



OPEN

CAN reveals rapid intensification features of spring cyclones over Mongolian plateau and Northeast China plain

Ruipeng Sun¹, Yina Diao²✉ & Jianping Li^{1,3}✉

Traditional composite analysis in cyclone studies, which compares variables at fixed grid points relative to the cyclone center, is limited by spatial misalignments caused by cyclone rotation. This misalignment blurs composite results and hinders accurate structural analysis. To address this issue, we propose the Cyclone Alignment Network (CAN) method. CAN aligns variables into a unified coordinate system by learning an affine transformation matrix, improving classification and composite results. Specifically designed for cyclones, CAN utilizes a Transformer structure with Rotary Position Embedding (RoPE) to effectively capture relative positional information, unlike typical Convolutional Neural Networks (CNNs). Its classification network, informed by cyclone development equations, concentrates coordinate transformation within the affine matrix. Our evaluation using a cyclone dataset shows that CAN-based composites outperform traditional methods, yielding more significant results and more coherent variable coupling. CAN reveals several key common features: (1) cyclone rapid intensification in spring is dominated by cold air activity; (2) topography significantly impacts intensification; and (3) downstream ridge structures potentially influence intensification by causing anomalous subsidence, which leads to low-level dynamic uplift and limited baroclinic energy release. CAN effectively analyzes cyclone circulation and structure.

Keywords Extra-tropical cyclones, Cyclone rapid intensification, Composite analysis, Machine learning

Extratropical cyclones are important synoptic-scale weather systems with impacts on global climate and weather; rapidly intensifying cyclones, in particular, can result in extreme storms and precipitation. Extratropical cyclones occurring frequently in winter and spring over Mongolia and Northeast China significantly impact northern China's weather. Statistical studies demonstrate a strong link between rapidly intensifying cyclones and spring storms. Long-lasting rapidly intensifying cyclones are a major cause of sandstorms in northern regions, especially amidst low precipitation and soil moisture^{1–4}. Beijing experienced its worst sandstorm in a decade on March 15, 2021, with PM10 levels exceeding 8,000 $\mu\text{g}/\text{m}^3$, which was associated with a rapidly intensifying cyclone in this region⁵.

Rapidly intensifying cyclones are typically defined as those exhibiting a maximum deepening rate over 1 Bergeron, or those with the upper percentiles of maximum deepening rates (e.g., the top 10% or top 5%). Previous studies on rapidly intensifying cyclones have largely focused on case studies, detailing their mechanisms. These studies have identified factors such as baroclinic instability^{6–11}, positive vorticity and temperature advections^{12–15}, latent heat release^{16–20}, superposition of upper- and lower-level vorticity maxima^{10,21–28}, downward propagation of upper-level momentum and dynamic forcing by upper-level jets²⁹, and tropopause folding^{22,30–32}. However, rapidly intensifying cyclones over land, particularly over Mongolia, are under-researched compared to oceanic cases. Some case studies suggest that the rapid development of rapidly intensifying cyclones over land may be related to jet enhancement, upper- and lower-level wind coupling, and the height of maximum heating rates^{19,33}. Baroclinic instability, anomalous distributions of tropospheric potential vorticity, downward propagation of upper-level potential vorticity, and interactions between tropopause folding and cold fronts are also important

¹Frontiers Science Center for Deep Ocean Multi-spheres and Earth System (DOMES), Key Laboratory of Physical Oceanography, Academy of Future Ocean, College of Oceanic and Atmospheric Sciences, Center for Ocean Carbon Neutrality, Ocean University of China, Qingdao 266100, China. ²College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao 266100, China. ³Laboratory for Ocean Dynamics and Climate, Qingdao Marine Science and Technology Center, Qingdao 266237, China. ✉email: diaoyin@ouc.edu.cn; ljip@ouc.edu.cn

mechanisms. The role of latent heat is considered limited in rapidly intensifying spring cyclones over land due to the scarcity of precipitation^{34–38}.

Individual case studies offer insights but are insufficient for generalized characteristics of rapid cyclogenesis; therefore, statistical analyses are essential. Although some statistical studies of rapid cyclogenesis in this region exist³⁹, a detailed analysis of their structural features remains lacking. Further investigation of their structural attributes is required to fully elucidate the mechanisms driving rapid intensification.

Traditional composite methods typically use the cyclone center as a reference to composite individual cyclones point-to-point on a latitude-longitude grid. In previous studies, composite analysis has been used to investigate the genesis, development, and decay of extratropical cyclones^{40–48}. In conjunction with subjective classification schemes, conceptual models of cyclone structure have been formulated, and the spatial distribution of meteorological variables along with the vertical coupling of upper- and lower-level features have been examined^{49–52}. Additionally, the distribution of associated cloud and precipitation and their potential future changes has been analyzed by incorporating model data^{53–61}.

However, the rotation of cyclones causes spatial misalignment of variables, particularly away from the cyclone center, blurring key mesoscale details such as frontal boundaries in traditional composite analyses⁶². This over-smoothing of spatial structures hinders the accurate representation of cyclone characteristics and limits the ability of traditional composite methods to reveal generalized features. While some studies have attempted rotational alignment using fronts^{36,37,63}, tracks^{33,64,65}, or other features^{66,67}, the complexity of cyclones means that single-feature alignment may mask other important information or introduce new biases.

To address the variable misalignment in traditional composite analysis, we propose a composite analysis method based on the Cyclone Alignment Network (CAN). Drawing inspiration from previous work on image registration^{68–73}, CAN aligns variables with physically analogous influences on cyclone activity by learning affine transformations, enhancing the effectiveness of compositing, clustering, and modal decomposition. In terms of network design, we adopt a Transformer structure with Rotary Position Embedding (RoPE)⁷⁴ to effectively capture relative positional information between variables, unlike commonly used CNNs. To concentrate coordinate transformation within the affine matrix and to integrate rapidly intensifying cyclone development mechanisms, we restricted the size, form, and depth of the convolutional layers in the classification network. This machine learning approach, guided by cyclone development equations, overcomes the limitations of traditional composite methods, thereby revealing key mechanisms of rapidly intensifying cyclogenesis and improving predictive capabilities, especially for rapidly intensifying cyclones over land. This study uses CAN and composite analysis to analyze rapidly intensifying spring cyclones over the Mongolian Plateau-Northeast China Plain, demonstrating CAN's advantages and providing new insights.

Results

Results on the MNIST dataset

To validate the effectiveness of the proposed method, we first conducted experiments using the classic MNIST⁷⁵ dataset (Modified National Institute of Standards and Technology database). Considering the wide use of CNNs in such problems, we constructed two networks: GAU-CAN (using GAUs, Gated Attention Units⁷⁶ and CNN-CAN (which replaced GAUs with CNNs). We evaluated the performance of GAUs by comparing the accuracy and clarity of the composite images generated by the two models on the MNIST classification task. To ensure fairness in the experiments, we adjusted the structure of CNN-CAN so that its number of parameters was the same as that of GAU-CAN (about 1.3 million). In this subsection, RAW represents the untransformed original images.

From the MNIST dataset, we randomly selected 3,000 samples per category, totaling 30,000 samples. Random rotations ranging from -15° to 15° were applied to these samples to construct the training set. Similarly, 2,000 samples were extracted for both the validation and test sets, ensuring no overlap among the three sets. The epoch with the highest accuracy on the validation set was selected as the final model. We averaged the raw/transformed images of each category in the test set to visualize the alignment effect of the affine transformation. The results are shown in Fig. 1. Both methods achieved an accuracy of nearly 99.4% on the test set.

Composite results of raw images (Fig. 1a) appear blurred, particularly at the edges, indicating a loss of detail when rotated images are directly composited. CNN-CAN (Fig. 1b) composites show improvement but still exhibit some edge blur. In contrast, GAU-CAN (Fig. 1c) composites display sharp digit contours, clear edges, and minimal blur. Figure 1d demonstrates GAU-CAN's effective correction of position and orientation for rotated digits, showcasing its feature alignment capability.

The experimental results show that the proposed CAN can effectively align images with rotational features, generating clearer composite images. Compared with CNNs, the GAU demonstrates stronger capabilities in handling complex transformations, making it more suitable for the alignment task presented in this paper.

Results on the cyclone dataset

Spatial distribution of cyclone activity and deepening rate distribution

To validate the effectiveness of CAN for compositing extratropical cyclones over the Mongolian Plateau and Northeast China Plain in boreal spring, we first constructed a dataset following the method described in Methods section. Figure 2a shows that the activity centers of these cyclones are primarily distributed on the leeward side of mountains, specifically to the east of the Altai, Tianshan, and Greater Khingan Mountains, as well as over the Mongolian Plateau and the Northeast China Plain. This suggests a close relationship between these cyclones and the leeward slope effect. To facilitate a classification task linking variable fields with deepening rates, samples were divided into six categories (Fig. 2b): weak, medium, and strong based on both positive and negative deepening rates, by equally dividing each into thirds.

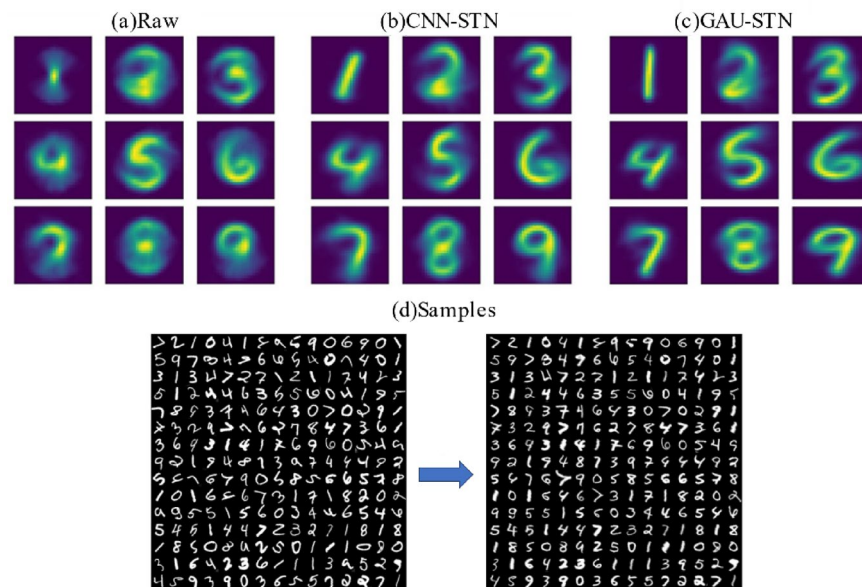


Fig. 1. The result of using a cyclone alignment network trained for distorted MNIST digit classification. **a** Average of the input data. The input to the cyclone alignment network are images of MNIST digit that are distorted with random rotation, to emulate rotation of cyclones. **b** Average of the output of the cyclone alignment network, after applying the transformation, but using CNN in transformer branch. **c** Average of the output of the cyclone alignment network, after applying the transformation, using GAU in transformer branch. **d** Some samples of inputs (left) and outputs (right) using GAU-CAN.

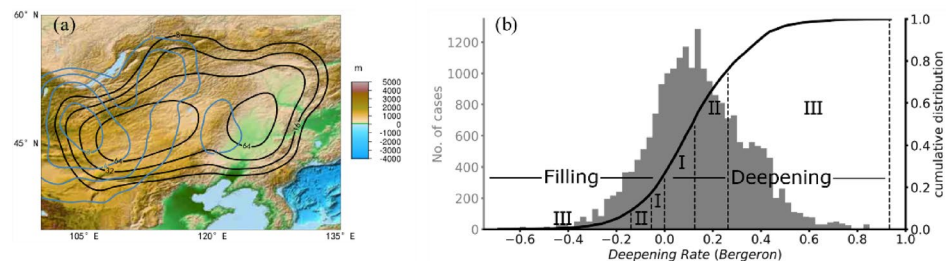


Fig. 2. Geographic locations and topography of Mongolia - Northeast China and distribution of 12-h deepening rate for all samples. **a** Geographic locations and topography of Mongolia - Northeast China (filled, unit: m), with data obtained from the ETOPO1, created using Python (version 3.9.5, <https://www.python.org/>) and the Matplotlib (version 3.4.2, <https://matplotlib.org/>) library with the 'GMT_globe' colormap (can be obtained from https://www.ncl.ucar.edu/Document/Graphics/ColorTables/GMT_globe.shtml or created manually.). Frequency (black contour, units: number per year), and genesis (blue contour, units: number per year) for all cyclones which reach their maximum deepening rate over Mongolia - Northeast China. **b** Distribution of 12-h deepening rate for all samples in the target domain (in Fig. 2a). The unit is number of samples. The three equal parts of deepened and filled samples are labeled, and this classification is done as the target of the CAN.

Results of classification

Since GAU showed better performance than CNN for this task, experiments on the classification task were conducted solely on GAU-CAN, hereinafter referred to as CAN. The experiment results are shown in Table 1. For distinction, the original variable fields are designated as 'raw fields', and those processed by the affine transformation branch are referred to as 'transformed fields'. For each sample x (raw field or transformed field), we calculate its similarity to the average feature μ_k of each category k in the training set. Here, category k is determined based on the true deepening rate, as shown in Fig. 2b. The average feature μ_k of category k is defined as:

$$\mu_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i, \quad k \in \{1, \dots, c\}, \quad (1)$$

Metrics	Acc_{raw}	Acc_{Trans}	Improvement rate
RMSE	26.8%	32.2%	20%
Cosine Similarity	26.6%	32.9%	24%
JSD	27.9%	34.1%	22%

Table 1. Accuracy of 6 classification of original and transformed fields with different metrics. Classification is performed using the nearest neighbor method centered on the mean field of each category.

Pressure Level	Raw fields	Transformed fields
850 hPa	59.4%	75.3%
500 hPa	82.1%	94.3%

Table 2. Percentage of the difference between the maximum cold advection of the mean field to the mean difference between the maximum cold advection of cases, considering the clusters of the fastest deepening and filling.

where C_k represents the set of all samples belonging to category k in the training set, $|C_k|$ is the number of samples in set C_k , N is the total number of samples, and $c=6$ is the total number of categories. Then, we select the category with the highest similarity as the predicted category of the sample, that is:

$$cluster = \underset{k}{\operatorname{argmax}} \operatorname{Similarity}(\mu_k, x), \quad k \in \{1, \dots, c\}. \quad (2)$$

Based on these classification results, we verify the superiority of CAN by comparing the classification accuracy of raw fields and transformed fields. The classification accuracy of the raw and transformed fields using different similarity measurements is shown in Table 1.

Table 1 shows that CAN achieves higher classification accuracy, indicating greater intra-class similarity and reduced smoothing in composites. CAN's enhanced classification accuracy (over 20% improvement across algorithms) demonstrates its effectiveness by better reflecting true cyclone structure in composite results. Furthermore, JSD-based (Jensen-Shannon Divergence) classification consistently outperforms RMSE and cosine similarity, suggesting that distributional differences are more crucial than numerical values in cyclone deepening.

Case studies indicate that the rapidly intensifying cyclones over the Mongolian Plateau-Northeast China Plain in spring are typically linked to cold air activity. Cold advection, particularly strong frontal advection, drives rapid cyclone development by inducing frontogenesis and vertical motion¹³. Assuming that this cold advection feature exhibits an approximate spatial distribution pattern during cyclone intensification, the intensity of maximum cold advection is a crucial characteristic related to rapidly intensifying cyclone development. To analyze whether the improvement in CAN's cyclone classification and composite results holds practical physical significance, we conducted a comparison. We focused on two key metrics: the Difference of Maximum Cold Advection based on Mean Fields (DMF, Eq. 3aa) and the Difference of Mean Maximum Cold Advection based on Individual Samples (DIS, Eq. 3ab). These metrics were compared between the raw and transformed fields for two types of events: fastest deepening (deepening III, characterized by a strong positive deepening rate, as shown in Fig. 2b) and fastest filling (filling III, with a strong negative deepening rate, also in Fig. 2b). In Eq. 3aa and 3b, C_- represents the set of fastest filling samples, C_+ is the set of fastest deepening samples, $|C|$ is the number of samples in the set, ω represents one sample, S is the latitude and longitude coordinates. Subsequently, we calculated the percentage of the composite value of the maximum temperature advection difference relative to the actual value, $\frac{Tadv_{DMF}}{Tadv_{DIS}}$, as an indicator to examine the ability of the composite analysis results to reveal the true characteristics of cyclones (Table 2). The results show that, whether at the lower or upper levels, the percentage of DMF to DIS in the transformed field is significantly higher than that in the raw field. Therefore, the composite results after CAN transformation are closer to the actual situation. This suggests that CAN transformation achieves a more reasonable alignment of variable fields compared to typical compositing. This enhanced alignment better reflects the evolving structural characteristics of cyclone development, as well as changes in temperature and vorticity advection, thus further revealing how thermodynamic and dynamic processes influence cyclone development.

$$Tadv_{DMF} = \min_{s \in S} \frac{1}{|C_+|} \sum_{\omega \in C_+} Tadv(\omega, s) - \min_{s \in S} \frac{1}{|C_-|} \sum_{\omega \in C_-} Tadv(\omega, s), \quad (3a)$$

$$Tadv_{DIS} = \frac{1}{|C_+|} \sum_{\omega \in C_+} \min_{s \in S} Tadv(\omega, s) - \frac{1}{|C_-|} \sum_{\omega \in C_-} \min_{s \in S} Tadv(\omega, s). \quad (3b)$$

To more clearly demonstrate the efficacy of the composite method based on CAN, we further compared the spatial distribution of composite fields of some variables on pressure levels between the raw and transformed

fields for deepening III and filling III events. The results indicate significant differences between the two approaches, evident in the following aspects (Figs. 3, 4, 5 and 6).

Figure 3 displays the composite results of vertical velocity, potential temperature, geopotential height, and topographic elevation at various pressure levels for both raw and transformed fields, along with the average 850 hPa front. Figure 4, in contrast, presents the average fields of vertical velocity, potential temperature, and geopotential height on pressure levels, alongside surface topographic elevation.

Comparing low-level composites, transformed fields (Figs. 3b1, b2) more effectively illustrate topography-wind interactions than raw fields (Figs. 3a1, a2). Transformed fields clearly show the coupling between topography and other variables, which is highly consistent with the typical characteristics of lee cyclones^{77–79}. Specifically, they capture upward motion on the windward slope behind the cold front and downward motion on the leeward slope ahead of it (near the intersection of AB and CD in Fig. 3b1). Further evidence for this topography-wind interactions is seen in the average fields (Fig. 4b1), which shows a northwest-to-southeast distribution of ascending-descending-ascending vertical motion near the topography, consistent with uplift on the windward side and descent on the leeward side. This characteristic is also evident in the cross-section perpendicular to the cold front (Fig. 8b4). This feature is not obvious in the raw fields, and the low-level ascent on the side of the cold front is not observed in Fig. 8a4. Additionally, an increased geopotential height is observed near the surface cold front. These are key features related to rapid cyclone intensification driven by strong cold air. These features are blurred or misaligned in raw fields, obscuring physical processes. Furthermore, transformed fields more accurately enhance climatological cold/warm regions west/south of the cyclone, as well as the upward motion consistent with cold air-induced lifting. In contrast, less intense cyclones (Figs. 3c1, c2) lack local topography anomalies and associated vertical motion anomalies, displaying typical frontal vertical motion instead. This highlights the crucial role of topography-meteorological element interaction in rapid cyclone intensification over the Mongolian Plateau-Northeast China Plain.

Upper-level composites reveal greater advantages of transformed fields. Raw field composites suggest cyclone intensification is driven by warm air due to large-scale positive anomalies in geopotential height and potential temperature (Figs. 3a3, a4). However, this contradicts synoptic observations, especially for rapidly intensifying spring cyclones over the Mongolian Plateau, where rapid development is linked to southward cold air movement.

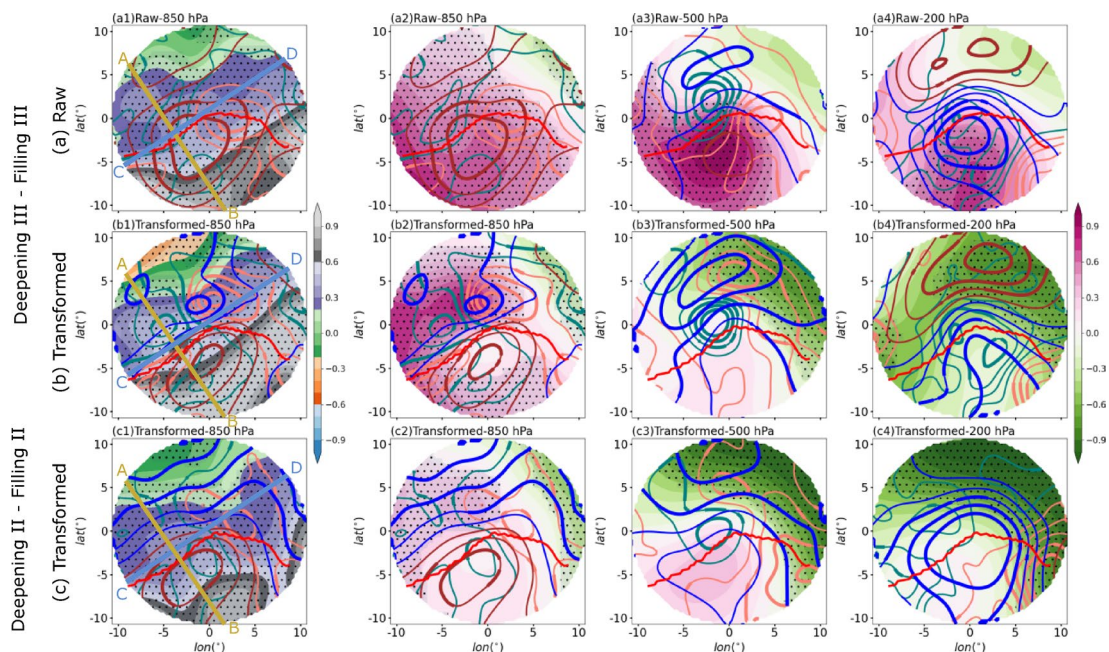


Fig. 3. Composites of vertical velocity, potential temperature and geopotential height on pressure levels, with surface topographic elevation and mean 850 hPa fronts. Composites (a1–a4 and b1–b4: deepening III – filling III, c1–c4: deepening II – filling II) of vertical velocity (salmon/teal contours indicate positive/negative values, thin/thick lines indicate non-significant/significant values respectively, unit: Pa/s, contour interval is 0.015 Pa/s), potential temperature (brown/blue contours for positive/negative values, thin/thick lines indicate non-significant/significant values respectively, unit: K, contour interval is 0.7 K), geopotential height (a2–a4, b2–b4, c2–c4, shaded, dotted areas indicate significance test passed, unit: 496 gpm) on pressure levels, topographic elevation (a1, b1 and c1, shaded, dotted areas indicate significance test passed, unit: 578 m), and mean 850 hPa fronts (red solid lines). 1–4 Correspond to levels of 850 hPa, 850 hPa, 500 hPa and 200 hPa, topographic elevation is filled in 1 and geopotential height is filled in 2, 3 and 4. The AB (yellow) and CD (steel blue) lines characterize the positions perpendicular and parallel to the 850 hPa cold front. All variables in the raw and transformed data were normalized by variable on the same scale. The shown anomalies are statistically significant at the 90% confidence level, based on a two-tailed Student's T test.

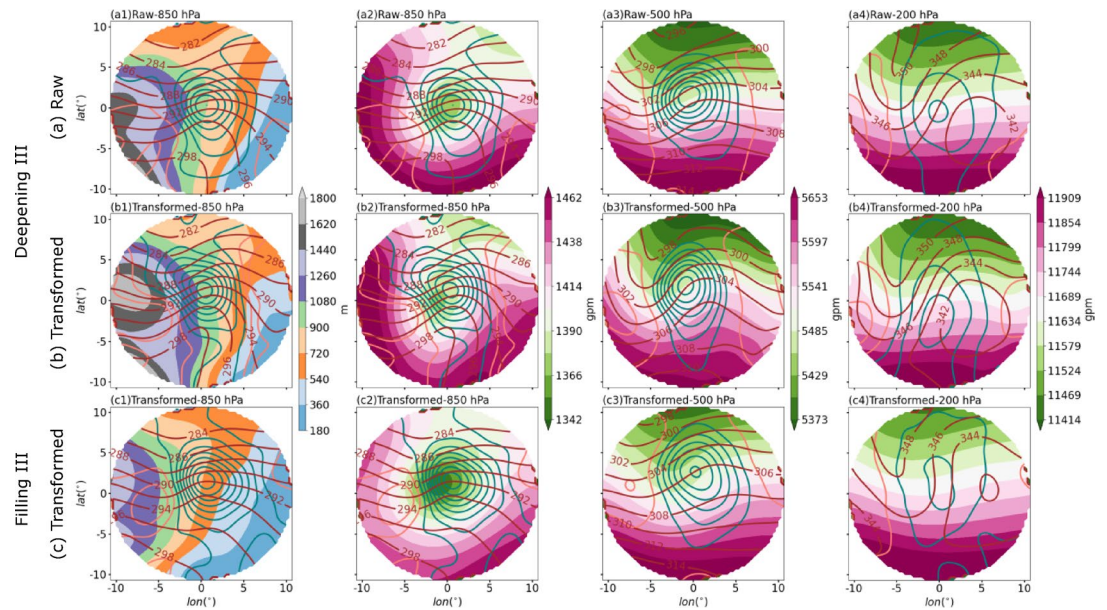


Fig. 4. Vertical velocity, potential temperature, and geopotential height averaged on pressure levels, with surface topographic elevation. Averages (a1-a4 and b1-b4: deepening III, c1-c4: filling III) of vertical velocity (salmon/teal contours indicate positive/negative values, unit: Pa/s, contour interval is 0.012 Pa/s for 850 hPa, 0.048 Pa/s for 500 hPa and 0.009 Pa/s for 200 hPa), potential temperature (brown contours, unit: K, contour interval is 2 K), geopotential height (a2-a4, b2-b4, c2-c4, shaded, unit: gpm) on pressure levels, and topographic elevation (a1, b1 and c1, shaded, unit: m). a1-a4 Averages of the raw fields. b1-b4 and c1-c4 Averages of the transformed fields. 1–4 Correspond to levels of 850 hPa, 850 hPa, 500 hPa and 200 hPa, topographic elevation is filled in 1 and geopotential height is filled in 2–4.

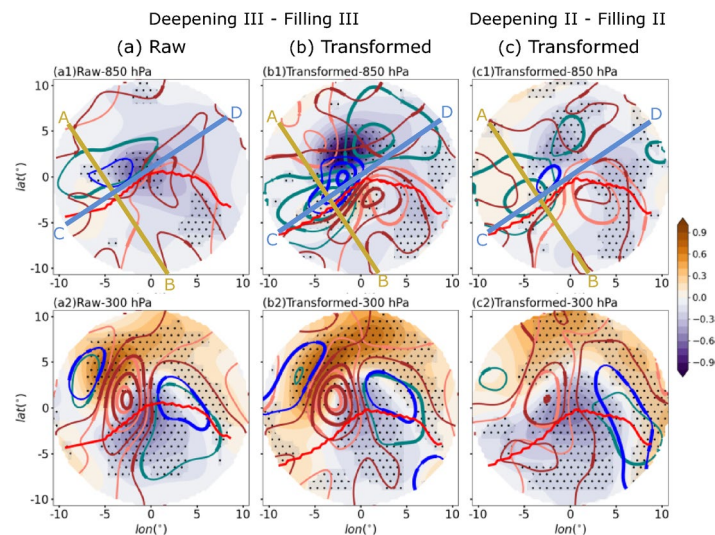


Fig. 5. Composites of temperature advection, vorticity advection and potential vorticity on pressure levels, and mean 850 hPa fronts. Composites (a1-a2 and b1-b2: deepening III – filling III, c1-c2: deepening II – filling II) of temperature advection (salmon/teal contours for positive/negative values, thin/thick for insignificant/significant values, contour interval is 5×10^{-9} K/s for 850 hPa and 10×10^{-9} K/s for 300 hPa), vorticity advection (brown/blue contours for positive/negative values, thin/thick for insignificant/significant values, contour interval is 10^{-9} s⁻² for 850 hPa and 5×10^{-9} s⁻² for 300 hPa) and potential vorticity (filled, hatches for significant values, unit: 0.25 PVU for 850 hPa and 1 PVU for 300 hPa) on pressure levels, and mean 850 hPa fronts (red solid lines). a1-a2 Composites of the raw data. b1-b2 and c1-c2 Composites of the transformed data. 1–2 Correspond to 850 hPa and 300 hPa. The AB (yellow) and CD (blue-gray) lines characterize the positions perpendicular and parallel to the 850 hPa cold front. All variables in the raw and transformed data were normalized by variable on the same scale. The shown anomalies are statistically significant at the 90% confidence level, based on a two-tailed Student's T test.

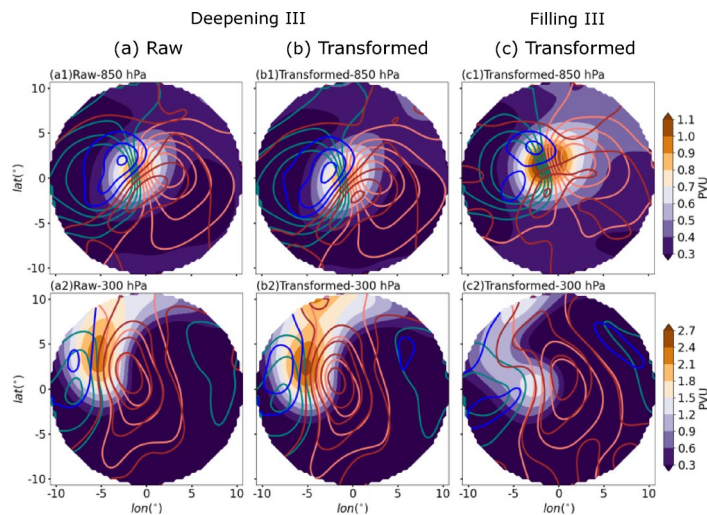


Fig. 6. Temperature advection, vorticity advection and potential vorticity averaged on pressure levels. Averages (a1-a2 and b1-b2: deepening III, c1-c2: filling III) of temperature advection (salmon/teal contours for positive/negative values, contour interval is $5 \times 10^{-9} \text{ K/s}$ for 850 hPa and $10 \times 10^{-9} \text{ K/s}$ for 300 hPa), vorticity advection (brown/blue contours for positive/negative values, contour interval is 10^{-9} s^{-2} for 850 hPa and $5 \times 10^{-9} \text{ s}^{-2}$ for 300 hPa) and potential vorticity (filled, unit: PVU) on pressure levels. a1-a2 Averages of the raw data. b1-b2 and c1-c2 Averages of the transformed data. 1–2 Correspond to 850 hPa and 300 hPa.

In contrast, transformed fields (Figs. 3b3, b4) show a reduced positive geopotential height anomaly to the south of the cyclones and an overall negative potential temperature anomaly, correctly indicating that strong cold air from the north drives intensification. This aligns with actual observations, highlighting CAN's ability to capture key physics.

Compared to rapidly intensifying cyclones, less intense cyclones (Figs. 3c3, c4) show a cold anomaly center at 500 hPa shifted northeast and a near absence of positive potential temperature anomaly at 200 hPa. These differences underscore that stronger cold air intrusion and significant tropopause folding are key characteristics of rapidly intensifying cyclones.

Evidence for stronger cold air intrusion is multifaceted. Average fields reveal that rapidly intensifying cyclones (Deepening III) have a deeper, north-south oriented 500 hPa temperature trough located northwest of the cyclone, whereas less intense cyclones (Filling III) show a shallower, northwest-southeast oriented trough closer to the center (Figs. 4b3, c3). Furthermore, analysis of the 500 hPa cold anomaly center in cross-sections shows its movement from C to D (northeastward) for less intense cyclones (Fig. 8b3, c3), a pattern also corroborated by advection centers (Fig. 10b3, c3). In both longitudinal cross-sections and those perpendicular to the front, rapidly intensifying cyclones display a stronger cold core with corresponding stronger subsidence (Figs. 8b1, b4, c1, c4), and a more vigorous, extensive cold advection on the cold front side (Figs. 10b1, b4, c1, c4), all indicating a more pronounced cold air intrusion.

For tropopause folding, a downward-extending region of high potential vorticity values is observed, which is significantly more pronounced in rapidly intensifying cyclones (Deepening III) than in less intense ones (Filling III) (Figs. 6b2, c2 and Figs. 10b1, b4, c1, c4). This clearly indicates more significant tropopause folding during rapid intensification.

Figure 5 shows the composite results of temperature advection, vorticity advection, and potential vorticity for both raw and transformed fields, along with the average 850 hPa front. Correspondingly, Fig. 6 presents the averaged fields of temperature advection, vorticity advection, and potential vorticity on pressure levels.

In low-level composites, raw fields show weak advection and potential vorticity anomaly signals, making it difficult to effectively identify cyclone-related dynamic and thermodynamic processes (Fig. 5a2). In contrast, transformed fields clearly reveal a significant positive potential vorticity anomaly (yellow shaded area near the cyclone center in Fig. 5b2), accompanied by strong vorticity and temperature advection. These signals are consistent with frontal distributions and classical cyclone theory, indicating strong cold air activity as a key driver for rapid intensification. Raw fields fail to capture these crucial low-level features, suggesting limitations in showing near-surface cyclone structure. These conclusions are also reflected in the average fields. Composites of less intense cyclones (Fig. 5c1) show weaker advection anomalies, indicating weaker cold air activity, further emphasizing strong cold air as crucial for rapid intensification.

At upper levels, although differences between raw and transformed fields are less stark than at lower levels, transformed fields show a larger, stronger positive potential vorticity anomaly (Figs. 5a1, b1). Downward propagation of upper-level positive potential vorticity is key to rapid cyclone intensification by strengthening low-level circulation. The enhanced positive potential vorticity anomaly in transformed fields demonstrates their better ability to capture upper-level dynamics, better reflecting cyclone vertical structure and development. These conclusions are also reflected in the average fields. Compared to less intense cyclones (Fig. 5c2), rapidly

intensifying cyclones exhibit more significant positive potential vorticity anomalies on the cold front side and vorticity advection near the center, highlighting these as important drivers for their intensification.

In summary, composites based on CAN not only demonstrate excellent performance in classification accuracy but also more accurately reflects the key physical characteristics of rapidly intensifying cyclones. In the lower levels, the transformed field clearly reveals the coupling relationship between cold air activity and cyclone development. For example, it clearly captures: (1) the topography-wind interaction consistent with characteristics consistent with lee cyclogenesis; (2) the enhancement of upward motion and temperature anomalies associated with the southward movement of strong cold air; and (3) the advection and potential vorticity anomalies consistent with the frontal distribution. Simultaneously, in the upper levels, the transformed field also better captures the potential temperature, geopotential height, and potential vorticity anomalies consistent with synoptic observations. These results indicate that the method proposed in this paper can capture the key elements influencing cyclone development, thereby demonstrating its potential in understanding rapidly intensifying cyclone development.

Composites of circulation characteristics

To better analyze the circulation characteristics of rapidly intensifying spring cyclones over the Mongolian Plateau-Northeast China Plain from the perspective of the transformed fields, we present the composite vertical profiles of cyclone circulation characteristics under the transformed fields (deepening III – filling III). These are then compared with the composites of the raw field, as shown in Fig. 7. Additionally, Fig. 8 provides vertical cross sections of anomaly geopotential height, potential temperature, and averaged vertical velocity.

Through comparison, it is evident that transformed fields exhibit clearer characteristics in several aspects. Firstly, transformed fields more effectively capture the key features related to upper-level cold lows. In the regional latitudinal cross-section (Figs. 7a1, b1), the composite of potential temperature in the transformed fields shows significant negative anomalies, coupled with negative mid-to-upper level geopotential height anomalies, which clearly indicates a stronger upper-level cold low system. This suggests that the rapid intensification of cyclones is closely related to the activity of upper-level cold lows. In contrast, the raw fields composite shows warmer potential temperature and higher geopotential height anomalies, failing to clearly reveal the upper-level cold lows.

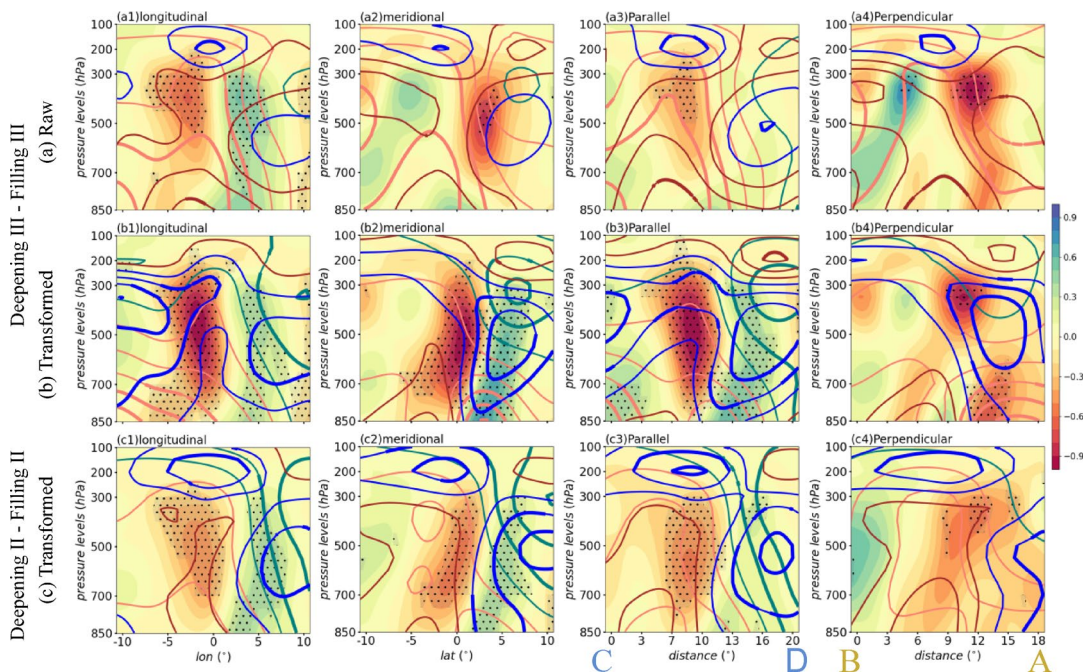


Fig. 7. Vertical cross sections of composites of geopotential height, potential temperature and vertical velocity. Vertical cross sections of composites (a1–a4 and b1–b4: deepening III – filling III, c1–c4: deepening II – filling II) of geopotential height (salmon/teal contours for positive/negative values, thin/thick for insignificant/significant values, contour interval is 166 gpm), potential temperature (brown/blue contours for positive/negative values, thin/thick for insignificant/significant values, contour interval is 0.8 K) and vertical velocity (filled, hatches for significant values, unit: 0.1 Pa/s). a1–a4 Composites of the raw data. b1–b4 and c1–c4 Composites of the transformed data. a1, b1 and c1 Longitudinal cross-section. a2, b2 and c2 Meridional cross-section. a3, b3 and c3 Cross-section along line CD (parallel to the 850 hPa cold front) in Fig. 3. a4, b4 and c4 Cross-section along line AB (perpendicular to the 850 hPa cold front) in Fig. 3. All variables in the raw and transformed data were normalized by variable on the same scale. The shown anomalies are statistically significant at the 90% confidence level, based on a two-tailed Student's T test.

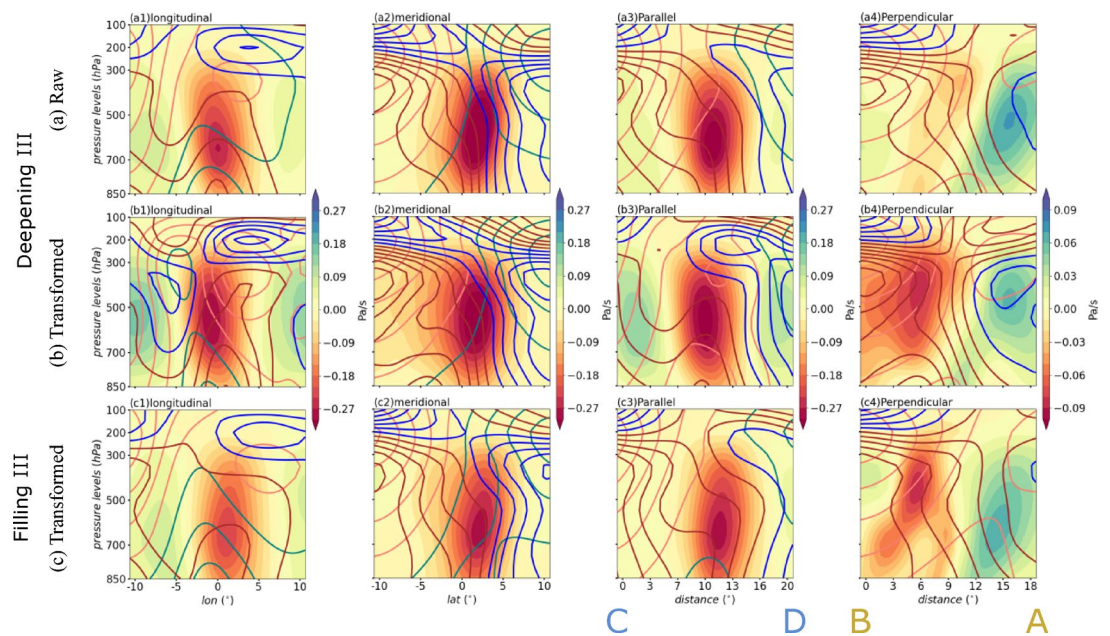


Fig. 8. Vertical cross sections of anomaly geopotential height, potential temperature and averaged vertical velocity. Vertical cross sections (a1-a4 and b1-b4: deepening III, c1-c4: filling III) of anomaly geopotential height (salmon/teal contours for positive/negative values, contour interval is 30 gpm), potential temperature (brown/blue contours for positive/negative values, contour interval is 2 K) and averaged vertical velocity (filled, unit: Pa/s). a1-a4 for the raw data. b1-b4 and c1-c4 for the transformed data. a1, b1 and c1 Longitudinal cross-section. a2, b2 and c2 Meridional cross-section. a3, b3 and c3 Cross-section along line CD (parallel to the 850 hPa cold front) in Fig. 3.3. a4, b4 and c4 Cross-section along line AB (perpendicular to the 850 hPa cold front) in Fig. 3.3.

Furthermore, transformed fields can more effectively capture the key features related to lower-level cold air activity and vertical circulation. In the transformed field, significant anomalies of low potential temperature and high geopotential height are evident on the west side of the cyclone at lower levels. This indicates cold air accumulation due to sinking air ahead of topography. The leading edge of this cold air spatially coincides with columnar ascending anomalies, associated with relatively high potential temperature and geopotential height anomalies near the cyclone center. This demonstrates that strong vertical circulation in rapidly intensifying cyclones is related to dynamic uplift from lower-level cold air, consistent with synoptic observations. In contrast, raw fields show warmer potential temperature and higher geopotential anomalies, along with unclear variable coupling, different from synoptic observations.

Moreover, transformed fields uniquely capture a cold sinking anomaly on the north side of the cyclones, absent in raw fields. The regional meridional cross-section (as shown in Figs. 7a2, b2) shows that the characteristics on the north side of the cyclone are particularly significant in the transformed fields: specifically, a sinking vertical velocity anomaly on the north side of cyclones, which coincides with colder potential temperature, upper-level low geopotential height anomalies, and lower-level high geopotential height anomalies. This indicates an anomalous cold sinking airflow. Combined with the distribution of the mid-to-upper level geopotential height field in Fig. 3b2, it can be inferred that this is likely related to the convergence, cold advection, and negative vorticity advection brought by the ridge area downstream of the upper-level trough. This ridge-like curvature has also been noted in previous studies³⁹, but the characteristics of this cold sinking anomaly are only clearly evident in the transformed field. In contrast, raw fields lack significant signals on the north side of cyclones. In contrast, raw fields lack significant signals on the north side of cyclones. Furthermore, the overlap of low potential temperature and low geopotential height anomalies with the ascending anomaly makes it difficult to provide a reasonable explanation.

Additionally, transformed fields better capture the feature characteristics near fronts. The cross-sections parallel to the low-level cold front are shown in Figs. 7a3, b3. In the transformed fields (Fig. 7b3), the circulation center exhibits an ascending anomaly, while both sides show sinking anomalies. These sinking regions correlate with colder temperatures, upper-level low, and lower-level high geopotential height anomalies, indicating cold sinking anomalies. Conversely, the ascending center shows a relatively warm signal, indicating warm ascending anomaly. This suggests a characteristic of rapidly intensifying cyclones in this region: stronger cold sinking on both sides of the cyclone converges at lower levels, forcing the central warm air to rise and forming a more intense but narrower ascending region. The raw fields (Fig. 7a3) failed to clearly show these characteristics.

The cross-sections perpendicular to the low-level cold front are shown in Figs. 7a4, b4. Transformed fields (Fig. 7b4) clearly show asymmetric frontal characteristics: a significant cold anomaly northwest of the cold front and a warm anomaly southeast. The cold side shows a lower-level high/upper-level low geopotential height configuration. At the leading edge of the cold high anomalies, a vertical velocity dipole (ascending-sinking)

and a strong potential temperature gradient are evident, demonstrating coupling with topography (Fig. 3b1). This suggests windward cold air accumulation forms a lower-level cold high. Southeast of the topography, the accumulated cold air sinks on the leeward slope, forcing the warm air ahead to rise. This process forms a strong temperature gradient on the leeward slope, thereby enhancing both the frontal and the cyclone's vertical circulation. In contrast, the composite signal of the original field (Fig. 7a4) is generally more inclined to a symmetrical structure, and the coupling characteristics of variables and topography are not clear.

In contrast, less intense cyclones (Figs. 7c1–c4) show a weaker cold potential temperature anomaly at the cold front, lack a deep downward anomaly northeast of the cyclone, and feature a broader upward area. This indicates that strong cold air activity and significant downward motion from the ridge downstream are significant features of rapidly intensifying cyclones. These features contribute to forming a relatively narrow but intense upward zone, which enables concentrated baroclinic energy release and promotes rapid cyclone development.

From the above comparative analysis, it is evident that transformed fields have significant advantages over raw fields in revealing the key characteristics of rapid cyclone intensification, especially in capturing the feature characteristics such as upper-level cold lows, lower-level cold air activity, topography, and fronts. Transformed fields can provide more significant and physically meaningful signals. Furthermore, transformed fields also reveal a characteristic of rapidly intensifying cyclones: narrow but intense upward regions (Figs. 8b1, 3, c1, 3), which raw fields struggle to effectively demonstrate.

Composites of advection characteristics

To better analyze the characteristics of rapidly intensifying cyclones over the Mongolian Plateau–Northeast China Plain in spring from the perspective of transformed fields, we composite the vertical profiles of cyclone potential vorticity, temperature, and vorticity advection under the transformed fields (deepening III – filling III), and compare them with the raw field composites, as shown in Fig. 9. Furthermore, Fig. 10 presents vertical cross sections of averaged temperature advection, vorticity advection, and potential vorticity.

Figures 9a1, b1, a2, and b2 display the regional latitudinal and longitudinal cross-sections. While transformed and raw fields show similarities in variable distributions, the transformed fields offer distinct advantages. Transformed fields show a stronger upper-level positive potential vorticity (PV) anomaly and a larger, more intense positive PV anomaly northwest of the cyclone, clearly indicating downward transport of upper-level PV during rapid intensification. Lower-level advection anomalies are stronger and spatially clearer, suggesting better

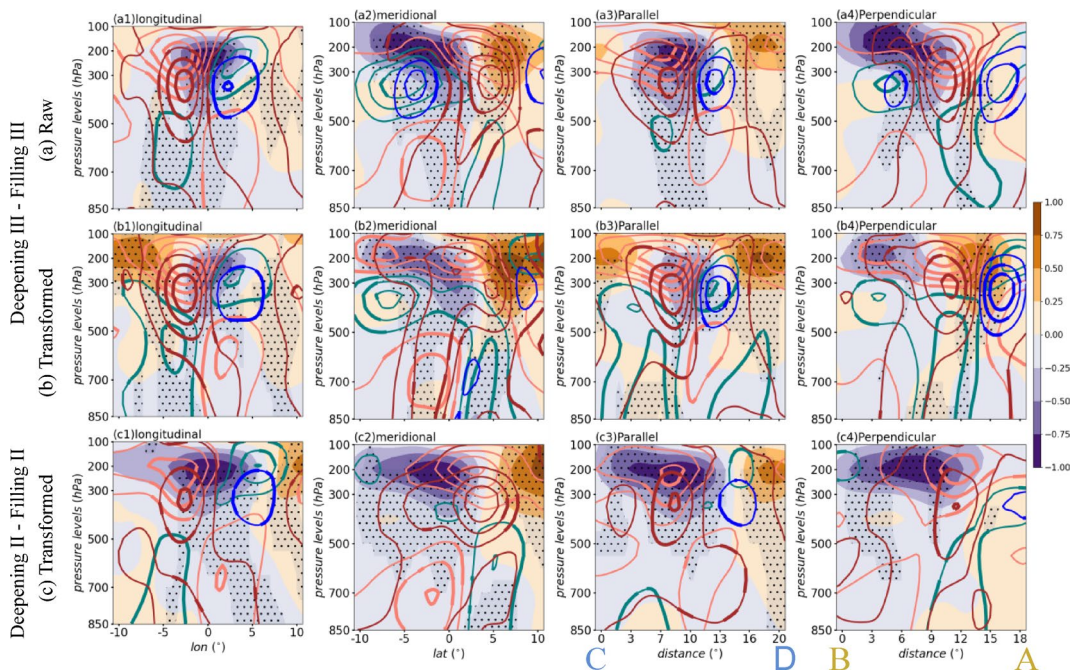


Fig. 9. Vertical cross sections of composites of temperature advection, vorticity advection and potential vorticity. Vertical cross sections of composites (a1–a4 and b1–b4: deepening III – filling III, c1–c4: deepening II – filling II) of temperature advection (salmon/teal contours for positive/negative values, thin/thick for insignificant/significant values, contour interval is 10^{-8} K/s), vorticity advection (brown/blue contours for positive/negative values, thin/thick for insignificant/significant values, contour interval is 5×10^{-9} s $^{-2}$) and potential vorticity (filled, unit: 1 PVU, hatches for significant values). a1–a4 Composites of the raw data. b1–b4 and c1–c4 Composites of the transformed data. a1, b1 and c1 Longitudinal cross-section. a2, b2 and c2 Meridional cross-section. a3, b3 and c3 Cross-section along line CD (parallel to the 850 hPa cold front) in Fig. 3. a4, b4 and c4 Cross-section along line AB (perpendicular to the 850 hPa cold front) in Fig. 3. All variables in the raw and transformed data except PV were normalized by variable on the same scale. The shown anomalies are statistically significant at the 90% confidence level, based on a two-tailed Student's T test.

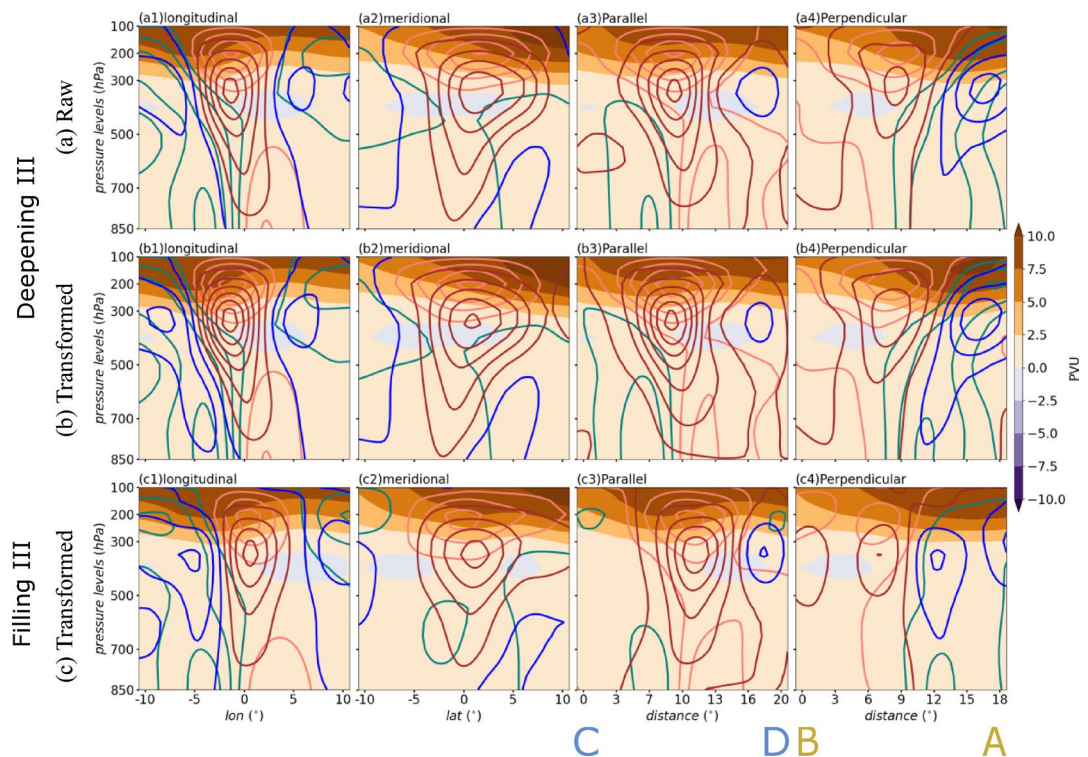


Fig. 10. Vertical cross sections of averaged temperature advection, vorticity advection and potential vorticity. Vertical cross sections (a1-a4 and b1-b4: deepening III, c1-c4: filling III) of averaged temperature advection (salmon/teal contours for positive/negative values, contour interval is $10^{-8} K/s$), vorticity advection (brown/blue contours for positive/negative values, contour interval is $5 \times 10^{-9} s^{-2}$) and potential vorticity (filled, unit: PVU). a1-a4 Averages of the raw data. b1-b4 and c1-c4 Averages of the transformed data. a1, b1 and c1 Longitudinal cross-section. a2, b2 and c2 Meridional cross-section. a3, b3 and c3 Cross-section along line CD (parallel to the 850 hPa cold front) in Fig. 3.3. a4, b4 and c4 Cross-section along line AB (perpendicular to the 850 hPa cold front) in Fig. 3.3.

capture of small-scale structures potentially important for rapid intensification. Furthermore, transformed fields exhibit more coherent physical fields coupling (Fig. 9b2). Positive (negative) temperature advection anomalies and positive (negative) vorticity advection anomalies correspond better with upward (downward) anomalies (Fig. 7b2), indicating improved representation of physical process interactions.

The cross-sections parallel to the lower-level cold front are shown in Fig. 9a3 and b3. At the lower level, the transformed fields and raw fields show significant differences. Raw fields show weak lower-level signals, but transformed fields exhibit a significant temperature advection anomaly at lower levels, spatially consistent with vertical motion anomalies in Fig. 7b3. At higher levels (around 300 hPa) in transformed fields, upward anomalies coincide with positive vorticity advection anomalies, indicating vorticity advection drives upper-level vertical motion. The spatial overlap of upper and lower-level upward areas contributes to the rapid intensification of the cyclone's vertical circulation.

The cross-sections perpendicular to the lower-level cold front are shown in Figs. 9a4 and b4. Overall, transformed fields more clearly show the tilted frontal structure and upper-lower level variable coupling. Compared to the raw fields, the transformed fields show a west-tilting positive PV anomaly from the lower to the upper levels and a corresponding negative temperature advection anomaly. Strong advection anomalies and PV anomalies may correspond to frontal intensification. Comparing Fig. 10b4 and c4, it can be observed that the advection in deepening III exhibits a tilted structure, while the advection in Filling III shows a relatively barotropic structure. This indicates that a strong, tilted front indeed exists in deepening III, while the front in Filling III is relatively insignificant. A significant upper-level positive PV anomaly at the frontal position further highlights the downward transport of upper-level PV in rapidly intensifying cyclones. At lower levels, temperature and vorticity advection anomalies show a west-tilting distribution consistent with frontal orientation, supporting frontal maintenance and development. While this lower-level west-tilting structure has been observed in synoptic analysis^{4,35,80}, raw field composites generally do not capture it. In contrast, our method distinctly demonstrates this as a significant characteristic of rapidly intensifying cyclones.

In contrast, composites of less intense cyclones (Figs. 9c1-c4) show weak cold air advection around the cyclone, with its presence on the cold front side limited to lower levels. The positive vorticity advection anomaly at 300 hPa near the cyclone center is also weak, and a positive potential vorticity anomaly is lacking on the cold front side. These differences indicate that several features are significant characteristics of rapidly intensifying cyclones: strong cold air advection resulting from intense cold air activity, and pronounced downward motion

in the periphery coupled with upward motion in the center, driven by strong upper-level positive vorticity advection. Combined with the downward transport of upper-level positive potential vorticity, these factors promote the enhancement and rapid development of the cyclone circulation.

The above comparative analysis clearly demonstrates that transformed fields have significant advantages over raw fields in revealing the key characteristics of rapid cyclone intensification, especially in their ability to highlight the upper-level positive potential vorticity anomaly and its downward transport, lower-level advection anomalies, and the representation of physical process interactions. Furthermore, the transformed fields effectively capture the tilted frontal structure and vertical coupling, which are not effectively captured by raw field analysis.

Discussion

Spatial misalignment from cyclone rotation limits traditional composite analysis. To address this, the Cyclone Alignment Network (CAN) is proposed, adaptively aligning features by learning spatial relationships. By reducing blurring caused by spatial misalignment, CAN better reveals cyclone structure. Tests on MNIST and cyclone datasets demonstrate that CAN improves classification accuracy and composite field quality compared to traditional methods. CAN's improved performance is likely due to its enhanced capture of spatial relationships among low-level variables such as topography and advection. These relationships are crucial for cyclone intensification over the Mongolian Plateau-Northeast China Plain in spring but are often poorly represented in traditional composites. These rapidly intensifying cyclones exhibit characteristics similar to those of lee cyclones. CAN effectively reveals the influence of cold air, topography, and other factors on cyclone development, which are often blurred in traditional composites. CAN analysis reveals these rapidly intensifying cyclones located ahead of high-curvature troughs with strong divergence. Downstream ridges with subsidence limit the upward extent, thereby concentrating baroclinic energy release. All these features are indistinct in traditional composites. Previous case studies have highlighted the importance of baroclinic distribution of low-level temperature/vorticity advection for rapid intensification, but this is often missed in typical composites. CAN clearly displays this phenomenon, confirming its importance for spring rapid intensification in this region.

In summary, CAN effectively displays characteristics of rapidly intensifying cyclones. Furthermore, it can analyze other cyclone features (e.g., movement, precipitation) by changing target variables. Future research could enhance CAN by increasing data resolution and expanding to 3D affine transformations to capture vertical cyclone structure.

Methods

Cyclone identification and cyclone deepening rates

We used an objective identification method⁶³ to identify extratropical cyclones over land in East Asia during boreal spring (1979–2023) and extracted their tracks and intensities. The primary advantage of this method is its sole reliance on sea level pressure data, which, after appropriate screening, yields relatively reliable cyclone track data⁸¹. Using these identified tracks, we subsequently obtained the distribution of various meteorological variables. These included surface variables (pressure, 10-m wind speed, and temperature) and variables at multiple pressure levels (850, 700, 500, 300, 200, and 100 hPa), such as temperature, humidity, geopotential height, horizontal wind, and vertical velocity. All data were extracted within a $20^\circ \times 20^\circ$ box centered on each cyclone.

The calculation of cyclone deepening rate (R) was calculated using a 12-hour time window and a standard latitude of 60°N :

$$R = \left[\frac{P_{t-6} - P_{t+6}}{12} \right] \times \left[\frac{\sin 60^\circ}{\sin((\phi_{t-6} + \phi_{t+6})/2)} \right], \quad (4)$$

where P represents the sea level pressure at the cyclone center, ϕ denotes the latitude of the cyclone center (unit: $^\circ$), and the subscripts $t - 6$ and $t + 6$ indicate the values before and after 6 h, respectively.

Thermal front parameter

As extratropical cyclones are the primary focus of this study, understanding frontal structure is essential. We use the thermal front parameter (TFP)⁸² to identify the location of fronts. The calculation of TFP is given by:

$$TFP = -\nabla \tau \cdot |\nabla \tau| \bullet \frac{\nabla \tau}{|\nabla \tau|}, \quad (5)$$

where τ represents any suitable thermodynamic parameter. For the purpose of this study, 850 hPa potential temperature is chosen for τ . Based on previous studies^{83–85}, the frontal location is then determined by the line connecting the positions of maximum TFP.

Cyclone alignment network (CAN)

Overview of CAN

We design a network capable of adaptive spatial alignment of cyclone features to better represent variables related to cyclone development in composite analysis. This architecture integrates a Transformer branch, an affine transformation branch, and a convolutional branch, achieving this alignment via a classification task (Fig. 11). To validate the method, we conducted experiments using a randomly rotated MNIST dataset and a dataset of spring extratropical cyclones over East Asian land areas. The extraction method for the cyclone dataset is detailed in Section Cyclone Identification and Cyclone Deepening Rates.

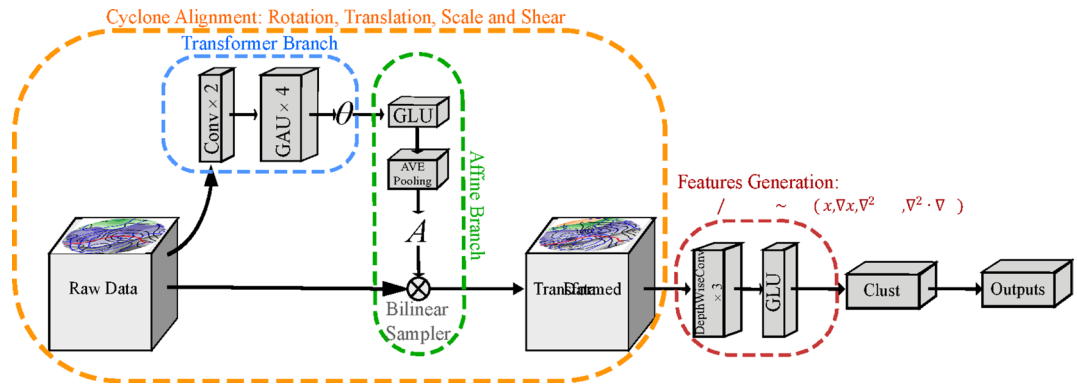


Fig. 11. The structure of Cyclone Alignment Network (CAN). The original variable field is input to the Transformer branch, which learns attention weights θ to capture relative positional information between variables. These weights are then passed to the affine branch, which generates an affine transformation matrix A to rotate, scale, translate, and shear the original field. The transformed field is then passed to a deep CNN. The CNN's kernel size and network depth, is designed based on constraints derived from cyclone development equations to extract physically consistent features for cyclone classification.

Transformer branch

The relative position and interaction of variable fields are important in cyclone circulation analysis. Therefore, the relative distribution of variables is a key factor to consider when aligning feature fields. However, identifying and characterizing the distribution of variables related to the target variable (the deepening rate) is a complex task. To address this challenge through adaptive spatial localization, we employ a neural network architecture. Recognizing the better ability of the Transformer⁸⁶ to capture relative positional information compared to traditional CNNs, we employ it as the core component for providing spatial localization cues.

In the Transformer branch, we first standardize each variable in the original fields and then extract local features using two 3×3 convolutional layers. The second convolutional layer's stride is set to 2, meaning the kernel shifts by two grid points in each dimension as it slides across the input feature map. This reduces the output feature map's dimensions by half, achieving downsampling and reducing computation. These feature maps are then input into an encoder consisting of four GAU modules. The GAU module combines self-attention⁸⁶ and a feedforward neural network (FNN) to effectively and efficiently capture long-range dependencies between variables. RoPE is applied to each GAU module to enhance the model's perception of relative position information. This allows the model to adaptively learn spatial relationships between variables, providing key cues for feature alignment.

Affine branch

Due to the rotation of cyclones, variables associated with cyclone activity can be spatially misaligned. For example, the position of features such as fronts may vary across different cyclone cases. To align variables with similar influences on cyclone activity to the same coordinate system, the original variable fields require an affine transformation. This transformation involves a rotation component to compensate for the cyclone's rotation, as well as potential minor shear, scaling, and translation adjustments for fine-tuning the alignment. The affine transformation branch is responsible for performing this alignment.

The affine transformation branch receives the output tensor from the Transformer branch. To apply the affine transformation to the original variable fields, this branch first generates parameters for the affine transformation matrix using a GLU (Gated Linear Unit⁸⁷) and a global average pooling layer. These parameters define a mapping from the original coordinate system to a new coordinate system. This affine transformation matrix is then used to transform the grid point coordinates of the original variable fields to new positions. Since the transformed coordinates are usually non-integers, we use bilinear interpolation⁶⁸ to compute the values at the new positions to ensure the continuity. Bilinear interpolation is a common interpolation method that estimates the variable values at the new positions by taking a weighted average of the values of the nearest four grid points, which can effectively handle spatial scaling and rotation.

In general, the coordinate transformation under the Cartesian coordinate system can be expressed as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = A \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (6)$$

where x, y are the original spatial coordinates, x', y' are the transformed spatial coordinates, and $A \in R^{3 \times 3}$ is the required affine transformation matrix. This matrix has 6 degrees of freedom and can represent scaling, rotation, translation, shear, and their combinations:

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

Directly learning these six parameters can lead to unpredictable or unstable transformations. For example, it is typically desirable to avoid excessive scaling or shearing and to constrain the rotation angle during transformation. To address these issues, we decompose the affine transformation into more basic transformations and control them separately.

Let A_{Sc} , A_R , A_T , and A_{Sh} denote the transformation matrices corresponding to the scaling, rotation, translation, and shear operations, respectively. Then:

$$A_{Sc} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (7a)$$

$$A_R = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (7b)$$

$$A_T = \begin{bmatrix} 1 & 0 & T_x \\ 0 & 1 & T_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (7c)$$

$$A_{Sh} = \begin{bmatrix} 1 & a & 0 \\ b & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (7d)$$

$$A = A_T \times A_{Sc} \times A_R \times A_{Sh}, \quad (8)$$

these transformations have seven degrees of freedom ($s_x, s_y, \theta, T_x, T_y, a, b$). In network construction, we constrain the rotation angle $\theta \in (-\pi, \pi)$ and apply L2 regularization to the other parameters to control the transformation magnitude. After compositing, the composite fields are rotated using the inverse of the average transform matrix, such that the composites of transformed fields have an orientation comparable to that of the raw fields.

Features extraction and classification

The vorticity and development equations suggest that the deepening rate is typically influenced by the spatial derivatives of input variables and their combinations, rather than the variables themselves. Therefore, we introduce a module to generate variables in this form. From a neural network perspective, this can be realized through multiple convolutional layers. However, convolutional layers inherently possess translational invariance and locality. This characteristic allows them to partially learn translation and rotation information directly from the input data, thereby weakening the corrective effect of the affine transformation. Our aim is to ensure that all coordinate transformation information is concentrated in the affine transformation matrix A for accurate correction.

Cyclone development is typically described by equations such as the potential vorticity tendency equation, the vorticity equation, and the development equation. Geopotential height, vertical motion, temperature advection, and vorticity advection play important roles in these equations. Fronts, key features of extratropical cyclones, whose strengths are closely related to potential temperature gradients, have a significantly influence on cyclone development. As described in Sect. 1, tropopause folding and upper-level potential vorticity intrusion are also important mechanisms for explosive cyclone development, facilitating the downward transport of high potential vorticity and momentum air from the upper levels, thereby strengthening low-level cyclonic circulation^{22,29–32}. Therefore, potential vorticity strongly influences cyclone development. Topography also plays a role by affecting airflow lifting, blocking, convergence, and divergence^{77–79}. Consequently, we select geopotential height, wind speed, vertical velocity, potential temperature, topography height and other relevant variables as model inputs.

Considering that these variables involve up to third-order horizontal spatial derivatives in the equations, we use three depthwise separable convolutional layers to approximate these derivatives. Depthwise separable convolution effectively extracts spatial features while reducing computation. Furthermore, to account for interactions between different variables—namely, product terms—and potential first-order vertical gradients present in the equations, we employ a GLU after the convolutional layers. The extracted features are denoted as $F_m \in R^{m \times n \times f}$, where m and n are the spatial dimensions, and f is the number of extracted feature variables.

Ideally, F_m under the new coordinate system should present a relatively fixed distribution, which is highly correlated with the target variable (cyclone deepening rate). This implies aligning a dominant structural feature influencing the deepening rate to a similar spatial orientation. This fixed distribution is assumed to be represented by $H_m \in R^{c \times m \times n \times f}$, where m, n and f are as defined for F_m , and c denotes the number of categories. Point-to-point similarity (Eq. 10) calculates the probability $p(k|x)$ of each sample belonging to each category, which serves as the network output (Eq. 11). The category with the highest probability is assigned as the sample's *cluster* (Eq. 11). This point-to-point similarity measure constrains F_m and the corresponding H_m to have similar spatial distributions.

$$S_i = \text{Similarity}(F_m, H_{m,i}), \quad i \in \{1, \dots, c\}, \quad (9)$$

$$p(k|x) = \frac{\exp(S_k)}{\sum_{i=1}^c \exp(S_i)}, \quad (10)$$

$$\text{cluster} = \underset{i}{\operatorname{argmax}}(i|x), \quad i \in \{1, \dots, c\}, \quad (11)$$

the similarity function is measured using the Jensen-Shannon Divergence (JSD):

$$JSD(P||Q) = \frac{1}{2} \sum (p_i \bullet \log(m_i)) + \frac{1}{2} \sum (q_i \bullet \log(m_i)), \quad M = (P + Q)/2, \quad (12)$$

$$\text{Similarity}(P, Q) = 1/JSD(\text{softmax}(P)||\text{softmax}(Q)). \quad (13)$$

Data and experiment settings

Data

For this study, hourly atmospheric variables were extracted from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA-5)⁸⁸. The extracted data, covering the period from 1979 to 2023, has a spatial resolution of $0.5^\circ \times 0.5^\circ$ and includes surface pressure, sea level pressure (SLP), 10-m wind, geopotential height, horizontal wind, vertical velocity, specific humidity, and temperature. Specifically, SLP is used to identify the extratropical cyclones for analysis.

And topography from ETOPO1⁸⁹ 1 Arc-Minute Global Relief Model is used to show the topography under the cyclones.

MNIST⁷⁵ dataset is used to test our CAN.

Experiment settings

Batch size, hardware and schedule We trained our models on one machine with an RTX 3080 Ti GPU and the batch size is set as 128. For our models using the hyperparameters described throughout the paper, each training step took about 0.5 s on MNIST dataset or 1.0 s on cyclone dataset. We trained the models for a total of 75,000 steps (750 epochs).

Optimizer We used the AdamW optimizer⁹⁰ with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-7}$ and $w = 10^{-5}$. We varied the learning rate (l_{rate}^i) over the course of training for each epoch (epoch_i), according to the formula:

$$l_{rate}^i = \begin{cases} \eta_{max} \left(\frac{\text{epoch}_i \bmod T}{\text{epoch}_{warmup}} \right), & \text{epoch}_i \bmod T < \text{epoch}_{warmup}, \\ \eta_{min} + 0.5 (\eta_{max} - \eta_{min}) \left(1 + \cos \left(\pi \frac{\text{epoch}_i \bmod T - \text{epoch}_{warmup}}{T - \text{epoch}_{warmup}} \right) \right), & \text{epoch}_i \bmod T \geq \text{epoch}_{warmup}. \end{cases}$$

This corresponds to increasing the learning rate linearly for the warm up training epochs, and decreasing it thereafter according to the cosine annealing⁹¹. We used $\eta_{max} = 10^{-4}$, $\eta_{min} = 0$, $T = 50$ and $\text{epoch}_{warmup} = 10$.

Regularization We employ two types of regularization during training:

Flooding we employed flooding⁹² of value $b = 1.1$. This keeps the training loss to stay around a small constant value, to avoid zero training loss. With flooding, the validation/test accuracy can be improved.

Label smoothing During training, we employed label smoothing⁹³ of value $\varepsilon_{ls} = 0.1$. This hurts perplexity, as the model learns to be more unsure, but improves accuracy.

Data availability

The data used in this article can be accessed at the following URL: 1. ERA-5 data: surface: <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download> pressure levels: <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels?tab=download>. 2. ETOPO1 data: <https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ngdc.mgg.dem:3163>. 3. MNIST dataset: <http://yann.lecun.com/exdb/mnist/or> <https://storage.googleapis.com/tensorflow/tf-keras-datasets/mnist.npz>.

Received: 14 March 2025; Accepted: 13 October 2025

Published online: 19 November 2025

References

1. Liu, J. T. & Zheng, M. Q. The Climatic characteristics of black storm in the Northern part of North China. *Meteorol. Mon.* **24**, 39–44 (1998).
2. Lu, M. & Pu, J. P. The analysis of two synoptic processes of sand storm in 2000. *Scientia Meteor. Sinica*. **22**, 210–217 (2002).
3. Liu, J. T. et al. A case study of a severe dust storm resulted from a explosive Mongolia cyclone. *Clim. Environ. Res.* **8**, 218–229 (2003).
4. Zhao, L. N. & Zhao, S. X. A diagnostic study of rapid developing cyclone in North China. *Chin. J. Atmos. Sci.* **28**, 722–735 (2004).
5. Kong, S. S. K. et al. Distinct transport mechanisms of East Asian dust and the impact on downwind marine and atmospheric environments. *Sci. Total Environ.* **827**, 154255 (2022).
6. Roebber, P. J. Statistical analysis and updated climatology of explosive cyclones. *Mon. Weather Rev.* **112**, 1577–1589 (1984).
7. Rogers, E. & Bosart, L. F. An investigation of explosively deepening oceanic cyclones. *Mon. Weather Rev.* **114**, 702–718 (1986).

8. Yi, Q. J. & Ding, Y. H. An analysis of the explosive cyclone over yellow sea and Bohai sea. *J. Appl. Meteorol. Sci.* **7**, 483–490 (1996).
9. Huang, L. W. et al. Numerical simulation and experiment study on explosive marine cyclones. *J. Meteorol. Res.* **57**, 410–427 (1999).
10. Yin, J. Y. et al. A diagnostic study of an intense developing extratropical cyclone over the Bohai and yellow sea. *Meteorol. Mon.* **37**, 1526–1533 (2011).
11. Kouroutzoglou, J. et al. On the vertical structure of mediterranean explosive cyclones. *Theor. Appl. Climatol.* **110**, 155–176 (2012).
12. Wash, C. H. et al. Diagnostic study of explosive cyclogenesis during FGGE. *Mon. Weather Rev.* **116**, 431–451 (1988).
13. Li, C. Q. & Ding, Y. H. A diagnostic study of an explosively deepening oceanic cyclone over the Northwest Pacific ocean. *J. Meteorol. Res.* **47**, 180–190 (1989).
14. Wang, J. S., Ding, Z. Y., He, J. H. & Chen, J. K. Diagnosis of an explosive cyclone with Zwack-Okossi equation. *Trans. Atmos. Sci.* **22**, 180–188 (1999).
15. Yoshida, A. & Asuma, Y. Structures and environment of explosively developing extratropical cyclones in the Northwestern Pacific region. *Mon. Weather Rev.* **132**, 1122–1142 (2004).
16. Gyakum, J. R. On the evolution of the QE II storm. Part I: synoptic aspects. *Mon. Weather Rev.* **111**, 1137–1155 (1983).
17. Chen, S. J. Dell'osso, L. A numerical case study of East Asian coastal cyclogenesis. *Mon. Weather Rev.* **115**, 477–487 (1987).
18. Kuo, Y., Low-Nam, S. & Reed, R. J. Effects of surface energy fluxes during the early development and rapid intensification stages of seven explosive cyclones in the Western Atlantic. *Mon. Weather Rev.* **119**, 457–476 (1991).
19. Ding, Y. H. & Zhu, T. Dynamical analysis and numerical experiments on the explosive development of a land cyclone. *Sci. China Ser. B*, **23**, 1226–1232 (1993).
20. Zhou, Y. et al. Potential vorticity invertibility diagnostic about the importance of latent heat in the cyclone rapid development process. *J. Meteorol. Sci.* **18** (1998).
21. Bosart, L. F. & Lin, S. C. A diagnostic analysis of the presidents' day storm of February 1979. *Mon. Weather Rev.* **112**, 2148–2177 (1984).
22. Uccellini, L. W. et al. The President's Day cyclone of 18–19 February 1979: Influence of upstream trough amplification and associated tropopause folding on rapid cyclogenesis. *Mon. Weather Rev.* **113**, 962–988 (1985).
23. Shou, S. W., Li, Y. H. & Fan, K. Isentropic potential vorticity analysis of the mesoscale cyclone development in a heavy rain process. *Acta Meteorol. Sin.* **59**, 560–568 (2001).
24. Wu, H. Y. & Shou, S. W. Potential vorticity disturbance and cyclone development. *J. Nanjing Inst. Meteorol.* **25**, 509–517 (2002).
25. Reader, M. & Moore, G. W. K. Stratosphere-troposphere interactions associated with a case of explosive cyclogenesis in the Labrador sea. *Tellus A*, **47**, 849–863 (2002).
26. Zehnder, J. & Keyser, D. The influence of interior gradients of potential vorticity on rapid cyclogenesis. *Tellus A*, **43**, 198–212 (1991).
27. Zhao, B. K., Wu, G. X. & Yao, X. P. A diagnostic analysis of potential vorticity associated with development of a strong cyclone during the Meiyu period of 2003. *Chin. J. Atmos. Sci.* **32**, 999–1014 (2008).
28. Cordeira, J. M. & Bosart, L. F. Cyclone interactions and evolutions during the perfect storms of late October and early November 1991. *Mon. Weather Rev.* **139**, 1683–1707 (2011).
29. Lackmann, G. *Midlatitude Synoptic Meteorology* (American Meteorological Society, 2011).
30. Bleck, R. Short-range prediction in isentropic coordinates with filtered and unfiltered numerical models. *Mon. Weather Rev.* **102**, 813–829 (1974).
31. Hoskins, B. J., McIntyre, M. E. & Robertson, A. W. On the use and significance of isentropic potential vorticity maps. *Quart. J. Roy Meteorol. Soc.* **111**, 877–946 (1985).
32. Lupo, A. R., Smith, P. J. & Zwack, P. A diagnosis of the explosive development of two extratropical cyclones. *Mon. Weather Rev.* **120**, 1490–1523 (1992).
33. Tian, S. C. & Liu, S. H. Diagnostic study of a rapidly developing cyclone in spring. *Acta Meteorol. Sin.* **46**, 285–293 (1988).
34. Ogura, Y. & Juang, H. M. H. A case study of rapid cyclogenesis over Canada. Part I: diagnostic study. *Mon. Weather Rev.* **118**, 655–672 (1990).
35. Juang, H. M. H. & Ogura, Y. A case study of rapid cyclogenesis over Canada. Part II: simulations. *Mon. Weather Rev.* **118**, 674–704 (1990).
36. Li, Z. J. & Zhao, S. X. Structure and dynamics of cold fronts observed in East Asia in spring. Part I: structure of strong spring cold fronts. *Chin. J. Atmos. Sci.* **20**, 662–672 (1996).
37. Li, Z. J. & Zhao, S. X. A structure and dynamics of cold fronts observed in East Asia during spring. Part II: dynamics of strong spring cold front. *Chin. J. Atmos. Sci.* **21**, 91–98 (1997).
38. Zhao, L. N. & Zhao, S. X. Simulation of rapid developing cyclone associated with strong dust storm in North China. *Clim. Environ. Res.* **9**, 116–126 (2004).
39. Kang, J. M., Lee, J., Son, S. W., Kim, J. & Chen, D. The rapid intensification of East Asian cyclones around the Korean Peninsula and their surface impacts. *J. Geophys. Res. Atmos.* **125**, e2019JD031874 (2020).
40. Pettersen, S., Bradbury, D. L. & Pedersen, K. The Norwegian cyclone models in relation to heat and cold sources. *Geophys. Norv.* **24**, 243–280 (1962).
41. Sanders, F. & Gyakum, J. R. Synoptic–dynamic climatology of the bomb. *Mon. Weather Rev.* **108**, 1589–1606 (1980).
42. Lackmann, G. M., Bosart, L. F. & Keyser, D. Planetary- and synoptic-scale characteristics of explosive wintertime cyclogenesis over the Western North Atlantic ocean. *Mon. Weather Rev.* **124**, 2672–2702 (1996).
43. Lackmann, G. M., Keyser, L. D. & Bosart, L. F. A characteristic life cycle of upper-tropospheric cyclogenetic precursors during the experiment on rapidly intensifying cyclones over the Atlantic (ERICA). *Mon. Weather Rev.* **125**, 2729–2758 (1997).
44. Martin, J. E., Grauman, R. D. & Marsili, N. Surface cyclolysis in the North Pacific Ocean. Part I: A synoptic climatology. *Mon. Weather Rev.* **129**, 748–765 (2001).
45. Martin, J. E. & Marsili, N. Surface cyclolysis in the North Pacific Ocean. Part II: piecewise potential vorticity analysis of a rapid cyclolysis event. *Mon. Weather Rev.* **130**, 1264–1281 (2002).
46. McLay, J. G. & Martin, J. E. Surface cyclolysis in the North Pacific Ocean. Part III: composite local energetics of tropospheric-deep cyclone decay associated with rapid surface cyclolysis. *Mon. Weather Rev.* **130**, 1282–1297 (2002).
47. Hart, R. E., Evans, J. L. & Evans, C. Synoptic composites of the extratropical transition life cycle of North Atlantic tropical cyclones: factors determining posttransition evolution. *Mon. Weather Rev.* **134**, 553–578 (2006).
48. Rudeva, I. & Gulev, S. K. Composite analysis of North Atlantic extratropical cyclones in NCEP–NCAR reanalysis data. *Mon. Weather Rev.* **139**, 1419–1446 (2011).
49. Sinclair, M. R. & Revell, M. J. Classification and composite diagnosis of extratropical cyclogenesis events in the Southwest Pacific. *Mon. Weather Rev.* **128**, 1089–1105 (2000).
50. Hanley, D., Molinari, J. & Keyser, D. A. Composite study of the interactions between tropical cyclones and Upper-Tropospheric troughs. *Mon. Weather Rev.* **129**, 2570–2584 (2001).
51. Wang, C. C. & Rogers, J. C. A composite study of explosive cyclogenesis in different sectors of the North Atlantic. Part I: cyclone structure and evolution. *Mon. Weather Rev.* **129**, 1481–1499 (2001).
52. Yuan, J. N. & Wang, D. X. Potential vorticity diagnosis of tropical cyclone USAGI (2001) genesis induced by a mid-level vortex in the South China sea. *Meteorol. Atmos. Phys.* **125**, 75–87 (2014).
53. Lau, N. C. & Crane, M. W. A satellite view of the synoptic-scale organization of cloud properties in midlatitude and tropical circulation systems. *Mon. Weather Rev.* **123**, 1984–2006 (1995).

54. Norris, J. R. & Iacobellis, S. F. North Pacific cloud feedbacks inferred from synoptic-scale dynamic and thermodynamic relationships. *J. Clim.* **18**, 4862–4878 (2005).
55. Bauer, M. & Del Genio, A. D. Composite analysis of winter cyclones in a GCM: influence on Climatological humidity. *J. Clim.* **19**, 1652–1672 (2006).
56. Chang, E. K. M. & Song, S. The seasonal cycles in the distribution of precipitation around cyclones in the Western North Pacific and Atlantic. *J. Atmos. Sci.* **63**, 815–839 (2006).
57. Field, P. R. & Wood, R. Precipitation and cloud structure in midlatitude cyclones. *J. Clim.* **20**, 233–254 (2007).
58. Field, P. R. et al. Midlatitude cyclone compositing to constrain climate model behavior using satellite observations. *J. Clim.* **21**, 5887–5903 (2008).
59. Boutle, I. A., Belcher, S. E. & Plant, R. S. Moisture transport in midlatitude cyclones. *Quart. J. Roy. Meteorol. Soc.* **137**, 360–373 (2011).
60. Booth, J. F., Wang, S. & Polvani, L. Midlatitude storms in a moister world: lessons from idealized baroclinic life cycle experiments. *Clim. Dyn.* **41**, 787–802 (2013).
61. Pfahl, S., O’Gorman, P. A. & Singh, M. S. Extratropical cyclones in idealized simulations of changed climates. *J. Clim.* **28**, 9373–9392 (2015).
62. Dacre, H. F., Hawcroft, M. K., Stringer, M. A. & Hodges, K. I. An extratropical cyclone atlas: A tool for illustrating cyclone structure and evolution characteristics. *Bull. Amer. Meteorol. Soc.* **93**, 1497–1502 (2012).
63. Zhao, Y., Fu, L., Yang, C. F. & Chen, X. F. Case study of a heavy snowstorm associated with an extratropical cyclone featuring a Back-Bent warm front structure. *Atmosphere* **11**, 1272 (2020).
64. Hart, R. E. A cyclone phase space derived from thermal wind and thermal asymmetry. *Mon. Weather Rev.* **131**, 585–616 (2003).
65. Catto, J. L., Shaffrey, L. C. & Hodges, K. I. Can climate models capture the structure of extratropical cyclones? *J. Clim.* **23**, 1621–1635 (2010).
66. Roebber, P. J. A diagnostic case study of Self-Development as an antecedent conditioning process in explosive cyclogenesis. *Mon. Weather Rev.* **121**, 976–1006 (1993).
67. Zhang, S., Tang, Y., Zhang, L., Liao, Q. & Zhang, T. Variations in key factors at different explosive development stages of an extreme explosive cyclone over the Japan sea. *Atmosphere* **14**, 1327 (2023).
68. Jaderberg, M., Simonyan, K., Zisserman, A. & Kavukcuoglu, K. Spatial transformer networks. *Proc. Mach. Learn. Res.* **28**, 2017–2025 (2015).
69. Liu, X. et al. Fully convolutional attention networks for fine-grained recognition. (2016). arXiv:1611.05244.
70. Cheng, D. S., Cristani, M., Stoppa, M., Bazzani, L. & Murino, V. Custom pictorial structures for re-identification. *Brit. Mach. Vis. Conf.* (2011).
71. Baltieri, D., Vezzani, R. & Cucchiara, R. Mapping appearance descriptors on 3d body models for people re-identification. *Int. J. Comput. Vis.* **111**, 345–364 (2015).
72. Zheng, L., Huang, Y., Lu, H. & Yang, Y. Pose invariant embedding for deep person reidentification. arXiv:1701.07732 (2017).
73. Zheng, Z., Zheng, L. & Yang, Y. Pedestrian alignment network for Large-scale person Re-Identification. *IEEE Trans. Circuits Syst. Video Technol.* **29**, 3037–3045 (2019).
74. Su, J., Lu, Y., Pan, S., Wen, B. & Liu, Y. RoFormer: Enhanced Transformer with Rotary Position Embedding. *Neurocomputing* **568**, C (2021).
75. Deng, L. The MNIST database of handwritten digit images for machine learning research. *IEEE Signal. Process. Mag.* **29**, 141–142 (2012).
76. Hua, W., Dai, Z., Liu, H. & Le, Q. Transformer quality in linear time. *Proc. Mach. Learn. Res.* **162**, 9099–9117 (2022).
77. Bleck, R. & Mattocks, C. A preliminary analysis of the role of potential vorticity in alpine Lee cyclogenesis. *Contrib. Phys.* **11**, 38–69 (1984).
78. Mattocks, C. & Bleck, R. Jet streak dynamics and geostrophic adjustment processes during the initial stages of Lee cyclogenesis. *Mon. Weather Rev.* **114**, 2033–2056 (1986).
79. Zupanski, M. & McGinley, J. A. Numerical analysis of the influence of jets, fronts, and mountain on alpine Lee cyclogenesis. *Mon. Weather Rev.* **117**, 154–176 (1989).
80. Zhao, L. N., Tu, N. N. & Jiao, M. Y. Comparison of two types duststorms weather on different dynamical and thermal background condition. *Clim. Environ. Res.* **12**, 329–338 (2007).
81. Ren, L. Y. & Dian, Y. N. Variability of winter extratropical cyclone activity in the North Pacific and its relationship with previous autumn Arctic sea ice. *Period. Ocean. Univ. China.* **48**, 20–29 (2018).
82. Renard, R. J. & Clarke, L. C. Experiments in numerical objective frontal analysis. *Mon. Weather Rev.* **93**, 547–556 (1965).
83. Berry, G., Reeder, M. J. & Jakob, C. A global climatology of atmospheric fronts. *Geophys. Res. Lett.* **38**, L04809 (2011).
84. Takahashi, N. An objective frontal data set to represent the seasonal and interannual variations in the frontal zone around Japan. *J. Meteorol. Soc. Japan.* **91**, 391–406 (2013).
85. Takahashi, N. A study of methods for making the frontal zone data around Japan. *Proc. Assoc. Japan. Geogr.* **2018s**, 196 (2018).
86. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, 6000–6010 (2017).
87. Dauphin, Y. N., Fan, A., Auli, M. & Grangier, D. Language modeling with gated convolutional networks. *Proc. Mach. Learn. Res.* **34**, 933–941 (2017).
88. Hersbach, H. et al. The ERA5 global reanalysis. *Quart. J. Roy. Meteorol. Soc.* **146**, 1999–2049 (2020).
89. NOAA National Centers for Environmental Information. *ETOPO 2022 15 Arc-Second Global Relief Model* (NOAA National Centers for Environmental Information, 2022).
90. Loshchilov, I. & Hutter, F. Fixing weight decay regularization in Adam. arXiv:1711.05101 (2017).
91. Loshchilov, I. & Hutter, F. SGDR: stochastic gradient descent with warm restarts. arXiv:1608.03983 (2016).
92. Ishida, T., Yamane, I., Sakai, T., Niu, G. & Sugiyama, M. Do we need zero training loss after achieving zero training error? *Proc. Int. Conf. Mach. Learn.* (2020).
93. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. *IEEE Conf. Comput. Vis. Pattern Recognit.* 2818–2826 (2016).

Acknowledgements

This work was jointly supported by National Natural Science Foundation of China (42288101), Shandong Natural Science Foundation Project (ZR2019ZD12) and National Natural Science Foundation of China (NSFC) Project (42075025).

Author contributions

J.P.L., Y.N.D. and R.P.S. contributed equally to this work. J.P.L. conceived the idea. R.P.S. performed all calculations and wrote the initial manuscript with the help of J.P.L. and Y.N.D. Both authors contributed to analyses, interpretation and writing of results.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.D. or J.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025