https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

Earth System Science Data
Open Access Discussions

An ERA5-derived mesovortex tracking framework for investigating tropical cyclogenesis

Jianqiao Fan[1], Liguang Wu[1,3], Haikun Zhao[2]， Zhenyuan Dong[1]

[1]Dept.of Atmospheric and Oceanic Sciences and Institute of Atmospheric Sciences, Fudan University, Shanghai, China

[2]Key Laboratory of Meteorological Disaster, Nanjing University of Information Science and Technology, Nanjing, China

[3]CMA-FDU Joint Laboratory of Marine Meteorology, Shanghai 200438, China

May 22, 2026

Submitted to *Earth System Science Data*

Corresponding author address: Dr. Liguang Wu
Dept. of Atmospheric and Oceanic Sciences &Institute of Atmospheric Sciences
Fudan University, Shanghai 200438, China
E-mail: liguangwu@fudan.edu.cn

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

Open Access Earth System Science Data Discussions

Abstract

Previous studies have suggested that the vertical mesovortex couple plays a critical role in tropical cyclogenesis. This couple consists of midlevel and low-level mesoscale vortices, whose formation is believed to result from different physical mechanisms. However, existing tropical cyclone datasets primarily document post-formation stages, and a substantial gap remains between the well-observed life cycles of mature tropical cyclones and the poorly documented mesoscale processes preceding formation. To address this gap, a flexible, objective detection and tracking framework was developed based on ERA5 reanalysis, termed the Mesovortex Analysis of Structure and Tracking (MAST), and a corresponding multi-decadal dataset was constructed for the western North Pacific (WNP) basin covering 1984–2023. Technical validation shows that MAST successfully detects 96% of the mesovortex couple associated with observed tropical cyclogenesis events at the formation time, with more than 50% of cases identifiable as early as 48 h prior to formation. A comparative analysis further reveals systematic differences in the vertical tilt evolution of the mesovortex couple: genesis cases exhibit a steady reduction in tilt with time, whereas non-genesis cases maintain a persistently large tilt of approximately 200 km. The MAST framework extends the long-term record of mesoscale vortex evolution during the pre-formation stage and provides a valuable resource for understanding tropical cyclogenesis, numerical model evaluation, and machine-learning-based genesis prediction.

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

1. Introduction

While large-scale conditions for tropical cyclogenesis (TCG) have long been established (Gray, 1968, 1998), TCG remains one of the least understood processes in atmospheric science and continues to pose a major challenge for prediction (Liang et al., 2021; Yamaguchi et al., 2015). TCG fundamentally involves multiscale interactions (Fang and Zhang, 2010; Simpson et al., 1997), and the final transition to a self-sustaining tropical cyclone (TC) is critically dependent upon mesoscale processes that concentrate vorticity and organize deep convection (Bister and Emanuel, 1997; Gray, 1968; Hendricks et al., 2004; Nam et al., 2023; Simpson et al., 1997). Accumulating evidence suggests that TCG is closely associated with the formation and evolution of a vertical mesovortex couple that consists of mesoscale vortices at low- and mid- levels (Nam et al., 2023; Wu and Fang, 2023).

Observational analyses and numerical simulations indicate two key mesoscale structures: the mid-level vortex (MV) and the low-level vortex (LV). The MV, which is often termed a mesoscale convective vortex (MCV), typically forms within the stratiform precipitation regions of larger mesoscale convective systems (MCSs). Numerous studies have demonstrated that an MV frequently serves as a precursor to TCG (Bister and Emanuel, 1997; Hendricks et al., 2004; Montgomery et al., 2006). It plays an essential role in dynamically and thermodynamically modifying the inner-core environment, thereby facilitating a transition toward bottom-heavy convection (Bell and Montgomery, 2019; Montgomery et al., 2006; Raymond et al., 2007; Teng and Tang, 2025).

The development of the LV is also a critical step towards TCG. It has been argued that the fundamental spin-up mechanism is the concentration of vorticity through mass convergence in the lower troposphere, often driven by deep convection or vortical hot towers (VHTs; Fang and Zhang, 2010; Hendricks et al., 2004; Montgomery et al., 2006; Montgomery and Enagonio, 1998). This process can lead to the formation and aggregation of LVs, moistening the mid-troposphere and facilitating the development of further deep convection (Montgomery et al., 2006; Nolan, 2007; Wang et al., 2012).

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

However, substantial uncertainty remains regarding the relative importance of the MV and LV, their precise formation timing, and the persistence of their spatial relationship during TCG. While some numerical studies suggest that environments characterized by a strong pre-existing LV exhibit higher genesis efficiency (Ge et al., 2013), others demonstrate that external modulations, such as the diurnal radiative cycle, can significantly reshape the relative evolution of the two vortices (Teng and Tang, 2025). In addition, statistical analyses of initial vortex structures reveal pronounced regional variability. For example, a recent study over the western North Pacific (WNP) indicates that nearly three-quarters of TCs originate from systems in which the LV is either the dominant initial feature or is already dynamically coupled with the MV (Wu and Fang, 2023). Despite the sequence of the formation of MV and LV, a central dynamical constraint on successful TC formation is the presence and alignment of the vertical mesovortex couple (Nam et al., 2023). Yet existing long-term datasets generally lack the explicit, continuous tracking of both MV and LV centers required to enable robust, large-sample statistical analyses of their structural evolution and coupling processes.

To address this gap, we developed a flexible mesoscale vortex detection and tracking framework, the Mesovortex Analysis of Structure and Tracking (MAST), and applied it to construct a new multi-decadal dataset based on the European Centre for Medium-Range Weather Forecasts (ECMWF) fifth-generation reanalysis, ERA5 (Hersbach et al., 2020). The resulting dataset provides the high spatial and temporal resolution required to detect, characterize, and continuously track multi-level mesoscale vortices throughout the TCG process.

MAST is designed as a highly customizable framework that allows users to flexibly adjust key detection parameters while simultaneously identifying vortex centers defined as extrema of the stream function currently at 850 and 500 hPa. This configuration enables the construction of continuous records of vertically coupled LV and MV pairs, thereby facilitating explicit and quantitative analyses of vertical tilt and their temporal evolution.

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

Unlike previous tracking approaches that primarily focus on mesoscale convective systems (MCSs; Feng et al., 2021; Hahn et al., 2025; Huang et al., 2018), MAST incorporates user-selectable connection strategies that robustly resolve vortex interaction, merger, and continuity. As a result, the derived dataset provides precise formation timing and detailed track evolution for both genesis and non-genesis vortex pairs. This capability opens two research avenues that have previously been constrained by data limitations: (1) systematic investigations of vertical alignment and tilt dynamics preceding successful TCG, and (2) statistical analyses of vortex formation timing, structural characteristics, and associated environmental conditions. The adaptable MAST framework and its accompanying dataset establish a valuable resource for advancing understanding of the mesoscale processes governing TCG, with broad applications such as numerical model evaluation and machine-learning–based genesis prediction.

The remainder of this paper is organized as follows. Section 2 describes the data sources and the algorithmic design of the MAST framework. Section 3 details the dataset structure and records. Section 4 presents the statistical results, technical validation, and potential applications.

2. Data and methods

LVs and MVs associated with TCG are identified using the ERA5 reanalysis for June–October during 1984–2023. The used wind data have a horizontal resolution of 0.25° × 0.25° and a temporal resolution of 3 hours. TC formation information in the WNP basin is obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) version 4 dataset (Knapp et al., 2010), defined as the first instance when the maximum sustained surface wind speed reaches or exceeds 35 knots. In this study, mesoscale vortices are detected at 500 hPa for the midlevel and 850 hPa for the low level.

This section describes the workflow of the MAST framework, from initial data preprocessing to the final matching with observed TC formation events. A schematic

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

overview of the MAST workflow is presented in Figure 1. The framework consists of five main steps: data preprocessing, mesoscale vortex identification, vortex tracking, pairing of LVs and MVs, and matching with observed TC formation events (Fig. 1). Each component of the workflow is described in detail below.

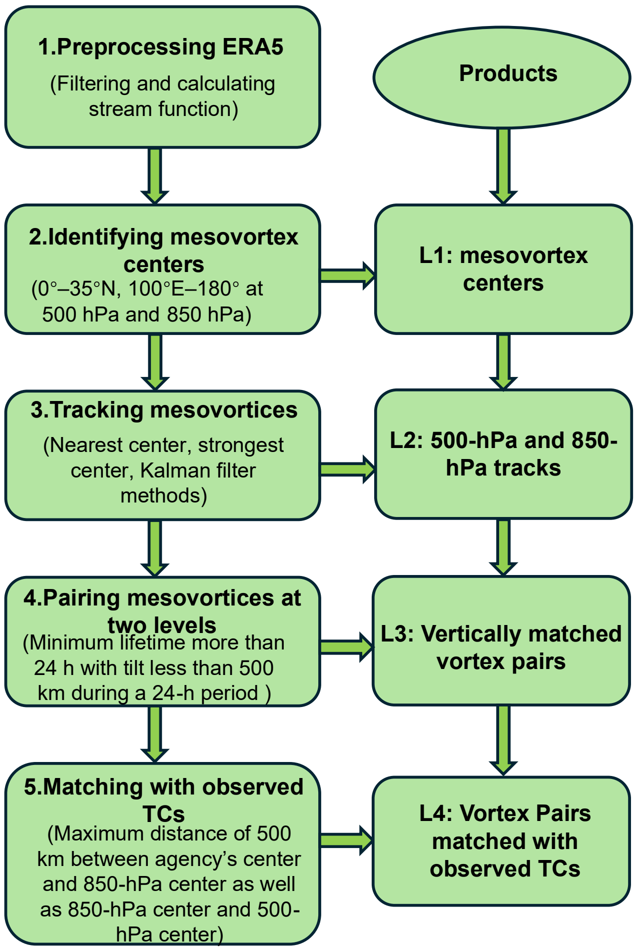

Figure 1: Schematic flowchart of the Mesovortex Analysis of Structure and Tracking (MAST) framework and the four levels of data products (L1–L4).

2.1 Data preprocessing

Although the critical role of the MV and LV in TCG has been well established, there is no universally accepted definition of their spatial scale, which varies considerably among previous studies. Observational and numerical analyses typically

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

identify mesoscale vortices with diameters of approximately 50–200 km (meso-β scale) (Houze et al., 2009; Montgomery et al., 2006; Reasor et al., 2005; Simpson et al., 1997; Teng and Tang, 2025), whereas other studies report larger vortices with diameters of about 200–800 km (meso-α scale) (Bell and Montgomery, 2019; Gray, 1998; Nam and Bell, 2021; Nolan, 2007; Ritchie and Holland, 1997). In this study, MVs and LVs are extracted using a Barnes filter with a 1000-km cutoff scale (Barnes, 1964; Doswell, 1977).

Figure 2 shows an example of the 500-hPa and 850-hPa mesoscale vortices extracted from the ERA5 reanalysis. The mesoscale vortices were associated with the formation of Super Typhoon Nepartak (2016). At 0000 UTC 1 July 2016, about 60 hours before the formation, the filtered fields revealed a nearly circular vortex with a diameter of approximately 400 km at 500 hPa (Fig. 2b) and another with a diameter of around 500 km at 850 hPa (Fig. 2d). In contrast, the unfiltered wind fields exhibited much weaker circulations and the MV at 500 hPa was barely discernible relative to the large-scale environmental flow (Fig. 2a) while no distinct vortex was evident at 850 hPa (Fig. 2c). This example demonstrates that the Barnes filter can effectively isolate mesoscale vortices associated with TCG from the large-scale background flow. The horizontal scales of these vortices are near the lower bound of the meso-α range, consistent with previous studies.

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

Open Access Earth System Science Data Discussions

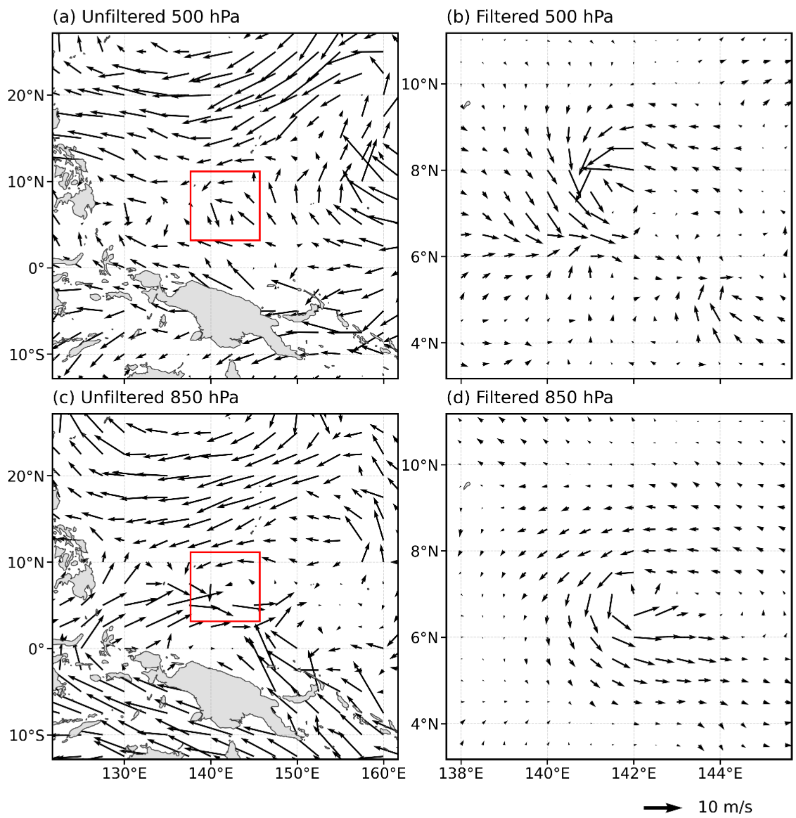

Figure 2: Mesoscale vortices during the genesis of Typhoon Nepartak (2016). (a, c) Unfiltered wind vectors at 500 hPa (a) and 850 hPa (c) at 0000 UTC on 1 July 2016. The red boxes (8° × 8°) denote the region of interest centered on the developing TC and correspond to the spatial domain shown in the right-hand panels. (b, d) Corresponding filtered wind fields at 500 hPa (b) and 850 hPa (d).

2.2 Identification of mesovortex centers

Using the filtered wind fields at 500 hPa and 850 hPa, the MV and LV centers are identified using the stream function within the domain 0°–35°N, 100°E–180°, according to the following criteria:

(1) The stream function value is less than -5×10^{5} m^2 s^{-1}, ensuring that the center is associated with a sufficiently strong cyclonic circulation.

(2) The grid point is a local minimum of the stream function within a 200 km radius, constraining the spatial scale of the vortex.

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

(3) The stream function values at the four surrounding grid points (1° east, west, north, and south) are all negative, ensuring a closed circulation.

The northern boundary of 35°N is chosen because TCG in the WNP basin predominantly occurs south of the latitude. The thresholds used here are our recommendations, 96 % of the TCs in the WNP basin could be detected using these thresholds at formation time. Users can adjust these parameters to suit specific research objectives.

Following the initial center identification, the exact center position of each mesovortex is determined through a three-step iterative cubic spline interpolation. The horizontal grid spacing is successively increased to 0.1°, 0.03°, and 0.01°. The first interpolation is performed within a 2° radius centered on each candidate. Before each subsequent iteration, the radius is reduced to half of its previous value (similar to Yang et al., 2020). After the final iteration, the 200 km separation criterion for local minima is re-applied to the refined set of candidates.

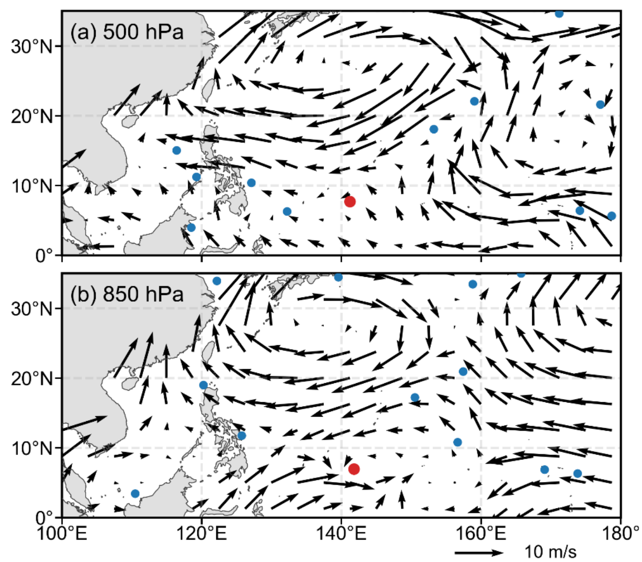

Figure 3: Spatial distribution of identified mesovortex centers (dots) and large-scale wind field (vectors) in the WNP basin at 0000 UTC on 1 July 2016 (a) at 500 hPa and (b) at 850 hPa. Blue dots denote identified mesovortex centers with the red dot for the precursor to Typhoon Nepartak (2016).

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

Figure 3 displayed the spatial distribution of identified mesovortices and the large-scale wind field by removing mesoscale-filtered components at 0000 UTC on 1 July 2016. Approximately 10 vortex centers were identified at both 500 hPa and 850 hPa levels. Notably, the precursor to Typhoon Nepartak (red dots) was clearly visible across the middle and lower levels approximately 60 hours prior to its genesis.

2.3 Mesovortex tracking

To obtain vortex trajectories, a spatial continuity condition is applied such that the displacement of a mesoscale vortex between two consecutive 3-hour intervals cannot exceed a specified maximum distance (d_{max}). In general, the movement of a continuously evolving mesoscale vortex does not exceed 200 km within three hours. However, given that mesoscale circulations often undergo deformation, the displacement between adjacent time steps may approach the characteristic vortex scale. Therefore, the maximum allowable displacement is set to 300 km.

To ensure that the tracked vortex center and any potential candidate at the next time step are within the same coherent cyclonic system, we implement a connectivity check during the tracking process. For each candidate connection, a straight line is constructed linking the established center at time t to the candidate center at time $t+\Delta t$. We sample 15 equidistant points along this connecting line, deriving their stream function values via bilinear interpolation. To account for the deformation of the vortex structure during the time interval, this check is performed using the stream function fields at both time steps. The candidate is accepted if the sampled points exhibit negative (cyclonic) stream function values at either time t or time $t+\Delta t$.

Note that multiple centers can be found in a larger mesoscale vortex. The resulting trajectory depends on how these centers are connected. Currently, three distinct connection methods are implemented within the MAST framework:

(1) Nearest-center method: The center nearest to the previous position is selected to connect the track.

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

(2) Strongest-center method: The center with the minimum stream function value (i.e., the strongest circulation) is chosen for connection.

(3) Kalman filter method: Unlike the above two methods, which consider only the current position, the Kalman filter predicts the expected vortex position at the next step based on its prior motion state. The observed center closest to the predicted position is then selected for connection. This approach has been widely used for tracking MCSs (Huang et al., 2018).

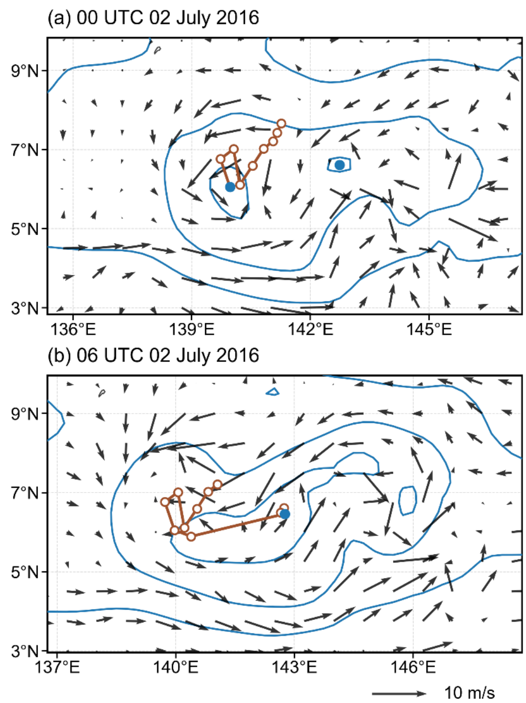

Figure 4: The filtered 500-hPa stream function (contours) and wind field (vectors) for the genesis of Typhoon Nepartak (2016). (a) 0000 UTC and (b) 0600 UTC on 2 July 2016. The current vortex center is marked by blue circles, and the brown line indicates the previous track connected with open circles at 3-hour intervals. The stream function contour interval is 5×10^5 m^2 s^{-1} and only non-positive values are displayed.

When multiple centers coexist, several distinct tracks may be identified. Figure 4 illustrates the merger of MVs during the genesis process of Typhoon Nepartak between

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

0000 UTC and 0600 UTC on 2 July 2016. The track shown in Fig. 4 was constructed using the nearest-center method. At 0000 UTC, the stream function and wind field revealed two centers at (6.05°N,139.97°E) and (6.61°N,142.74°E), respectively(Fig. 4a). By 0600 UTC, these centers merged into a single vortex center at (6.46°N, 142.76° E) (Fig. 4b). Since there is little change in the identified genesis process between $d_{\max} = 300$ km and $d_{\max} = 400$ km, we adopt $d_{\max} = 300$ km in this study. Users can alter $d_{\max}$ threshold according to their own purpose. A smaller $d_{\max}$ focuses on continuous movement, a larger $d_{\max}$ will capture more potential merger processes.

2.4 Pairing of MVs and LVs

Because TCG occurs exclusively over oceans, vortex centers located over land may be connected to existing tracks but are not permitted to initiate new tracks. Approximately 25 % of the tracks identified under the above thresholds consist of a single time step and are therefore excluded. In this study, both the 850-hPa and 500-hPa vortex tracks must have a minimum lifetime of 24 h. To qualify as a vertical mesovortex couple, the 850-hPa and 500-hPa vortex centers must maintain a horizontal separation (tilt) within 500 km continuously for at least 24 h. Also, users are allowed to change these thresholds for their own purpose. It is worth noting that while some of these paired vortices eventually developed into TCs, others did not. In the next step, we isolate the vortex pairs associated with the observed formation events from the full set of identified pairs.

2.5 Matching with observed formation

To match vertical mesovortex couples with the observed TCG event, a two-step procedure is employed. Using data from the IBTrACS dataset, only LV and MV tracks that persist until the formation time are considered.

Step 1: Matching LVs with TC formation positions. For each TCG event, all active 850-hPa vortex tracks at the formation time are examined. Among the vortices located within a 500-km radius of the official formation position, the one closest to the best-

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

track formation point is selected as the corresponding LV track. If no vortices are identified at formation time, a relaxed matching is carried out, allowing for a ±6-h time window. The spatial distance of 500 km is similar to the criterion of Liang et al. (2021), but the temporal restriction is much stricter (They use a time window of 24 hours).

Step 2: Matching MVs with identified LVs. Using the position of the identified LVs at the formation time as a reference, the algorithm then searches for MVs within a 500-km radius.

Once the vortex pair(s) at the match time are determined, a backward search can be performed to find all potential vortex pairs related to a particular TC case, including those that continue to the match time and those that terminated before the match time.

Figure 5 illustrates an example of the matching at the formation time of Typhoon Nepartak (2016). Following the above steps, the algorithm first matches LVs with the official TC formation position. Within a 500-km radius of the official TC formation position, the 850-hPa vortex track closest to the formation center was selected as the corresponding LV (Step 1). Subsequently, using this identified 850-hPa center as a reference, the algorithm searches for MVs within another 500-km radius (Step 2). In this case, a well-defined MV was successfully identified as it was located well within the 500-km search radius of the LV.

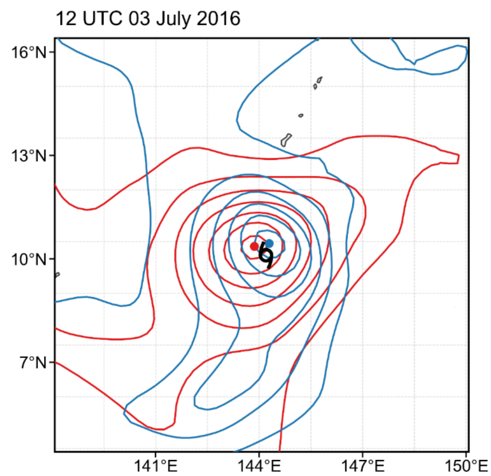

Figure 5: The filtered 850-hPa stream function (red contours) and 500 hPa stream function (blue contours) at 1200 UTC 03 July, the formation time of Typhoon Nepartak (2016). The 850-hPa vortex center (red circle) and 500-hPa vortex center (blue circle)

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

are also shown. The black TC symbol represents the position in IBTrACS. The contour interval is 5×10^5 m^2 s^{-1}, and only non-positive values are plotted.

After the above steps the LV and MV continuous track pairs that are associated with the observed TCs. Up to 96 % of the TC cases are matched with vertical mesovortex couples at formation time. Since possible track merge processes are taken into account during track connection, more than one mesovortex couples may be associated with a single formation case.

3. Dataset overview and descriptions

The data products are organized into four levels (Level 1 to Level 4), representing the sequential stages of the vortex identification, tracking, vertical coupling, and the events associated with the observed TCs (Fig. 1). All data files are provided in plain text (.txt and .csv) format for broad accessibility. The directory structure is strictly organized by basin, year, and the specific strategies used for connection and matching, ensuring full traceability of the workflow.

3.1 Level 1: Vortex detection and center refinement

This component contains the instantaneous locations of potential vortex centers identified from the ERA5 reanalysis dataset. It includes both the preliminary detection results and the refined coordinates after cubic spline interpolation.

- Location:
  - Preliminary: level1/[Basin]_vortices_[Parameters]_preliminary/[Year]/
  - Refined: level1/[Basin]_vortices_[Parameters]_refined/[Year]/
- File Structure: Files are organized by timestamp, named [YYYYMMDDHH]UTC_[Level]hPa[_refined].txt.
- File Format: Each row represents a detected vortex center with the following columns:
  - System_ID: Local index within the time step.
  - Lat: Latitude of the vortex center.
  - Lon: Longitude of the vortex center.

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

- SF_Value: Stream function intensity value.

3.2 Level 2: Single-level vortex tracks

This component provides the tracks of vortices at 850 hPa and 500 hPa, respectively, using the Level 1 product.

- Location: level2/[Basin]_tracks_[Parameters]/Strategy_[ID]/[Year]/
- File Structure: Each tracked system is stored as a unique file: system_[Year]_[Level]hPa_[ID].txt.
- Summary File: tracking_statistics_summary.csv provides annual statistics on track continuity and single-point systems.
- Track File Format: Each row corresponds to a time step along the track:
  - Year, Month, Day, Hour: Timestamp.
  - Lat, Lon: Vortex location.
  - SF_Value: Intensity.
  - Score: Matching score from the tracking algorithm.
  - Strategy: Connection method used.
  - Duration: Total duration of the system in hours.

3.3 Level 3: Vertically matched mesoscale vortex pairs

This level contains 850 hPa and 500 hPa vortex pairs that satisfy the vertical coupling criteria (minimum persistence of 24 hours with tilt ≤ 500 km during the interaction period). These pairs represent the full lifespans of the coupled systems.

- Location: level3/vertically_matched_systems_[Basin]_[Parameters]/strategy_[ID]/
- File Structure:
  - Summary: matched_summary_[Year].csv lists all identified pairs and their coupling duration.
  - Tracks: Subdirectories matched_tracks/[Year]/pair_[ID850] _[ID500]/ contain two files:
    1. track_850hPa_[ID].txt

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

   2. track_500hPa_[ID].txt

- Track File Format: Space-separated columns: Time (YYYY-MM-DD HH:MM:SS), Lat, Lon, SF.

3.4 Level 4: TC-associated mesoscale vortex pairs (Genesis cases)

This component filters the Level 3 pairs to retain only those associated with observed TC genesis events (based on IBTrACS). It introduces topological status flags (Continuous, Merger, Interrupted) to describe the vortex evolution relative to the TC.

- Location: level4/TC_matches_from_pairs_[Basin]/[Parameters]/connect[ID]/[Year]/[TC_ATCF_ID]/
- File Structure:
    - Network Flags: network_flags.csv describes the topological status of all nodes in the vortex network.
    - Tracks Directory: Contains individual track files for each associated pair, named Pair_[ID850] _[ID500] _[Level]hPa_[ID].txt.
- Track File Format: Each row includes status flags:
    - Time, Lat, Lon, SF.
    - Cont (Continuous), Merg (Merger), Intr (Interrupted): Binary flags (0/1) indicating the system's status at that time step.

4. Results

By comparing with previous studies, this section presents some statistical results and discussion for the MAST framework. All results shown here are based on the default thresholds defined in the previous section. It should be emphasized that these parameters can be flexibly adjusted by users within the MAST framework to accommodate specific research objectives or to conduct sensitivity tests. Unless otherwise specified, the representative tracks and statistical analyses (e.g., detection rates and tilt evolution) presented in this section are derived using the nearest-center connection method. Robustness tests using the strongest-center and Kalman filter methods yield highly similar statistical patterns, indicating that the core features of the

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

dataset discussed below are not sensitive to the choice of connection strategy. While adjusting these tracking thresholds and strategies may lead to minor quantitative variations, such modifications do not alter the fundamental physical insights regarding mesoscale vortex evolution during TCG.

4.1 Validation of mesosale processes in ERA5

We first demonstrate that the ERA5 reanalysis can effectively capture the evolution of LVs and MVs during TCG. Using the WRF model, Wu and Fang (2019) simulated the mesoscale circulations prior to the formation of Typhoon Nepartak (2016). Although their simulation was initialized with 0.25°×0.25° NCEP FNL analysis data at 6-hour intervals, the evolution of the LVs and MVs derived from the ERA5 reanalysis with our tracking framework exhibited strong consistency with their simulation.

Their simulation revealed that a LV first emerged at 0800 UTC on 30 June, followed by the appearance of a mid-level counterpart around 0000 UTC on 1 July. While the initial tilt between these vortices was about 150 km, it subsequently increased to 300 km by 1100 UTC on 1 July. Around 0000 UTC on 2 July, there were multiple MVs. The vertical tilt was reduced to less than 100 km by 1200 UTC on 2 July, leading to the formation of Typhoon Nepartak at 1200 UTC on 3 July.

In the ERA5 reanalysis, based on the tracking framework, the 850-hPa mesoscale vortex formed at 0000 UTC on 30 June (Fig. 6a), 8 hours earlier than in the simulation. The MV in ERA5 emerged at 0000 UTC on 1 July, consistent with the simulation. At this stage, the reanalysis indicated a tilt of 50–100 km between the MVs and LVs (Fig. 6b). Similar to the simulation, the tilt increased to approximately 300 km by 1200 UTC on 1 July. By 0000 UTC on 2 July (Fig. 6c), the ERA5 reanalysis also exhibited two MVs (Fig. 6d). Subsequently, the tilt decreased to less than 100 km around 1200 UTC on 2 July (Fig. 6e). We can see that the evolution of LVs and MVs during TCG has been effectively captured in the ERA5 reanalysis.

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

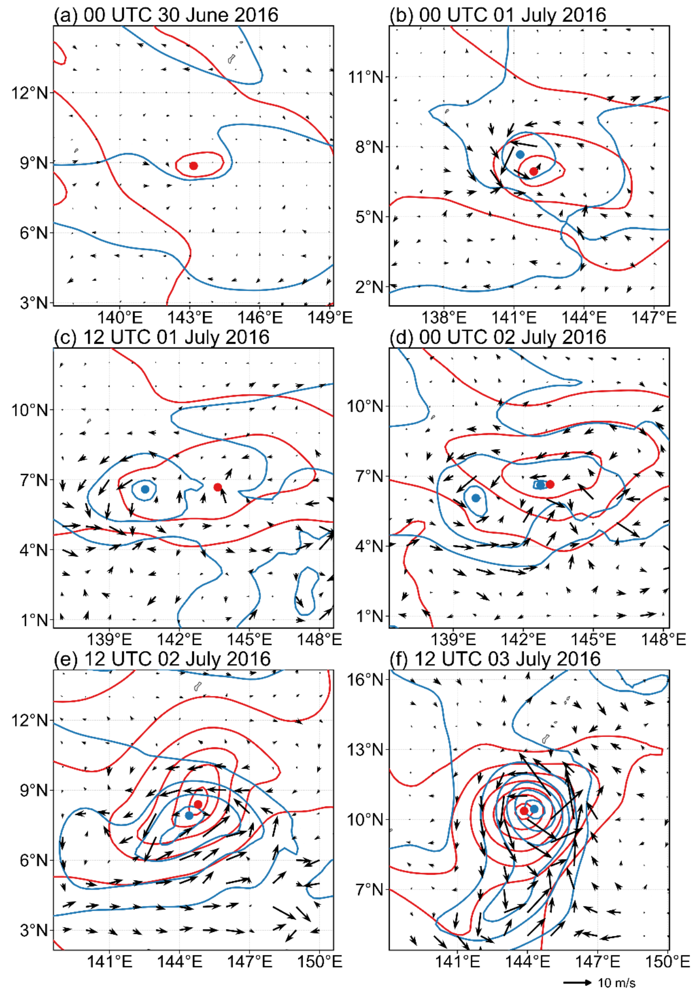

Figure 6: The evolution processes of the precursor mesoscale vortex for Typhoon Nepartak (2016). Panels show the evolution of the filtered stream function (contours), the identified vortex centers (dots) and the filtered 500-hPa wind field (vectors) at (a) 0000 UTC 30 June, (b) 0000 UTC 01 July, (c) 1200 UTC 01 July, (d) 0000 UTC 02 July, (e) 1200 UTC 02 July, and (f) 1200 UTC 03 July 2016. Red contours and the red circle represent the stream function and center at 850 hPa, while blue contours and the blue circle represent those at 500 hPa. The contour interval is 5×10^5 m^2 s^{-1} and only non-positive (cyclonic) values are plotted.

We further evaluated the framework by characterizing the development pathways of mesoscale vortices during TCG. Following Wu and Fang (2023), cases were categorized based on the time lag between the initial detection of LVs and MVs. Development is classified as “Simultaneous” if the two vortices form within 6 h, “Low-

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

level First" if the 850-hPa vortex precedes the 500-hPa vortex by more than 6 h, and "Mid-level First" for the reverse sequence. Consistent with Wu and Fang (2023), our tracking framework reveals that TCG over the WNP is dominated by the "Low-level First" pathway, which accounts for 59.3 % (451 cases) of all events. The "Mid-level First" and "Simultaneous" pathways account for 29.0 % (220 cases) and 11.7 % (89 cases), respectively.

4.2 Comparison of tilt evolution between genesis and non-genesis cases

Previous studies have identified sustained vertical vortex alignment as a critical bifurcation separating successful from failed tropical cyclogenesis (Nam et al., 2023). Figure 7 illustrates the temporal evolution of the vertical tilt between the mesoscale vortex centers at 500 and 850 hPa. For genesis cases, time is normalized relative to the formation time, whereas for non-genesis cases the sequence is referenced to the final time step when both LVs and MVs are simultaneously detected. In the genesis group (Fig. 7a), both the mean and median tilt decrease steadily with time, indicating progressive vertical alignment prior to cyclone formation. In sharp contrast, non-genesis cases (Fig. 7b) exhibit no systematic alignment trend; instead, the tilt remains persistently large, on the order of approximately 200 km, highlighting the failure to establish a coherent vertically aligned vortex.

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

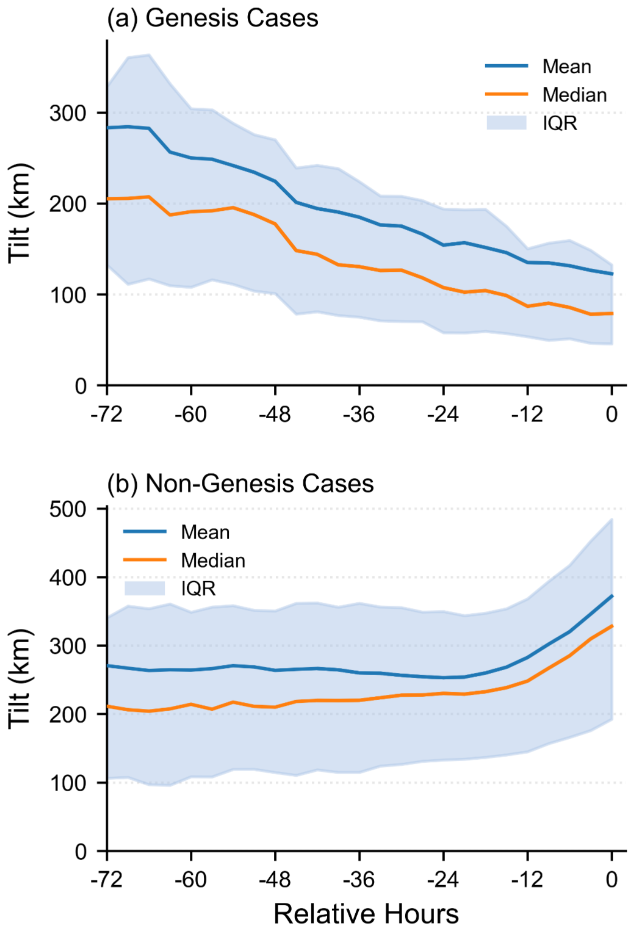

Figure 7: Comparison of tilt evolution between genesis (a) and non-genesis (b) cases. The light blue areas are the interquartile range (IQR) between 25th and 75th percentiles. The dark blue lines are average values, and orange lines are median values.

4.3 Basic performance of the framework

Figure 8 demonstrates the performance of the tracking framework by depicting the detection rate of TCs as a function of lead time prior to formation. Here, a case is considered successfully tracked when both LVs and MVs are simultaneously identified. At the formation time, 96 % of TC cases are successfully detected, and notably, approximately 50 % of the cases can be concurrently tracked as early as 48 h before formation. When the matching criteria are relaxed to allow a ±6-h window around the formation time, the tracking success rate further reaches 98 %. The small number of

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

remaining unmatched cases is mainly associated with systems near the domain boundaries (35°N and 180°) or with poorly defined vortex structures at the time of formation in the reanalysis dataset.

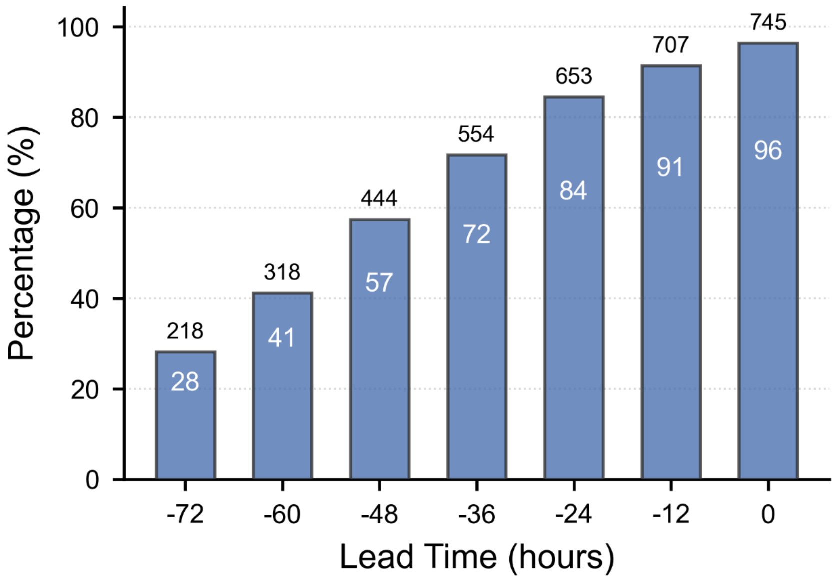

Figure 8: Percentage of TC cases detected by the MAST framework as a function of lead time relative to the TC formation time (0 h). The total number of detected cases is indicated above each bar, and the corresponding percentage is labeled inside the bar.

5. Conclusions and discussion

In this study, we developed the Mesovortex Analysis of Structure and Tracking (MAST) framework to address the long-standing data gap in documenting mesoscale vortex evolution during TCG. Applied to the ERA5 reanalysis over the western North Pacific from 1984 to 2023, the framework demonstrates high technical reliability, successfully detecting 96 % of observed TCs at their formation time and capturing over 50 % of genesis cases as early as 48 hours prior to formation. The multi-level tracking capability of MAST enables a quantitative characterization of the vertical tilt evolution, revealing a distinctive and steady reduction in tilt for developing systems compared to the persistently large tilt in non-genesis cases.

To facilitate the effective use of the MAST framework and the derived dataset, this

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

section provides additional information regarding algorithm limitations, strategy selection, data interpretation, and potential applications.

The current tracking algorithm (Level 2) tracks vortex mergers by allowing multiple tracks to connect to a single point at the subsequent time step. However, it does not explicitly account for vortex splitting processes, which may lead to track discontinuation in complex divergence scenarios. Furthermore, the 0.25° horizontal grid spacing of the ERA5 input data restricts the algorithm's detection capabilities primarily to meso-α scale systems (approximately 200–1000 km). The identification of smaller meso-β scale vortices may be limited by this resolution.

To offer flexibility for different research needs, the dataset provides outputs derived from three connection strategies (for single-level tracking). It is worth noting that while the statistical characteristics of the dataset (e.g., annual frequency, spatial distribution) remain largely consistent across different strategy combinations, significant deviations may occur in the trajectories of individual cases. Users should select the appropriate strategy based on their specific research objectives. For instance, the strongest-center method is suitable for tracking the evolution of the strongest core, while the Kalman filter method is preferable for kinematic analyses in complex flow.

The dataset was generated using specific threshold parameters which serve as recommended default values for mesoscale vortex studies: a minimum vortex lifespan of 24 hours and a maximum tilt of 500 km (during the core interaction period). Users running the provided source code may adjust these thresholds to suit specific sensitivity tests. In terms of data interpretation, users should note the strict definitions of "Genesis" and "Non-genesis" cases used in this dataset. "Genesis cases" refer exclusively to vortex pairs where both the low-level and mid-level vortices maintain continuity and successfully match with a confirmed TC at its formation time. In contrast, "non-genesis cases" are defined as vertically coupled vortex pairs where neither the low-level nor the mid-level vortex persists to match with any observed TC genesis event. It is important to emphasize that partial matches—where one vortex layer successfully persists to the TC formation time while the other does not—are explicitly excluded from both the

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

genesis and non-genesis categories. This deliberate exclusion is designed to filter out track fragments generated during the tracking of actual developing TCs. Classifying an incomplete vortex pair belonging to an eventual TC as a "non-genesis" event would be physically unreasonable and could skew statistical comparisons.

Finally, this dataset is primarily suitable for three key applications: (a) statistical analyses of the climatology of the TCG process; (b) analysis of individual genesis cases and the validation of numerical simulations; and (c) serving as a training dataset for artificial intelligence (AI) models.

Code and data availability

All code and data have been published in the form of txt files and are made publicly available through the Zenodo repository at https://doi.org/10.5281/zenodo.18145741 (Fan et al., 2026).

Author Contributions

LW designed the study and wrote the manuscript. JF performed data processing, designed the framework and drafted the manuscript. ZD joined the initial experiment of this study.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

This study is supported by the National Natural Science Foundation of China (No. 42192551).

References

Barnes, S. L.: A Technique for Maximizing Details in Numerical Weather Map Analysis,

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

J. Appl. Meteor., 3, 396–409, https://doi.org/10.1175/1520-0450(1964)003<0396:ATFMDI>2.0.CO;2, 1964.

Bell, M. M. and Montgomery, M. T.: Mesoscale Processes during the Genesis of Hurricane Karl (2010), J. Atmos. Sci., 76, 2235–2255, https://doi.org/10.1175/JAS-D-18-0161.1, 2019.

Bister, M. and Emanuel, K. A.: The Genesis of Hurricane Guillermo: TEXMEX Analyses and a Modeling Study, Mon. Wea. Rev., 125, 2662–2682, https://doi.org/10.1175/1520-0493(1997)125<2662:TGOHGT>2.0.CO;2, 1997.

Doswell, C. A.: Obtaining Meteorologically Significant Surface Divergence Fields Through the Filtering Property of Objective Analysis, Mon. Wea. Rev., 105, 885–892, https://doi.org/10.1175/1520-0493(1977)105<0885:OMSSDF>2.0.CO;2, 1977.

Fan, J., Wu, L., and Dong, Z.: Source code and benchmark dataset for an ERA5-derived mesovortex tracking framework for investigating tropical cyclogenesis, Zenodo [software and data set], https://doi.org/10.5281/zenodo.18145741, 2026.

Fang, J. and Zhang, F.: Initial Development and Genesis of Hurricane Dolly (2008), J. Atmos. Sci., 67, 655–672, https://doi.org/10.1175/2009JAS3115.1, 2010.

Feng, X., Klingaman, N. P., and Hodges, K. I.: Poleward migration of western North Pacific tropical cyclones related to changes in cyclone seasonality, Nat. Commun., 12, 6210, https://doi.org/10.1038/s41467-021-26369-7, 2021.

Ge, X., Li, T., and Peng, M. S.: Tropical cyclone genesis efficiency: Mid-level versus bottom vortex, J. Trop. Meteorol., 19, 197–213, https://doi.org/10.16555/j.1006-8775.2013.03.001, 2013.

Gray, W. M.: Global view of the origin of tropical disturbances and storms, Mon. Wea. Rev., 96, 669–700, https://doi.org/10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2, 1968.

Gray, W. M.: The formation of tropical cyclones, Meteorol. Atmos. Phys., 67, 37–69, https://doi.org/10.1007/BF01277501, 1998.

Hahn, T., Weiner, H., Brooks, C., Li, J. X., Gupta, S., and Wang, D.: CoCoMET v1.0:

A Unified Open-Source Toolkit for Atmospheric Object Tracking and Analysis, Geosci. Model Dev., 18, 5971–5996, https://doi.org/10.5194/gmd-18-5971-2025, 2025.

Hendricks, E. A., Montgomery, M. T., and Davis, C. A.: 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, 2004.

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., and Thépaut, J.: The ERA5 global reanalysis, Quart. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Houze, R. A., Lee, W.-C., and Bell, M. M.: Convective Contribution to the Genesis of Hurricane Ophelia (2005), Mon. Wea. Rev., 137, 2778–2800, https://doi.org/10.1175/2009MWR2727.1, 2009.

Huang, X., Hu, C., Huang, X., Chu, Y., Tseng, Y., Zhang, G. J., and Lin, Y.: A long-term tropical mesoscale convective systems dataset based on a novel objective automatic tracking algorithm, Clim. Dyn., 51, 3145–3159, https://doi.org/10.1007/s00382-018-4071-0, 2018.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J.: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data, Bull. Amer. Meteor. Soc., 91, 363–376, https://doi.org/10.1175/2009BAMS2755.1, 2010.

Liang, M., Chan, J. C. L., Xu, J., and Yamaguchi, M.: Numerical prediction of tropical cyclogenesis part I: Evaluation of model performance, Quart. J. Roy. Meteor. Soc.,

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

147, 1626–1641, https://doi.org/10.1002/qj.3987, 2021.

Montgomery, M. T. and Enagonio, J.: Tropical Cyclogenesis via Convectively Forced Vortex Rossby Waves in a Three-Dimensional Quasigeostrophic Model, J. Atmos. Sci., 55, 3176–3207, https://doi.org/10.1175/1520-0469(1998)055<3176:TCVCFV>2.0.CO;2, 1998.

Montgomery, M. T., Nicholls, M. E., Cram, T. A., and Saunders, A. B.: A Vortical Hot Tower Route to Tropical Cyclogenesis, J. Atmos. Sci., 63, 355–386, https://doi.org/10.1175/JAS3604.1, 2006.

Nam, C. C. and Bell, M. M.: Multiscale Shear Impacts during the Genesis of Hagupit (2008), Mon. Wea. Rev., 149, 551–569, https://doi.org/10.1175/MWR-D-20-0133.1, 2021.

Nam, C. C., Bell, M. M., and Tao, D.: Bifurcation Points for Tropical Cyclone Genesis and Intensification in Sheared and Dry Environments, J. Atmos. Sci., 80, 2239–2259, https://doi.org/10.1175/JAS-D-22-0100.1, 2023.

Nolan, D. S.: What is the trigger for tropical cyclogenesis, Aust. Meteorol. Mag., 56, 241–266, 2007.

Raymond, D. J., Sessions, S. L., and Fuchs, Ž.: A theory for the spinup of tropical depressions, Quart. J. Roy. Meteor. Soc., 133, 1743–1754, https://doi.org/10.1002/qj.125, 2007.

Reasor, P. D., Montgomery, M. T., and Bosart, L. F.: Mesoscale Observations of the Genesis of Hurricane Dolly (1996), J. Atmos. Sci., 62, 3151–3171, https://doi.org/10.1175/JAS3540.1, 2005.

Ritchie, E. A. and Holland, G. J.: Scale Interactions during the Formation of Typhoon Irving, Mon. Wea. Rev., 125, 1377–1396, https://doi.org/10.1175/1520-0493(1997)125<1377:SIDTFO>2.0.CO;2, 1997.

Simpson, J., Ritchie, E., Holland, G. J., Halverson, J., and Stewart, S.: Mesoscale Interactions in Tropical Cyclone Genesis, Mon. Wea. Rev., 125, 2643–2661, https://doi.org/10.1175/1520-0493(1997)125<2643:MIITCG>2.0.CO;2, 1997.

Teng, Y. and Tang, X.: Relative Importance of Mid-Level and Low-Level Vortices in

https://doi.org/10.5194/essd-2026-388
Preprint. Discussion started: 14 July 2026

Tropical Cyclogenesis Inferred From Experiments on Sensitivity to Radiation, J. Geophys. Res. Atmos., 130, e2024JD042153, https://doi.org/10.1029/2024JD042153, 2025.

Wang, Z., Dunkerton, T. J., and Montgomery, M. T.: Application of the Marsupial Paradigm to Tropical Cyclone Formation from Northwestward-Propagating Disturbances, Mon. Wea. Rev., 140, 66–76, https://doi.org/10.1175/2011MWR3604.1, 2012.

Wu, S. and Fang, J.: 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, 2019.

Wu, S. and Fang, J.: The Initial Mesoscale Vortexes Leading to the Formation of Tropical Cyclones in the Western North Pacific, Adv. Atmos. Sci., 40, 804–823, https://doi.org/10.1007/s00376-022-2029-y, 2023.

Yamaguchi, M., Vitart, F., Lang, S. T. K., Magnusson, L., Elsberry, R. L., Elliott, G., Kyouda, M., and Nakazawa, T.: Global Distribution of the Skill of Tropical Cyclone Activity Forecasts on Short- to Medium-Range Time Scales, Wea. Forecasting, 30, 1695–1709, https://doi.org/10.1175/WAF-D-14-00136.1, 2015.