Geometric Clustering Analysis of Typhoon Track and Its Impact on Northwest Pacific Countries

Yuei-An Liou and Truong-Vinh Le

February 22, 2022

Geometric clustering analysis of typhoon track and its impact on Northwest Pacific countries

Yuei-An Liou [1] and Truong-Vinh Le [1,2]
[1] Center for Space and Remote Sensing Research, National Central University, Taoyuan, Taiwan, 320317
[2] Taiwan International Graduate Program – Earth System Science Program, Academia Sinica, Taipei, Taiwan, 11529
Email: yueian@csrsr.ncu.edu.tw (Corresponding author)

***Abstract*+—Tropical cyclones (TCs) are among the most dangerous meteorological phenomena with the power to cause catastrophic damages to human lives, societies, and properties. Their activities and occurrences have been considerably altered as a consequence of climate change. This study investigates the impact of TC tracks on the Northwest Pacific (NWP) nations by using Unsupervised Machine Learning (UML) K-mean clustering. The results indicated that the optimal number for clustering K-mean analysis of TCs is three. In addition, the risk of each NWP nation to the clustered TCs was investigated. It is found that most countries are vulnerable to cluster no. 2 TCs, whereas China and Vietnam are highly prone to cluster no. 3 events. Also, the geometric clustering analysis is a potentially useful technique to redefine the forecast trajectories and interpret their influence on the NWP countries.**

***Keywords*—machine learning, clustering, tropical cyclone, NWP**

I. Introduction

Tropical cyclones (TCs) have been always a major concern of meteorologists and warning centers all over the world. The rising power of typhoons as a possible consequence of climate change raises the awareness of global researchers on the tropical storms [1]-[3]. Many researchers emphasized the origin, frequency, intensity, and lifetime of TCs in the ocean basins [4], [5]. In the recent studies, clustering methods were applied to categorize the TCs based on their features, such as form and length. For example, Chand and Walsh [6] and Bell et al. [7] used the curve-clustering approach to classify TC tracks. Camargo et al. [8], [9] developed a special probabilistic clustering algorithm relying on a regression mixture model to cluster the TC tracks. Since K-means is the most basic and widely used clustering method for dividing a dataset into a collection of "k" groups [10], [11] so that this study investigates the use of Unsupervised Machine Learning (UML) K-means clustering method to categorize the TC tracks into subgroups by their characteristics and evaluates their impacts on the Northwest Pacific (NWP). The input data consists of TC number, length, lifetime, wind speed, and central pressure. In addition, the impact of TC clustered tracks is evaluated.

II. Data and Methods

A. TC best-track dataset

The original TC best-track data of Regional Specialized Meteorological Centers (RSMC) – Tokyo Typhoon Center were collected from the International Best Track Archive for Climate Stewardship Version 4 (IBTrACS V4). The TC information includes its center location (longitude and latitude), classification, maximum wind speed (kt), and minimum central pressure (hPa) with 6-hour interval [12]. This work only explores the mature typhoon phases (i.e., when the maximum wind speed reaches 35 kt) and excludes the milder depression phases. All typhoon tracks crossing the NWP are represented by the grey color lines (Figure 1).

B. Methods

The TCs IBTrACS V4 dataset was transformed into a readable format. Then, the number of TCs that passed across each national border was automatically counted (see orange lines in Figure 1). Each TC track length was then calculated by using the Vincenty's geodesic distance [13]. In addition, the TC lifetime that affects the NWP was computed by using the basic length and velocity concept.

To cluster the typhoon data, we follow four major steps: (1) preprocess data (normalization, scaling, and feature data transformation), (2) create comparison metric, (3) execute the K-means clustering; and (4) interpret the outcomes and modify the clustering. The basic idea behind the K-means clustering is to move each data point to its closest center based on a predefined starting number of clusters, upgrade the clustering center by calculating the viciousness of the target part, and repeat the relocating-and-updating procedure until the focalized criteria (such as a predefined number of iterations or the difference within the confidence of the mutilation work) are fulfilled [11]. Three criteria were used to determine the most acceptable number of clusters (or the optimal k number): average Silhouette width, gap statistic, and Elbow method.

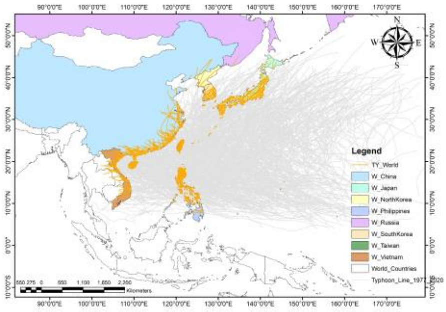

Figure 1. TC tracks mapping (grey lines) in the NWP and their landfall to islands and nations (orange lines) with country boundaries highlighted.

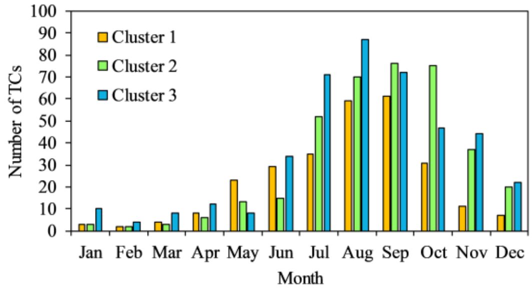

Figure 3. Seasonal variation of cluster number of TCs in the NWP basin.

III. Results And Discussions

In this section, the results show the cluster analysis, monthly-wise TC number variations and vulnerability to each country and island in the NWP.

A. TCs clustering for the NWP and seasonality variation

The TCs are divided into three groups as the optimal number of clusters retrieved from the K-mean method. The distribution of clustered TCs is presented in Figure 2. The cluster study revealed that different TC group originates and moves in different parts of the NWP basin. The most popular clusters no. 2 and no. 3, which account for around 75% of all NWP TCs, are located from the north to south of the NWP, respectively. Cluster no. 1 TCs (representing 25%) are mostly located in the northern region.

In the NWP, the typhoon season varies with location. However, most of the typhoons are active from May to October [5]. In the winter season months (e.g., October), cluster no. 1 has lower TC count, whereas cluster no. 2 and no. 3, especially cluster no. 2, have a higher value in (Figure 3). TCs of the 1st and 2nd clusters appear the highest in September, while the 3rd cluster TCs in August.

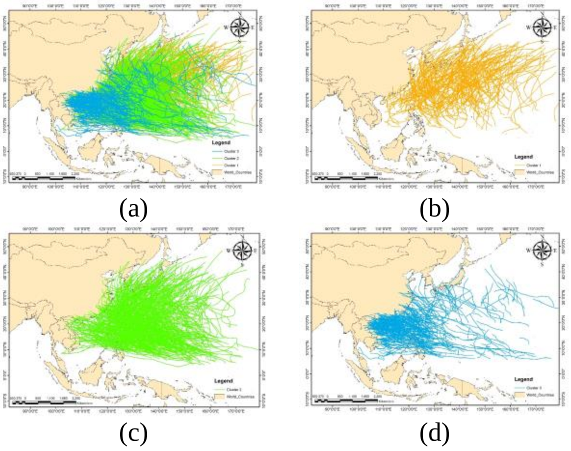

Figure 2. Distribution of the clustered tracks according to three groups over the past 44 years in the NWP ocean: (a) all clusters, (b) cluster 1, (c) cluster 2, (d) cluster 3.

Cluster no. 2 TCs, which have stronger intensities and sustain longer, often occur from September to October. The previous research found slightly different outcomes from ours, likely related to the TCs in this cluster. These findings are summarized as follows: Winter super-typhoons have a wider genesis angle and turning point angle than the corresponding variables for the summer super-typhoons [14] and they have more energy than summertime super-typhoons [4].

B. Impact of clustered tracks on island and nations in NWP

China is the most vulnerable country to TCs over the past 44 years with 246 TCs, followed by the Philippines and Japan with 182 and 160, respectively. Figure 4 shows the quantity of TCs from every cluster on each territory. The total number of TCs from north to south of the NWP varies significantly among different countries.

Most of the countries are prone to the threat of cluster no. 2 events, which have characteristics of longer lifetime, extensive tracking, and higher intensity, including Russia, North Korea, South Korea, Japan, Taiwan, and the Philippines. Cluster no. 3 contributes the highest number of TCs to hit China and Vietnam. The TCs in cluster no. 3 have a shorter pathway and smaller intensity in ocean before landing (Figure 4) in comparison with cluster no. 2 TCs. Even so, we still observed extreme events making landfall over Vietnam and China.

The above mentioned findings show that geometric clustering analysis is an essential technique to redefine the TCs projected trajectories and evaluate their influence on the NWP nations.

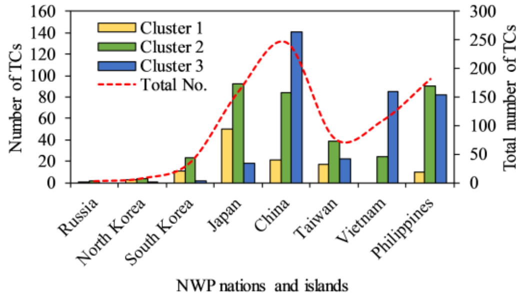

Figure 4. Cluster number of TCs in the NWP basin.

IV. Conclusions

This study applied the UML K-means clustering to classify the TCs in the NWP into three groups. Clusters no. 2 and no. 3 are primarily located in the north and south of the NWP, respectively; cluster no. 1 TCs (representing 25.5%) are mostly located in the northern region. The seasonality of clustered TCs is different for each group. Cluster no. 1 has a lower TCs count in the winter season like October, whereas clusters no. 2 and 3 have greater numbers in the summertime, especially cluster no. 2.

The total number of typhoons is the highest in China over the past 40 years, followed by the Philippines and Japan. Most of the countries are at risk during cluster no. 2 events. China and Vietnam are predominantly vulnerable to Cluster no. 3 TCs.

Acknowledgment

The authors appreciate the financial support from Taiwan's Ministry of Science and Technology (MOST) under project code 108-2923-M-008-002-MY3.

References

[1] Y.-A. Liou, J.-C. Liu, M.-X. Wu, Y.-J. Lee, C.-H. Cheng, C.-P. Kuei, and R.-M. Hong, "Generalized Empirical Formulas of Threshold Distance to Characterize Cyclone–Cyclone Interactions," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 6, pp. 3502–3512, 2016.

[2] Y.-A. Liou, J.-C. Liu, C. P. Liu, and C.-C. Liu, "Season-Dependent Distributions and Profiles of Seven Super-Typhoons (2014) in the Northwestern Pacific Ocean From Satellite Cloud Images," IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 5, pp. 2949–2957, 2018.

[3] R. S. Pandey and Y.-A. Liou, "Decadal behaviors of tropical storm tracks in the North West Pacific Ocean," Atmospheric Research, vol. 246, p. 105143, 2020.

[4] P. J. Webster, "Changes in Tropical Cyclone Number, Duration, and Intensity in a Warming Environment," Science, vol. 309, no. 5742, pp. 1844–1846, 2005.

[5] K. Emanuel, "100 Years of Progress in Tropical Cyclone Research," Meteorological Monographs, vol. 59, 2018.

[6] S. S. Chand and K. J. E. Walsh, "The Influence of the Madden–Julian Oscillation on Tropical Cyclone Activity in the Fiji Region," Journal of Climate, vol. 23, no. 4, pp. 868–886, 2010.

[7] S. S. Bell, S. S. Chand, K. J. Tory, A. J. Dowdy, C. Turville, and H. Ye, "Projections of southern hemisphere tropical cyclone track density using CMIP5 models," Climate Dynamics, vol. 52, no. 9-10, pp. 6065–6079, 2018.

[8] S. J. Camargo, A. W. Robertson, S. J. Gaffney, P. Smyth, and M. Ghil, "Cluster Analysis of Typhoon Tracks. Part I: General Properties," Journal of Climate, vol. 20, no. 14, pp. 3635–3653, 2007.

[9] S. J. Camargo, A. W. Robertson, S. J. Gaffney, P. Smyth, and M. Ghil, "Cluster Analysis of Typhoon Tracks. Part II: Large-Scale Circulation and ENSO," Journal of Climate, vol. 20, no. 14, pp. 3654–3676, 2007.

[10] S. Lloyd, "Least squares quantization in PCM," IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129–137, 1982.

[11] C. Sammut and G. I. Webb, Encyclopedia of Machine Learning, New York, NY, USA:Springer, 2011.

[12] K. R. Knapp, M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, "The International Best Track Archive for Climate Stewardship (IBTrACS)," Bulletin of the American Meteorological Society, vol. 91, no. 3, pp. 363–376, 2010.

[13] T. Vincenty, "Direct And Inverse Solutions Of Geodesics On The Ellipsoid With Application Of Nested Equations," Survey Review, vol. 23, no. 176, pp. 88–93, 1975.

[14] R. S. Pandey, Y.-A. Liou, and J.-C. Liu, "Season-dependent variability and influential environmental factors of super-typhoons in the Northwest Pacific basin during 2013–2017," Weather and Climate Extremes, vol. 31, p. 100307, 2021.