atmosphere

MDPI

Article

Comparison of the Applicability of Mainstream Objective Circulation Type Classification Methods in China

Minjin Ma *, Ran Chen and Xingyu Zhang

College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China; chenr2023@lzu.edu.cn (R.C.); zhxingyu2024@lzu.edu.cn (X.Z.)

* Correspondence: minjinma@lzu.edu.cn

Abstract

Circulation type classification (CTC) is an important method in atmospheric sciences, which reveals the relationship between atmospheric circulation and regional weather and climate. Accurate circulation classification helps to improve weather forecasting accuracy and supports climate change research. China has complex topography and significant spatiotemporal variability in its circulation patterns, making the study of circulation type classification in this region highly significant. This study aims to evaluate the applicability of several mainstream objective CTC methods in the China region. We applied methods including T-mode principal component analysis (PCT), Ward linkage, K-means, and Self-Organizing Maps (SOM) to classify the sea-level pressure daily mean fields from 1993 to 2023 in the study area, and compared the classification results in terms of internal metrics, continuity, seasonal variation, separability of related meteorological variables (e.g., temperature, precipitation), and stability to spatiotemporal resolution. The results show that each method has its advantages in different contexts, with the K-means method showing the best overall performance. Additionally, an optimized approach combining PCT and K-means is proposed.

Keywords: objective circulation type classification; T-mode principal component analysis; hierarchical clustering; K-means clustering; self-organizing maps; algorithm comparison

check for updates

Received: 17 September 2025
Revised: 12 October 2025
Accepted: 21 October 2025
Published: 24 October 2025

Citation: Ma, M.; Chen, R.; Zhang, X. Comparison of the Applicability of Mainstream Objective Circulation Type Classification Methods in China. *Atmosphere* **2025**, *16*, 1231. https://doi.org/10.3390/atmos16111231

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Atmospheric circulation patterns and their variations have significant impacts on regional weather and climate. Therefore, accurately classifying and characterizing these patterns is crucial in weather analysis and climate research [1–4]. Circulation Type Classification (CTC) divides the circulation patterns into several Circulation Types (CTs) with significant differences, helping researchers identify key features of circulation patterns associated with weather events and climate change. CTC methods are typically categorized into three categories: subjective classification, objective classification, and hybrid classification [5]. Subjective classification relies on meteorologists' knowledge and experience, typically having strong specificity and being inefficient and unsystematic in handling large amounts of data. Hybrid classification attempts to automate the classification process based on predefined standards, offering some general applicability but often leading to redundant CTs [6,7]. In contrast, objective classification methods use algorithms to automatically analyze and classify atmospheric circulation patterns. enabling the systematic and adaptive capture of the main features and variations in circulation patterns, reducing human subjectivity. These methods typically offer better general applicability [8–10]. With the growth of

meteorological data and advancements in computer technology, various objective methods have emerged. Currently, mainstream objective classification methods include T-mode rotated principal component analysis (PCT), hierarchical clustering, K-means clustering, and self-organizing maps (SOM), etc. [11–14].

Different objective CTC methods attempt to extract the main features of circulation patterns through various metrics and strategies, aiming to obtain accurate and reasonable types that reflect circulation differences. There is a wide variety of objective CTC methods, each with its own advantages and disadvantages, making the selection of the most suitable method a challenge. Huth [15] conducted an early and relatively comprehensive comparative study of objective CTC methods, comparing the effectiveness of five classification methods (correlation method, sums-of-squares method, average linkage, K-means, and PCT) in classifying 700 hPa geopotential heights in the European and adjacent Atlantic regions. The results showed that the CTC based on K-means clustering had stronger separability compared to other classification methods, but as the data dimensionality increased, distance metrics gradually became ineffective, making it difficult to find valid clusters, a phenomenon known as the "curse of dimensionality" [16], which led to insufficient stability and reliability of the classification results [14,17]. Additionally, PCT has been shown to have advantages in reproducing predefined CTs, but it is more sensitive to outliers and can be influenced by extreme cases, causing shifts in the direction of the principal components [18,19]. These studies indicate that no single classification method performs best in all aspects (the no free lunch theorem [20]). Furthermore, the European COST733 project (European Cooperation in Science and Technology) provided a more systematic evaluation of different CTC methods and, by establishing a consistent classification catalog, offered important references for weather and climate research in Europe [5,21].

China is located in the eastern part of Asia, with vast territory and complex geographical conditions, resulting in significant spatiotemporal variability in its circulation patterns. In the study of regional circulation patterns in China, some research based on objective classification methods has been conducted. For example, Liu et al. [22] used SOM to classify the 500 hPa geopotential height over the Tibetan Plateau and correlated it with regional precipitation, revealing the impact of circulation changes on regional precipitation patterns from 1961 to 2010. Sun et al. [23] applied hierarchical clustering to classify sea-level pressure, 850 hPa relative humidity, and wind fields during the formation and dissipation of pollution weather in the Sichuan Basin, revealing the influence of different meteorological patterns (such as high-pressure systems and weak high-pressure systems) on pollution events. Yi et al. [6] used the K-means method to classify dust storm weather in northern China and found that certain cyclone patterns were closely related to the occurrence of dust storms. Liu et al. [24] used SOM and K-means to classify sea-level pressure during ozone pollution events in the Guangzhou region from 2015 to 2022, finding strong consistency between the two methods in classification results. However, there is relatively little research on the applicability of CTC methods in China, and a systematic evaluation and comparison of the advantages and disadvantages of different methods has not been conducted. Choosing the appropriate CTC method to achieve more accurate classification has become a key issue in current research.

This study employs several mainstream objective CTC methods (PCT, Ward linkage, K-means, and SOM) to classify sea-level pressure fields in the China region, evaluating the applicability of each method from multiple perspectives. Additionally, an integrated classification approach combining PCT and K-means is proposed. The aim is to provide valuable guidance for selecting suitable CTC methods in future weather and climate research, both in China and globally.

2. Materials and Methods

2.1. Reanalysis Data

The data used in this study for classification are from the ERA5 reanalysis sea level pressure data provided by the European Centre for Medium-Range Weather Forecasts [25]. The spatial coverage is from 15° N to 60° N and from 70° E to 140° E (as shown in Figure 1), with a resolution of 0.25° × 0.25°. The time span is from 1 January 1993 to 31 December 2023, with a daily resolution. In addition, the average 2 m temperature and total precipitation reanalysis data from the central region of the study area (36° N–39° N, 103° E–107° E, as indicated by the red area in Figure 1) were used to conduct an effectiveness test of the related meteorological elements in the classification results.

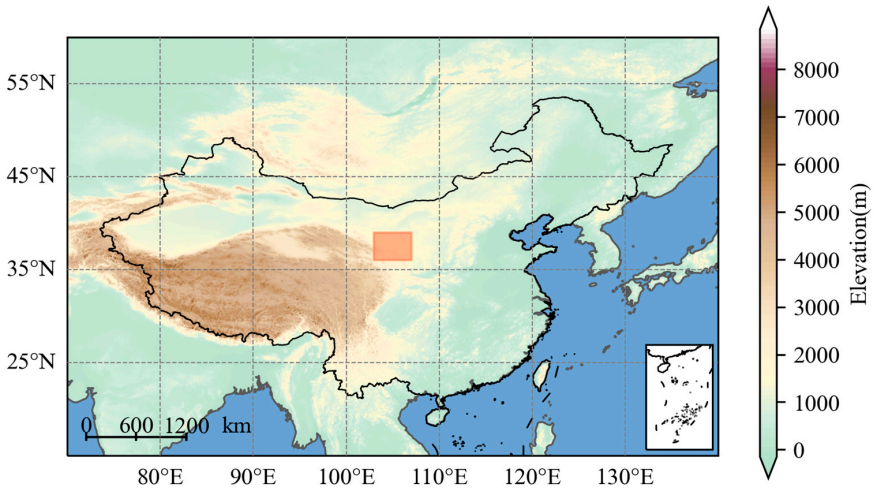


Figure 1. Topographic map of the study area (the red area is the region for testing the separability of meteorological variables).

2.2. Objective Circulation Type Classification Methods

The objective CTC methods compared in this study include the T-mode rotated principal component method (PCT), Ward Linkage, K-means, and Self-Organizing Maps (SOM). Among them, PCT compared the classification effects of orthogonal rotation (varimax) and oblique rotation (oblimin).

2.2.1. T-Mode Rotated Principal Component Analysis

Principal Component Analysis is a method that maps data from a high-dimensional space to a lower-dimensional space, retaining as much variability as possible while reducing dimensions [19]. In PCT, time points are treated as variables, and spatial points are treated as samples, with the goal of identifying virtual time points (i.e., principal components) that have significant variability in spatial modes [11]. The specific method is as follows:

(1) Standardize the data matrix Z:

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \tag{1}$$

where X_{ij} represents the element in the i-th row and j-th column of the original data matrix, μ_j is the mean of the j-th column, σ_j is the standard deviation of the j-th column, and Z_{ij} is the element in the i-th row and j-th column of the standardized data matrix.

(2) Calculate the covariance matrix C:

$$C = \frac{1}{n-1} Z^T Z \tag{2}$$

where Z is the standardized data matrix, and n is the number of samples.

(3) Solve for the eigenvalue λ and eigenvector v:

$$Cv = \lambda v \tag{3}$$

where v represents the direction of the new time points, and λ is the variance of all spatial points projected onto the corresponding v (i.e., spatial variability). The top k eigenvectors corresponding to the largest λ are selected as the principal components to retain the majority of the variability. The elbow method or cumulative variance explained is generally used to help determine the value of k.

(4) Calculate the loading matrix of the principal components L:

$$L = \sqrt{\lambda}v \tag{4}$$

the values in the load matrix L reflect the correlation between the spatial modes at each time point in the original data and the projection of all spatial points onto the principal components (i.e., the principal component scores).

(5) Rotation of loadings: Since loadings represent the correlation between each time point and the scores of each principal component, time points can be assigned to different principal components based on the absolute values of the loadings. However, during the previous process of solving for the principal components, which only focuses on maximizing the projection variance, the obtained loadings are often entangled (i.e., a time point has high loadings on multiple principal components), making classification difficult. Therefore, it is necessary to rotate the principal components to polarize the absolute values of the loadings. Common rotation methods include varimax and oblimin rotations. The former is an orthogonal rotation, where the rotated principal components remain orthogonal (i.e., uncorrelated), while the latter is an oblique rotation, which does not constrain the principal components to remain orthogonal. We will refer to the PCT with varimax rotation as PCTV and to that with oblimin rotation as PCTO. Previous studies have shown that the principal component scores obtained through oblique rotation are closer to the true spatial modes of the original data [26], but there is no conclusive evidence regarding which rotation is more advantageous for instance assignment.

(6) Classify based on the rotated loadings: Assign each time point to the principal component with the largest absolute loading, and divide into positive and negative categories based on the loading sign.

2.2.2. Ward Linkage

Hierarchical clustering is a method that constructs a tree-like structure by gradually merging data points or clusters. During this process, different linkage methods are used to determine how to merge the clusters. Common linkage methods include single linkage, complete linkage, average linkage, centroid linkage, and Ward linkage, each using different criteria for measuring the distance between clusters, which in turn affects the clustering results. Among these, Ward linkage minimizes the increase in within-cluster variance, effectively reducing the risk of excessive merging of clusters and ensuring better internal consistency. This helps avoid the snowball effect [27] that can occur with other linkage methods, where most of the data points are merged into one large cluster while outlier data points are placed in small, separate clusters. Therefore, in this study, we selected Ward linkage as the preferred linkage method.

(1) Initialize clusters: Each sample (time point) initially acts as an independent cluster.

(2) Merge clusters based on the increment of within-cluster variance: For the cluster C_k with center μ_k, its within-cluster variance is:

$$V_k = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{5}$$

the initial sum of variances is 0. For each pair of clusters C_i and C_j, the variance increment after their merger is:

$$\Delta V = \frac{|C_i||C_j|}{|C_i| + |C_j|} \cdot \|\mu_i - \mu_j\|^2 \tag{6}$$

where $|C_i|$ and $|C_j|$ are the number of samples in clusters C_i and C_j, respectively, μ_i and μ_j are their cluster centers, and $\|\mu_i - \mu_j\|$ is the Euclidean distance between the two cluster centers. The clusters with the smallest variance increment are chosen for merging.

(3) Repeat step (2) until all samples are merged into a single cluster or the predetermined number of clusters is reached.

2.2.3. K-Means

K-means is an iterative clustering algorithm. Its core idea is to partition the samples (time points) in the data into k clusters in such a way that the sum of the distances from each sample to its assigned cluster center (centroid) is as small as possible.

(1) Initialize k cluster centers C: At the beginning of the algorithm, k samples are randomly selected as the initial cluster centers. The choice of these initial cluster centers can have a certain impact on the final clustering result [28].

(2) Assign samples to the nearest cluster center: For each sample point, calculate its distance to each cluster center and assign it to the cluster center that is closest. Euclidean distance is used as the distance metric here:

$$d(x_i, \mu_k) = \|x_i - \mu_k\| = \sqrt{\sum_{j=1}^{d} \left(x_{ij} - \mu_{kj}\right)^2} \tag{7}$$

where x_i is the i-th sample, μ_k is the k-th cluster center, d is the dimensionality of the sample (number of space dimensions), and x_i and μ_{kj} are the values of the j-th dimension of x_i and μ_k, respectively.

(3) Update cluster centers: Recalculate the cluster centers for each cluster:

$$\mu_k^{new} = \frac{1}{|C_k|} \sum_{x_i \in S_k} x_i \tag{8}$$

where $|C_k|$ represents the number of samples in cluster C_k.

(4) Iterate and optimize: Repeat steps (2) and (3) until the cluster centers converge.

2.2.4. Self-Organizing Map

SOM is a type of artificial neural network, commonly used for data dimensionality reduction, clustering, and visualization [29]. Unlike the currently popular deep neural networks optimized using backpropagation, SOM gradually optimizes the network through a competitive learning strategy (where neurons compete with each other), and it has only two layers: the input layer and the competition layer. As a clustering algorithm, the basic idea of SOM is similar to K-means, both aiming to group samples that are close in distance, but the specific optimization methods differ.

(1) Build model: The number of neurons in the input layer is determined by the data dimension. The competition layer, also known as the output layer, usually has a

lower dimensional distribution (1–2 dimensions), with the number of neurons equal to the number of clusters (k). Each neuron has a weight vector of the same dimension as the sample, so each neuron can be considered as a "cluster center." The learning process of SOM involves mapping each sample to a neuron in the competition layer, and ensuring that samples closer in distance are mapped to neurons in the competition layer that are also geometrically close.

(2) Initializing node weights: The weights of the competition layer neurons are initialized with small random values or random samples. In cyclic clustering, initializing with small random values can lead to situations where certain neurons are unable to win during the competition process, resulting in instability in the number of cyclic clusters. Therefore, random sample initialization is typically used.

(3) Training model: In each training iteration, a sample (time point) data x_i is randomly selected, and the distance (Euclidean distance in this case) between it and the weights of each neuron W_j is calculated. The neuron with the smallest distance is chosen as the Best Matching Unit (BMU), and the weights of the BMU and its neighboring neurons are updated. The update rule is as follows:

$$W_j(t+1) = W_j(t) + \eta(t) \cdot h_{j,BMU}(t) \cdot \left(x_i - W_j(t)\right) \tag{9}$$

where $\eta(t) \in [0,1]$ is the learning rate, which determines the step size for weight updates, and $h_{j,BMU}(t)$ is the neighborhood function that controls the decay of weight updates with respect to the neighborhood radius. Common neighborhood functions include Gaussian and Bubble. In this study, a smoother Gaussian function is used as the neighborhood function:

$$h_{j,BMU}(t) = e^{-\frac{\|r_j - r_{BMU}\|^2}{2\sigma^2(t)}} \tag{10}$$

where r_j and r_{BMU} are the positions of neuron j and the BMU, respectively, and $\sigma(t)$ is the neighborhood radius. This process continues until the entire dataset is traversed, which is called one epoch. To improve the convergence speed and stability of the training process, as the number of training epochs increases, $\eta(t)$ and $\sigma(t)$ gradually decay, helping the model to learn quickly in the early stages and fine-tune in the later stages. The decay strategy used in this study is progressive decay:

$$\eta(t) = \eta_0 \left(1 + \frac{2t}{max_epoch}\right)^{-1} \tag{11}$$

$$\sigma(t) = \sigma_0 \left(1 + \frac{2t}{max_epoch}\right)^{-1} \tag{12}$$

where η_0 and σ_0 are the initial learning rate and neighborhood radius, respectively, and *max_epoch* is the maximum number of epochs, ensuring that the learning rate and neighborhood radius gradually approach one-third of their initial values.

(4) Activate mapping: Use the trained SOM to map the samples to their respective BMUs, completing the clustering process.

2.3. Evaluation Indicators

In CTC, it is difficult to provide an absolutely correct classification standard [30]. Therefore, internal evaluation metrics based on the structure of the clustering results themselves are used to assess the classification performance of various methods, without the need for true labels or external standards.

Common internal evaluation metrics for assessing clustering performance include Explained Variance (EV), Calinski–Harabasz index (CH), Silhouette Coefficient (SC), and

Pattern Correlation Ratio (PCR). The first three are metrics related to intra-cluster compactness and inter-cluster separation based on Euclidean distance measures. EV only focuses on intra-cluster compactness, while SC places too much emphasis on inter-cluster separation. Although CH balances both, its value range is undefined, and the value itself has no intrinsic meaning, only comparative significance. Therefore, this study proposes a Euclidean Distance Ratio (*EDR*) to evaluate the intra-cluster compactness and inter-cluster separation of the results:

$$EDR = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\overline{d(x_i, C_{in})}}{\overline{d(x_i, C_{out})}}\right) \tag{13}$$

where $\overline{d(x_i, C_{in})}$ represents the average Euclidean distance between sample x_i and samples within the same cluster, while $\overline{d(x_i, C_{out})}$ represents the average Euclidean distance between sample x_i and samples from different clusters. *PCR* is a metric based on linear relationships, considering both intra-cluster positive correlation and inter-cluster null/negative correlation. However, its numerical change direction is opposite to that of *EDR* [15]. To enhance the readability of this study, we have made certain improvements based on *PCR*:

$$PCR = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{1 - \overline{r(x_i, C_{in})}}{1 - \overline{r(x_i, C_{out})}}\right) \tag{14}$$

where $\overline{r(x_i, C_{in})}$ represents the average Pearson correlation coefficient between sample x_i and the samples within the same cluster, while $\overline{r(x_i, C_{out})}$ represents the average Pearson correlation coefficient between sample x_i and the samples from different clusters. The upper bound of both *EDR* and *PCR* is 1. When the value is greater than 0, it indicates that the algorithm has achieved effective clustering; the larger the value, the better the clustering performance of the algorithm.

In addition to the above internal evaluation metrics, the classification results of each method were also evaluated from the aspects of same-type duration, differences in related meteorological factors, and seasonal variation.

3. Results

3.1. Determination of the Number of Types K

Before performing classification using PCTV, PCTO, Ward linkage, K-means, and SOM on the daily average sea-level pressure data of China from 1 January 1993 to 31 December 2023, it is necessary to manually set the number of CTs, K, i.e., the number of clusters. The elbow method based on the sum of squared errors is a common auxiliary approach for determining the number of clusters, but its effectiveness depends on the data distribution. When the data points are evenly distributed and the differences between clusters are clear, the elbow is easily identifiable. In contrast, when the data distribution is complex or contains outliers, the elbow may be unclear or even absent. In our study, due to the large spatial range, long time span (i.e., high feature dimensionality and many data points), and complex data distribution with some outliers, the elbow method fails. This makes it difficult to determine the optimal K value. However, since the focus of our study is to compare the effectiveness of different classification methods, the optimal K value is not necessary for our research. It is sufficient to determine a consistent K value, and we can then compare how different methods recognize the data structure and handle noise and outliers. Therefore, we chose K = 12, which strikes a balance between ensuring rich clustering and minimizing computational resource requirements, while also explaining over 70% of the variance in PCT. Additionally, the sea-level pressure data underwent anomaly pattern adjustment during preprocessing. This adjustment does not affect the classification results of Ward linkage, K-means, or SOM, which are based on Euclidean distance. However,

for PCT, which uses a linear relationship metric, there is a significant difference between the results when using the anomaly pattern as opposed to the original pattern [11]. The primary issue with using the original pattern in PCT is that it is influenced by the climatic mean state, causing most of the loadings to concentrate on the same principal component, leading to very few instances within certain CTs. This seems more like a detection method rather than a good classification method. In contrast, when using the anomaly pattern, the influence of the climatic mean state is removed, resulting in a more even distribution of loadings and more balanced results. This ensures consistency in the number of types across all methods, which is the main reason for using the anomaly pattern in PCT in this study.

3.2. Internal Metrics

Internal evaluation metrics do not rely on external labels or prior knowledge. They judge whether the clustering algorithm can reasonably partition the data into clusters with high similarity and distinguishable differences between clusters, based solely on the clustering results and the data itself. The effectiveness of internal evaluation metrics is crucial for CTC as a method within the clustering analysis framework. The two internal evaluation metrics (EDR and PCR) for the classification results of different methods are shown in Figure 2. It can be observed that the PCT-based classification method consistently achieves higher PCR than the other methods, while the opposite is true for EDR. This indicates that PCT is stronger in capturing the "shape" characteristics but weaker in capturing the "value" characteristics, whereas Ward linkage, K-means, and SOM perform the opposite. This is not surprising, as we know from the method explanation that PCT is based on linear relationship metrics, while Ward linkage, K-means, and SOM are based on Euclidean distance metrics. This suggests that, in specific studies, the appropriate classification method should be chosen based on the focus of the research (whether on the correlation or intensity of circulation).

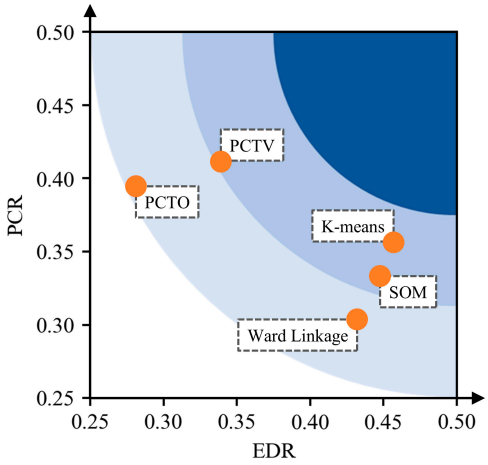


Figure 2. EDR and PCR of the circulation type classification methods.

Between the two rotation methods in PCT, the orthogonal rotation method Varimax outperforms the oblique rotation method Oblimin in both PCR and EDR, which seems to contradict existing studies (which suggest that oblique rotations capture more realistic spatial modes). However, we still attempt to provide a reasonable explanation for this result: oblique rotations do not constrain the orthogonality of the principal components, meaning that after rotation, the principal component scores (PC scores) are correlated with each other. Additionally, oblique rotations tend to bring as many sample points as possible closer to the principal components, which results in more balanced variance of the scores, as shown in Figure 3. This is why oblique rotations can capture more realistic spatial modes (which are correlated and more balanced in reality). However, capturing more realistic

modes does not necessarily lead to better clustering performance. We know that loadings represent the correlation between the spatial modes of the original sample points and the scores of the principal components. The more similar the scores of the principal components are, the more likely it is that the original sample points become indistinguishable (at least in the strategy of using the maximum absolute loading for classification). Classification requires not only more realistic but also more separable principal component projections. The figure demonstrates the correlation between the principal component scores for each rotation method, which to some extent supports this argument.

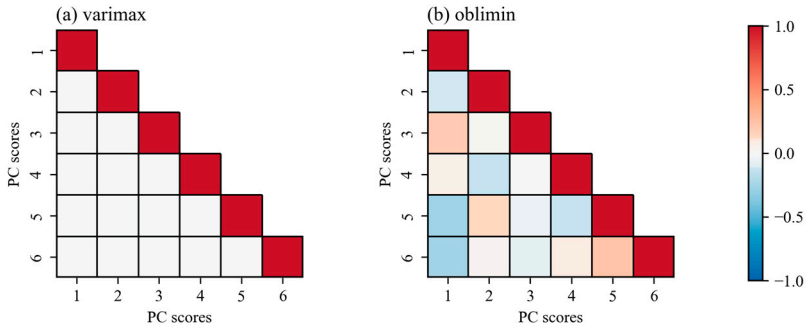


Figure 3. The correlation between the principal component scores of 2 rotation methods: (**a**) varimax; (**b**) oblimin.

Ward linkage, K-means, and SOM are all methods based on Euclidean distance metrics for CTC. Among them, Ward linkage shows poorer performance in both EDR and PCR, but because it does not involve iterative optimization, it has a result stability that the other two methods do not have. In other words, given a fixed dataset, the classification results are also fixed. The EDR and PCR values for K-means and SOM in Figure 2 represent the average of 100 runs. K-means selects random samples as the initial centroids and sets a convergence threshold of 0.001 or a maximum of 300 iterations. SOM selects random samples as the initial weights for the neurons, with the initial learning rate randomly set within the range of [0.01, 0.99], and the initial neighborhood radius randomly set within the range of [1, 3], training for 10 epochs. It can be seen that both methods have similar average EDR and PCR values, but SOM has greater parameter tuning difficulty and a more complex training process. Meanwhile, K-means requires the least computational resources among the three methods.

3.3. Continuity

Large-scale weather patterns tend to exhibit both continuous and stable evolution processes, as well as rapid transitional changes. This pattern aligns with the geostrophic adjustment theory and practical experience. Therefore, good classification results should demonstrate more continuity and fewer isolated patterns. Table 1 shows the proportion of different duration patterns in the classification process, including isolated events lasting only 1 day, short events lasting 2–3 days, medium events lasting 4–7 days, long events lasting 8–15 days, and super-long events lasting more than 16 days. From the table, it can be observed that short-term events account for the highest proportion across all methods, with the PCTO method exhibiting a particularly large proportion of short-term events in its classification results. At the same time, PCTO shows the most isolated events and the fewest medium, long, and super-long events, indicating weaker classification continuity. In contrast, the K-means method exhibits the fewest isolated events and performs relatively well in terms of medium and long-term events. However, since the true classification result cannot be determined, the possibility of false super-long events cannot be ruled out.

Nonetheless, from the perspective of minimizing isolated events, K-means demonstrates the most reasonable continuity in its classification structure.

Table 1. The proportion of the same circulation type sustained over multiple consecutive days.

Method	1 d (%)	2–3 d (%)	4–7 d (%)	8–15 d (%)	16 d+ (%)	Average Duration (d)
PCTV	19.35	34.42	23.06	9.48	13.63	7.75
PCTO	24.86	45.40	23.79	5.80	0.14	3.11
Ward Linkage	15.24	31.00	25.73	14.11	13.90	8.54
K-means	14.05	35.44	28.03	11.48	10.99	7.09
SOM	15.04	34.04	26.35	10.95	13.62	8.71

3.4. Seasonal Variability

The dominant circulation patterns typically vary across different seasons, and seasonal variability becomes an important indicator for evaluating the effectiveness of these methods. By analyzing the variability of CTs across different seasons, we can identify which methods reflect seasonal circulation pattern changes and which methods are unaffected by seasonal factors, leading to suboptimal classification results.

Figure 4 shows the seasonal variability of the classification results for five methods. The variability can be roughly categorized into three levels. Methods with higher variability include Ward linkage, K-means, and SOM, where the CTs in summer and winter do not overlap, and certain CTs appear only in specific seasons, showing significant seasonal variation. Next are PCT-V, where there is some overlap of CTs between winter and summer, but it still shows some seasonal variation. The methods with lower variability are PCTO, where the differences between CTs across seasons are minimal, and most CTs appear in all seasons, with the results generally being more uniform. It can be observed that classification methods based on Euclidean distance measurements are better at capturing seasonal variability compared to those based on linear relationship measurements.

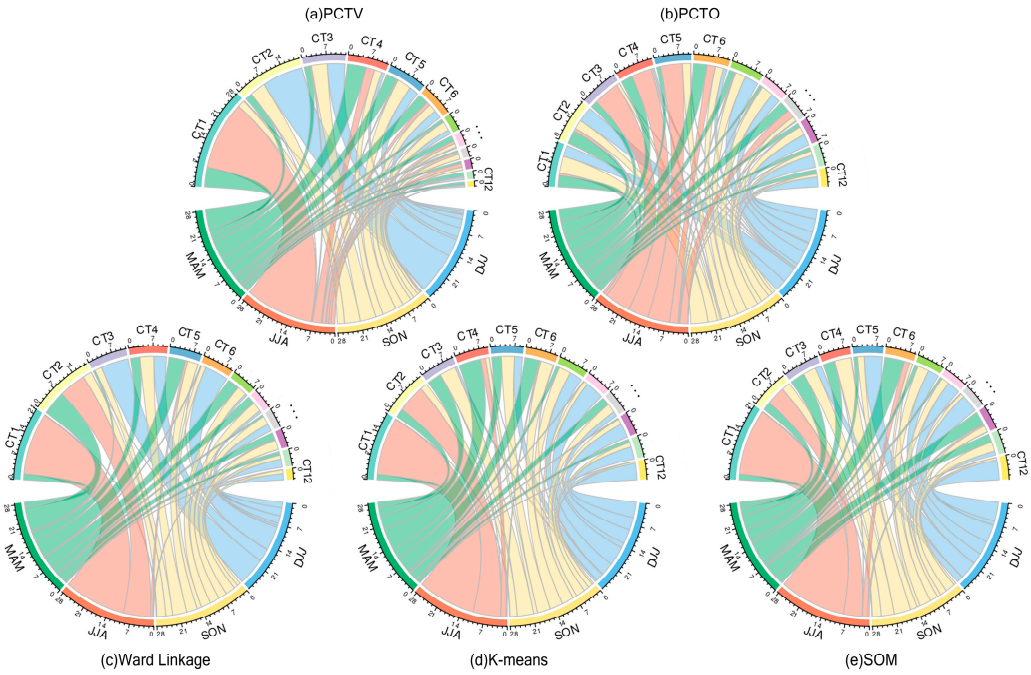


Figure 4. The seasonal variability of the classification types of 5 circulation type classification methods: (**a**) PCTV; (**b**) PCTO; (**c**) Ward Linkage; (**d**) K-means; (**e**) SOM.

3.5. Separability of Meteorological Elements

The effectiveness of CTC methods is not only reflected in their ability to accurately capture the differences between circulation patterns themselves but also in whether the classification results (CTs) implicitly capture the differences in certain related meteorological variables. To evaluate the quality of CTC methods, it's necessary to assess whether there are significant differences in the relevant meteorological variables between the obtained CTs, especially the separability of key meteorological variables such as temperature and precipitation. If a CTC method produces CTs that clearly show changes in certain meteorological variables, the classification performance of the method is considered superior. On the other hand, if a method fails to effectively distinguish these meteorological elements, resulting in overlap or confusion between different CTs, its classification accuracy and practical value will be significantly reduced.

3.5.1. Temperature

Temperature, as a sensitive variable responding to circulation patterns, often exhibits significant differences under different circulation patterns. For example, when a region is under the control of the westerlies, the temperature may be relatively mild, while in high-pressure ridge areas or surface high-pressure systems, the temperature is typically lower. By examining whether there are significant temperature differences in the CTs, we can assess the usability of the CTs. Figure 5 displays the distribution characteristics of the 2 m temperature mean in the central region of circulation fields (36° N–39° N, 103° E–107° E) belonging to different CTs from different CTC methods. Ideal CTs should have the following characteristics: the temperature distribution under the same CT should be highly concentrated, while the distributions of different CTs should be clearly distinguished, with minimal overlap. Therefore, the box plot should show relatively short boxes (indicating stronger concentration), and the overlap between the boxes of different CTs should be as small as possible, reflecting the superiority of CTs in capturing temperature differences. It can be observed that the methods such as Ward linkage, K-means, and SOM generally have shorter boxes and more dispersed box distributions compared to methods based on PCT.

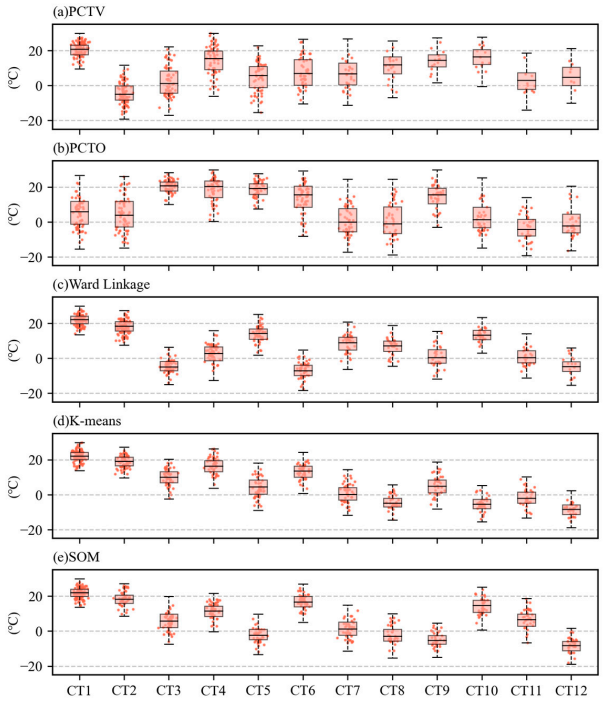


Figure 5. The 2 m temperature differences in the classification types of 5 circulation type classification methods: (**a**) PCTV; (**b**) PCTO; (**c**) Ward Linkage; (**d**) K-means; (**e**) SOM.

3.5.2. Precipitation

In addition to examining the distribution characteristics of temperature, the separability of precipitation is also an important evaluation metric. As a key manifestation of weather systems, precipitation's spatiotemporal distribution is significantly influenced by circulation patterns. Similarly to temperature fields, in an ideal CTC method, different CTs should effectively separate distinct precipitation characteristics. Specifically, precipitation distributions under the same CT should exhibit high similarity, while precipitation patterns under different CTs should show significant differences. Figure 6 shows the distribution characteristics of the precipitation mean in the central region of circulation fields (36° N–39° N, 103° E–107° E) belonging to different CTs from different CTC methods. Ideal CTs should have the same characteristics as those in the temperature field. However, in practice, there is generally overlap between the boxes of all methods in the low-value region. This is because, compared to temperature, precipitation is a variable with a practical lower boundary, and this drop can be easily and commonly reached. Therefore, it is difficult to judge the effectiveness of classification methods simply based on box overlap. However, we can still assess whether the precipitation distribution under the same CT is concentrated by examining the length of the boxes. It can be observed that methods such as Ward linkage, K-means, and SOM generally have more extremely short boxes compared to methods based on PCT, indicating that these methods have better precipitation separability (at least to some extent, reflecting whether precipitation occurs).

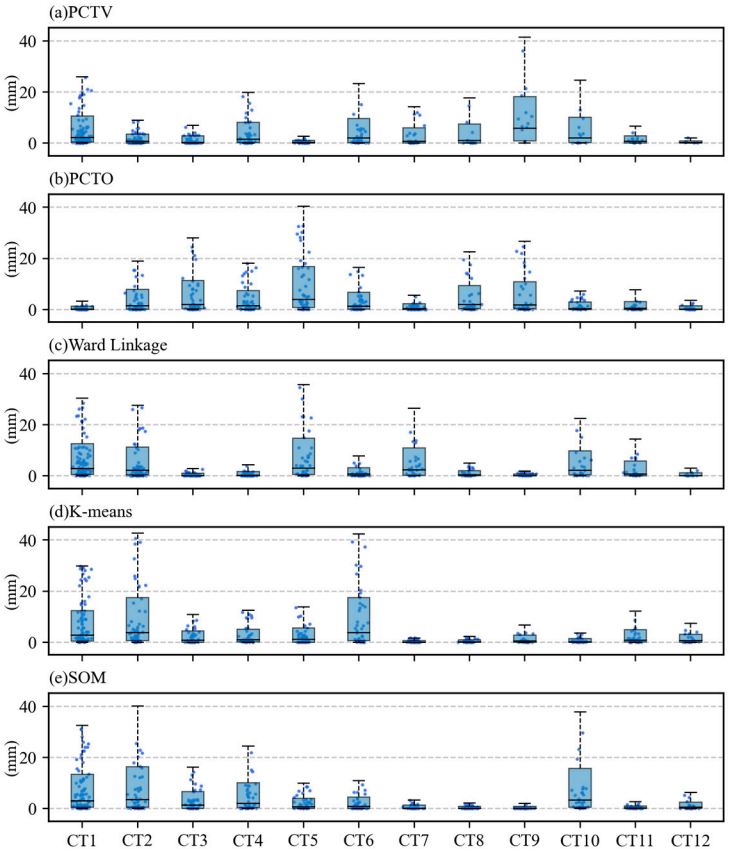


Figure 6. The precipitation differences in the classification types of 5 circulation type classification methods: (**a**) PCTV; (**b**) PCTO; (**c**) Ward Linkage; (**d**) K-means; (**e**) SOM.

3.6. Sensitivity of Temporal and Spatial Resolution

In order to assess the stability of different CTC methods, this study conducts sensitivity experiments on the spatial and temporal resolutions of the data. The aim is to investigate whether these changes significantly affect the stability and consistency of the classification

results. If the results of a classification method show minimal variation under different resolutions, it indicates that the method has strong stability with respect to spatiotemporal resolution. Conversely, significant changes in results may suggest that the method is sensitive to variations in spatiotemporal resolution and has poor stability.

3.6.1. Spatial Resolution

The accuracy and reliability of classification results are often significantly influenced by the spatial resolution of the data. Spatial resolution largely determines whether the classification method can accurately capture the details of circulation features. Higher spatial resolution typically reveals more refined circulation structures, whereas lower resolution may result in the loss of local variations, thereby affecting the accuracy and precision of the classification results. Therefore, this study reduces the spatial resolution of the data from $0.25^\circ \times 0.25^\circ$ to $1^\circ \times 1^\circ$, evaluating the performance of different CTC methods at lower spatial resolutions to assess their stability and reliability under various resolution conditions.

Figure 7 shows the normalized confusion matrix, where the classification results at the original resolution (Control group, Ctrl) are considered the true values, and the results at the lower resolution (Experimental group, Exp) are considered the predicted values. The percentage in the upper-right corner represents the macro-average recall, which is the mean of the main diagonal elements. The results indicate that the macro-average recall is higher for the PCT method, especially for PCT-V. Among the three methods based on Euclidean distance metrics, Ward Linkage excessively relies on the Euclidean distance between the data and lacks an adjustment or optimization process, leading to a lower macro-average recall. In contrast, K-means and SOM, although using an iterative strategy, result in some fluctuation in the classification outcomes. However, the fluctuation is relatively small, and they still exhibit better spatial stability.

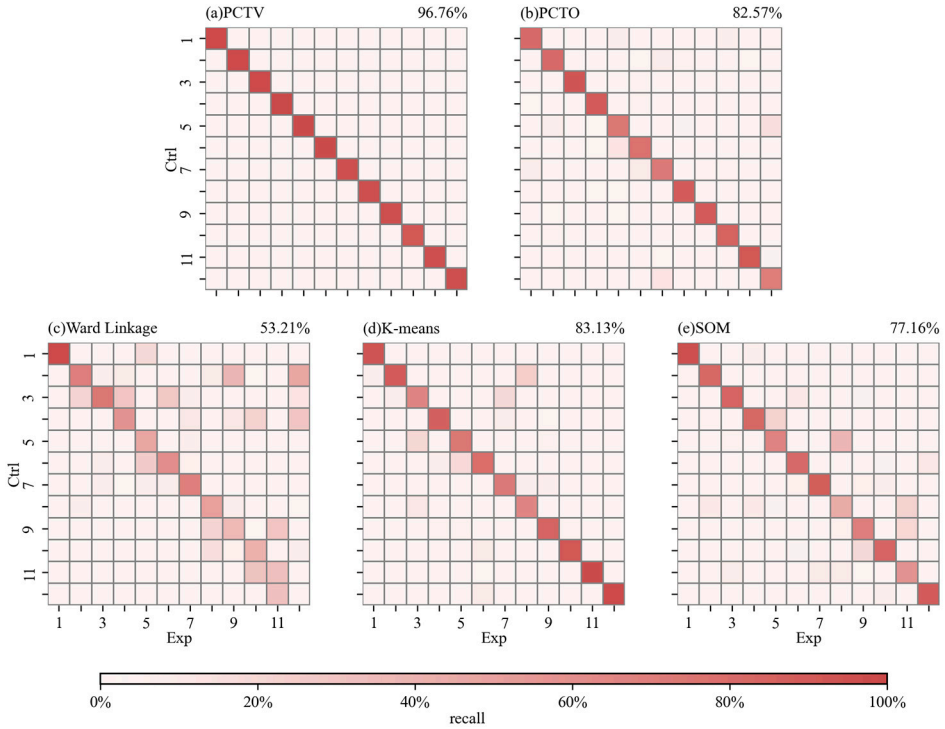


Figure 7. The sensitivity of 5 circulation type classification methods to spatial resolution: (**a**) PCTV; (**b**) PCTO; (**c**) Ward Linkage; (**d**) K-means; (**e**) SOM. (the control group represents the classification results of the original data, while the experimental group represents the classification results of the data after reducing spatial resolution).

3.6.2. Temporal Resolution

Since CTC essentially involves clustering at time points, time resolution has a more direct impact on the classification results. Within a fixed period, the size of the time resolution determines the number of circulation patterns. Fewer circulation patterns may lead to incomplete identification or increased errors in recognizing circulation patterns due to insufficient temporal information. On the other hand, more circulation patterns help improve classification accuracy by capturing richer details of circulation variations. Based on this, this study reduces the time resolution from 1 day to 4 days to investigate the stability of circulation classification under different numbers of circulation patterns.

Figure 8 is similar to Figure 7 but focuses on time resolution sensitivity. It can be observed that, compared to spatial resolution, these 9 classification methods are more sensitive to time resolution. Even PCT-V, which performs excellently in terms of spatial resolution, experiences a significant reduction here but still achieves the highest recall rate. Among the three methods based on Euclidean distance metrics, Ward Linkage still exhibits the worst time stability. This hierarchical clustering result, which gradually merges the data, is highly dependent on the data, making the results fully reproducible on a fixed dataset, but seemingly also increasing the data sensitivity.

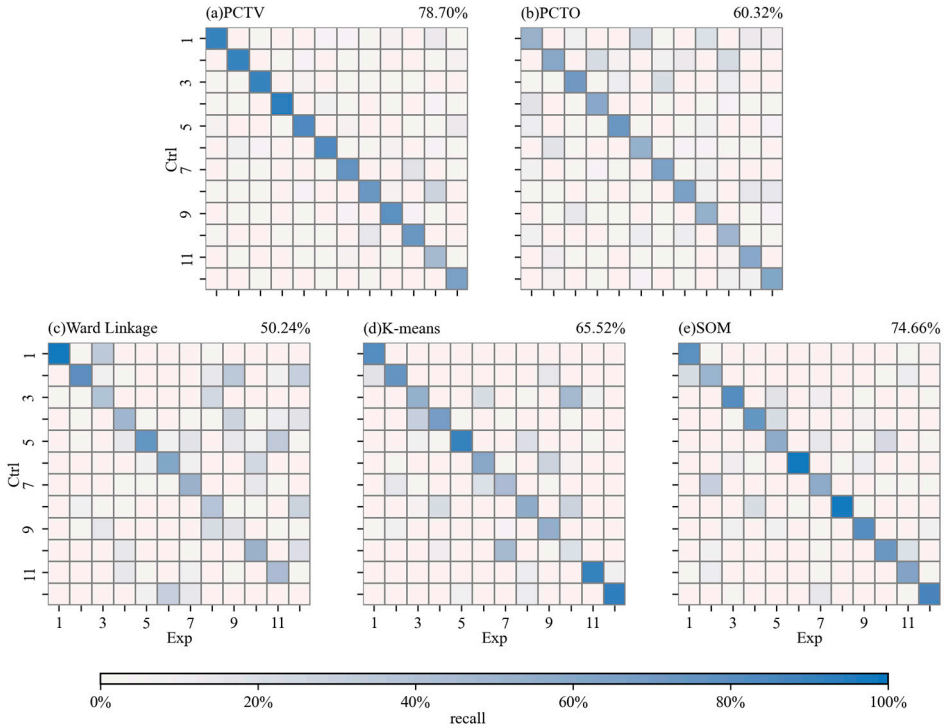


Figure 8. The sensitivity of 5 circulation type classification methods to temporal resolution: (**a**) PCTV; (**b**) PCTO; (**c**) Ward Linkage; (**d**) K-means; (**e**) SOM. (the control group represents the classification results of the original data, while the experimental group represents the classification results of the data after reducing spatial resolution).

3.7. Method Optimization

As mentioned in Section 2, PCT classifies based solely on the maximum absolute load at each time point, ignoring the loadings of other principal components. While this method may filter out some noise to a certain extent, it also loses a significant amount of information. To address this, the study proposes using K-means to perform more refined classification of principal component loadings, treating PCT as a feature extractor. Table 2 shows the changes in the internal evaluation metrics for the improved PCTV and PCTO methods, where both EDR and PCR show considerable increases.

Table 2. The EDR and PCR of the optimized PCTV and PCTO.

Method	EDR (Increase)	PCR (Increase)
PCTV	0.3564 (0.0174)	0.4405 (0.0285)
PCTO	0.3337 (0.0527)	0.4311 (0.0357)

In iterative optimization algorithms such as K-means and SOM, the initialization of centroids and neurons is crucial. This study proposes the use of principal component scores extracted by PCT as the initial centroids and neuron weights for these algorithms. We use normalized principal component scores and their negative modes superimposed with the climate mean state as the initialization centroids and neurons for K-means and SOM. For K-means, initialization using PCTV and PCTO outperformed the average results (after 100 runs) in terms of both EDR and PCR, with the final results using PCTV slightly higher than PCTO. For SOM, the results from initializing with PCTV and PCTO showed no difference, indicating that SOM is less sensitive to initialization than K-means. However, this does not mean that SOM is more stable; rather, SOM shifts its instability to being more sensitive to other hyperparameters (such as learning rate and neighborhood radius) and the training process (such as sample order and decay functions). Although these initialization methods did not lead to significant performance improvements, they helped stabilize the algorithm's results and ensure reproducibility in situations where other hyperparameters remain unchanged.

4. Discussion

This paper applies several commonly used objective CTC methods to classify sea level pressure fields in the China region, evaluating the performance of different methods in terms of internal metrics, persistence, seasonal variability, separability of related meteorological elements, and spatiotemporal stability. Although our study is based on the China region, the study area primarily serves as the weather and climate context for evaluation, rather than the main focus of the research itself. Therefore, our conclusions are also of certain general applicability and provide valuable insights.

PCT, Ward linkage, K-means, and SOM can be categorized into two groups based on their metric criteria: linear relationship metrics and Euclidean distance metrics. Methods based on linear relationships (PCT) tend to group more correlated circulation patterns together, which better captures the "shape similarity" of the circulation patterns. This is more important when conducting studies on larger-scale circulation features. In contrast, methods based on Euclidean distance (Ward linkage, K-means, and SOM) focus more on the overall intensity of the circulation patterns, which better captures the "value proximity" of the patterns, which is more advantageous for studies of medium and small-scale circulation features. Further research results indicate that classifications based on Euclidean distance better align with atmospheric patterns, such as having longer persistence, larger seasonal variability, and better differentiation of certain meteorological elements. This suggests that when considering factors like seasonal variability, methods such as K-means would be a good choice. Conversely, PCT would be a better choice for studies that do not require such considerations. In the PCT method, although oblique rotation can capture more realistic circulation patterns, it does not seem to offer an advantage in subsequent classifications. In fact, the correlation of principal components may lead to suboptimal classification results, and the spatiotemporal stability of the classification decreases.

Among Ward linkage, K-means, and SOM, all of which are based on the Euclidean distance metric, Ward linkage provides stable and reproducible results with fixed datasets, while the latter two methods, due to their iterative optimization process, produce slightly different results each time. Nevertheless, Ward linkage is more sensitive to spatial and

temporal resolution, and varying data resolutions can significantly affect its results. In contrast, the latter two methods maintain relatively high spatiotemporal stability as long as the convergence conditions are consistent. Among these, K-means, as a simple and effective classification method, has advantages in both classification results and computational efficiency. While SOM produces classification results similar to K-means, it involves a more complex parameter tuning process.

Additionally, the study considers the combined application of multiple methods to leverage their respective advantages and address the limitations of individual methods. In this research, PCT and K-means are combined. One strategy is to use K-means to help PCT cluster the loadings, thereby capturing detailed information. From the internal metrics of the classification results, this strategy improves the intra-cluster tightness and inter-cluster separation. Another strategy is to use the circulation patterns captured by PCT to replace the initial centroids in K-means, which makes the classification results reproducible while maintaining the original performance. A similar approach is applied to SOM, where the circulation patterns captured by PCT replace the initial neurons in SOM, making the results more stable. However, consistency cannot be fully achieved as it is in K-means, because SOM is also more influenced by the learning rate, neighborhood radius, and the order of training samples.

It is hoped that the comparison of these methods will provide theoretical and technical support for future climate research and extreme weather forecasting in China and similar regions globally. Future research can further explore the combined application of different classification methods to better serve climate prediction and the development of climate change adaptation strategies for the China region.

Author Contributions: Conceptualization, M.M. and R.C.; methodology, M.M.; software, R.C.; validation, M.M., R.C. and X.Z.; formal analysis, M.M.; investigation, R.C.; resources, M.M.; data curation, X.Z.; writing—original draft preparation, R.C. and M.M.; writing—review and editing, M.M. and X.Z.; visualization, R.C.; supervision, M.M.; project administration, M.M.; funding acquisition, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Gansu Provincial Science and Technology Plan Funding, grant number 25JRRA1129.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available at https://cds.climate.copernicus.eu/ (accessed on 16 January 2025).

Acknowledgments: We would like to thank the developers of scikit-learn for providing the machine learning algorithms used in this study. Their open-source library greatly contributed to the implementation and analysis of our models. The work is supported by Supercomputing Center of Lanzhou University. We also thank the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ERA-5 reanalysis data, which were essential for our analysis. During the preparation of this manuscript, the authors used ChatGPT (version 4.0) and Deepseek (v3.0) for translation and text editing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CTC	Circulation type classification
CT	Circulation type
PCT	T-mode principal component analysis
PCTV	T-mode principal component analysis with varimax rotation
PCTO	T-mode principal component analysis with oblimin rotation
SOM	Self-organizing map

References

1. García-Valero, J.A.; Montavez, J.P.; Jerez, S.; Gómez-Navarro, J.J.; Lorente-Plazas, R.; Jiménez-Guerrero, P. A seasonal study of the atmospheric dynamics over the Iberian Peninsula based on circulation types. *Theor. Appl. Climatol.* **2012**, *110*, 291–310. [CrossRef]
2. Lebedeva, M.G.; Lupo, A.R.; Chendev, Y.G.; Krymskaya, O.V.; Solovyev, A.B. Changes in the Atmospheric Circulation Conditions and Regional Climatic Characteristics in Two Remote Regions Since the Mid-20th Century. *Atmosphere* **2019**, *10*, 11. [CrossRef]
3. Szyga-Pluta, K. Large Day-to-Day Variability of Extreme Air Temperatures in Poland and Its Dependency on Atmospheric Circulation. *Atmosphere* **2021**, *12*, 80. [CrossRef]
4. Brázdil, R.; Zahradníček, P.; Dobrovolný, P.; Řehoř, J.; Trnka, M.; Lhotka, O.; Štěpánek, P. Circulation and Climate Variability in the Czech Republic between 1961 and 2020: A Comparison of Changes for Two "Normal" Periods. *Atmosphere* **2022**, *13*, 137. [CrossRef]
5. Huth, R.; Beck, C.; Philipp, A.; Demuzere, M.; Ustrnul, Z.; Cahynová, M.; Kyselý, J.; Tveito, O.E. Classifications of Atmospheric Circulation Patterns. *Ann. N. Y. Acad. Sci.* **2008**, *1146*, 105–152. [CrossRef]
6. Yi, Z.; Wang, Y.; Chen, W.; Guo, B.; Zhang, B.; Che, H.; Zhang, X. Classification of the Circulation Patterns Related to Strong Dust Weather in China Using a Combination of the Lamb–Jenkinson and k-Means Clustering Methods. *Atmosphere* **2021**, *12*, 1545. [CrossRef]
7. Chen, X.; Wang, N.; Wang, G.; Wang, Z.; Chen, H.; Cheng, C.; Li, M.; Zheng, L.; Wu, L.; Zhang, Q.; et al. The Influence of Synoptic Weather Patterns on Spatiotemporal Characteristics of Ozone Pollution Across Pearl River Delta of Southern China. *J. Geophys. Res. Atmos.* **2022**, *127*, e2022JD037121. [CrossRef]
8. Kyselý, J.; Huth, R. Changes in atmospheric circulation over Europe detected by objective and subjective methods. *Theor. Appl. Climatol.* **2006**, *85*, 19–36. [CrossRef]
9. Bodenheimer, S.; Nirel, R.; Lensky, I.M.; Dayan, U. Relationship between AOD and synoptic circulation over the Eastern Mediterranean: A comparison between subjective and objective classifications. *Atmos. Environ.* **2018**, *177*, 253–261. [CrossRef]
10. Beck, C.; Philipp, A. Evaluation and comparison of circulation type classifications for the European domain. *Phys. Chem. Earth* **2010**, *35*, 374–387. [CrossRef]
11. Huth, R. Properties of the circulation classification scheme based on the rotated principal component analysis. *Meteorol. Atmos. Phys.* **1996**, *59*, 217–233. [CrossRef]
12. Vesanto, J.; Alhoniemi, E. Clustering of the self-organizing map. *IEEE Trans. Neural Netw.* **2000**, *11*, 586–600. [CrossRef] [PubMed]
13. Philipp, A.; Bartholy, J.; Beck, C.; Erpicum, M.; Esteban, P.; Fettweis, X.; Huth, R.; James, P.; Jourdain, S.; Kreienkamp, F.; et al. Cost733cat—A database of weather and circulation type classifications. *Phys. Chem. Earth* **2010**, *35*, 360–373. [CrossRef]
14. Ahmed, M.; Seraj, R.; Islam, S.M. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics* **2020**, *9*, 1295. [CrossRef]
15. Huth, R. An Intercomparison of computer-assisted circulation classification methods. *Int. J. Climatol.* **1996**, *16*, 893–922. [CrossRef]
16. Assent, I. Clustering high dimensional data. *WIREs Data Min. Knowl. Discov.* **2012**, *2*, 340–350. [CrossRef]
17. Christiansen, B. Atmospheric Circulation Regimes: Can Cluster Analysis Provide the Number? *J. Clim.* **2007**, *20*, 2229–2250. [CrossRef]
18. Compagnucci, R.H.; Richman, M.B. Can principal component analysis provide atmospheric circulation or teleconnection patterns? *Int. J. Climatol.* **2008**, *28*, 703–726. [CrossRef]
19. Maćkiewicz, A.; Ratajczak, W. Principal components analysis (PCA). *Comput. Geosci.* **1993**, *19*, 303–342. [CrossRef]
20. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [CrossRef]
21. Demuzere, M.; Kassomenos, P.; Philipp, A. The COST733 circulation type classification software: An example for surface ozone concentrations in Central Europe. *Theor. Appl. Climatol.* **2011**, *105*, 143–166. [CrossRef]
22. Liu, W.; Wang, L.; Chen, D.; Tu, K.; Ruan, C.; Hu, Z. Large-scale circulation classification and its links to observed precipitation in the eastern and central Tibetan Plateau. *Clim. Dyn.* **2016**, *46*, 3481–3497. [CrossRef]

23. Sun, Y.; Niu, T.; He, J.; Ma, Z.; Liu, P.; Xiao, D.; Hu, J.; Yang, J.; Yan, X. Classification of circulation patterns during the formation and dissipation of continuous pollution weather over the Sichuan Basin, China. *Atmos. Environ.* **2020**, *223*, 117244. [CrossRef]
24. Liu, N.; He, G.; Wang, H.; He, C.; Wang, H.; Liu, C.; Wang, Y.; Wang, H.; Li, L.; Lu, X.; et al. Rising frequency of ozone-favorable synoptic weather patterns contributes to 2015–2022 ozone increase in Guangzhou. *J. Environ. Sci.* **2025**, *148*, 502–514. [CrossRef]
25. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
26. Ibebuchi, C.C.; Richman, M.B. Circulation typing with fuzzy rotated T-mode principal component analysis: Methodological considerations. *Theor. Appl. Climatol.* **2023**, *153*, 495–523. [CrossRef]
27. Huth, R.; Nemesova, I.; Klimperová, N. Weather categorization based on the average linkage clustering technique: An application to European mid-latitudes. *Int. J. Climatol.* **1993**, *13*, 817–835. [CrossRef]
28. Arthur, D.; Vassilvitskii, S. k-means++: The advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, LA, USA, 7–9 January 2007; pp. 1027–1035.
29. Kohonen, T. Essentials of the self-organizing map. *Neural Netw.* **2013**, *37*, 52–65. [CrossRef]
30. Hansen, F.; Belušić, D. Tailoring circulation type classification outcomes. *Int. J. Climatol.* **2021**, *41*, 6145–6161. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.