Weather and Climate Extremes 50 (2025) 100804

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Discernability of the vertical vortex structure of pre-existing disturbances and their implication for tropical cyclone formation

Hung Ming Cheung [a,b], Jung-Eun Chu [a,*]

[a] School of Energy and Environment, City University of Hong Kong, Kowloon, Hong Kong, China
[b] Department of Climate and Energy Systems Engineering, Ewha Womans University, Seoul, South Korea

ARTICLE INFO

Keywords:
Tropical cyclone
Pre-existing disturbance
Tropical cyclogenesis
Vortex structure
Machine learning

ABSTRACT

The formation of a tropical cyclone (TC) is often rooted in a pre-existing disturbance, yet our understanding of their structural differences and evolution into TCs remains limited. To bridge the knowledge gap, we examine tropical disturbances and depressions in the western North Pacific during the period 2004–2021 from a best-track dataset. Here we show four discernible structures of pre-existing disturbances in terms of their vertical and radial extents: broad vortex dominated by lower-tropospheric vorticity (Cluster 1), narrow vortex with its vorticity maximum in the lower troposphere (Cluster 2), broad and deep vortex spanning most of the troposphere (Cluster 3), and narrow vortex dominated by upper-tropospheric vorticity (Cluster 4), by applying unsupervised machine learning techniques. Out of the 2014 samples analyzed, almost 80 % exhibit vorticity maximum in the lower troposphere, while the others peak aloft. While these different structures have varying implications for stratiform and convective precipitations, there is no clear preference for specific vortex structures in pre-existing disturbances for TC genesis in the next 6 h. On the other hand, the time it takes for TC genesis or the intensification rate is more closely related to the upper-level extent of relative vorticity rather than the local maximum magnitude or radial size of the vortices. Despite the uncertainty concerning the data during the earlier lifetime, the study introduces a systematic approach to categorizing the vortex structures of pre-existing disturbances which provides new insights into their role in TC formation.

1. Introduction

Tropical cyclone (TC) genesis often involves the intensification of pre-existing disturbances (disturbances that precede tropical cyclone formation) and the subsequent formation of mature TCs (Fig. 1). Despite their seemingly low-impact nature, these early-stage phenomena can evolve into powerful storms under suitable environmental conditions (Gray, 1998), causing widespread destruction and loss. In earlier studies, the precursors of TCs are alternatively referred to as TC seeds (Hsieh et al., 2020; Ikehata and Satoh, 2021; Sugi et al., 2020), tropical cloud clusters (Hennon et al., 2013; Kerns and Chen, 2013; Teng et al., 2019), tropical disturbances (Gray, 1968, 1998; Gu et al., 2022; Zawislak, 2020), or tropical depressions (Murthy and Boos, 2019). Identification of these pre-existing disturbances poses a challenge due to the absence of a universally accepted scientific definition. Typically, their identification involves examining lower-tropospheric relative vorticity on gridded atmospheric reanalysis or numerical model output (Gu et al., 2022; Hsieh et al., 2020; Ikehata and Satoh, 2021; Kerns and Chen, 2013; Sugi et al., 2020; Zarzycki, 2022). Additionally, atmospheric convection (Hennon et al., 2013; Matsuoka et al., 2018) or sea-level pressure (Yamada et al., 2021) have been used to characterize these disturbances. In the western North Pacific, pre-existing disturbances can be regarded as storms with 1-min maximum sustained wind speed lower than 34 knots according to the operational practice of the Joint Typhoon Warning Center (JTWC).

Advancing the understanding of pre-existing disturbances has profound scientific implications for TC genesis and frequency, as it is commonly recognized that TCs originate from disturbances in the tropics. Many research efforts have attempted to understand how large-scale atmospheric and oceanic conditions influence the TC genesis (Fu et al., 2012; Kerns and Zipser, 2009; Kerns and Chen, 2013; McBride and Zehr, 1981; Ritchie and Holland, 1999; Teng et al., 2014, 2019). In recent years, considerable work has been dedicated to investigating the modulation of TC frequency by TC seeds, which can impact TC frequency through changes in seed frequency and survival rate (the proportion of seeds that develop into TCs) (Hsieh et al., 2020; Ikehata and

* Corresponding author.
E-mail address: jungeun.chu@cityu.edu.hk (J.-E. Chu).

https://doi.org/10.1016/j.wace.2025.100804
Received 3 October 2024; Received in revised form 16 July 2025; Accepted 4 September 2025
Available online 5 September 2025
2212-0947/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

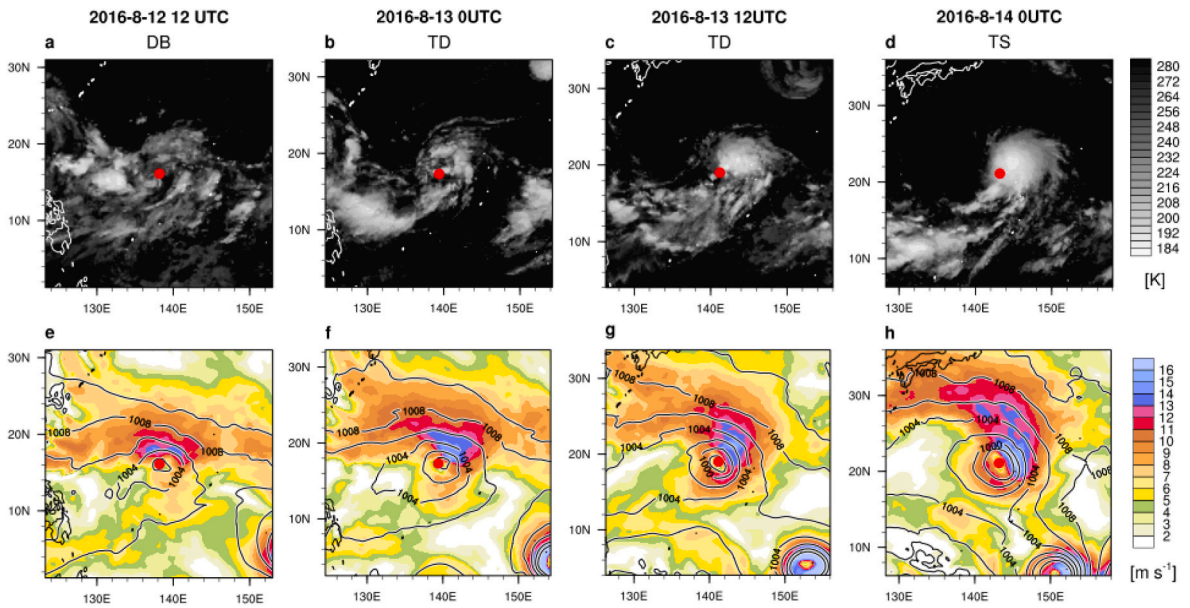


Fig. 1. An example of the evolution of a pre-existing disturbance associated with Severe Tropical Storm Chanthu (2016). (a–d) Brightness temperature of the infrared window channel, (e–h) mean sea-level pressure (contour line), and 10-m wind speed (shading) of a pre-existing disturbance during its tropical disturbance stage (DB), tropical depression stage (TD), and tropical storm stage (TS). The red dots indicate the storm center.

Satoh, 2021; Sobel et al., 2021; Sugi et al., 2020). Several modeling studies have provided evidence that, under warming conditions, seed frequency has a greater impact on TC frequency compared to the survival rate (Hsieh et al., 2020; Vecchi et al., 2019; Yamada et al., 2021). On the contrary, another study that utilized a reanalysis dataset found that both the interannual variation and spatial distribution of TC frequency are more strongly correlated with the survival rate of TC seeds than seed frequency (Ikehata and Satoh, 2021). These studies also demonstrated that large-scale environmental variables play a role in shaping both seed frequency and survival rate. Other studies have sought to comprehend how pre-existing disturbances can evolve into TCs (Gray, 1998; Zehr, 1992), the climatology of these disturbances (Hennon et al., 2013; Kerns et al., 2008), and the distinctions between developing and non-developing disturbances (Chang et al., 2017; Fu et al., 2012; Kerns and Chen, 2013). Despite the significant research efforts on pre-existing disturbances, our understanding of their structural differences and evolution to TCs remains limited (Murthy and Boos, 2019).

Several observational approaches have been made to overcome the limitations, mostly using field experiments (Bister and Emanuel, 1997; McBride, 1981; Lussier III et al., 2014). Such observations lead to a discussion on the pathway of TC genesis, focusing on the vertical evolution of the vortices. Two contrasting viewpoints exist regarding the development of TCs: the top-down theory and the bottom-up theory. The top-down theory highlights the importance of a mesoscale convective vortex or mid-level vortex. This theory suggests that TC formation requires the merging of two mid-level vortices originating from the stratiform region of mesoscale convective system (Ritchie and Holland, 1997), or the generation of a mid-level vortex through diabatic heating and evaporative cooling from stratiform precipitation (Bister and Emanuel, 1997). In both cases, there is a downward growth of cyclonic vorticity, which facilitates the spin-up of the vortex on the surface. On the other hand, a bottom-up theory suggests that the presence of vortical hot towers, which are small-scale deep cumulonimbus convection (Hendricks et al., 2004; Montgomery et al., 2006), and their mergers can generate vorticity near the surface. While the role of the mid-level vortex in TC genesis has been supported by some studies (Bister and Emanuel, 1997; Ritchie and Holland, 1997; Simpson et al., 1997), a few studies have raised doubts on this viewpoint (Kilroy et al., 2017, 2018; Wu and Fang, 2019). Therefore, further investigation is needed to determine what are the characteristics of the pre-existing vortex structure and how they influence TC genesis using comprehensive datasets.

Here, we investigate the vertical structures of the pre-existing disturbances in one of the most active convective regions, the western North Pacific, by using the two stages prior to a tropical storm: tropical disturbance and tropical depression recorded by JTWC and the Hurricane Databases (HURDAT; for the TCs formed in the Eastern Pacific and later entered the western North Pacific) archived in the International Best Track Archive for Climate Stewardship (IBTrACS) best-track dataset (Knapp et al., 2010). These systems are characterized by intensities of less than 34 knots. We consider tropical disturbances and tropical depressions as pre-existing disturbances because they are weaker storms that eventually complete genesis (when intensity category becomes tropical storm). Relative vorticity is used for characterizing pre-existing disturbances, which has also been used in other studies for similar purposes (Gu et al., 2022; Ikehata and Satoh, 2021; Kerns and Chen, 2013; Zawislak and Zipser, 2010). The study includes the pre-existing disturbances that eventually develop into TCs. We first classify azimuthally averaged relative vorticity at each 6-h time step using unsupervised machine learning technique. We then discern the different vortex structures and their associated thermodynamic characteristics, specifically focusing on their relevance to tropical cyclogenesis. The findings of this study contribute to a deeper understanding of pre-existing disturbances and their role in the development of TCs.

2. Methods

2.1. Historical TC and gridded datasets

We used the best track of IBTrACS of version v04r00. This dataset contains observations from the forecast agencies in their respective areas of responsibility at a six-hourly interval. We selected the reporting agency from the United States for the TCs in the region [0°–30°N, 110°–190°E], which are the JTWC and the HURDAT. The study domain mainly covers the western North Pacific, so most of the TCs are recorded by JTWC, and some of them are formed in the eastern North Pacific. Categories including tropical disturbances and depressions (which have labels DB and TD in the variable USA_STATUS) are regarded as pre-

existing disturbances in the present study. A TC is designated as tropical depression when its intensity is less than 34 knots. On the other hand, JTWC initiates warning on tropical disturbance in the western North Pacific when any of the four criteria is met: (1) the estimated maximum sustained wind speeds within a closed tropical circulation reach or surpass 25 knots, or (2) are expected to increase to at least 35 knots within 48 h; (3) a TC could pose a threat to life and/or property within the next three days; and (4) they are instructed to initiate a warning (Francis and Strahl, 2021). It implies that tropical disturbances are not necessarily always defined based on objective criteria. Only the samples during the tropical cyclone season (June–October) in the period 2004–2021, not interpolated, developed to at least tropical storms are included. Since the label DB in the data file is not available before the year 2004, the study period starts from 2004.

To represent the vortex structure and the associated characteristics of the pre-existing disturbances, we used the parameters (relative vorticity, potential vorticity, vertical velocity, air temperature, 10-m wind speed, and mean sea-level pressure) from the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5) (Hersbach et al., 2020). This is a gridded dataset with a horizontal resolution of 0.25° × 0.25° on a rectilinear grid. The data at 00, 06, 12, 18 UTC at each 50 hPa from 200 to 1 000 hPa (except for 10-m wind speed and mean sea-level pressure) is used for this study. For precipitation, we used Climate Prediction Center morphing method (CMORPH) dataset (Joyce et al., 2004). CMORPH estimates precipitation by combining satellite observations from the microwave and infrared channel. The spatial and temporal resolution of CMORPH is 0.07277° and 30 min, respectively. In addition, the brightness temperature of the infrared window channel from Gridded Satellite (GridSat-B1) data (Knapp et al., 2011), which is derived from International Satellite Cloud Climatology Project (ISCCP) B1 dataset, is utilized. The GridSat-B1 dataset has global coverage (from 70°N to 70°S) and is available at a horizontal resolution of 0.07° × 0.07° for every 3 h. Data from multiple geostationary satellites is included for deriving a homogeneous dataset.

2.2. Regridding of reanalysis data

For the convenience of computation, parameters from reanalysis dataset for each pre-existing disturbance are gridded to a curvilinear grid. The grid's center aligns with the storm center location obtained from the best-track dataset. We performed the regridding by bilinear interpolation using NCAR Command Language (NCL; version 6.6.2), developed by the National Center for Atmospheric Research. The curvilinear grid has a horizontal resolution of 0.25° in the radial direction and 1° in the azimuthal direction.

2.3. Clustering algorithms

Self-organizing map (SOM) and *k*-means clustering are applied in sequence for classifying vortex structures. Several studies applied this hybrid technique for their classification task (Brentan et al., 2018; Li et al., 2009; Qian et al., 2019; Van Laerhoven, 2001; Yang et al., 2024). SOM (Kohonen, 1990) is an unsupervised learning method for reducing the dimensionality of high-dimensional data to a low-dimensional space. In other words, the algorithm groups similar data samples in a large dataset together, a process which is analogous to clustering. This algorithm can be regarded as a type of artificial neural network that only has an input and an output layer. The output layer is a two-dimensional map that contains a predefined number of nodes, and each node contains the weight vector that is a representative of a specific group of data samples. A weight vector $W(t)$ has the same dimension as an input data sample X. To train a SOM, the weight vector of each node is initialized randomly, and is updated based on formula $W(t+1) = W(t) + \varepsilon(t)\, h(t)\, [X–W(t)]$, where t is iteration, and $\varepsilon(t)$ and $h(t)$ are the learning rate and neighborhood function, respectively, at iteration t. Both $\varepsilon(t)$ and $h(t)$ decay with t. The node with the weight vector closest to an input sample in terms of Euclidean distance, defined as the square root of the sum of squared differences between corresponding elements, is assigned to be the best matching unit of the input sample, and the weight vector of this node and the neighboring nodes are adjusted to be closer to the input sample. The amount of adjustment depends on the learning rate and the distance between the node and the best matching unit on an SOM. A greater adjustment can be found in the earlier epochs and nodes closer to the best matching unit. At the end of training, we can obtain a map of patterns which denotes classification of input samples, and is organized in a way such that closer nodes show higher similarity. Each input sample is assigned a coordinate on the SOM that represents the location of the node on the map. In typical clustering methods (e.g., *k*-means clustering), a data sample only exerts influence on the group it belongs to, so the groups are independent of each other. This is the point at which SOM is different from other clustering methods. SOM has been applied in various topics in atmospheric sciences, such as synoptic climatology (Hewitson and Crane, 2002; Sheridan and Lee, 2011) and the evolution of monsoon (Chu et al., 2012).

Similarly, the objective of *k*-means clustering, which is also an unsupervised learning algorithm, is to partition data samples into groups so that samples within each group are more similar to each other and distinct from the samples in other groups. A label for each input sample is created such that they are assigned to each of the pre-defined number of groups. The first step of clustering is to initialize a centroid for each of the k clusters. Each input sample is assigned to the cluster which has a centroid closest to it. The cluster centroid is updated by computing the average coordinate of the samples in the corresponding cluster. The above steps are repeated until a specified number of iterations is completed or the error is below a tolerance value.

Before the vortex structure of the pre-existing disturbances are classified, the relative vorticity fields associated with each of them were first interpolated to a storm-centered curvilinear grid and averaged azimuthally. Each sample is then a 2-D array with 17 grid points in the vertical direction (from 200 to 1 000 hPa with an increment of 50 hPa) and 11 grid points in the radial direction (from 0° to 2.5° from the storm center with an increment of 0.25°). We set the outermost radius of samples at 2.5° because the standard deviation of relative vorticity is negligible at greater distance (Fig. S1). To ensure the clustering algorithms classify data samples based on vortex structure rather than magnitude in relative vorticity, we performed data normalization separately for each sample: instead of dividing the value of each grid point by the range (i.e., maximum minus minimum) of the entire dataset, we divided it by the range of the corresponding sample. Therefore, the maximum and minimum values of each sample are always 1 and 0, respectively. Since relative vorticity is a signed quantity, we used absolute maximum and absolute minimum in the denominator for the computation of normalization.

For the SOM in the present study, the optimal number of nodes is determined such that a greater number of vortex structures can be discovered and the number of samples of each vortex structure is not too small. For simplicity, we choose rectangular topography with the same length on both axes (i.e., $N \times N$ nodes). The number of iterations was set to 2014, which is the number of samples. The neighborhood function is a Gaussian function ($h(t) = \exp[-||r_i - r_{BMU}||^2/(2\sigma(t)^2)]$) with its initial sigma value equal N divided by 3, which was chosen heuristically. Here, r_i represents the position of neuron i in the map grid, and r_{BMU} represents the position of the best matching unit. The learning rate $\varepsilon(t)$ is 0.08 and is sufficiently small to allow convergence. Both the sigma value $\sigma(t)$ and learning rate $\varepsilon(t)$ decay with an asymptotic function. We justify the choice of a topography with 5 × 5 nodes using a Sammon map (Sammon, 1969). Sammon map can be used for showing the relative distance between nodes in a 2-dimensional space. It simplifies the visualization of the relationships between the weight vectors. Unstable learning can also be detected if a Sammon map folds over on itself. From Fig. S2, we can see that the Sammon map for the SOM is relatively flat and the data points, which represent the SOM nodes, are somewhat evenly

Table 1
The number of tropical disturbances, tropical depression, and the total number in each cluster. The percent in the last column denotes the percentage of samples in each cluster.

	Tropical disturbance	Tropical depression	Total
C1	231	123	354 (17.6 %)
C2	310	292	602 (29.9 %)
C3	274	365	639 (31.7 %)
C4	207	212	419 (20.8 %)
Total	1 022	992	2014

distributed. The learning process is also stable as no folding is observed. For *k*-means clustering applied to the 25 weight vectors resulted from the SOM algorithm, we test the outcomes for 2 to 7 clusters. The optimal number of clusters was determined by the elbow method (Alpaydin, 2020) using inertia (sum of squared distance), and we decided to select four clusters as the reduction in inertia with increased number of clusters drops when it is greater than four. The maximum number of iterations, tolerance, and random state are 600, 10^{-5}, and 0, respectively. The primary motivation for utilizing SOM analysis followed by *k*-means clustering is to obtain both a condensed, interpretable grouping (four clusters) and detailed information about the contributors to these clusters. While *k*-means clustering or 2 × 2 SOM analysis alone would give us 4 clusters, we would lose the ability to understand the internal structure and nonlinear variability within each cluster. As SOM preserves the topological relationships of the complex data, the 5 × 5 SOM provides a preliminary structure that can guide the clustering process. Then, *k*-means clustering can refine these initial clusters by optimizing the centroid placement based on the SOM output, ensuring that clusters are well-defined and distinct. The combined clustering methods enhance the analysis of complex datasets, improving both the clarity and dependability of clustering results. Greedy *k*-means++ algorithm is used for initialization of centroids. We used the Python package MiniSOM (https://github.com/JustGlowing/minisom) to implement SOM, and scikit-learn for *k*-means clustering (https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html). Sammon map is implemented using the code downloaded from https://github.com/tompollard/sammon.

2.4. Characterizing vortex asymmetry

We use one-way analysis of variance (ANOVA) for checking whether a vortex is symmetric or not by comparing the means at the same grid point across the four quadrants (front-left, front-right, rear-left, and rear-right) relative to the direction of storm motion. ANOVA is a statistical method that determines whether there are statistically significant differences between the means of three or more groups (Wilks, 2019). The significance level is set to 0.05. We designate a grid point as asymmetric when the null hypothesis of the test (there is no difference among group means) is rejected. It means that even if the mean values at a grid point of three out of four quadrants are not different from each other, the vortex is still considered as asymmetric at this grid point.

3. Results

3.1. Distinct categorization of vortex structure

Each six-hourly record in the IBTrACS best-track dataset during the tropical cyclone season in the western North Pacific (June–October) in the period 2004–2021 that is designated as a tropical disturbance or tropical depression is used in this study. Out of the total 2014 samples, 1022 are tropical disturbances, and 992 are tropical depressions (Table 1). The 3-D relative vorticity field of each sample is regridded to a storm-centered curvilinear grid and averaged azimuthally. Afterwards, all the 2014 samples of tropical disturbance and tropical depression are pooled and classified by SOM and then *k*-means clustering (Fig. 2). We utilize the azimuthal average to depict the vertical structure of the vortex in a simplified manner. It's important to note that the horizontal structure of pre-existing disturbances is often asymmetric. Therefore, our use of the azimuthal average is to indicate the location at which

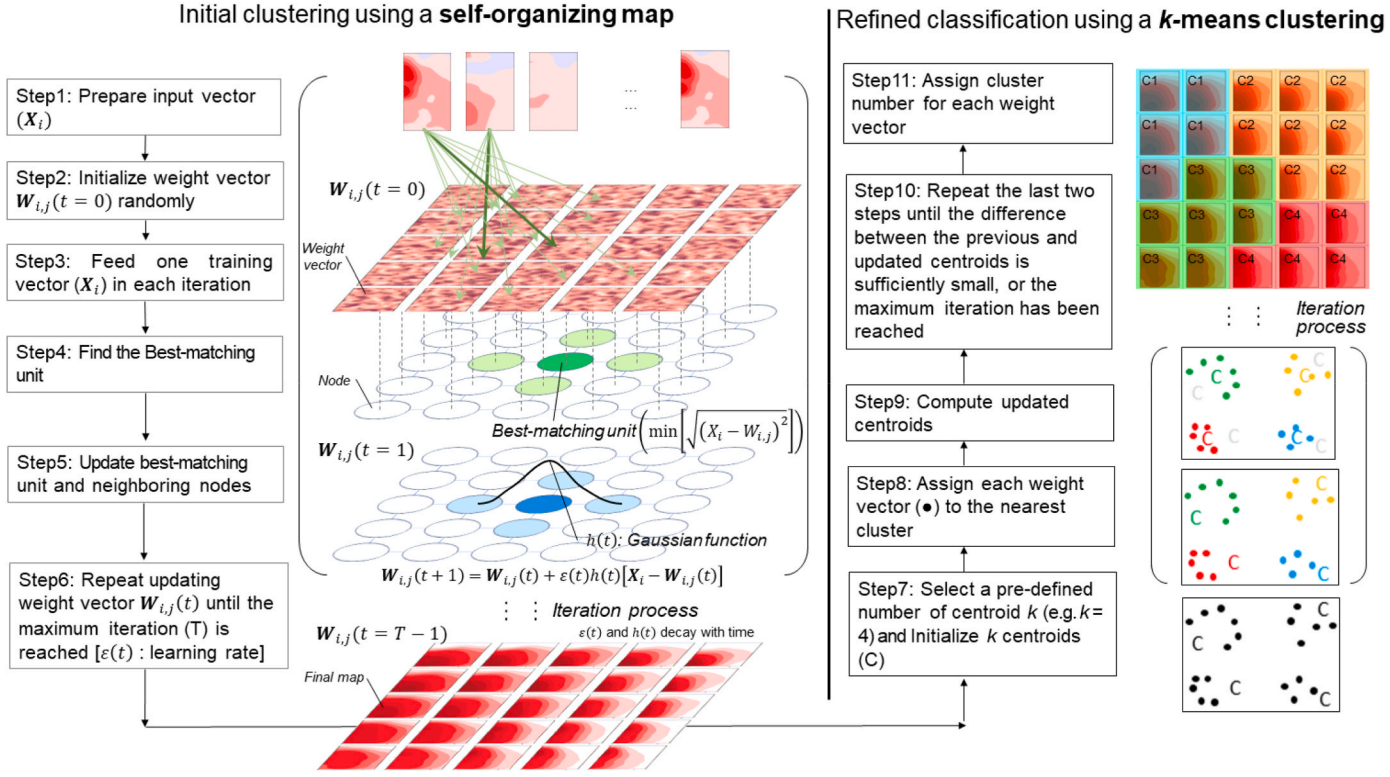


Fig. 2. Illustration of the classification of the vertical structure characterized by relative vorticity.

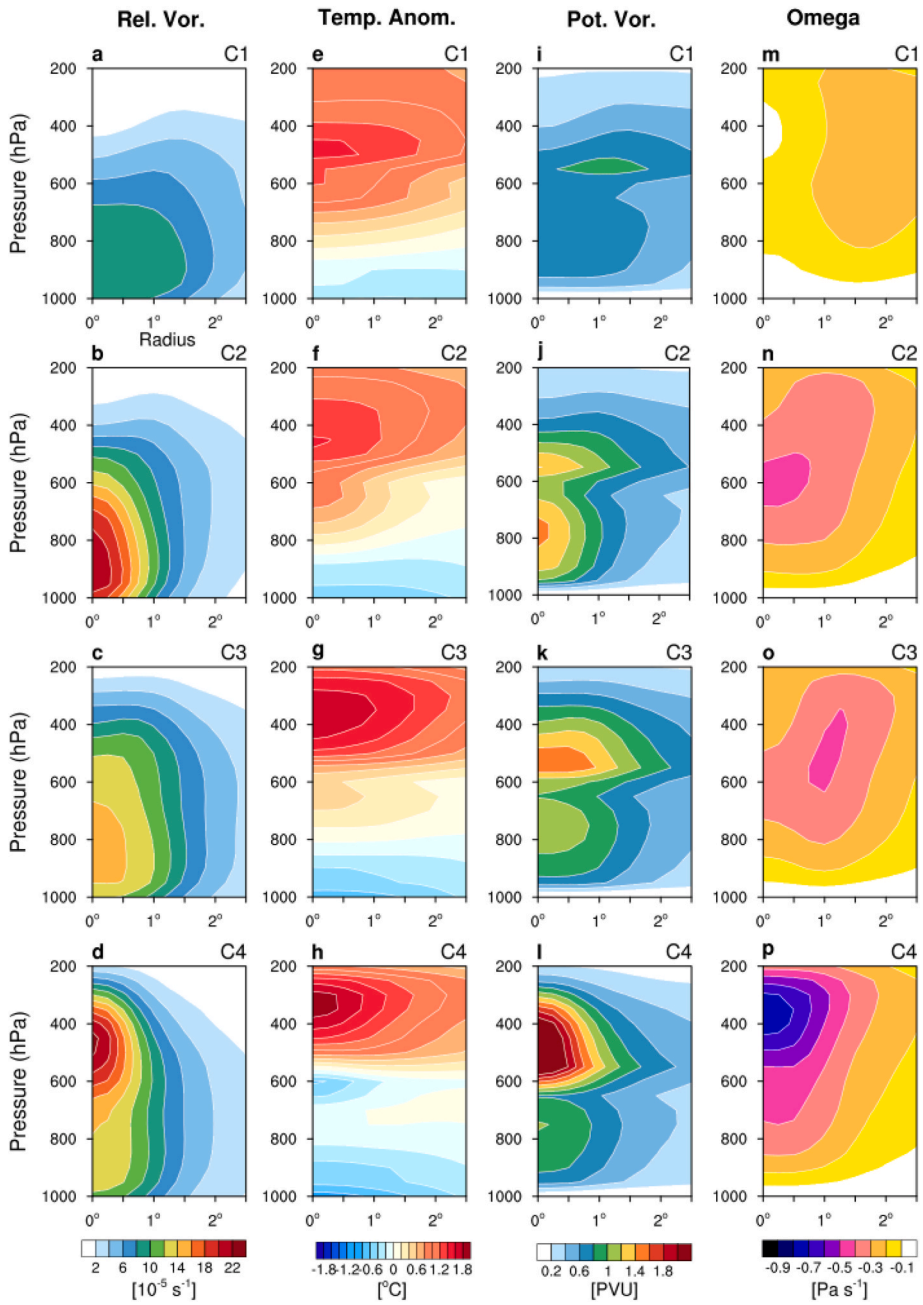


Fig. 3. The composite of azimuthally averaged vortex structure and associated characteristics of the four clusters of pre-existing disturbances. (a–d) Relative vorticity (10^{-5} s^{-1}); (e–h) temperature perturbation (K); (i–l) potential vorticity (PVU, or 10^{-6} m^2 s^{-1} K kg^{-1}); (m–p) omega (vertical velocity; Pa s^{-1}).

coherent signals of vorticity can be detected. The asymmetric characteristics will be further discussed in Section 3.3. Instead of performing *k*-means clustering alone, the advantages of using both of the above algorithms successively are three-fold: (1) vortex structure of a pre-existing disturbance can undergo changes over its lifetime, implying that each structure is not independent and justifying the suitability of SOM to capture this connectivity; (2) SOM organize the centroids such that similar ones are closer to each other due to its topology preserving property, and it allows us to learn the relationships and structures within the data; and (3) a finer grouping with *k*-means clustering allows for the identification of distinct groups and an easier interpretation.

The pre-existing disturbances are grouped into four clusters denoted as C1, C2, C3, and C4 (Fig. 3), where 'C' stands for 'cluster', after splitting into 25 groups by SOM. The four clusters exhibit distinctive structures in terms of the vertical extent (shallower and deeper) and radial extent (broader and narrower) (Fig. 3a–d). C1 is characterized by pre-existing disturbances with a weak and broader low-level vortex, where the maximum value of relative vorticity is concentrated primarily in the lower troposphere below 700 hPa (shallower vortex) and extends farther from the center (broader vortex). The C1 cluster has the smallest magnitude of maximum relative vorticity among the four clusters, measuring around 10×10^{-5} s^{-1}. C2 is marked by pre-existing disturbances with a strong and narrow low-level vortex with the maximum relative vorticity of about 20×10^{-5} s^{-1}. While both C2 and C1 both have relative vorticity concentrated in the lower troposphere, C2 exhibits a stronger radial gradient of relative vorticity compared to C1. C3 displays a weak and broader vortex with low-level maximum similar to C1, but it is featured a more evenly distributed maximum vorticity that extends throughout a deeper column within the troposphere. C4 is linked with pre-existing disturbances that exhibit stronger radial

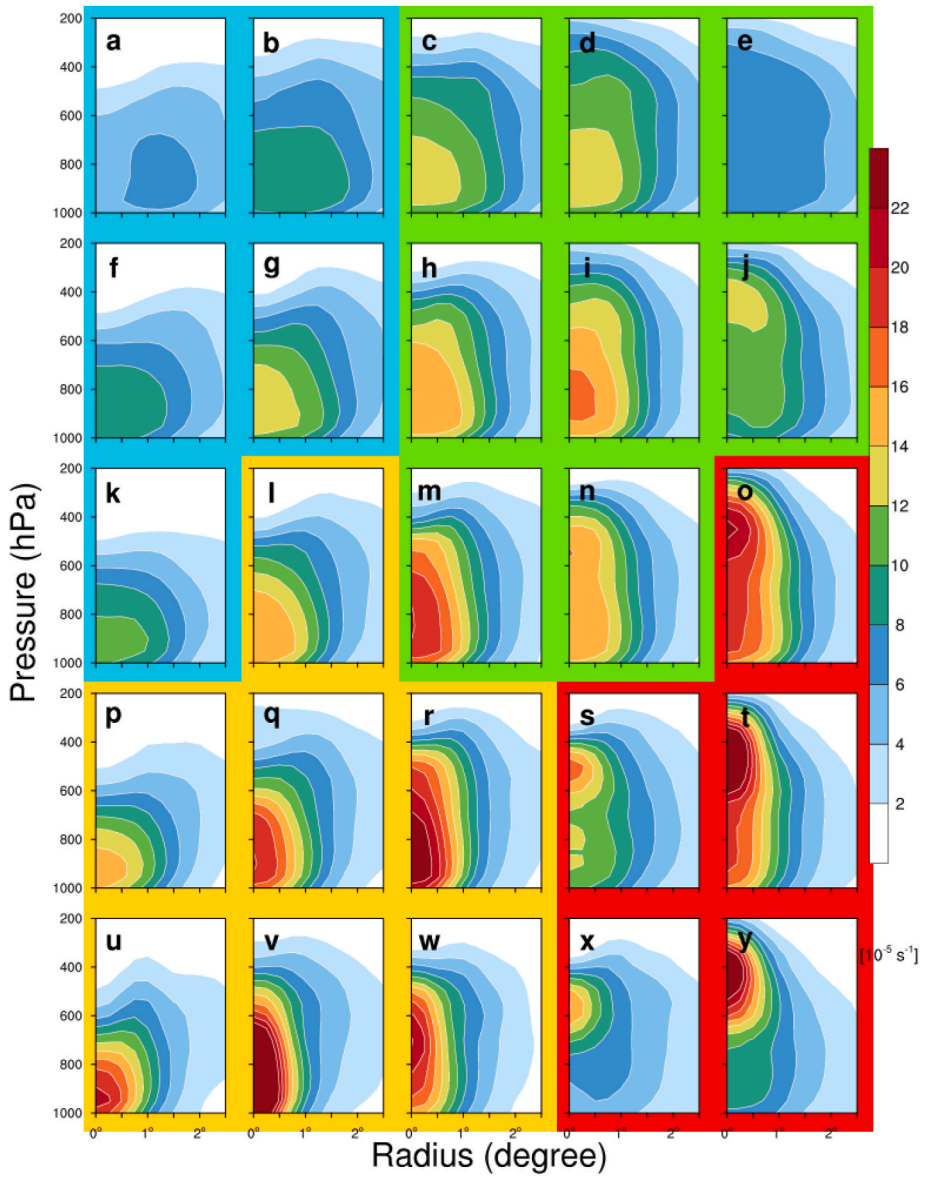


Fig. 4. The composite of relative vorticity of all nodes on the SOM. The colors in the background indicate the cluster in which a node belongs to (blue for C1; orange for C2; green for C3; red for C4).

gradients, and the maximum relative vorticity is found in the mid-to-upper troposphere between 400 hPa and 600 hPa. Among the four clusters, C4 has the highest magnitude of maximum relative vorticity, exceeding 20×10^{-5} s^{-1}.

Table 1 provides a summary of the sample distribution among the clusters. C2 and C3 are the predominant cluster among others, representing 29.9 % and 31.7 % of the total samples, respectively. The remaining clusters have similar percentages, with approximately 20 % of the samples. Among them, C1 has the smallest number of samples, comprising 17.6 %. The SOM-based composite patterns in C1 (blue background), C2 (orange background), and C4 (red background) each consistently exhibit a vorticity maximum either in the lower or upper troposphere (Fig. 4). In contrast, C3 (green background) is more heterogeneous: its eight associated SOM nodes display both broader and narrower vortices, and the vorticity maxima vary between the lower and upper troposphere. This variability suggests that C3 includes a wider range of vortex structures, characterized mainly by moderate relative vorticity extending over a deeper vertical layer, compared to the other clusters. C1, C2, and C3 have bottom-heavy patterns in relative vorticity, meaning the relative vorticity is concentrated more in the lower troposphere, and share a high level of resemblance, except for three nodes in C3 (Fig. 4e–j, n). This similarity results from the nature of the SOM, where similar nodes are positioned in close proximity to each other (Fig. 4). These bottom-heavy vortices, which add up to 79.2 % of the samples (68.9 % when the nodes in Fig. 4e–j, n are excluded), are also commonly observed in previous studies (Ditchek et al., 2017; Kerns and Zipser, 2009; Murthy and Boos, 2019; Ritchie and Holland, 1999; Zehr, 1976).

The structural difference between C1 and C2 can be interpreted through the conservation of angular momentum. Although the clustering is based on snapshots of vortex structures and does not consider time evolution, the spatial structure of the vortex circulation still provides insight. When air moves radially inwards, the vortex must spin faster to conserve angular momentum, thus relative vorticity increases (Kilroy and Smith, 2017). It helps explain why the broader vortices (C1) have weaker relative vorticity than the narrower ones (C2). C4 can be associated with pre-existing disturbances with a mid-level vortex, which has maximum relative vorticity in the mid-troposphere. This structure commonly coincides with stratiform precipitation: latent heat release in the stratiform anvil and evaporative cooling below the melting layer

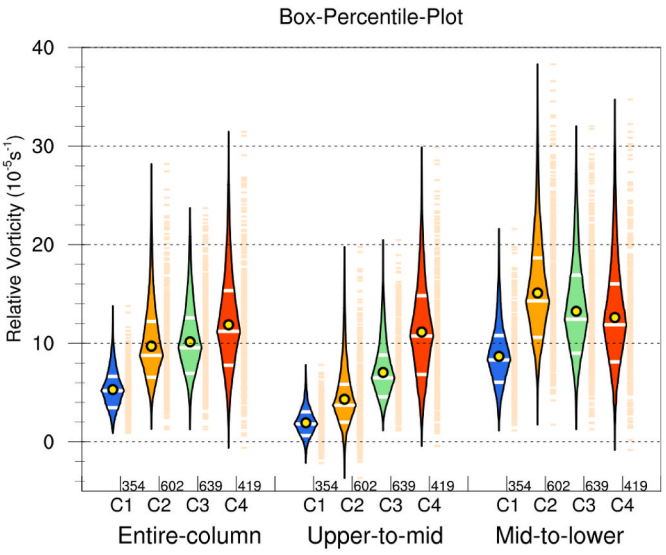


Fig. 5. The boxplot-percentile plot of relative vorticity (10^{-5} s^{-1}) averaged within 1° from the storm center in the entire (200–950 hPa), mid-to-lower (600–950 hPa), and upper-to-mid troposphere (200–550 hPa) for each cluster. The yellow circle inside a box represents the mean value. The three white horizontal lines in a box represent the lower quartile, median, and upper quartile. The yellow dashes on the right-hand side of a box denote the magnitude of samples associated with the corresponding cluster. The number of samples for each cluster is marked along the x-axis.

sharpen the mid-tropospheric potential-temperature gradient, generating a PV maximum that spins up the mid-level vortex (Bister and Emanuel, 1997; Ritchie and Holland, 1997; Simpson et al., 1997). C3 is mainly associated with nodes that have a weak gradient of relative vorticity along the vertical axis. Unlike other clusters, C3 does not have an apparent maximum in relative vorticity which can be caused by a lack of dominant heating by stratiform or deep convection. This property will be further investigated in the next subsection.

To quantitatively compare the magnitude of relative vorticity at different levels, the average relative vorticity within a 1° radius from the storm center is shown in Fig. 5 for the entire troposphere (200–950 hPa), upper-to-mid troposphere (200–550 hPa), and mid-to-lower troposphere (600–950 hPa). For the entire-tropospheric average, C4 has the highest relative vorticity, followed by C3, C2, and then C1, primarily influenced by the upper-to-mid relative vorticity. For most pre-existing disturbances, the average relative vorticity is greater in the lower than the upper troposphere. The largest difference in vertical levels can be found for C2, where the vorticity is most pronounced at the mid-to-lower level, while the minimal difference was found at C4. The relationship between clusters remains similar when the average within 2° from the center is calculated (Fig. S3). This result shows that the vertical vortex structure of pre-existing disturbances is discernible in terms of the magnitude of vorticity (i.e., radial gradient) and their vertical extent. Among the samples in the western North Pacific, 79.2 % of the samples (those in C1, C2, and C3) show a bottom-heavy vorticity maximum, while 20.8 % display a top-heavy structure (i.e., with vorticity concentrated in the upper troposphere).

3.2. Thermodynamic structure for different vortices

To elucidate each vortex structure, we investigate the composite of temperature perturbation (Fig. 3e–h), potential vorticity (PV; Fig. 3i–l), and omega (i.e., vertical velocity in pressure coordinates; Fig. 3m–p). There is a similarity in the temperature perturbation field, which is defined as the deviation of temperature from the background value (average between 5° and 10° from the center) on the same pressure level, among all four clusters. All of them exhibit warm anomalies above 500 hPa. In the composite of C1, C2, and C3 (Fig. 3e, f, g), the warm anomalies reach their peak intensity between 300 hPa and 500 hPa near the center and exhibit a downward extension. The warm anomalies associated with C1 extend radially away from the center in a broader manner compared to C2. The maximum of warm anomaly in C3 is stronger and extends over a thicker vertical layer where the temperature anomaly exceeds 1.2 K, and it also reaches a higher pressure level than in C1 and C2. A relatively cold anomaly can be found at pressure levels below 800 hPa in all four clusters. Upon closer examination of the composite of SOM nodes, we identify that some of them have a secondary warm anomaly at ~700 hPa (Figs. S4d, i, r) or a primary warm anomaly at ~600 hPa (Fig. S4k and p, u). It suggests that the lower portion of the warm anomaly (between 600 and 800 hPa) in the composite of C1 and C2 can be attributed to two separate warm anomalies at two vertical levels that partially overlap. The two warm anomalies we find closely match the heating profiles shown by Hartmann et al. (1984) and Houze (1982): the 600-hPa peak corresponds to deep convective towers at earlier stage alone, while the 400-hPa peak corresponds to mature tropical cloud clusters combining convective tower and stratiform clouds.

In the case of C4 (Fig. 3h), the peak intensity of the warm anomaly is located at a higher level between 200 hPa and 400 hPa compared to C1 and C2, and it does not have a secondary low-level peak. Additionally, the maximum intensity of warm anomaly in C4 is stronger (greater than +2 K), and the intensity of cold anomaly in the lower troposphere is also stronger (about −1.0 K). Furthermore, there is a cold anomaly in the mid-level around ~600 hPa, which does not exist in other clusters. Some studies have reported the co-existence of a maximum in relative vorticity in the mid-troposphere and the presence of a cold anomaly directly beneath the warm anomaly (Bartels and Maddox, 1991; Ge et al., 2013), which can also be observed in the composite of C4. These nascent storms with a mid-level vortex can have a cold anomaly at a higher altitude than those with a bottom-heavy vorticity field (Kushwaha et al., 2021). This warm-over-cold pattern has been commonly observed in pre-existing disturbances or the formation of mesoscale convective vortices by a combination of condensation heating above and evaporative cooling below in stratiform anvils (Bister and Emanuel, 1997; Lussier III et al., 2014; Wang et al., 2010; Wu and Fang, 2019). This thermal structure induces vertical shear in tangential wind via the thermal wind balance: $\partial V_t/\partial z = (g/fT)\,(\partial T/\partial r)$, where V_t is the tangential wind speed, z is the vertical coordinate, g is the gravitational constant, f is the Coriolis parameter, T is the temperature, and r is the radial distance from the vortex center. According to the thermal wind relation, the lower-level cold anomaly ($\partial T/\partial r > 0$) corresponds to a positive vertical shear in the tangential wind ($\partial V_t/\partial z > 0$) that causes the wind to strengthen with height. Conversely, the upper-level warm anomaly ($\partial T/\partial r < 0$) implies a negative shear ($\partial V_t/\partial z < \mathrm{e}T/\partial r = 0$) is also the level of zero vertical wind shear ($\partial V_t/\partial z = 0$). This altitude marks the location of the maximum tangential wind and therefore the strongest vertical component of relative vorticity. The temperature structures of all clusters agree well with the previous description of developing disturbances measured by radiosonde in the western North Pacific: a maximum warm anomaly at ~250 hPa, and a cold anomaly below ~850 hPa (Zehr, 1976).

PV is a diagnostic variable that combines absolute vorticity and stratification ($1/\rho \times \zeta_a \cdot \nabla\theta$), where ρ is the air density, ζ_a is the absolute vorticity vector, and $\nabla\theta$ is the gradient of potential temperature. The composite PV of C1 has a maximum (0.8–1.0 PVU) at ~550 hPa, but it only occupies a shallow layer (Fig. 3i). For C2, there is also a maximum of 1.2–1.4 PVU at ~550 hPa, but another maximum of slightly greater magnitude that encompasses 650–950 hPa also exists (Fig. 3j). The pattern of PV of C3 is similar to C1 with a primary maximum in the mid-troposphere, but the magnitude is greater throughout the atmosphere (Fig. 3k). C4 has the strongest PV maximum (greater than 2.5 PVU) amongst all clusters which is located at 500 hPa (Fig. 3l). A secondary maximum located at ~850 hPa is also observed. Since latent heat release

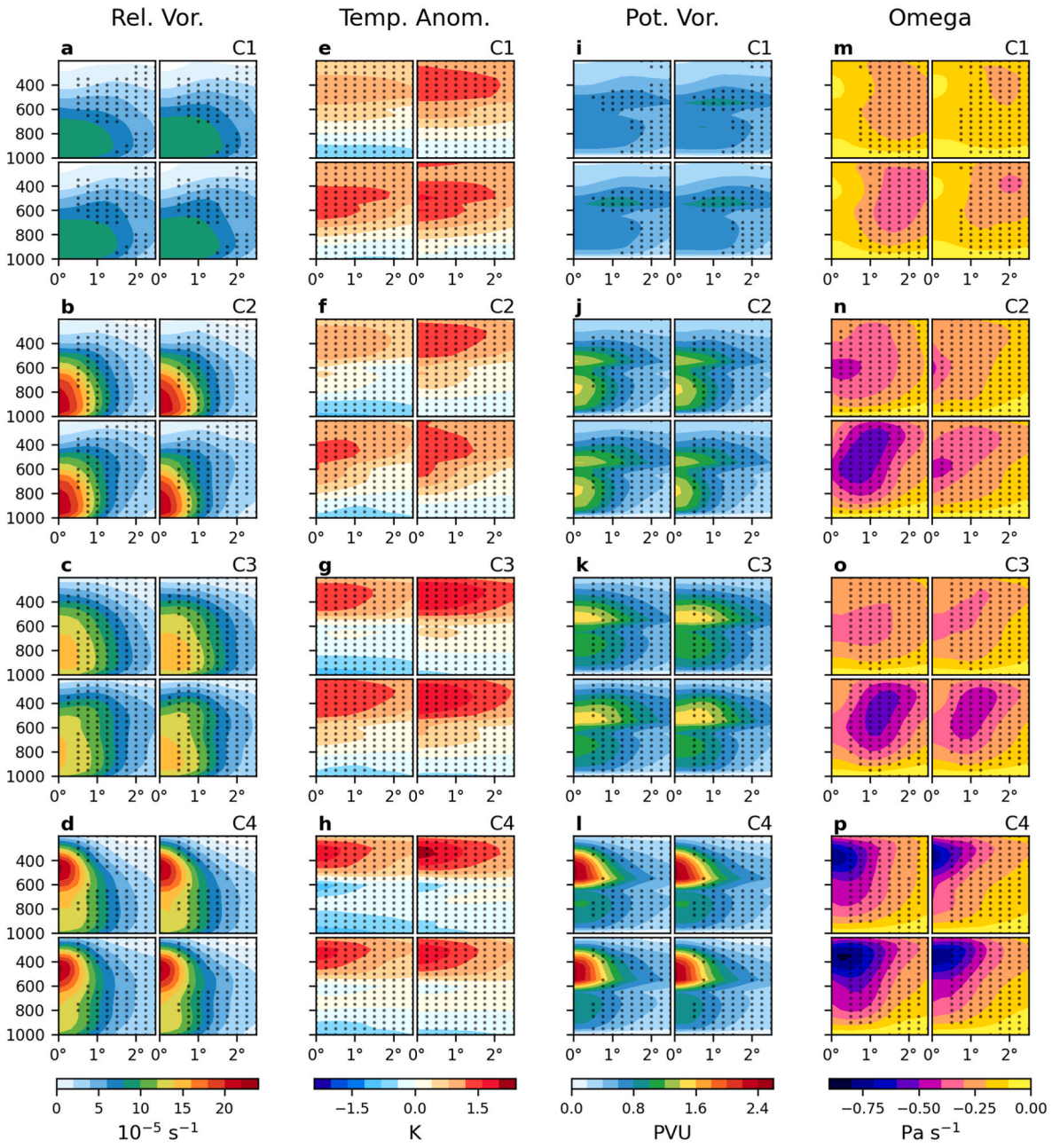


Fig. 6. The composite of quadrant averaged vortex structure and associated characteristics of the four clusters of pre-existing disturbances. (a–d) Relative vorticity (10^{-5} s^{-1}); (e–h) temperature perturbation (K); (i–l) potential vorticity (PVU, or 10^{-6} m^2 s^{-1} K kg^{-1}); (m–p) omega (vertical velocity; Pa s^{-1}). Stippling indicates grid points where the each variable differs significantly among quadrants at the 95% confidence level based on one-way ANOVA.

in the pre-existing disturbances breaks the conservation of PV, the stratiform and deep convective heating each produces its own PV maximum. Murthy and Boos (2019) showed that deep convective heating is responsible for the PV maximum in the lower troposphere, whereas stratiform heating is responsible for the mid-troposphere. Therefore, the shallower vortices (C1 and C2) lack strong mid-level heating and have weaker PV above 400 hPa, while the deeper vortices (C3 and C4) exhibit pronounced mid-level PV. On the other hand, broader vortices (C1 and C3) tend to have a weaker secondary PV peak in the lower troposphere than the narrower ones (C2 and C4). Furthermore, C1 has a weak lower-tropospheric PV maximum, which indicates that deep convective heating is neither strong nor persistent. Consequently, this weak convection cannot sustain a continuous stratiform anvil in the mid-troposphere. As a result, latent heating in the mid-troposphere generated by the stratiform anvil becomes intermittent and disorganized and leads to off-center PV generation therein.

For omega, the maximum of C1 is not confined near the center but is located 1° away from the center (Fig. 3m). The maximum rising motion of C1 is the weakest among the four clusters (−0.2 to −0.3 Pa s^{-1}). The spatial distribution of omega of C2 and C3 are similar, although the maximum upward motion is found near the center and at 1° radius from the center, respectively (Fig. 3n and o). The greatest upward motion at the center of C2 and C3 is located at 600 hPa. The difference between the two clusters is more evident from the node composites (Fig. S6). The maximum magnitude of omega in nodes associated with C3 (Fig. 3o) does not show much variation. However, there is a greater contrast in magnitude between the nodes that belong to C2. Some nodes in C2 have weaker omega than those in C3, but some of them are stronger. On the other hand, C4 exhibits a maximum at ~350 hPa near the center (Fig. 3p and Fig. S6). Overall, C4 exhibits the strongest rising motion among the

four clusters (<−0.8 Pa s^{-1}). Pre-existing disturbances with greater radial gradient or greater depth in relative vorticity generally have a strong rising motion near the storm center, which can be seen from the bottom row and rightmost column in Fig. S6. The composite omega profile of developing tropical cloud clusters from rawinsonde data demonstrated a maximum rising motion spanning 300–500 hPa (Lee, 1989; McBride, 1981; McBride and Zehr, 1981). It implies that such a vertical omega profile is mostly dominated by nodes exhibiting strong mid-tropospheric ascent (~400 hPa; e.g., Figs. S6j, o, t, x, y), such that nodes with weak ascent contribute little to the composite cluster pattern. The observation from the Global Atmospheric Research Program's Atlantic Tropical Experiment showed that the vertical profile of omega changed with the heating profile (Houze, 1982): a low-level maximum in omega existed in the earlier stage of development when the heating is purely convection, and the maximum shifted to upper level when the cloud cluster intensified. The two major peaks of rising motion can be deduced from the PV maxima, since each PV maximum marks the level of strongest latent heating (stratiform aloft and convective below) that drives the corresponding ascent. When there is a sufficiently strong and dominating PV maximum, which is usually detected in the mid-troposphere at around 300–600 hPa (except for some nodes in C2), a maximum in omega can also be seen at a similar pressure level, and when the two maxima in PV have comparable magnitude, only one maximum of omega is observed at a pressure level between two PV maxima are located. Therefore, the upper and lower maxima in omega correspond to stratiform and deep convective heating, respectively, as a result of the buoyancy generated by the latent heat release.

The overall vortex properties of each cluster can be summarized as follows: C1 displays a broad bottom-heavy vortex; C2 is a narrow bottom-heavy vortex; C3 exhibits a broad deep vortex, with strong warm anomaly and upward wind in the upper levels; and C4 is a narrow top-heavy vortex. Comparatively, a broader distribution of relative vorticity (C1 and C3) can lead to a more widespread area of convection (Kilroy and Smith, 2017), which in turn causes a greater radial extent of rising motion and warm anomaly. The maximum magnitude of temperature perturbation of C2 is similar to that of C1, but more confined to the center. Potential vorticity over the center in the lower troposphere of C2 is stronger than that of C1 and C3, and the rising motion of C2 is similar to that of C3. For C4, it has a maximum of relative vorticity in the mid-troposphere and a stronger magnitude near the storm center compared to other clusters. The warm anomaly in the upper troposphere and the cold anomaly in the lower troposphere, the maximum PV in the mid-troposphere, and the rising motion in the upper and mid-troposphere are all stronger than those of other clusters.

3.3. Vortex axisymmetry

Given that TCs during their early intensification stages tend to exhibit highly asymmetric structures, we further examine the quadrant average of the four clusters (Fig. 6). Before calculating the average value of each quadrant for different variables, we rotate each sample so that the motion is aligned to the same direction. Adopting quadrant average allows us to retain the asymmetry of horizontal storm structure without excessively increasing the complexity of the analysis. While individual cases may exhibit varying axisymmetric characteristics, the composite relative vorticity near the storm center (i.e., within 0.5° from the storm center) displays a similar distribution across all four quadrants. This is also indicated by insignificant differences in relative vorticity between the four quadrants, for all four clusters (Fig. 6a–d). However, at a distance 0.5° away from the storm center, relative vorticity is largely asymmetric. Similar discrepancy in symmetry is also observed in the composite of potential vorticity (Fig. 6i–l) and omega (Fig. 6m–p). On the other hand, temperature anomaly exhibits asymmetries for all radii (Fig. 6e–h). Therefore, the average structure of the four quadrants does not differ too much within 0.5° from the storm center for relative vorticity and potential vorticity.

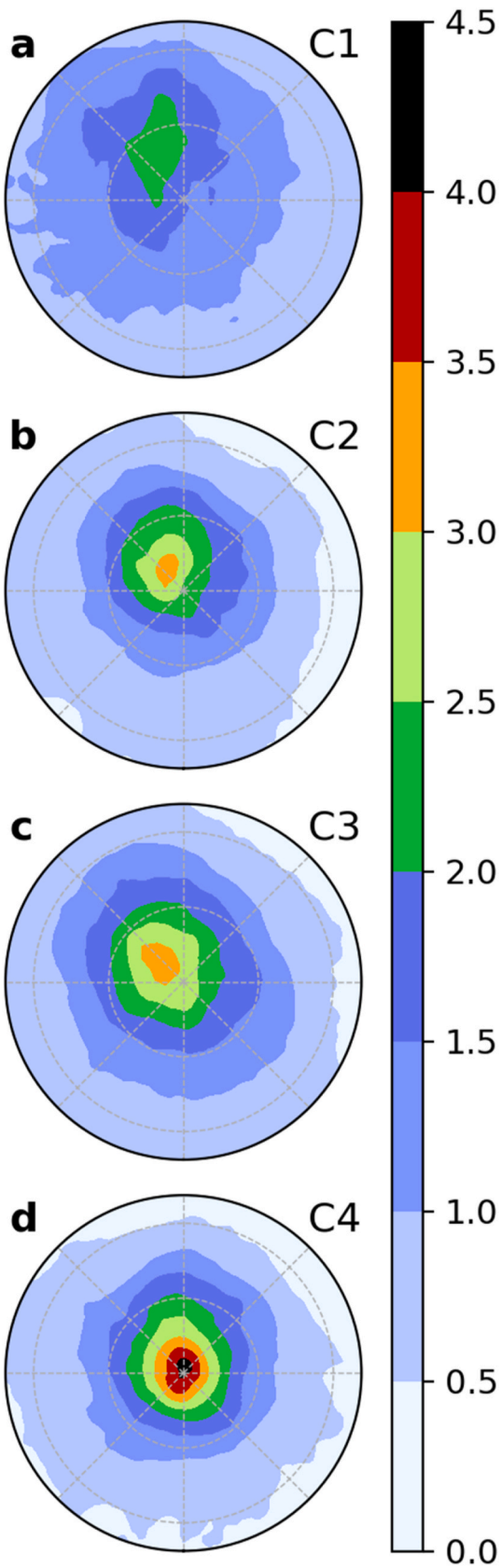


Fig. 7. The composite of precipitation field of the four clusters of pre-existing disturbances. The unit of rain rate is mm hr^{-1}.

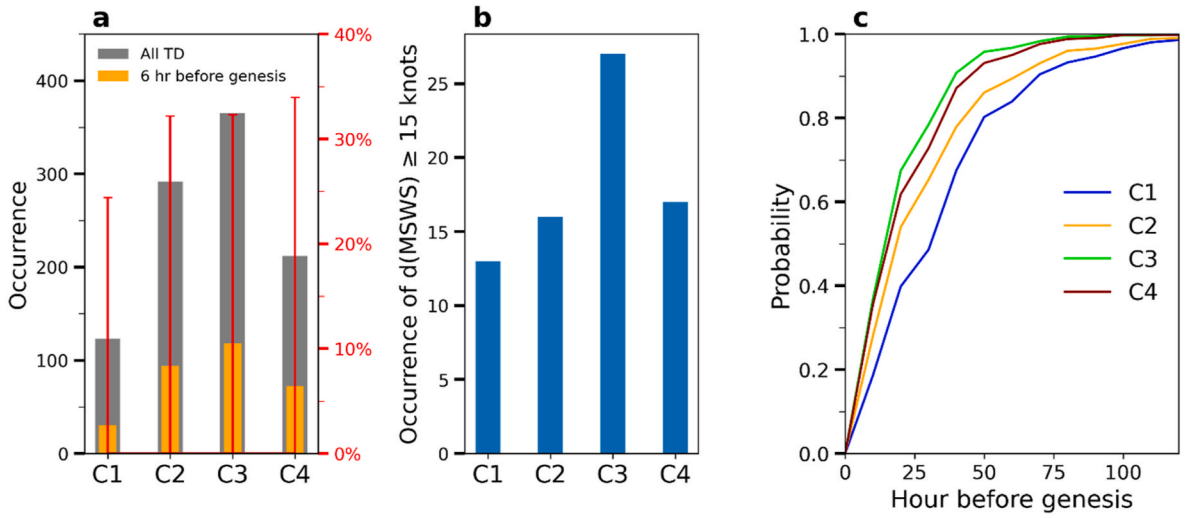


Fig. 8. Comparison in properties associated with the completion of genesis. (a) Black and orange bars represent tropical depressions and the subset that were present 6 h before the completion of tropical cyclogenesis, respectively. Red lines denotes the percentage of tropical depressions that occur 6 h prior to tropical cyclone genesis. (b) The occurrence of pre-existing disturbances that experience an increase in maximum sustained wind speed greater or equal to 15 knots within 24 h in each cluster. (c) The cumulative distribution function of the time to genesis of all pre-existing disturbances for each cluster.

In addition, we inspect the composite precipitation field of each cluster (Fig. 7). C1 has the lowest maximum rain rate amongst the four clusters (Fig. 7a), while C2 and C3 show similar maximum rain rates (Fig. 7b and c). The precipitation fields for these three clusters tend to have a maximum in the left front quadrant. On the other hand, C4 has the greatest maximum rain rate and possesses the highest degree of rain rate symmetry (Fig. 7d). The intensity of precipitation generally corresponds well to the magnitude of temperature anomalies (Fig. 3e–h) and omega (Fig. 3m–p). It is important to note that the asymmetric distribution of rainfall in pre-existing disturbances can be influenced by various environmental factors, including vertical wind shear, relative humidity, and the size and intensity of the storm. This study primarily focuses on vortex structure, so further analysis will be necessary to fully understand the relationship between rainfall asymmetry and pre-existing disturbances.

3.4. The pathways to the TC genesis

A question that remains unanswered is whether the vertical and radial structure of TC precursors represent different stages of TC development and distinct pathways to TC genesis. It should be mentioned that changes in TC intensity involve multiscale processes (Chen et al., 2023; Hendricks, 2012; Rogers, 2021). The interaction between convective-scale processes with the TC vortex and its surrounding environment determines the evolution of the storm's intensity. Here we further examine the number of cases and duration of each cluster resulting in the TC genesis in the subsequent timestep (i.e., in the next 6 h). Because JTWC does not always issue warnings for tropical disturbances depending on purely objective criteria, tropical disturbances can add considerable variations in the duration data. Therefore, we exclude tropical disturbances and only consider tropical depressions in this analysis. The number of samples that occur 6 h prior to TC genesis is 30, 94, 118, and 72 for C1, C2, C3, and C4, respectively (Fig. 8a). The corresponding percentages are 24.4 %, 32.2 %, 32.3 %, and 34.0 % of the total tropical depression sample. C1 has the lowest number and percentage of tropical depressions leading to genesis in the next 6 h among all clusters. There are more tropical depressions in C3, but C2, C3, and C4 have a similar percentage (32 %–34 %) in each of them that complete genesis 6 h later. This observation indicates that, except for very weak vortex in C1, which has a maximum relative vorticity of less than 10×10^{-5} s^{-1}, the three other types of vortex structures (C2, C3, and C4) exhibit similar rates of genesis completion.

We also investigate whether any type of vertical structure can be favorable for fast increase in storm intensity (indicated by changes in 10-m maximum sustained wind speed within 24 h) of pre-existing disturbances. Only the period in which a storm remains a pre-existing disturbance is considered. 912 of them have a value in the range from −10 knots to +20 knots at an interval of 5 knots (Fig. S7). We select +15 knots, which is the 92-th percentile of the data, as the threshold for relatively rapid intensification. The occurrence of storms with maximum sustained wind speed greater than or equal to 15 knots for C1, C2, C3, and C4 are 13, 16, 27, and 17 (Fig. 8b), respectively, which correspond to 18 %, 22 %, 37 %, and 23 %. C3, the broad deep vortex, are more associated with fast intensifying pre-existing disturbances, while C1 is the least associated. Furthermore, the time to genesis, which refers to the duration it takes for a pre-existing disturbance from the point of record to develop into a tropical storm, is also inspected (Fig. 8c). Note that a pre-existing disturbance does not necessarily remain in the same cluster before the completion of genesis, but exhibit jump from one cluster to another. The average duration for a pre-existing disturbance to transition into tropical storm intensity is 29.1 h. C3 (green) and C4 (red), characterized by broad deep or narrow top-heavy vortices, exhibit a higher proportion of cases closer to genesis, whereas C1 (blue) and C2 (orange) display a relatively smaller proportion (Fig. 8c). C3 contains both the greatest number and proportion of samples that have a shorter time to genesis. It may imply that sufficient relative vorticity penetrating a greater depth can facilitate more rapid TC intensification. The results also align with the ratio of the number of tropical depressions to tropical disturbances (Table 1): a greater ratio corresponds to a greater proportion of samples with a shorter time before genesis. In summary, the vertical extent of relative vorticity to upper level is a more important factor of the formation of tropical cyclones compared to the other aspects of vortex structures, given sufficient relative vorticity is present. Our result is also in line with a recent study that suggested TCs of greater height, as defined using tangential wind speed, possess greater intensity (DesRosiers et al., 2023).

3.5. Difference between tropical disturbances and tropical depressions

From satellite imagery in the infrared channel (indicated the top of a convective storm) of a selected sample, both tropical disturbance and tropical depression can consist of loosely organized cloud clusters

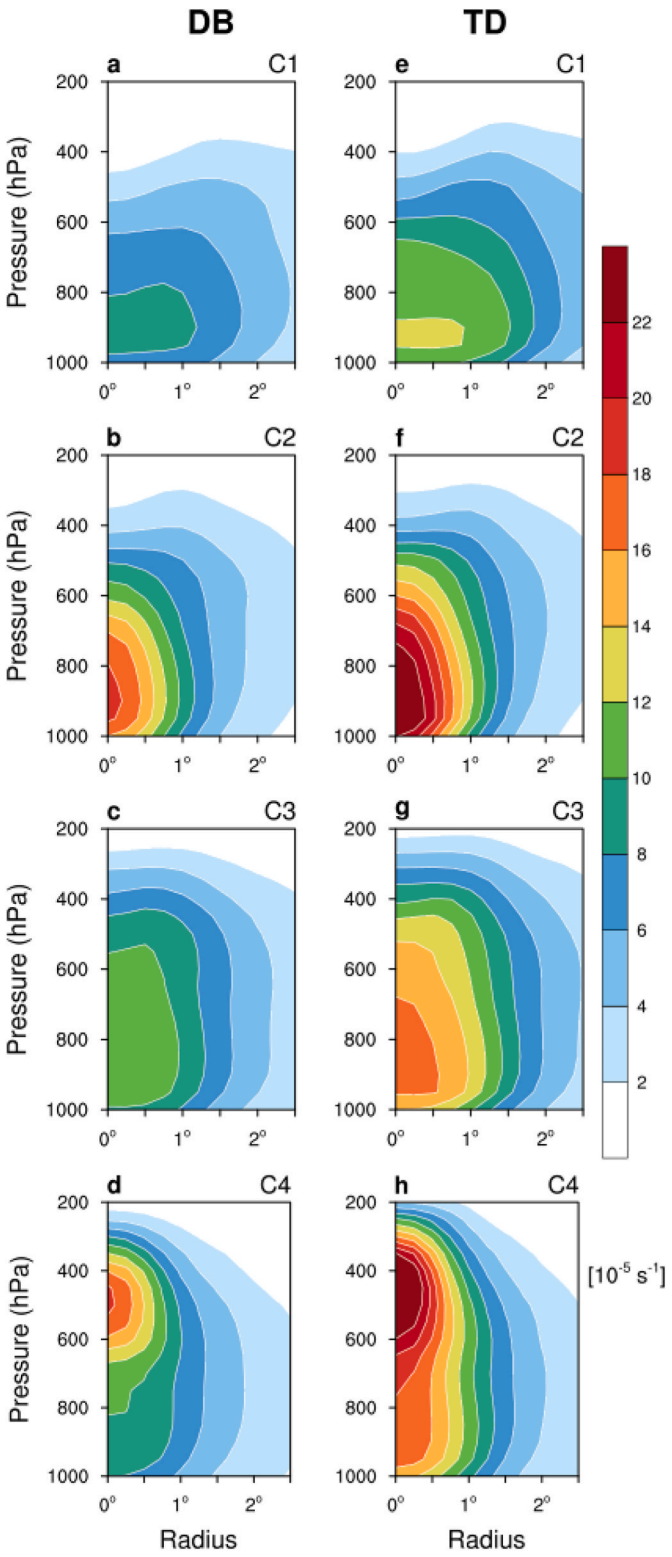


Fig. 9. The composite of azimuthally averaged relative vorticity (10^{-5} s^{-1}) of the four clusters of pre-existing disturbances. Left and right column are the cluster composites for tropical disturbances (DB) and tropical depressions (TD), respectively.

(Fig. 1a–c). However, tropical depression shows lower minimum surface pressure, stronger surface pressure gradient, and thus higher 10-m wind speed than tropical disturbances (Fig. 1e–g). This is consistent with operational practice, where forecast centers use either 10-m wind speed or minimum sea-level pressure to indicate storm intensity. To understand the relationship between surface storm intensity and vertical vortex structure, we compare each cluster between tropical disturbances and tropical depressions (Fig. 9).

The vertical structure characteristics of the four clusters remain largely consistent. However, the magnitude of relative vorticity in tropical disturbances is lower than that in tropical depressions across all clusters. Additionally, there is a higher occurrence of tropical disturbances compared to tropical depressions in the weakest vortex cluster, C1 (Table 1). Nonetheless, the difference in occurrence is less pronounced in other clusters. These two results suggest that tropical disturbances not only exhibit a lower intensity near the surface but also possess a smaller magnitude of relative vorticity throughout the troposphere. The composite of tropical disturbances and tropical depressions varies mainly in magnitude for different variables as well, without pronounced discrepancy in structure (temperature perturbation, potential vorticity, and omega; Fig. S8–10).

4. Conclusion and discussion

We identify four predominant structures of the pre-existing disturbances in the western North Pacific by applying unsupervised machine learning techniques, including SOM and *k*-means clustering. These classified four clusters exhibit discernible structures in relative vorticity in terms of vertical and radial characteristics: C1 has a weak, broad, and bottom-heavy vortex; C2 is a strong, narrow, and bottom-heavy vortex; C3 is a moderate and broad but vertically deep vortex; C4 is characterized by strong, narrow, and top-heavy vortex.

The distinct characteristics of the vortex structures across clusters can be linked to differences in latent heating associated with stratiform and convective precipitation. Stratiform precipitation tends to release latent heat in the upper troposphere that contributes to the development of a mid-level vortex, a positive PV maximum in the mid-troposphere, and upward motion around the 300–500 hPa. In contrast, convective precipitation primarily releases latent heat in the lower to mid-troposphere, thus contributing to a warm anomaly at around 600 hPa, a PV maximum in the lower troposphere, and rising motion in the layer between 500 and 700 hPa. When stratiform and convective precipitation combine, they can induce a warm anomaly at approximately 400 hPa. In addition, temperature perturbation, PV, and rising motion show variations among clusters. The substantial difference among the four clusters confirms that the classification by the selected algorithms is effective. Furthermore, C2, C3, and C4 account for a similar fraction of tropical depressions that eventually intensify into tropical storms. It implies that each of these three clusters has an equal likelihood of TC formation. Lastly, the major difference between tropical disturbances and tropical depressions mainly lies in their intensity rather than their vortex structure.

However, our study has several limitations. Firstly, the reliance on the best-track dataset may induce uncertainty in the analyses. In general, the uncertainty in storm center position decreases with the storm intensity (Torn and Snyder, 2012). The lack of eye or symmetric low-level circulation pattern makes the determination of weaker storms relatively more difficult. Therefore, we can expect the results in the analyses to inherit a certain amount of uncertainty, although computing cluster composite can offset the impact of the uncertainty in storm position, as random errors tend to average out and reduce their influence on the composite structure. In addition, a subgrid-scale feature of a pre-existing disturbance cannot be resolved if such a feature is smaller than the grid size of the reanalysis dataset. It can also be another source of uncertainty in the present study. On the other hand, we did not include non-developing disturbances in our analyses for consistency. Although non-developing disturbances are archived in the best-track dataset, they only constitute a small subset of all possible candidates. Therefore, our results do not reflect the general properties of all possible TC precursors.

From the node composite of relative vorticity and temperature

anomaly, it is typical for pre-existing disturbances to exhibit a certain amount of relative vorticity in the lower troposphere (Fig. 4), a warm anomaly in the upper-to-mid troposphere, and a cold anomaly in the lower troposphere (Fig. S4). This information suggests that these variables can be utilized to help identify the presence of pre-existing disturbances objectively in gridded datasets.

Our analysis identifies four clusters of distinct vortex structures that align closely to the vortex structures noted in the bottom-up and top-down paradigms. However, because our study focuses on snapshots of pre-existing disturbances rather than their temporal trajectories, we do not examine the processes by which a pre-existing disturbance transitions along a top-down or bottom-up pathway to tropical cyclone formation. For example, our findings cannot answer whether a mid-level vortex is the pre-requisite for tropical cyclogenesis. Consequently, while our findings confirm that all types of vortex structures noted in these theories are present in the western North Pacific, they cannot validate the proposed mechanisms. Relating our findings to these paradigms for tropical cyclogenesis requires a follow-up study that examines the transition and maintenance among vortex structures of pre-existing disturbances.

While our study focuses on vortex structure and associated properties of pre-existing disturbances, their evolution during intensification warrants further investigation. In particular, examining the transition and maintenance among vortex structures of pre-existing disturbances can enhance our understanding of the developmental pathway leading to the eventual completion of genesis. To translate these findings into improvement for tropical cyclogenesis, future work must apply the same classification method to a broader dataset that includes non-developing pre-existing disturbances to quantify which vortex structures and developmental pathways differentiate systems that do or do not undergo tropical cyclogenesis. Despite the uncertainties stemming from the reanalysis and best-track datasets, the classification resulting from the present study offers a new perspective. It can be effectively applied for studying the evolution of vortex structure and shed light on the tropical cyclogenesis process.

CRediT authorship contribution statement

Hung Ming Cheung: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Jung-Eun Chu:** Writing – review & editing, Visualization, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Hong Kong Research Grants Council Early Career Scheme (9048309).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wace.2025.100804.

Data availability

The best track data of the IBTrACS are available at https://www.ncei.noaa.gov/products/international-best-track-archive. The ERA5 reanalysis dataset can be downloaded at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels and https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels. CMORPH dataset can be obtained at https://www.ncei.noaa.gov/products/climate-data-records/precipitation-cmorph. GridSat-B1 dataset is retrieved from https://www.ncei.noaa.gov/thredds/catalog/cdr/gridsat/catalog.html.

References

Alpaydin, E., 2020. Introduction to machine learning, fourth. Adaptive Computation and Machine Learning Series. MIT Press, Cambridge, Massachusetts.
Bartels, D.L., Maddox, R.A., 1991. Midlevel cyclonic vortices generated by mesoscale convective systems. Mon. Weather Rev. 119, 104–118. https://doi.org/10.1175/1520-0493(1991)119<0104:MCVGBM>2.0.CO;2.
Bister, M., Emanuel, K.A., 1997. The genesis of Hurricane Guillermo: TEXMEX analyses and a modeling study. Mon. Weather Rev. 125, 2662–2682. https://doi.org/10.1175/1520-0493(1997)125<2662:TGOHGT>2.0.CO;2.
Brentan, B., Meirelles, G., Luvizotto, E., Izquierdo, J., 2018. Hybrid SOM+*k*-Means clustering to improve planning, operation and management in water distribution systems. Environ. Model. Software 106, 77–88. https://doi.org/10.1016/j.envsoft.2018.02.013. Special Issue on Environmental Data Science. Applications to Air quality and Water cycle.
Chang, M., Ho, C., Park, M., Kim, J., Ahn, M., 2017. Multiday evolution of convective bursts during Western North Pacific tropical cyclone development and nondevelopment using geostationary satellite measurements. J. Geophys. Res. Atmos. 122, 1635–1649. https://doi.org/10.1002/2016JD025535.
Chen, X., Rozoff, C.M., Rogers, R.F., Corbosiero, K.L., Tao, D., Gu, J.-F., Judt, F., Hendricks, E.A., Wang, Y., Bell, M.M., Stern, D.P., Musgrave, K.D., Knaff, J.A., Kaplan, J., 2023. Research advances on internal processes affecting tropical cyclone intensity change from 2018–2022. Trop. Cycl. Res. Rev. 12, 10–29. https://doi.org/10.1016/j.tcrr.2023.05.001.
Chu, J.-E., Hameed, S.N., Ha, K.-J., 2012. Nonlinear, intraseasonal phases of the East Asian summer monsoon: extraction and analysis using self-organizing maps. J. Clim. 25, 6975–6988. https://doi.org/10.1175/JCLI-D-11-00512.1.
DesRosiers, A.J., Bell, M.M., Klotzbach, P.J., Fischer, M.S., Reasor, P.D., 2023. Observed relationships between tropical cyclone vortex height, intensity, and intensification rate. Geophys. Res. Lett. 50, e2022GL101877. https://doi.org/10.1029/2022GL101877.
Ditchek, S.D., Nelson, T.C., Rosenmayer, M., Corbosiero, K.L., 2017. The relationship between tropical cyclones at genesis and their maximum attained intensity. J. Clim. 30, 4897–4913. https://doi.org/10.1175/JCLI-D-16-0554.1.
Francis, A.S., Strahl, B.R., 2021. Annual tropical cyclone report 2020. Joint Typhoon Warning Center.
Fu, B., Peng, M.S., Li, T., Stevens, D.E., 2012. Developing versus nondeveloping disturbances for tropical cyclone formation. Part II: western North Pacific. Mon. Weather Rev. 140, 1067–1080. https://doi.org/10.1175/2011MWR3618.1.
Ge, X., Li, T., Peng, M.S., 2013. Tropical cyclone genesis efficiency: mid-level versus bottom vortex. J. Trop. Meteorol. 19, 197–213.
Gray, W.M., 1998. The formation of tropical cyclones. Meteorol. Atmos. Phys. 67, 37–69. https://doi.org/10.1007/BF01277501.
Gray, W.M., 1968. Global view of the origin of tropical disturbances and storms. Mon. Weather Rev. 96, 669–700. https://doi.org/10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2.
Gu, Y., Wu, L., Zhan, R., Wen, Z., 2022. Climatology of developing and nondeveloping disturbances for tropical cyclone genesis over the Western North Pacific. Terr. Atmos. Ocean Sci. 33, 13. https://doi.org/10.1007/s44195-022-00012-4.
Hartmann, D.L., Hendon, H.H., Houze, R.A., 1984. Some implications of the mesoscale circulations in tropical cloud clusters for large-scale dynamics and climate. J. Atmos. Sci. 41, 113–121. https://doi.org/10.1175/1520-0469(1984)041<0113:SIOTMC>2.0.CO;2.
Hendricks, E.A., 2012. Internal dynamical control on tropical cyclone intensity variability – a review. Trop. Cycl. Res. Rev. 1, 97–105. https://doi.org/10.6057/2012TCRR01.11.
Hendricks, E.A., Montgomery, M.T., Davis, C.A., 2004. The role of "vortical" hot towers in the formation of tropical cyclone diana (1984). J. Atmos. Sci. 61, 1209–1232. https://doi.org/10.1175/1520-0469(2004)061<1209:TROVHT>2.0.CO;2.
Hennon, C.C., Papin, P.P., Zarzar, C.M., Michael, J.R., Caudill, J.A., Douglas, C.R., Groetsema, W.C., Lacy, J.H., Maye, Z.D., Reid, J.L., Scales, M.A., Talley, M.D., Helms, C.N., 2013. Tropical cloud cluster climatology, variability, and genesis productivity. J. Clim. 26, 3046–3066. https://doi.org/10.1175/JCLI-D-12-00387.1.
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803.
Hewitson, B., Crane, R., 2002. Self-organizing maps: applications to synoptic climatology. Climate Res 22, 13–26. https://doi.org/10.3354/cr022013.
Houze, R.A., 1982. Cloud clusters and large-scale vertical motions in the tropics. J. Meteorol. Soc. Jpn. 60, 396–410. https://doi.org/10.2151/jmsj1965.60.1_396.
Hsieh, T.-L., Vecchi, G.A., Yang, W., Held, I.M., Garner, S.T., 2020. Large-scale control on the frequency of tropical cyclones and seeds: a consistent relationship across a

hierarchy of global atmospheric models. Clim. Dyn. 55, 3177–3196. https://doi.org/10.1007/s00382-020-05446-5.
Ikehata, K., Satoh, M., 2021. Climatology of tropical cyclone seed frequency and survival rate in tropical cyclones. Geophys. Res. Lett. 48. https://doi.org/10.1029/2021GL093626.
Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. J. Hydrometeorol. 5, 487–503.
Kerns, B., Greene, K., Zipser, E., 2008. Four years of tropical ERA-40 vorticity maxima tracks. Part I: climatology and vertical vorticity structure. Mon. Weather Rev. 136, 4301–4319. https://doi.org/10.1175/2008MWR2390.1.
Kerns, B., Zipser, E., 2009. Four years of tropical ERA-40 vorticity maxima tracks. Part II: differences between developing and nondeveloping disturbances. Mon. Weather Rev. 137 (1), 2576–2591. https://doi.org/10.1175/2008MWR2545.
Kerns, B.W., Chen, S.S., 2013. Cloud clusters and tropical cyclogenesis: developing and nondeveloping systems and their large-scale environment. Mon. Weather Rev. 141, 192–210. https://doi.org/10.1175/MWR-D-11-00239.1.
Kilroy, G., Smith, R.K., 2017. The effects of initial vortex size on tropical cyclogenesis and intensification. Q. J. R. Meteorol. Soc. 143, 2832–2845. https://doi.org/10.1002/qj.3134.
Kilroy, G., Smith, R.K., Montgomery, M.T., 2018. The role of heating and cooling associated with ice processes on tropical cyclogenesis and intensification. Q. J. R. Meteorol. Soc. 144, 99–114. https://doi.org/10.1002/qj.3187.
Kilroy, G., Smith, R.K., Montgomery, M.T., 2017. A unified view of tropical cyclogenesis and intensification. Q. J. R. Meteorol. Soc. 143, 450–462. https://doi.org/10.1002/qj.2934.
Knapp, K.R., Ansari, S., Bain, C.L., Bourassa, M.A., Dickinson, M.J., Funk, C., Helms, C. N., Hennon, C.C., Holmes, C.D., Huffman, G.J., Kossin, J.P., Lee, H.-T., Loew, A., Magnusdottir, G., 2011. Globally gridded satellite observations for climate studies. Bull. Am. Meteorol. Soc. 92, 893–907. https://doi.org/10.1175/2011BAMS3039.1.
Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J., Neumann, C.J., 2010. The international best track archive for climate stewardship (IBTrACS): unifying tropical cyclone data. Bull. Am. Meteorol. Soc. 91, 363–376. https://doi.org/10.1175/2009BAMS2755.1.
Kohonen, T., 1990. The self-organizing map. Proc. IEEE 78, 1464–1480. https://doi.org/10.1109/5.58325.
Kushwaha, P., Sukhatme, J., Nanjundiah, R., 2021. A global tropical survey of mid-tropospheric cyclones. Mon. Weather Rev. https://doi.org/10.1175/MWR-D-20-0222.1.
Lee, C.S., 1989. Observational analysis of tropical cyclogenesis in the Western North Pacific. Part I: structural evolution of cloud clusters. J. Atmos. Sci. 46, 2580–2598. https://doi.org/10.1175/1520-0469(1989)046<2580:OAOTCI>2.0.CO;2.
Li, Y., Wu, Y., Lin, F., 2009. Research on customer segmentation based on a two-stage SOM clustering algorithm. In: 2009 International Conference on Management and Service Science. Presented at the 2009. International Conference on Management and Service Science, pp. 1–4. https://doi.org/10.1109/ICMSS.2009.5302076.
Lussier III, L.L., Montgomery, M.T., Bell, M.M., 2014. The genesis of typhoon Nuri as observed during the tropical cyclone structure 2008 (TCS-08) field experiment – part 3: dynamics of low-level spin-up during the genesis. Atmos. Chem. Phys. 14, 8795–8812. https://doi.org/10.5194/acp-14-8795-2014.
Matsuoka, D., Nakano, M., Sugiyama, D., Uchida, S., 2018. Deep learning approach for detecting tropical cyclones and their precursors in the simulation by a cloud-resolving global nonhydrostatic atmospheric model. Prog. Earth Planet. Sci. 5, 80. https://doi.org/10.1186/s40645-018-0245-y.
McBride, J.L., 1981. Observational analysis of tropical cyclone formation. Part I: basic description of data sets. J. Atmos. Sci. 38, 1117–1131. https://doi.org/10.1175/1520-0469(1981)038<1117:OAOTCF>2.0.CO;2.
McBride, J.L., Zehr, R., 1981. Observational analysis of tropical cyclone formation. Part II: comparison of non-developing versus developing systems. J. Atmos. Sci. 38, 1132–1151. https://doi.org/10.1175/1520-0469(1981)038<1132:OAOTCF>2.0.CO;2.
Montgomery, M.T., Nicholls, M.E., Cram, T.A., Saunders, A.B., 2006. A vortical hot tower route to tropical cyclogenesis. J. Atmos. Sci. 63, 355–386. https://doi.org/10.1175/JAS3604.1.
Murthy, V.S., Boos, W.R., 2019. Understanding the vertical structure of potential vorticity in tropical depressions. Q. J. R. Meteorol. Soc. 145, 1968–1991. https://doi.org/10.1002/qj.3539.
Qian, J., Nguyen, N.P., Oya, Y., Kikugawa, G., Okabe, T., Huang, Y., Ohuchi, F.S., 2019. Introducing self-organized maps (SOM) as a visualization tool for materials research and education. Results Mater. 4, 100020. https://doi.org/10.1016/j.rinma.2019.100020.
Ritchie, E.A., Holland, G.J., 1999. Large-scale patterns associated with tropical cyclogenesis in the Western Pacific. Mon. Weather Rev. 127, 2027–2043. https://doi.org/10.1175/1520-0493(1999)127<2027:LSPAWT>2.0.CO;2.
Ritchie, E.A., Holland, G.J., 1997. Scale interactions during the formation of Typhoon Irving. Mon. Weather Rev. 125, 1377–1396. https://doi.org/10.1175/1520-0493(1997)125<1377:SIDTFO>2.0.CO;2.
Rogers, R.F., 2021. Recent advances in our understanding of tropical cyclone intensity change processes from airborne observations. Atmosphere 12, 650. https://doi.org/10.3390/atmos12050650.
Sammon, J.W., 1969. A nonlinear mapping for data structure analysis. IEEE Trans. Comput. C–18, 401–409. https://doi.org/10.1109/T-C.1969.222678.
Sheridan, S.C., Lee, C.C., 2011. The self-organizing map in synoptic climatological research. Prog. Phys. Geogr. 35, 109–119. https://doi.org/10.1177/0309133310397582.
Simpson, J., Ritchie, E., Holland, G.J., Halverson, J., Stewart, S., 1997. Mesoscale interactions in tropical cyclone genesis. Mon. Weather Rev. 125, 2643–2661. https://doi.org/10.1175/1520-0493(1997)125<2643:MIITCG>2.0.CO;2.
Sobel, A.H., Wing, A.A., Camargo, S.J., Patricola, C.M., Vecchi, G.A., Lee, C., Tippett, M. K., 2021. Tropical cyclone frequency. Earths Future 9. https://doi.org/10.1029/2021EF002275.
Sugi, M., Yamada, Y., Yoshida, K., Mizuta, R., Nakano, M., Kodama, C., Satoh, M., 2020. Future changes in the global frequency of tropical cyclone seeds. SOLA 16, 70–74. https://doi.org/10.2151/sola.2020-012.
Teng, H., Lee, C., Hsu, H., 2014. Influence of ENSO on formation of tropical cloud clusters and their development into tropical cyclones in the Western North Pacific. Geophys. Res. Lett. 41, 9120–9126. https://doi.org/10.1002/2014GL061823.
Teng, H.-F., Lee, C.-S., Hsu, H.-H., Done, J.M., Holland, G.J., 2019. Tropical cloud cluster environments and their importance for tropical cyclone formation. J. Clim. 32, 4069–4088. https://doi.org/10.1175/JCLI-D-18-0679.1.
Torn, R.D., Snyder, C., 2012. Uncertainty of tropical cyclone best-track information. Weather Forecast. 27, 715–729. https://doi.org/10.1175/WAF-D-11-00085.1.
Van Laerhoven, K., 2001. Combining the self-organizing map and K-Means clustering for On-Line classification of sensor data. In: Dorffner, G., Bischof, H., Hornik, K. (Eds.), Artificial Neural Networks — ICANN 2001. Springer, Berlin, Heidelberg, pp. 464–469. https://doi.org/10.1007/3-540-44668-0_65.
Vecchi, G.A., Delworth, T.L., Murakami, H., Underwood, S.D., Wittenberg, A.T., Zeng, F., Zhang, W., Baldwin, J.W., Bhatia, K.T., Cooke, W., He, J., Kapnick, S.B., Knutson, T. R., Villarini, G., van der Wiel, K., Anderson, W., Balaji, V., Chen, J., Dixon, K.W., Gudgel, R., Harris, L.M., Jia, L., Johnson, N.C., Lin, S.-J., Liu, M., Ng, C.H.J., Rosati, A., Smith, J.A., Yang, X., 2019. Tropical cyclone sensitivities to CO2 doubling: roles of atmospheric resolution, synoptic variability and background climate changes. Clim. Dyn. 53, 5999–6033. https://doi.org/10.1007/s00382-019-04913-y.
Wang, Z., Montgomery, M.T., Dunkerton, T.J., 2010. Genesis of pre–hurricane felix (2007). Part II: warm core formation, precipitation evolution, and predictability. J. Atmos. Sci. 67, 1730–1744. https://doi.org/10.1175/2010JAS3435.1.
Wilks, D.S., 2019. Statistical forecasting. In: Wilks, D.S. (Ed.), Statistical Methods in the Atmospheric Sciences, fourth ed. Elsevier, pp. 235–312. https://doi.org/10.1016/B978-0-12-815823-4.00007-9.
Wu, S., Fang, J., 2019. The evolution and role of midtropospheric cyclonic vortex in the formation of super typhoon nepartak (2016). J. Geophys. Res. Atmos. 124, 9277–9298. https://doi.org/10.1029/2019JD030631.
Yamada, Y., Kodama, C., Satoh, M., Sugi, M., Roberts, M.J., Mizuta, R., Noda, A.T., Nasuno, T., Nakano, M., Vidale, P.L., 2021. Evaluation of the contribution of tropical cyclone seeds to changes in tropical cyclone frequency due to global warming in high-resolution multi-model ensemble simulations. Prog. Earth Planet. Sci. 8, 11. https://doi.org/10.1186/s40645-020-00397-1.
Yang, W.-C., Lai, J.-P., Liu, Y.-H., Lin, Y.-L., Hou, H.-P., Pai, P.-F., 2024. Using medical data and clustering techniques for a smart healthcare system. Electronics 13, 140. https://doi.org/10.3390/electronics13010140.
Zarzycki, C.M., 2022. Sowing storms: how model timestep can control tropical cyclone frequency in a GCM. J. Adv. Model. Earth Syst. 14, e2021MS002791. https://doi.org/10.1029/2021MS002791.
Zawislak, J., 2020. Global survey of precipitation properties observed during tropical cyclogenesis and their differences compared to nondeveloping disturbances. Mon. Weather Rev. 148, 1585–1606. https://doi.org/10.1175/MWR-D-18-0407.1.
Zawislak, J., Zipser, E.J., 2010. Observations of seven African easterly waves in the East Atlantic during 2006. J. Atmos. Sci. 67, 26–43. https://doi.org/10.1175/2009JAS3118.1.
Zehr, R., 1976. Tropical Disturbance Intensification. Colorado State University, Fort Collins, Colorado.
Zehr, R.M., 1992. Tropical Cyclogenesis in the Western North Pacific (No. Technical Report NESDIS 61). National Oceanic and Atmospheric Administration.