in section 2. In section 3, the degree of inter-cluster divergence and intra-cluster cohesiveness are calculated with the CV and NSE. The comparison and analysis of TCs' attributes in terms of space-time variation characteristics such as trajectory, lifespan, intensity and PDI are discussed in section 4. Conclusions follow in section 5.

2 DATA AND METHOD

2.1 *Data*

The TCs studied in this investigation over the rectangle zone (0°–55°N, 100°E–180°) cover the WNP and the SCS region. The TC best-track datasets for the WNP are usually issued by the China Meteorological Administration (CMA), the Joint Typhoon Warning Center (JTWC), the Regional Specialized Meteorological Center (RSMC), and the Hong Kong Observatory (HKO). Amongst the four agencies, CMA provides the relatively accurate and complete datasets over the offshore and mainland of China (Ying et al.[19]). Consequently, the best-track dataset obtained from CMA over the period of 1949—2013 is used for analysis, including TC ID, data and time, category, longitude and latitude coordinates, maximum sustained wind (MSW) obtained from 2-min mean and minimum sea level pressure (MSLP).

The TC sub-center, a circulation center with warm core, usually splits from its parent TC when encountering topography (such as mountain and island) or other synoptic systems, and its information has also been recorded in the best-track dataset of CMA (Bao et al.[20]; Ying et al.[19]). As the sub-centers' information is meaningless for the classification of TCs, this part of data is deleted from the whole WNP dataset. In total, 2079 TC samples are ultimately used here, whose strength have reached the level of tropical depressions (TD) and whose lifespan have lasted for 24 hours and even longer.

2.2 *Quantification of TCs*

The genesis of a TC is defined when its strength has reached the level of TD for the first time. The dissipation of a TC is defined when the strength is under the TD level constantly. Allowing for the trajectory, time and power, nine physical indexes are suggested to describe a TC, which comprise of genesis location (latitude and longitude coordinates), dissipation location (latitude and longitude coordinates), trajectory

length, trajectory direction, mean deflection angle, lifespan, seasonality, intensity and PDI. The definition of these physical indexes is presented in Table 1. These indexes could comprehensively describe the traits of TCs. Therefore, in order to obtain their cybotaxis and space-time principles objectively, the nine indexes need to be introduced and analyzed as crucial factors. To quantify the TC applied to the clustering analysis, the z-score normalization method is used to convert the overall information into a standardized 11-dimensional vector. Furthermore, all the indexes are given an even weight to avoid the subjective deviation.

Table 1. Physical indexes.

Attributes	Physical indexes	Circumscription
Trajectory attributes	(1) Genesis location	The latitude and longitude coordinates of TC genesis location
	(2) Dissipation location	The latitude and longitude coordinates of TC dissipation location
	(3) Trajectory length	The trajectory length between genesis and dissipation locations. Unit: km
	(4) Trajectory direction	The angle between the equator and the line that connects the genesis and dissipation locations. Unit: °
	(5) Mean deflection angle	The average of the deflection angle of each point on the TC trajectory with respect to the previous point. Unit: °
Time attributes	(1) Lifespan	The length of time for which a TC exists. Unit: d
	(2) Seasonality	The month of the TC genesis, indicated by 1 to 12
Power attributes	(1) Intensity	The maximum wind speed that appears in lifespan. Unit: ms^{-1}
	(2) PDI	$PDI=\int_0^n v^3 dt$. Unit: m^3s^{-2} where n is the lifespan of a TC and v is the wind speed

2.3 *K–means clustering algorithm*

K-means clustering algorithm is widely utilized in TC data mining due to its simplicity, maturity and stabilization. Within the algorithm, the distance between the two TC samples is determined by the Euclidean distance, and "silhouette" (S_i) values serve as the indicator to determine the optimal cluster number (Nakamura et al.[13]). The definition of S_i is

$$S_i = \frac{\min(b_i) - a_i}{\max[a_i, \min(b_i)]} \quad (1)$$

where a_i is the mean distance between the TC sample i and the other intra-cluster samples, and b_i is the mean distance from the TC sample i to the other inter-cluster samples. S_i values range from −1 to 1. A TC sample with a high S_i value indicates that it is cohesive to the other samples within the cluster. A negative S_i value indicates that the TC sample is possibly misclassified.

Figure 1 shows the mean S_i value and the number of negative S_i values for variable cluster number K ranging from 3 to 10. For the total TC samples, the number K corresponding to a high mean S_i value and less negative S_i, maybe a good choice of optimal cluster number when the algorithm converges. From Fig.1, it is obvious that mean S_i reaches its peak (0.335) and the number of negative S_i values bottoms simultaneously (95) when K=5. It indicates that the optimal cluster number is 5. In order to facilitate the distinction, the calculated five clusters are labeled with A, B, C, D and E. The quantity of TCs and the mean values of the nine physical indexes for each cluster and all TCs are shown in Table 2, from which it is easy to recognize the separation among different clusters clearly.

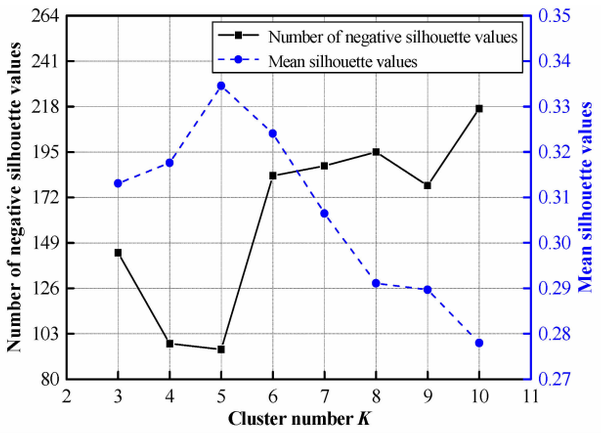

Figure 1. Mean silhouette and number of negative silhouette values.

3 DEGREE OF INTER –CLUSTER DIVERGENCE AND INTRA –CLUSTER COHESIVENESS

For the sake of the evaluation of the clustering method, the inter-cluster divergence and intra-cluster cohesiveness are introduced in this study. The degree of inter-cluster divergence is a measure of how well the clusters are separated, and the degree of intra-cluster cohesiveness is a measure of how cohesive each cluster is. The satisfactory classification should be with high degrees of inter-cluster divergence and intra-cluster cohesiveness. In this section, the CV and NSE are employed as indicators to assess the five clusters' divergence and cohesiveness. Simultaneously the

clusters classified by mass moments (CCMM) using K-means clustering method (Nakamura et al.[13]) are cited as a comparison here. The detailed result of the CCMM calculated by the modified method of Yu et al. [14] with an even weight coefficient 0.2 is exhibited in Table 3.

The quantity, mean centroid location and variance values of the CCMM are presented (see Table 3). According to Nakamura et al. [13], the TC trajectory, which is considered as a spatial open curve with the information of wind speed, can be defined by two mass moments. The first mass moment is the centroid location of a TC trajectory, i.e. the latitude and longitude coordinates (x, y). The second mass moment contains three directional variances (V_x, V_y and V_{xy}). The optimal cluster number K for the CCMM is determined by S_i values as well, and K=5. The five clusters are tagged with Ⅰ, Ⅱ, Ⅲ, Ⅳ and Ⅴ to distinguish from the clusters classified by physical indexes (CCPI).

Table 2. Quantity and mean values of physical indexes for each cluster and for all TCs.

Cluster	Quantity	Genesis		Dissipation		Seasonality	Trajectory length (km)	Trajectory direction (°)	Mean deflection angle (°)	Lifespan (d)	Intensity (ms^{-1})	PDI ($\times10^{-10}m^3s^{-2}$)
		Lat (°N)	Lon(°E)	Lat (°N)	Lon(°E)							
A	590	18.76	143.17	37.43	150.80	7.87	3381.56	104.83	14.29	5.98	32.50	1.09
B	222	16.26	123.69	19.89	119.45	8.36	1282.91	46.65	32.18	3.75	21.57	0.29
C	337	11.04	142.90	20.98	116.05	8.61	3818.50	24.24	13.28	8.74	49.09	3.64
D	597	14.08	127.66	18.56	117.25	7.75	1541.11	27.43	12.65	3.61	24.17	0.35
E	333	11.63	150.98	39.23	151.59	8.51	5609.08	90.27	12.75	10.22	61.11	7.51
All	2079	14.75	137.84	27.76	132.31	7.87	3056.58	61.00	15.32	6.19	36.21	2.23

Table 3. Mean centroid and variance values for the CCMM and for all TCs.

Cluster	Quantity	Centroid x (°E)	Centroid y (°N)	Variance x	Variance y	Variance xy
Ⅰ	349	151.34	27.92	52.10	56.69	27.78
Ⅱ	255	143.57	14.29	21.01	16.49	−3.34
Ⅲ	219	133.88	16.40	131.96	26.72	−38.89
Ⅳ	756	117.58	16.59	21.84	5.12	−5.20
Ⅴ	500	133.40	24.92	28.38	40.92	−1.94
All	2079	131.96	20.19	39.99	26.06	−2.20

3.1 *Degree of inter-cluster divergence*

CV, defined as the ratio of standard deviation to the mean, is usually used to characterize the dispersion of frequency distribution. To estimate the degree of inter-cluster divergence, the CV values of mean lifespan, trajectory length, intensity and PDI for the CCPI and CCMM are listed in Table 4. A high CV value always indicates that the calculated clusters are highly dispersive mutually.

In terms of lifespan, there is an apparent dissimilarity of the mean values between the CCPI and the CCMM. The mean values of CCPI's lifespan fall into the range of 3.61 d to 10.22 d, which is wider than the lifespan of CCMM ranging from 4.66 d to 9.81 d. What's more, the distinctive CV values indicate that, the mean lifespan of CCPI (CV=0.41) is more dispersive than that of CCMM (CV=0.27). By examining the data in Table 4, an approximate pattern occurs in trajectory length, intensity and PDI as well. That is to say, the CCPI has a higher degree of inter-cluster divergence than CCMM.

Table 4. CV values of mean lifespan, trajectory length, intensity and PDI for the CCPI and CCMM.

Physical index	Method	Cluster A/Ⅰ	Cluster B/Ⅱ	Cluster C/Ⅲ	Cluster D/Ⅳ	Cluster E/Ⅴ	CV
Lifespan (d)	CCPI	5.98	3.75	8.74	3.61	10.22	0.41
	CCMM	6.90	5.21	9.81	4.66	6.92	0.27
Trajectory length (km)	CCPI	3381.56	1282.91	3818.50	1541.11	5609.08	0.51
	CCMM	4181.99	2468.82	5050.86	1839.38	3537.71	0.34
Intensity (ms^{-1})	CCPI	32.50	21.57	49.09	24.17	61.11	0.40
	CCMM	38.77	32.20	56.82	28.20	39.57	0.25
PDI ($\times10^{-10}$ m^3s^{-2})	CCPI	1.09	0.29	3.64	0.35	7.51	1.07
	CCMM	2.42	1.79	6.51	0.81	2.62	0.69

3.2 *Intra-cluster cohesiveness*

In order to determine the intra-cluster cohesiveness of the CCPI, the NSE is adopted here. The NSE is a normalized index of relative magnitude of the simulated numerical data compared with observed field data in hydrological models (Nash and Sutcliffe [18]), which is calculated as follows:

$$\mathrm{NSE}=1-\frac{\sqrt{\sum_{i=1}^{n}(\mathrm{CV}_{Mi}-\mathrm{CV}_{Pi})^2}}{\sqrt{\sum_{i=1}^{n}(\mathrm{CV}_{Mi}-\mathrm{CV}_{M\mathrm{mean}})^2}} \quad (2)$$

where n is the number of clusters, CV_{Mi} is the CV value of TC samples within the ith cluster that is classified by mass moments, $\mathrm{CV}_{M\mathrm{mean}}$ is the mean CV_M, CV_{Pi} is the CV value of TC samples within the ith cluster that is classified by physical indexes. NSE values range from -∞ to 1. For instance, the NSE value of 1 corresponds to an excellent similarity between CV_M and CV_P. Contrarily, the value of NSE less than 0.5 means that the CV_P is quite different from CV_M in general (Moriasi et al.[21]). The NSE values for lifespan, trajectory length, intensity and PDI are calculated and listed in Table 5.

For the lifespan, the NSE is 0.21, the mean CV_P is smaller than CV_M, and the result means that the CV value of TC samples in CCPI is generally smaller than that in CCMM. That is, TCs in CCPI generally have a higher degree of intra-cluster cohesiveness than CCMM in terms of lifespan. The approximate pattern also happens in the trajectory length, intensity and PDI. It can be concluded that the intra-cluster cohesiveness of CCPI is better than that of CCMM. The comparison and analysis show that the physical index system could serve as a feasible method of TCs classification.

Table 5. CV values of CCPI and CCMM and NSE values in lifespan, trajectory length, intensity and PDI.

Physical index	CV	Cluster A/Ⅰ	Cluster B/Ⅱ	Cluster C/Ⅲ	Cluster D/Ⅳ	Cluster E/Ⅴ	Mean	NSE
Lifespan (d)	CV_P	0.42	0.58	0.28	0.47	0.25	0.40	0.21
	CV_M	0.45	0.59	0.27	0.58	0.47	0.47	
Trajectory length (km)	CV_P	0.39	0.60	0.26	0.52	0.21	0.40	0.32
	CV_M	0.42	0.66	0.25	0.58	0.45	0.47	
Intensity (ms^{-1})	CV_P	0.33	0.38	0.28	0.38	0.22	0.32	−0.06
	CV_M	0.41	0.57	0.29	0.46	0.44	0.43	
PDI ($\times10^{-10}$ $\mathrm{m^3s^{-2}}$)	CV_P	0.95	1.32	0.80	1.20	0.67	0.99	0.05
	CV_M	1.33	1.76	0.85	1.77	1.28	1.40	

3.3 *Theoretical analysis*

The approach applied in this paper had the TCs be quantified into 11-dimensional vectors, while the TCs' information in the CCMM were simplified into five dimensions. It is a common view that much helpful and essential information would be omitted or neglected when converted into low dimension. As for the TCs classification, once the cluster space was condensed from eleven dimensions into five dimensions, the TC samples' information would be distorted at the meantime, which can be demonstrated by a simplified model as follows. In Fig.2, there are six vectors with equal length and named ***a***, ***b***, ***c***, ***d***, ***e***, and ***f*** separately. If analyzed in a 2-dimensional perspective, the vectors would be divided into three clusters (i.e., ***a*** and ***f***, ***b*** and ***c***, ***d*** and ***e***) owing to their parallel orientations. However, if analyzed in single dimensional perspective alone, for example considering horizontal axis merely and neglecting vertical axis, the vectors would be divided into three clusters (i.e., ***a*** and ***c***, ***b*** and ***f***, ***d*** and ***e***). It is obvious to know that the classification results of the two methods are different and the 1-dimensional classification result is unreasonable due to incomplete information.

Although CCMM has taken the trajectory and wind speed into consideration, the method neglected the temporal and power-related attributes, which lowered the dimension of the TCs' information and could possibly lead to a distortion in classification. By the physical index system, CCPI expanded the cluster space into eleven dimensions as much as possible, allowing for the comprehensive attributes of TCs' trajectory, time and power, therefore the resultant TCs clusters behaved better intra-cluster cohesiveness and inter-cluster divergence in separate attribute.

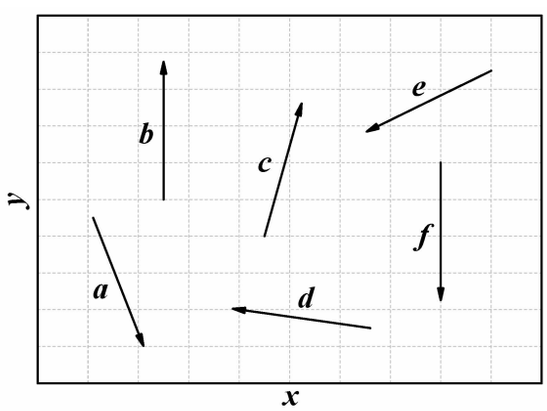

Figure 2. Vectors ***a***, ***b***, ***c***, ***d***, ***e*** and ***f***.

4 TC CLUSTERS

4.1 *Trajectory attributes*

The trajectories of TCs in CCPI and all TCs are

shown in Fig.3. Fig.4 (a) exhibits the distribution of trajectory length of CCPI and all TCs. In general, the WNP TC trajectories can be summarized into straight-moving (cluster B, C and D) and recurving types (cluster A and E).

Cluster A contains 590 TCs (see Table 2) with an average trajectory length of 3381.56 km, slightly longer than the gross average length of 3056.58 km. TCs in cluster A are active in most of concerned marine regions, nevertheless only a few TCs make landfall in the east coast of China literally. TCs in cluster B are short straight-moving trajectories with the minimum average length of 1282.91 km. The majority of active TCs in cluster B are usually restricted in the region of the SCS and impinge on the southeast coast of China. The longest straight-moving trajectories are found in cluster C with an average length of 3818.50 km. TCs in cluster C move west after genesis and mainly affect the southeast coast of China, Taiwan Island and the Philippines. The average trajectory length of cluster D is 1541.11 km, which is slightly longer than that of cluster B. TCs in cluster D follow straight trajectories across the Philippines and the SCS after genesis. The influenced scope of cluster D resembles to cluster B and cluster C. Cluster E is recurving, accounting for 16.02% of all TCs, with the longest average trajectory length of 5609.08 km. The trajectories in cluster E are characterized by a typical parabolic shape, which turns from northwest to northeast in the seas east of Taiwan Island. The activity of TCs in cluster E is similar to that of cluster A, while only a few TCs land in the coastal areas of China.

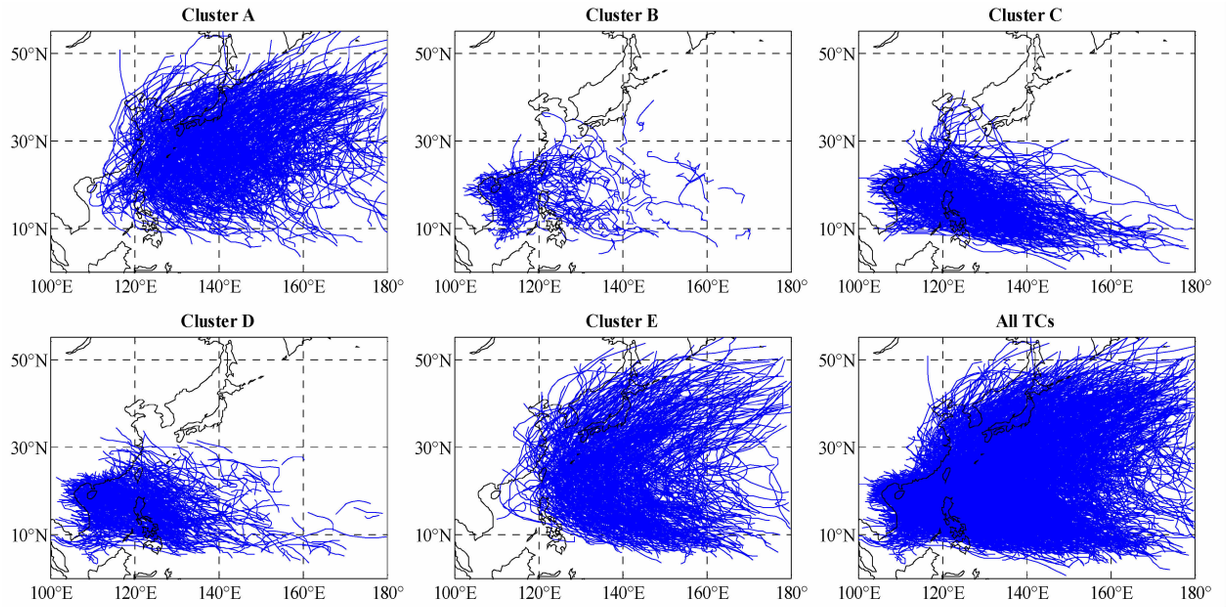

Figure 3. Trajectories for the CCPI and all TCs.

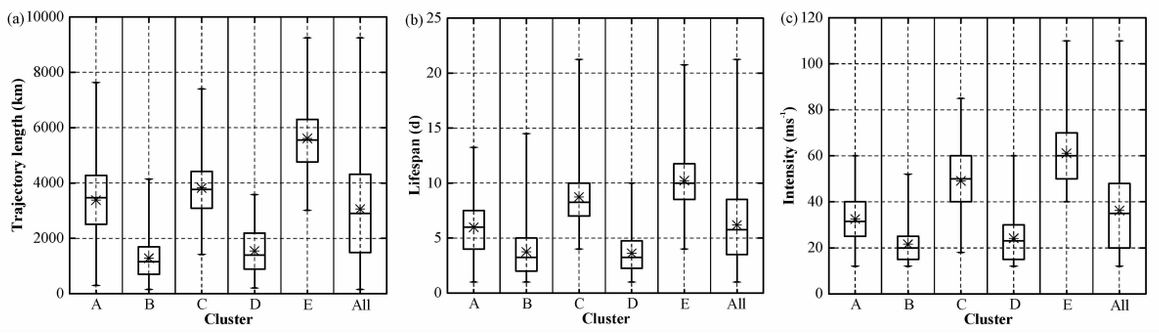

Figure 4. Distribution of trajectory length (a), lifespan (b) and intensity (c) for the CCPI and all TCs. The boxes show the 75th and 25th percentiles (upper and lower bounds of the box), the median (bar in middle), the mean (asterisk), maximum and minimum (dashed line) of the distribution.

4.2 *Time attributes*

The distribution of lifespan for the CCPI and all TCs is shown in Fig.4(b). Fig.5 illustrates the monthly average occurrence frequency of TCs. By contrasting the distribution of lifespan and seasonality for the CCPI, it is easy to know that the five clusters differ in temporal characteristics.

In general, the average lifespan of the WNP TCs is 6.19 d, and the main active period varies from June to November, especially in August. TCs in Cluster A, B and D occur most actively in August. Among the three, cluster A has the similar pattern of active period with

the overall TCs, while the active period of cluster B is prior to cluster A, and cluster D has a relatively unapparent active period. However, the lifespan of the three clusters is shorter than the average lifespan. The lifespan of cluster D (3.61 d) is the shortest among five clusters. TCs in cluster C have a longer active period (from July to November) with two peaks in August and October separately. TCs in cluster E have the latest active period with highest occurrence frequency in October. TCs in the two clusters have a longer lifespan which is obviously higher than the average level. The lifespan of cluster E (10.22 d) is the longest among the five clusters. The analysis of the five clusters' time attributes demonstrates that TCs with a late active period usually have a longer lifespan, such as the cluster E and C.

The interannual variation of TCs has been analyzed as well. The annual gross TCs number and the quadratic fitting curve for the CCPI and all TCs are shown in Fig. 6. From 1949 to 2013, there is an average of 31.98 TCs happened in the WNP every year. The occurrence frequency of the WNP TCs presents a trend of decrease

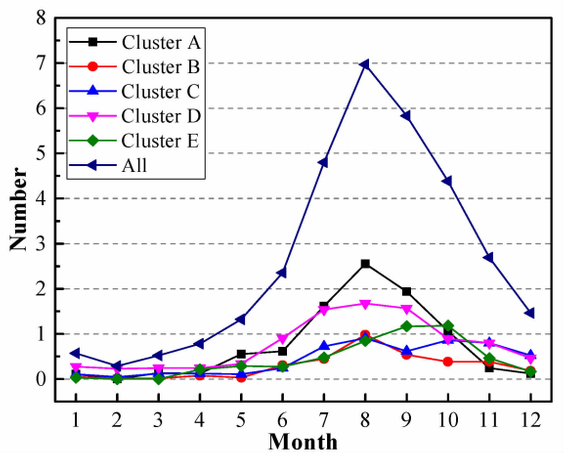

Figure 5. Monthly average number of TCs for the CCPI and all TCs.

in the last 40 years. Most of the clusters (cluster B, C, D, and E) keep a similar decreasing trend from 1970s to 2013. However, the number of TCs of cluster A has increased slightly in general.

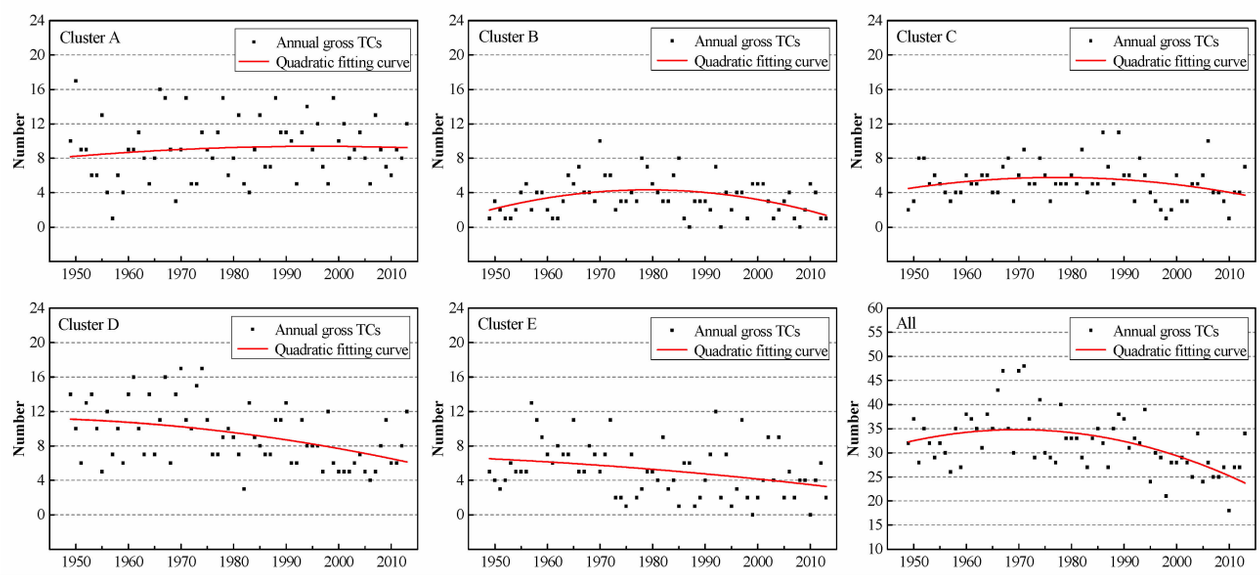

Figure 6. Annual TC number and its quadratic fitting curve for the CCPI and all TCs.

4.3 *Power attributes*

According to the Classification of Tropical Cyclones Standardization (GB/T 19201–2006) released by Standardization Administration of the People's Republic of China, TCs are divided into six stages: tropical depression (TD, 10.8 $ms^{-1} \leq MSW \leq 17.1\ ms^{-1}$), tropical storm (TS, 17.2 $ms^{-1} \leq MSW \leq 24.4\ ms^{-1}$), severe tropical storm (STS, 24.5 $ms^{-1} \leq MSW \leq 32.6\ ms^{-1}$), typhoon (TY, 32.7 $ms^{-1} \leq MSW \leq 41.4\ ms^{-1}$), severe typhoon (STY, 41.5 $ms^{-1} \leq MSW \leq 50.9\ ms^{-1}$) and super typhoon (Super TY, 51.0 $ms^{-1} \leq MSW$). The percentage of TCs with TD, TS, STS, TY, STY and Super TY strength is given in Fig.7. The distribution of intensity for the CCPI and all TCs is illustrated in Fig.4(c).

From Fig.7, it is easy to see that cluster E has selected most of the Super TYs from the total TCs and the Super TY are dominant in cluster E. Therefore, TCs in cluster E are the strongest among all clusters, with an average intensity as high as 61.11 ms^{-1}. Unlike the cluster E, the TYs and TSs account for most of TCs in cluster A, whose intensity is nearly the same to the overall level. The majority of TCs (87.84%) in cluster B usually reserve in the stage of STS, and only a small quantity of them deteriorate into TYs. Similarly, only a few TCs (19.77%) of cluster D have developed into TYs. The intensity of the two clusters is comparatively weak, both lower than 25 ms^{-1}. Contrarily, most of the TCs (89.61%) in cluster C are TYs, STYs or Super

TYs. As a result, TCs' intensity (49.09 ms^{-1}) in cluster C is quite much higher than the average intensity.

The PDI (mentioned in section 2) is used as a comprehensive index of lifespan and intensity. A higher PDI value indicates the region is relatively more severely affected by the TCs. Fig.8 exhibits PDI distribution of TCs in each cluster and TCs in total. The blue areas illustrate the gross influence sphere of TCs. And the red areas represent the most gravely impacted extents.

From the calculation results, apparently each cluster shows individual characteristics in effect on specific region. The two recurving clusters (cluster A and E) exhibit similar influence sphere, but the TCs in cluster E have a more severe impact region, which is located in the seas on the northeastern of the Philippines and the eastern of Taiwan Island. The region affected by TCs of cluster A is the eastern of Taiwan Island and southern Japan. The three straight-moving clusters (cluster B, C and D), have a slightly narrower influence sphere compared with the two recurving clusters, moreover TCs in cluster C have a strong effect on the seas between the Philippines and Guam, and the most seriously impacted extents of cluster D lie in the northern part of the SCS. In general, the recurving WNP TCs have severe influence on the seas on the eastern of Taiwan Island and the Philippines.

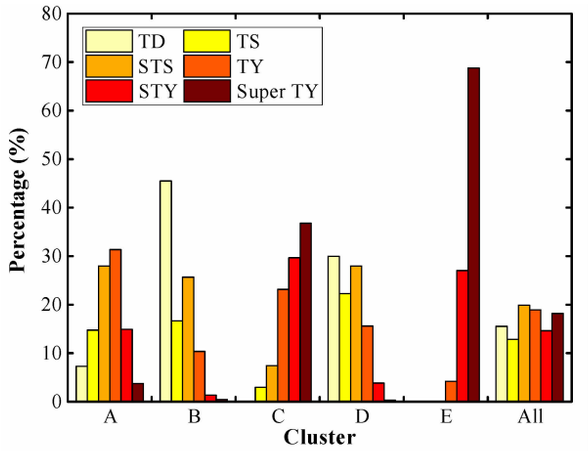

Figure 7. Percentage of TCs with different strengths for the CCPI and all TCs.

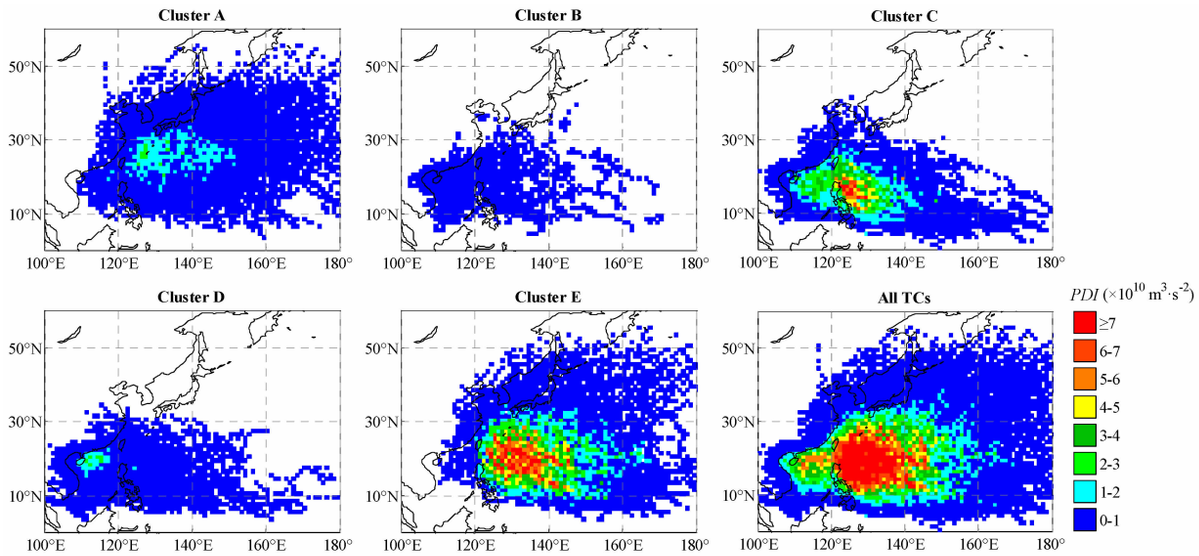

Figure 8. Distribution of the PDI for the CCPI and all TCs.

5 CONCLUSIONS AND DISCUSSION

A method for classification of TCs by using K-means clustering algorithm with the physical indexes is developed. This method emphasizes the trajectory, time and power attributes of TCs. The specific information by this method is converted into vectors composed of nine physical indexes. The TC best-track dataset of the WNP (including the SCS) during 1949—2013 obtained from the CMA is analyzed and compared in order to comprehend the space-time and power properties. The applied WNP TCs have been categorized into five clusters and the conclusions are as follows.

(1) Each TC can be quantified by using an 11-dimensional vector, including genesis location (latitude and longitude coordinates), dissipation location (latitude and longitude coordinates), trajectory length, trajectory direction, mean deflection angle, lifespan, seasonality, intensity and PDI. The indexes describe the TCs objectively and elaborately, and it proves to be an effective and comprehensive way to quantify TCs' attributes.

(2) By the comparison with the classification results of the mass moments, the high CV values and NSE values of CCPI indicate that the classification of physical index system has a higher degree of intra-cluster cohesiveness and inter-cluster divergence, and it verifies that this is a feasible method to classify TCs.

(3) Two clusters of recurvature (cluster A and E) and three clusters of straight-movement (cluster B, C and D) are distinguished. The three straight-moving clusters have roughly the same motion pattern: move west after genesis and then dissipate in southeastern China and the west of the SCS. Both clusters A and E have a relatively large scope of motion, but only a few TCs impact the eastern part of Chinese mainland.

(4) Most of the clusters have an active period in August, and the lifespan of a cluster gets longer when its active period is relatively late. From the long-term scale analysis, the TCs in the WNP show a declining trend in the last 40 years, except for cluster A.

(5) The TC composition of every cluster differs from each other, and each cluster has a distinctive impact region, that is represented by the PDI distribution. Amongst the five clusters, TCs in cluster E is the strongest, and has caused severe influence on the seas of the northeastern of the Philippines and waters on the eastern of Taiwan Island.

This approach has a potential advantage of compatibility, which means that the physical index system is extensible. The heavy rainfall caused by TCs, as a significant inducing factor of TC disasters, often leads to urban flood and waterlogging and brings about enormous economic loss and casualty. Supposing that the TC rainfall temporal-spatial process could be quantified and imported into the physical index system, the prevention and estimation of TC disasters would be much more reliable and exhaustive.

Further work will explore the landfall probability distributions and possible landing intensities of different clusters. The El Niño-Southern Oscillation (ENSO) event will lead to global climate anomalies that make TC activity abnormal. In recent years, many scholars have studied the impact of different types of ENSO events on TCs activities in the Pacific. However, there are few studies on ENSO events based on TC classification. As TCs properties are based on the meteorological conditions, the relationship between ENSO and each cluster of TCs will be investigated in the further work.

REFERENCES:

[1] WANG Xiao-ling, WANG Yong-mei, REN Fu-min, et al. Interdecadal variations in frequencies of typhoon affecting China during 1951-2004 [J]. Adv Clim Change Res, 2006, 3(3): 135-138 (in Chinese).

[2] XIAO Feng-jin, XIAO Zi-niu. Characteristics of tropical cyclones in China and their impacts analysis [J]. Nat Hazards, 2010, 54(3): 827-837.

[3] ZHANG L J, ZHU H Y, SUN X J. China's tropical cyclone disaster risk source analysis based on the gray density clustering [J]. Nat Hazards, 2014, 71 (2): 1053-1065.

[4] DESSAI S. The special climate change fund: origins and prioritisation assessment [J]. Clim Policy, 2003, 3 (3): 295-302.

[5] WU Li-guang, WANG Bin, GENG Shu-qin. Growing typhoon influence on East Asia [J]. Geophys Res Lett, 2005, 32(18): 109-127.

[6] LANDER M A. Specific tropical cyclone track types and unusual tropical cyclone motions associated with a reverse-oriented monsoon trough in the western North Pacific [J]. Wea Forecasting, 1996, 11(2): 170-186.

[7] HORVATH K, LIN Y L, IVANČAN-PICEK B. Classification of cyclone tracks over the Apennines and the Adriatic Sea [J]. Mon Wea Rev, 2008, 136(6): 2210-2227.

[8] HODANISH S, GRAY W M. An observational analysis of tropical cyclone recurvature [J]. Mon Wea Rev, 1991, 121 (10): 2665-2689.

[9] CAMARGO S J, ROBERTSON A W, GAFFNEY S J, et al. Cluster analysis of typhoon tracks. Part I: general properties [J]. J Climate, 2007, 20(14): 3635-3653.

[10] KIM H S, KIM J H, HO C H, et al. Pattern classification of typhoon tracks using the fuzzy c-means clustering method [J]. J Climate, 2011, 24(2): 488-508.

[11] ZHANG W, LEUNG Y, CHAN J C L. The analysis of tropical cyclone tracks in the western North Pacific through data mining. Part I: tropical cyclone recurvature [J]. J Appl Meteor Clim, 2013, 52(6): 1394-1416.

[12] ELSNER J B, LIU K B. Examining the ENSO-typhoon hypothesis [J]. Clim Res, 2003, 25(1): 43-54.

[13] NAKAMURA J, LALL U, KUSHNIR Y, et al. Classifying North Atlantic tropical cyclone tracks by mass moments [J]. J Climate, 2009, 22(20): 5481-5494.

[14] YU Jin-hua, ZHENG Ying-qing, WU Qi-shu, et al. K-means clustering for classification of the northwestern Pacific tropical cyclone tracks [J]. J Trop Meteor, 2016, 22(2): 127-135.

[15] ELSNER J B. Tracking hurricanes [J]. Bull Amer Meteor Soc, 2003, 84(3): 353-356.

[16] LIN Zhi-qiang, BIANBA Zha-xi, WEN Sheng-jun, et al. The objective clustering of tropical storms tracks over the Bay of Bengal [J]. J Trop Meteor, 2013, 29(6): 973-983 (in Chinese).

[17] EMANUEL K. Increasing destructiveness of tropical cyclones over the past 30 years [J]. Nature, 2005, 436 (7051): 686-688.

[18] NASH J E, SUTCLIFFE J V. River flow forecasting through conceptual models part I - A discussion of principles [J]. J Hydrol, 1970, 10(3): 282-290.

[19] YING Ming, ZHANG Wei, YU Hui, et al. An overview of the China Meteorological Administration tropical cyclone database [J]. J Atmos Oceanic Technol, 2014, 31 (2): 287-301.

[20] BAO Xu-wei, LEI Xiao-tu, TANG Bi, et al. An analysis on the features of secondary center produced by the tropical cyclones crossing Taiwan [J]. J Trop Meteor, 2013, 29(5): 717-726 (in Chinese).

[21] MORIASI D N, ARNOLD J G, VAN LIEW M W, et al. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations [J]. Trans ASABE, 2007, 50(3): 885-900.

Citation: YAN Dong-yi, XU Kui, MA Chao, et al. Classifying western north Pacific tropical cyclones by physical index system [J]. J Trop Meteor, 2018, 24(2): 142-150.