

RESEARCH ARTICLE

Open Access



Update of global maps of Alisov's climate classification

Ryu Shimabukuro^{1*} , Tomohiko Tomita² and Ken-ichi Fukui³

Abstract

Proposed in 1954, Alisov's climate classification (CC) focuses on climatic changes observed in January–July in large-scale air mass zones and their fronts. Herein, data clustering by machine learning was applied to global reanalysis data to quantitatively and objectively determine air mass zones, which were then used to classify the global climate. The differences in air mass zones between two half-year seasons were used to determine climatic zones, which were then subdivided into continental or maritime climatic regions or according to east–west climatic differences. This study renews Alisov's CC for the first time in almost 70 years and employs data-driven machine learning to establish a standard for causal CC based on air masses.

Keywords Alisov's climate classification, Air mass, Front, Data clustering by machine learning

1 Introduction

Climate classification (CC) divides the Earth's surface into regions based on the similarity of climatic features. The shift of climatology from the classical approach of describing the characteristics of climatic elements to the modern approach of explaining the formation of climatic phenomena is reflected in CC (Yazawa 1980). In other words, CC methods can be divided into two categories: resultant CC based on classical climatology and causal CC based on modern climatology.

With the development of synoptic meteorology and weather forecasting, modern climatology has been able to reveal the physical processes behind and causal relationships among climatic phenomena. Bergeron (1930) first introduced the concept of an “air mass,” which has since been refined into air mass climatology (Fukui 1962).

An air mass is a large-scale atmospheric volume with uniform temperature and humidity that forms over vast oceanic or continental surfaces until their properties reach near-equilibrium. An air mass tends to form with a large-scale stationary anticyclone. The boundary between different air masses is defined as a transition zone or front, where cyclonic disturbances such as extratropical cyclones frequently occur, develop, and move eastward. Because the weather and climate are inherent to an air mass, an air mass can give comprehensive information on a climate and the area covered by the climate (Fukui 1962). However, air mass climatology has two drawbacks (Yoshino 1978). First, determining the area of an air mass in a globally applicable manner is difficult. A global classification tends to ignore local climatic features, but a local classification would divide Earth's surface into many small areas. Second, the difficulty in determining air masses can result in an arbitrary determination of fronts. To overcome these two difficulties, an objective and quantitative approach to determining air masses is required.

Figure 1 shows Alisov's CC (ACC) (Alisov 1936, 1954), which aims to understand global climatic features based on air masses. While the Köppen–Geiger CC (KGCC) (Köppen 1918; Köppen and Geiger 1936)

*Correspondence:

Ryu Shimabukuro
ryujin.mabuya.1@gmail.com

¹ Graduate School of Science and Technology, Kumamoto University, Kumamoto, Japan

² Faculty of Advanced Science and Technology, Kumamoto University, Kumamoto, Japan

³ The Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan

is a representative resultant CC system based on differences in the vegetation landscape due to climate, the ACC is a representative causal CC system based on the mechanisms and physical processes caused by air mass zones and their fronts. The mean positions of large-scale air mass zones and fronts usually shift by season due to seasonal changes in general atmospheric circulations. The north–south variations of air mass zones and fronts can be used to divide climatic zones into two categories: those that remain year round, and those that change seasonally. The ACC (1954) divides the global climate into four air mass zones according to temperature (i.e., latitude): equatorial, tropical, polar, and Arctic/Antarctic. Then, the differences between air masses in January and July are used to determine seven climatic zones. In this study, the air mass zone was defined as the distribution of each air mass in the summer and winter seasons, and the climatic zone was defined as the spatial extent of a climate superimposed on a single map accounting for seasonal changes. However, the fronts at both latitudinal edges do not correspond to large precipitation lines, which made Suzuki (1961) question whether the locations of the fronts were correctly determined. Wadachi (1997) argued that the ACC could not identify areas with a prevailing subtropical high. Despite some shortcomings, the ACC demonstrates the most crucial circulation

processes in different climatic zones and may be used as a basis to explain global climatic types (Khlebnikova 2009).

Many causal CCs have the disadvantage of sometimes not corresponding to actual climatic conditions (Nishina 2019). The subtropical and polar climatic zones in the ACC have been criticized for their low correspondence with the vegetation landscape of arid and temperate climates in the KGCC (Khlebnikova 2009). Although the KGCC is based on vegetation rather than the actual climate, it is widely used to explain agricultural and cultural regions as one of the CC systems that best reflects real climate differences (Nishina 2019). The latest revisions to the KGCC (Kottek et al. 2006; Peel et al. 2007; Kriticos et al. 2012; Beck et al. 2018) refer to terrestrial climatic variables (i.e., monthly temperature and precipitation) to approximate the vegetation distribution on land but still exclude climatic zones over the ocean. Because the KGCC is based on vegetation distribution, it is inherently unable to classify the climate over oceans.

Alisov (1954) attempted to subdivide the seven major climatic zones (Fig. 1) into 22 smaller regions using surface conditions such as continent vs. ocean, eastern vs. western, and plains vs. mountains. However, he failed to reveal the figures with methodology behind such detailed types (Mizukoshi and Yamashita 1985), which is probably because data with sufficiently long periods and high spatial resolution were unavailable in the 1950s. The

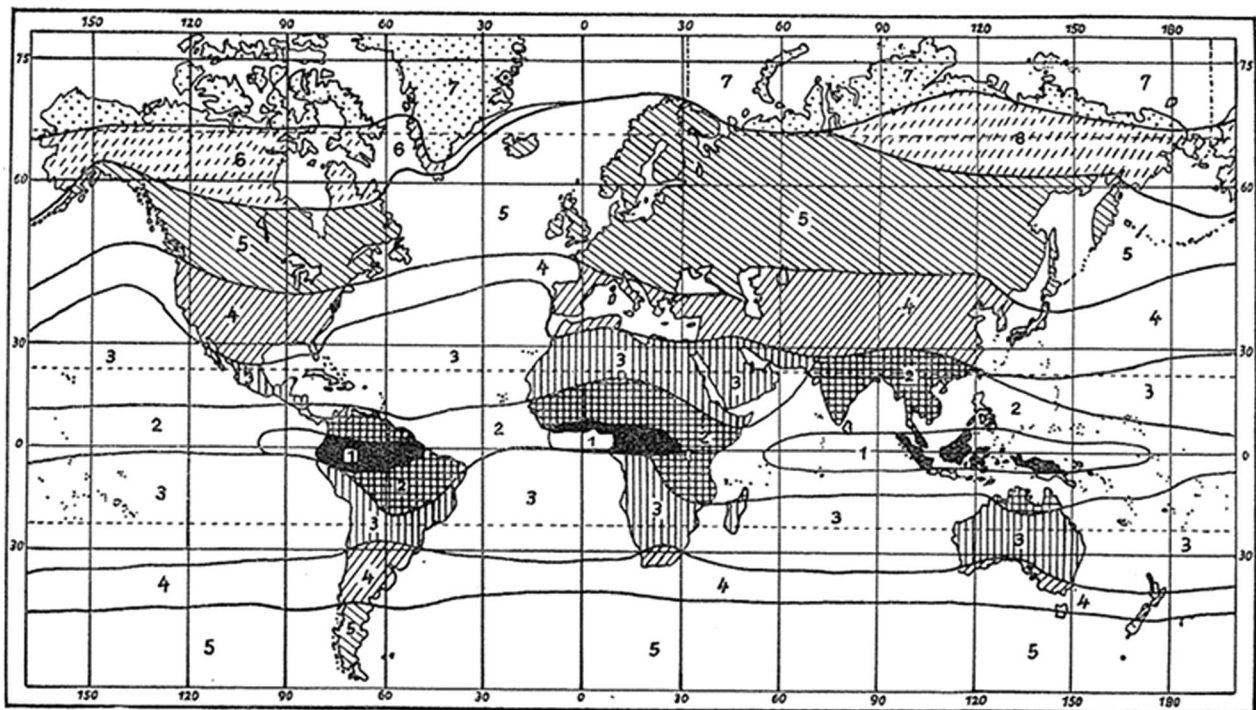


Fig. 1 Alisov's seven climatic zones (Alisov 1954). 1: Equatorial zone, 2: Subequatorial zone, 3: Tropical zone, 4: Subtropical zone, 5: Polar zone, 6: Subarctic zone, 7: Arctic/Antarctic zone

ACC, thus, does not consider east–west climatic differences in a climatic zone (e.g., Tokyo with a large amount of summer precipitation and Rome with the slight one are included in the same climatic zone (Nishina 2019)). Such subdivision is now possible with the presently available sophisticated data. Brunschweiler (1957) attempted to divide air masses into continental and maritime types and investigated the monthly occurrence frequency of individual air masses based on analysis of daily data at significant sites in the Northern Hemisphere. Then, he used the monthly changes in air mass areas to redefine the annual mean distribution of air masses in the Northern Hemisphere. Although Brunschweiler's methodology is clear, some researchers have questioned the rationale of using air mass frequencies of 80%, 50%, and 20% (Suzuki 1961). Oliver (1970) classified Australia in a similar manner using the annual prevailing air masses.

In recent years, weather observatories have been exploited globally, and several studies have tried applying machine learning to clustering such data for global CC (Mahlstein and Knutti 2010; Zscheischler et al. 2012; Metzger et al. 2013; Zhang and Yan 2014; Rohli et al. 2015; Netzel and Stepinski 2016; Sathiaraj et al. 2019; Gardner et al. 2020). Most of these studies focused on reproducing or comparing their work with the revised KGCC (Kottek et al. 2006; Peel et al. 2007). Rohli et al. (2015) used global reanalysis data to extend the KGCC over the whole Earth. Netzel and Stepinski (2016) showed that an information-theoretic measure of clustering called the V-measure (Rosenberg and Hirschberg 2007) could be used to quantitatively assess the homogeneity within a climatic type and differences between climatic types. Sathiaraj et al. (2019) compared three clustering techniques at their ability to identify climatic types in the United States: K-means (MacQueen 1967), DBCAN (Ester et al. 1996), and BIRCH (Zhang et al. 1996). Other studies have used clustering techniques to classify upper-level air masses (Vrac et al. 2012; Pernin et al. 2016; Watanabe et al. 2020). However, the above studies subjectively determined the number of clusters or air masses for classification.

In this study, we applied a technique of data clustering by machine learning to global reanalysis data in an attempt to objectively determine the global air mass distribution and develop a revised causal CC based on air masses. Because reliable objective reanalysis data with high spatiotemporal resolution are now available for Earth over a period of more than 40 years, it can be used to revise the causal CC system to reflect actual climatic conditions more accurately. This will allow us to classify the global climate in a data-driven manner by focusing on climatic causes (i.e., air masses and fronts) instead of observed responses such as vegetation. Thus, this study

renews the classical ACC for the first time in almost 70 years. We subdivide air mass zones into regions that distinguish between land and ocean and that reflect east–west differences. Objective estimation of the global distribution of large-scale air masses, including over the oceans, will lead to better understanding of future climate change and will be helpful for determining appropriate mitigation and adaptation measures (Mahlstein and Knutti 2010).

The remainder of this paper is organized as follows. Section 2 explains the data and methodology. Section 3 presents the global distribution of air mass zones after the optimal number of clusters is determined and compares the causal CC of this study with the conventional ACC. Section 4 summarizes the findings of this study and briefly discusses remaining issues.

2 Data and methodology

2.1 Data

We used Fifth-Generation ECMWF Atmospheric Reanalysis of the Global Climate (ERA5) data (Hersbach et al. 2020) compiled by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 data are available at high quality for more than 70 years with a spatial resolution of $0.25^\circ \times 0.25^\circ$ (Hersbach et al. 2020). The air mass analysis was based on monthly mean data for the temperature (T) and specific humidity (q) over the 40-year period of 1981–2020 at altitudes of 925–775 hPa (i.e., seven pressure surfaces), which corresponds to the lower troposphere. To correctly follow the definition of air mass, this study chose T and q, not employing precipitation. Although the vertical extent of air masses has not been clearly determined, we defined 775 hPa (i.e., about 2000 m) as the maximum altitude of air masses in the lower troposphere. For instance, the Siberian High forms from wintertime cold air in the boundary layer as a typical continental high-pressure system, and the inversion layer often appears around an altitude of 2000 m. The downward flow of the Hadley circulation forms a subtropical high, and the trade wind inversion forms around an altitude of 1000–2000 m. Referring to these information, we set the altitude for an air mass to be affected by the surface as 2000 m, although an atmospheric boundary layer typically has an altitude of around 1500 m, which is lower than in the tropics.

To evaluate the locations of fronts, we employed monthly mean surface precipitation data from the Global Precipitation Climatology Project (GPCP) Version 2.3 (Adler et al. 2018) on a 2.5° global grid for the 40-year period of 1981–2020. To account for seasonal differences, a year was divided into the October–March (O–M) and April–September (A–S) half-years.

2.2 Methodology

2.2.1 Clustering process

An air mass can be defined as a large-scale three-dimensional air volume with homogeneous temperature and humidity. To identify an air mass, we took the vertical mean of an atmospheric variable in the lower troposphere from 925 to 775 hPa:

$$\bar{x} \equiv \frac{1}{g} \int_{775 \text{ hPa}}^{925 \text{ hPa}} x \, dp / \frac{1}{g} \int_{775 \text{ hPa}}^{925 \text{ hPa}} dp, \quad (1)$$

where x is an atmospheric variable, the overbar ($\bar{}$) indicates the vertical mean, g is the acceleration of gravity, and p is the atmospheric pressure. In this study, we substituted the air temperature T (K) and specific humidity q (g kg^{-1}) into x . Equation (1) defines the mass-weighted vertical mean of x . To remove small-scale disturbances, \bar{x} was calculated for the O–M and A–S half-years (i.e., 6-month means), as shown in Fig. 2. To align the scales of the climatic variables (\bar{T} and \bar{q}), we applied z-score normalization (Raschka and Mirjalili 2020) to the global horizontal data of both half-years using the means and standard deviations.

To unify the clusters (i.e., air mass zones) identified in the two half-years, the normalized data of both half-years were simultaneously employed as input data. Multiplying the number of features, number of grid points in the global reanalysis data, and number of half-years resulted in input data with an array size of $2 (\bar{T}, \bar{q}) \times 1440 (\text{longitudes}) \times 721 (\text{latitudes}) \times 2 (\text{half-years}) = 4,152,960$. We then applied K-means++ clustering (Arthur and Vassilvitskii 2007) to the input data. Clustering is a statistical technique of grouping samples based on the similarity of features. K-means clustering is a well-known non-hierarchical clustering technique (MacQueen 1967) that is widely used in industry and academia. K-means clustering is suitable for analyzing big data, because it requires less computation of distances than hierarchical clustering techniques, in which close clusters are merged successively. In addition, K-means++ clustering has been reported to be more effective and more consistent than conventional K-means clustering, where the classification results depend on the position of the initial centroid (Arthur and Vassilvitskii 2007; Raschka and Mirjalili 2020). Thus, we used the K-means++ clustering technique to overcome the initial value dependency problem of conventional non-hierarchical clustering techniques such as K-means or K-medoids clustering.

K-means++ clustering creates clusters from the K points located farthest each other as the initial centroid position based on the probabilistic distribution of the samples. Then, samples are clustered by iteratively minimizing the sum of squared errors (SSE), which is estimated as follows:

$$\text{SSE} = \sum_{i=1}^K \sum_{a_n \in C_i} \|a_n - c_i\|^2, \quad (2)$$

where i is the number of a cluster from 1 to K , c_i is the centroid of i -th cluster (i.e., C_i), and a_n is a sample belonging to C_i . The SSE is the sum of the squared Euclidean distances between a_n and c_i for all clusters (i.e., the total variance of the clusters). A smaller SSE indicates that clusters are more compact in the feature space, and the SSE can be used to quantitatively evaluate the clustering performance. To further reduce the dependency on the initial position, we performed K-means++ clustering 20 times using different initial centroid positions, and the clustering results with the smallest SSE were selected as the best.

One advantage of using K-means++ clustering rather than hierarchical and density-based clustering techniques is that the former retains the coordinates of the centroid's final position in feature space, which allows different datasets (i.e., air temperature and specific humidity) to be clustered using the same coordinates. We performed K-means++ clustering by using the scikit-learn Python packages (Pedregosa et al. 2011).

2.2.2 Finding the appropriate number of clusters

Clustering techniques, including K-means++ clustering, requires setting the number of clusters a priori. To determine the appropriate number of clusters, we utilized four quantitative evaluation indices.

The first evaluation index was the Davies–Bouldin index (DBI) (Davies and Bouldin 1979):

$$\text{DBI} = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \left(\frac{s_i + s_j}{d_{i,j}} \right), \quad (3)$$

where s_i is the average distance between all samples belonging to a cluster C_i , and $d_{i,j} = \|c_i - c_j\|_2$ is the L2 norm or Euclidean distance between the centroids of two clusters C_i and C_j . The DBI represents the average similarity between a cluster C_i and its nearest cluster C_j for $i, j = 1, \dots, K$. A lower DBI indicates a greater distance separating the nearest cluster pair C_i and C_j . Thus, a small DBI indicates that the number of clusters is appropriate.

The second evaluation index was the silhouette coefficient (SC) (Rousseeuw 1987), calculated for a single sample a_n :

$$\text{SC}_n = \frac{e_n - f_n}{\max(e_n, f_n)}, \quad (4)$$

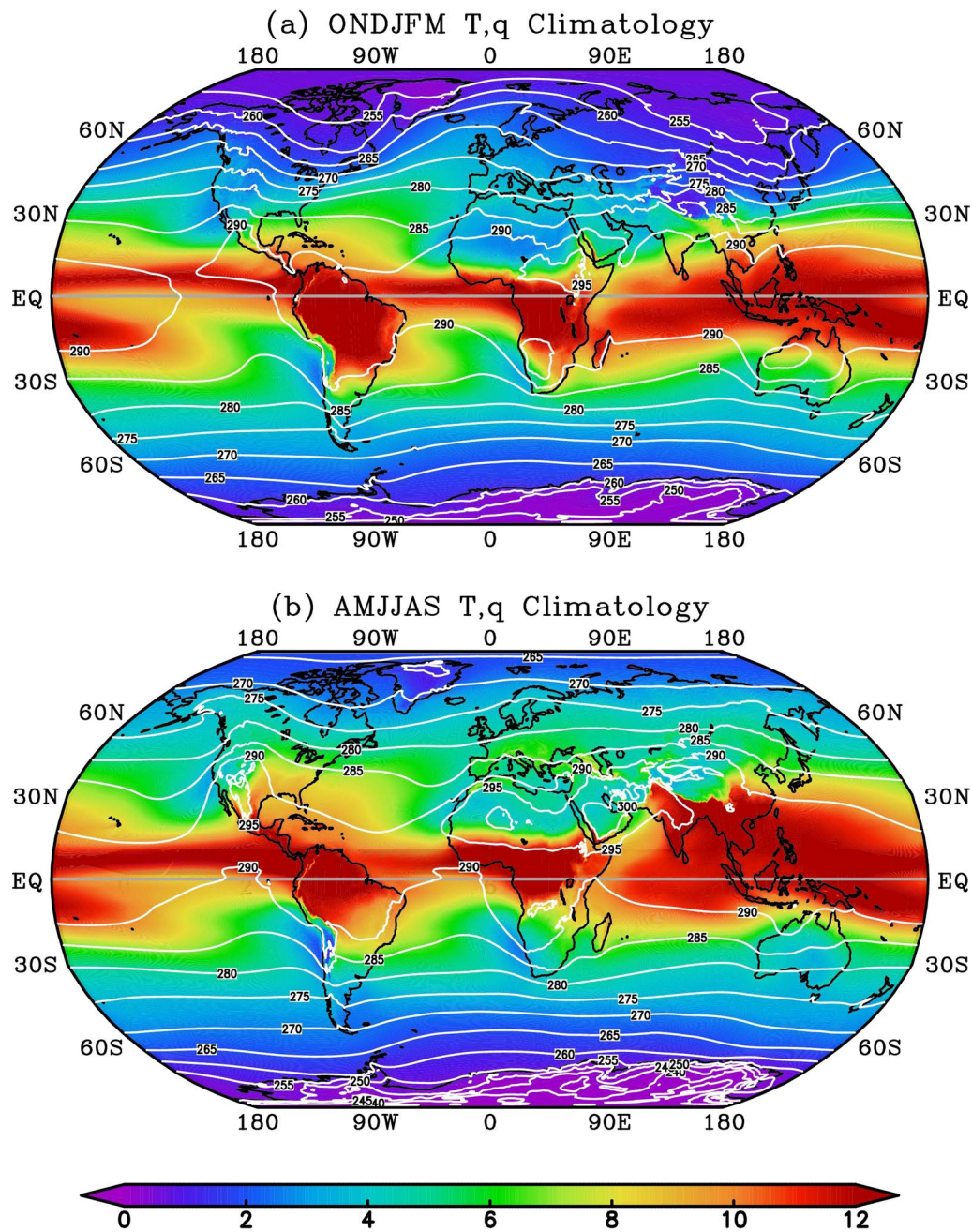


Fig. 2 Vertical mean distributions (925–775 hPa) of the temperature (K; contour) and specific humidity (g kg^{-1} ; shading) in the (a) O–M half-year and (b) A–S half-year for 1981–2020. The contour interval is 5 K. The scale of shading is shown at the bottom, and the warm and cold colors indicate relatively high and low specific humidity, respectively

where f_n is the average distance of sample a_n to all other samples in the same cluster, and e_n is the average distance of sample a_n to samples in the next nearest cluster. The SC_n value varies between -1 and 1 , with a value near 1 indicating that the clustering result is satisfactory on the sample a_n . The average of SC_n for a_n in the entire dataset,

$SC = \frac{1}{N} \sum_{n=1}^N SC_n$ (N is the number of all samples), can be used to evaluate the overall performance of the clustering results.

The third evaluation index was the Cailnski–Harabasz index (CHI) (Caliński and Harabasz 1974):

$$\text{CHI} = \frac{\text{trace}(B_K)}{\text{trace}(W_K)} \times \frac{N - K}{K - 1}, \quad (5)$$

where $W_K = \sum_{i=1}^K \sum_{a_n \in \mathcal{C}_i} (a_n - c_i)(a_n - c_i)^t$ is a variance matrix within clusters, and $B_K = \sum_{i=1}^K n_i(c_i - c_o)(c_i - c_o)^t$ is a variance matrix between clusters (n_i is the number of samples belonging to cluster \mathcal{C}_i and c_o is the centroid of all samples). The superscript t denotes transposition. A larger value for CHI indicates a denser sample distribution in each cluster and greater distance between clusters. In other words, a large CHI indicates that the number of clusters is appropriate.

The fourth evaluation index was the Bayesian information criterion (BIC) (Schwarz 1978), which is used to select models from a finite set of models. Under the assumption that the samples in cluster \mathcal{C}_i follow a Gaussian distribution, BIC can be defined as follows:

$$\text{BIC} = \sum_{i=1}^K \left[\log L(\theta_i; a_n \in \mathcal{C}_i) - \frac{D}{2} \log n_i \right], \quad (6)$$

where L is the likelihood function of the model, θ_i is the set of parameters of the likelihood function, and D is the number of parameters of the model with K components. The second term represents a penalty term imposed to prevent model overfitting due to an increase in the number of parameters. Note that we calculated BIC with the opposite sign of the usual definition. Thus, a larger value for the BIC indicates a better model prediction of the sample data and true unknown distribution.

2.2.3 Sea–land contrast and east–west climatic differences

Figures 2 and 3 show the distributions of the temperature and specific humidity in the lower troposphere and their anomalies from the zonal means in the O–M and A–S half-years, respectively. According to the ACC, latitudinal differences in the downward surface solar radiation flux determine the global surface temperature. Therefore, the number of latitudinal air mass zones may be increased by simply increasing the number of clusters (Fig. 2). Thus, we also considered anomalies from the zonal (0° – 360° E) mean at each latitude to clarify the sea–land or east–west distribution of air masses (Fig. 3). In addition, z-score normalization shown in Sect. 2.2.1 is applied to global anomalies in the mean data of the O–M and A–S half-years. The normalized anomalies were used to classify air mass zones into two for both half-years to represent dry and moist air masses.

3 Results and discussion

3.1 Determination of number of clusters

Figure 4 shows the values of the evaluation indices for the number of clusters. The DBI (blue) significantly decreased as the number of clusters was increased from two to four and reached its minimum at four, after which it rapidly increased with more clusters. This suggests that four clusters are optimal. The SC (red) decreased as the number of clusters increased from two to six, and it increased from six to eight and decreased again after eight. The decreasing trend was partially mitigated from three to four. Therefore, we judged that four clusters were applicable.

The CHI (green) and BIC (orange) both increased as the number of clusters was increased from two to ten. Both indices rapidly increased in value from two to four clusters but increased gradually from four to ten clusters. Based on these results, we determined the statistically optimal number of clusters to be four, which indicates that the global climate can be divided into four air mass zones. Thus, Alisov's four air mass zones from the 1950s are supported from a data-driven perspective using high-quality global reanalysis data.

3.2 Global distributions of air mass zones

Figure 5 shows the global distributions of the four air mass zones in the O–M and A–S half-years from low to high latitudes: tropical (red), subtropical (yellow), polar (light blue), and Arctic/Antarctic (purple). The tropical and subtropical air mass zones correspond to the equatorial and tropical air mass zones in the ACC (not shown). The seasonal changes in the general atmospheric circulations caused the air mass zones to move toward the poles in the summer hemisphere and toward the equator in the winter hemisphere (Figs. 5a and b). This indicates that the north–south seasonal shift was accurately captured. The Arctic air mass zone was shown to dominate from the North Pole to northeastern Siberia and Canada in the O–M half-year (Fig. 5a), but it disappeared in the A–S half-year (Fig. 5b).

We then investigated the sensitivity of the air mass zone distribution to differences in altitude by comparing the distributions estimated using the vertical means over 925–775 hPa and 925–850 hPa. The overall distribution of the four air mass zones showed almost no changes. Using the 925–850 hPa mean somewhat expanded the distribution of the tropical air mass zone over the ocean and shrank the distribution of the subtropical air mass zone (not shown). We attributed this to the reduced influence of trade wind inversion in and around subtropical anticyclones. Thus, the air mass zone distribution seemed to be robust against differences in the altitude range for the lower troposphere.

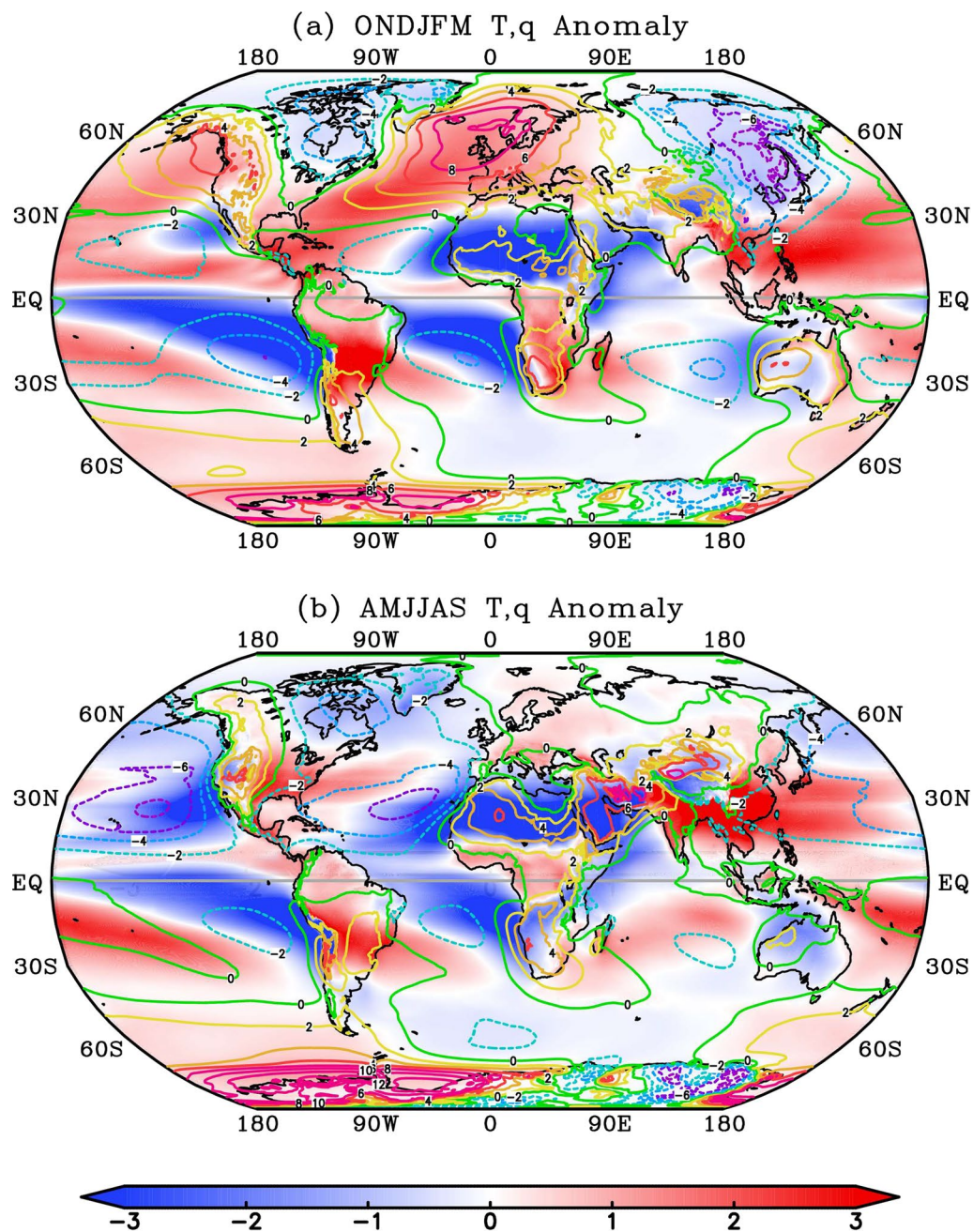


Fig. 3 Distributions of temperature (K; contour) and specific humidity (g kg^{-1} ; shading) anomalies from the zonal means in the (a) O–M half-year and (b) A–S half-year. The contour interval is 2 K. The warm solid and cold dashed contours indicate relatively high and low temperatures, respectively. The scale of shading is shown at the bottom, and the red and blue colors indicate positive and negative anomalies, respectively of the specific humidity

The white contours in Fig. 5 indicate areas where the half-year mean precipitation was greater than 3 mm day^{-1} . The large rain bands or fronts in the mid-latitudes are located near the boundary between the subtropical (yellow) and polar (light blue) air mass zones in the Northern Hemisphere during the O–M half-year

(Fig. 5a) and near the boundary between tropical (red) and subtropical (yellow) air mass zones in the Northern Hemisphere during the A–S half-year (Fig. 5b). The Atlantic polar front in boreal winter (Fig. 5a) is a transition zone with high baroclinicity where a subtropical air mass of the Azores high meets a polar air mass

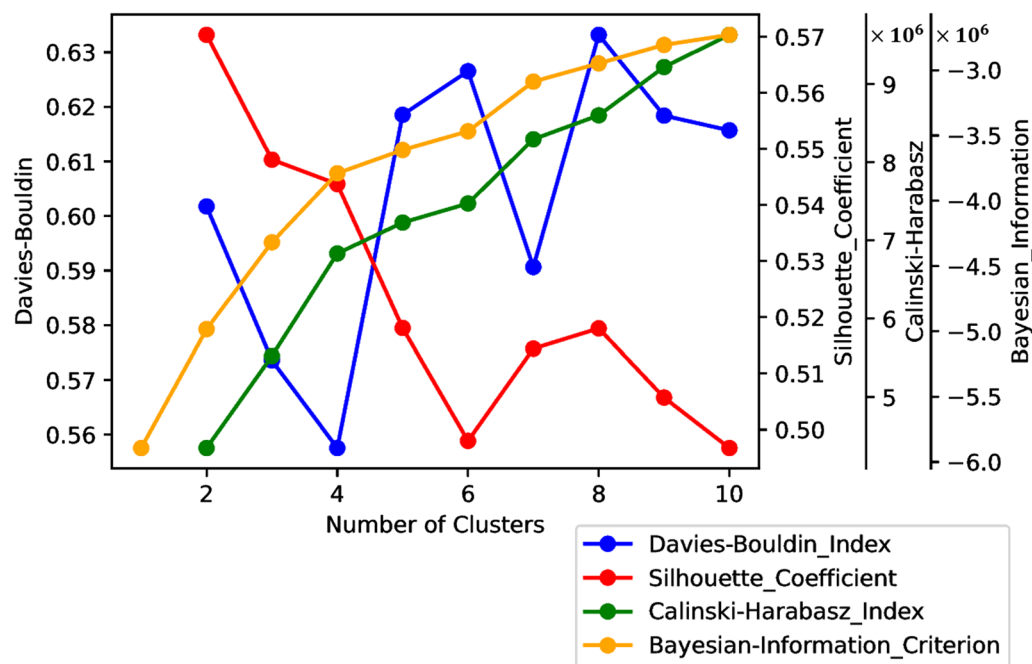


Fig. 4 Evaluation indices used to quantitatively determine the optimal number of clusters: DBI (blue), SC (red), CHI (green), and BIC (orange). All indices are dimensionless

extending southward. The high baroclinicity induces the development of extratropical cyclones, whose eastward movement brings precipitation over Europe. Thus, the clustering results seem to capture intense mid-latitude rain bands as fronts with baroclinic instability. Meanwhile, the subtropical front in the north Pacific crosses the Yangtze River basin and Japanese archipelago in boreal summer and represents the Meiyu–Baiu front, which brings significant precipitation to East Asia (Fig. 5b). In other words, the Meiyu–Baiu front is captured as a pronounced subtropical front between the tropical (red) and subtropical (yellow) air mass zones, which is consistent with Ninomiya’s results (1984). Thus, the clustering results accurately captured frontal precipitation between air mass zones in the mid- and high latitudes. The heavy precipitation in the tropics and weak precipitation around 60°S seem unrelated to the boundaries between air mass zones.

Figure 6 shows the sample statistics of the four air mass zones (i.e., clusters) in each half-year. In Fig. 6a, the centroids of the clusters (black stars) are almost equidistant, which ensures good clustering. The distribution pattern may reflect the Clausius–Clapeyron relation at the left edge line, where the amount of water vapor increases exponentially with increasing temperature. The centroid of each cluster is located near the left edge of the feature space because the samples were concentrated near

the edge, although colors appear equal. The subtropical air mass zone (yellow) had the largest area in the feature space, which may reflect the large temperature and humidity ranges of this zone. Figure 6b shows the sample means (error bars: standard deviations) of \bar{T} and \bar{q} for each air mass zone in the Northern Hemisphere. The fronts can be characterized according to differences between the atmospheric characteristics of two adjacent air mass zones. Because the latitude of each air mass zone shifts north or south according to the season, \bar{T} and \bar{q} of each zone did not differ significantly between half-years. The tropical and subtropical air mass zones had a relatively small temperature difference of less than 7 K but a large humidity difference of about 5.4 g kg⁻¹. The polar air mass zone had a lower temperature and humidity than the subtropical air mass zone, and the temperature and humidity differences were about 16 K and 2.7 g kg⁻¹, respectively. These results suggest that the subtropical front is characterized by a large humidity gradient while the polar front is characterized by both temperature and humidity fronts. The Arctic air mass zone had lower temperature and humidity than the polar air mass zone in the O–M half-year with temperature and humidity differences of about 14 K and 1.5 g kg⁻¹, respectively. These features were almost the same in the Southern Hemisphere (not shown).

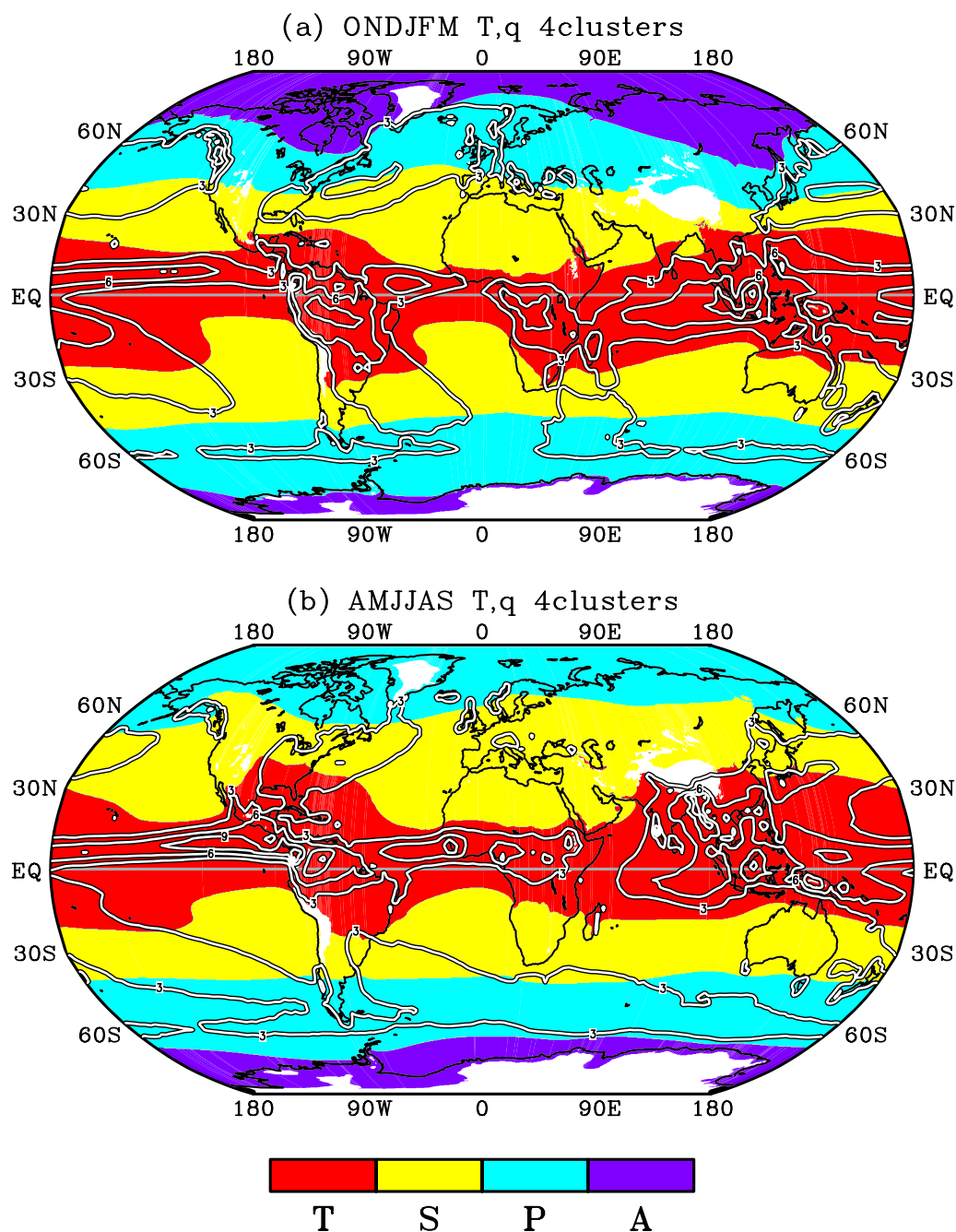


Fig. 5 Global distributions of four air mass zones (shading) and surface precipitation greater than 3 mm day^{-1} (white contour) in the (a) O–M half-year and (b) A–S half-year. The contour interval is 3 mm day^{-1} . The colors are explained at the bottom. Blank areas indicate regions where the altitude was greater than 775 hPa (about 2000 m). T: Tropical, S: Subtropical, P: Polar, A: Arctic/Antarctic

3.3 Global distributions of climatic zones and comparison with classical ACC

Figure 7 shows a world map of the climatic zones, which was created by overlapping the air mass zones of the O–M and A–S half-years (Fig. 5), and represents the renewed ACC. The hatched areas indicate regions where the air mass zones alternate between half-years. Table 1

summarizes the features of the climatic zones. The symbols used in Fig. 7 and Table 1 are discussed below. Note that the seasonal differences between the two hemispheres during a half-year were considered.

The global climate can be classified into four stable zones that are dominated by the same air mass zones throughout the year (TT, SS, PP, AA) and three zones

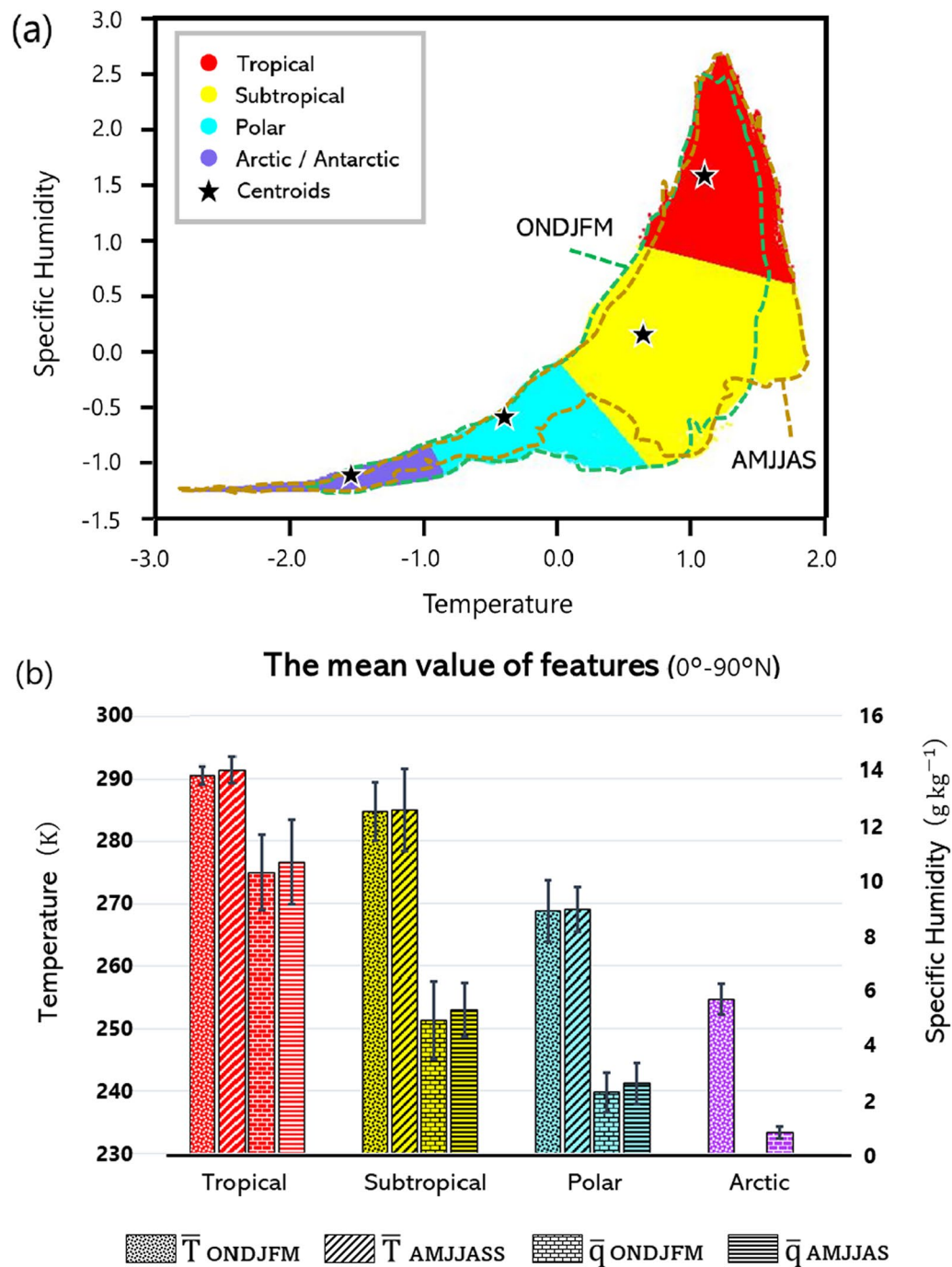


Fig. 6 (a) Distribution and centroid in feature space of the global air mass zones, and (b) the mean values (error bars: standard deviations) of the temperature and specific humidity in each air mass zone for the Northern Hemisphere (0° - 90° N) in the O-M and A-S half-years

with a one-class change each half-year (ST, PS, AP). This is similar to the classical ACC (Fig. 1). In addition, some regions have two-class changes (AS, PT) such as the eastern Siberia (ultramarine), central North America (black), Yangtze River basin (black), and western

Japan (black) (Fig. 7). These regions experience significant changes in climate throughout the year.

The most significant difference between the renewed ACC (Fig. 7) and classical ACC (Fig. 1) is the large width of the tropical climatic zone (TT) in the renewed ACC, which is the equatorial climatic zone in the classical

ACC. Alisov (1954) probably estimated the location of the inter-tropical convergence zone empirically because meteorological data, especially humidity, were scarce at that time, particularly over oceans. In addition, Alisov (1954) seemed to have arbitrarily determined the locations of fronts using only the surface temperature.

Figure 8 shows the global distributions of the four air mass zones in January and July estimated only by the temperature like Alisov (1954). The distributions of the Arctic/Antarctic air mass zones are similar to those in Fig. 5 except for the Antarctic zone in January. In addition, the width of the tropical air mass zone increased markedly from January to July. In July (Fig. 8b), the northern edge of the tropical air mass zone reached about 60° N over Eurasia and North America, and the polar air mass zone disappeared in the high latitudes of the Northern Hemisphere.

The white contours in Fig. 8 indicate areas with precipitation of greater than 3 mm day⁻¹. Although the monthly mean precipitation in January and July was greater than the seasonal means of the O–M and A–S half-years (Fig. 5), the distributions of the monthly and seasonal means did not show significant differences for the cold and warm periods. Figure 8b shows a misalignment between the subtropical front and intense rain

Table 1 Nine climatic zones

Symbol	Climatic zones	Winter / Summer	Behavior
TT	Tropical	Tropical / Tropical	stable
ST	Quasi-tropical	Subtropical / Tropical	one-class change
SS	Subtropical	Subtropical / Subtropical	stable
PS	Temperate	Polar / Subtropical	one-class change
PP	Polar	Polar / Polar	stable
AP	Subarctic	Arctic / Polar	one-class change
AA	Arctic /Antarctic	Arctic / Arctic	stable
AS	Subarctic	Arctic / Subtropical	two-class change
PT	Temperate	Polar / Tropical	two-class change

Nine climatic zones based on the dominant air mass zones during the winter and summer half-years in each hemisphere and their relationship with the behavior of the air mass zones

bands of greater than 6 mm day⁻¹ in July and a generally poor representation of the subtropical to temperate climates. This suggests that the humidity gradient is more critical than the temperature gradient for precipitation from the Meiyu–Baiu front. Overall, the renewed ACC (Fig. 7) better explains climatic phenomena in the Northern Hemisphere based on the concepts of air masses and

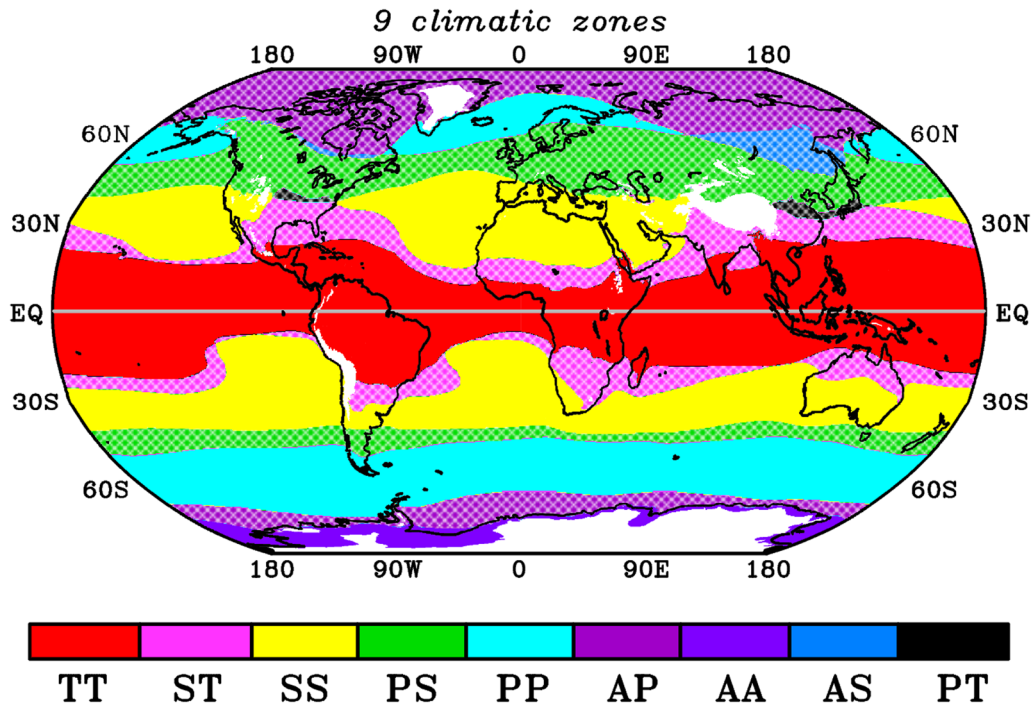


Fig. 7 Global climatic zones obtained by overlapping the air mass zones of the winter and summer half-years in each hemisphere (Fig. 5; see Table 1 for symbols). The hatched areas indicate the regions where the air mass zones alternate with the season. The blank areas mean the same as in Fig. 5

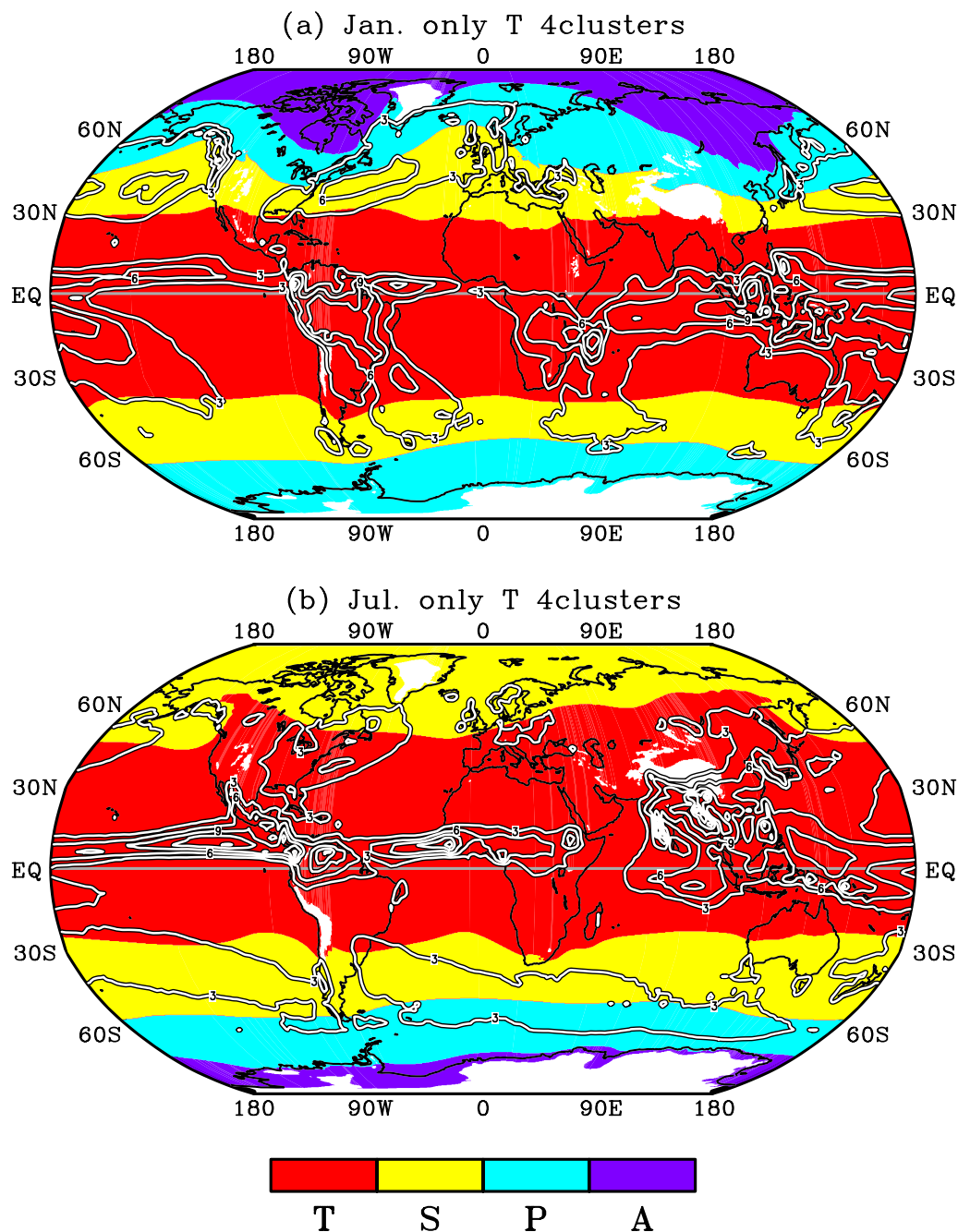


Fig. 8 Global distributions of four air mass zones (shading) and surface precipitation greater than 3 mm day^{-1} (white contour) in (a) January and (b) July. The contour interval is 3 mm day^{-1} . The blank areas mean the same as in Fig. 5. T: Tropical, S: Subtropical, P: Polar, A: Arctic/Antarctic

fronts than the classical ACC (Fig. 8) because the former considers both the temperature and specific humidity while the latter only considers the temperature.

3.4 Subdivision of climatic zones

We compared the renewed ACC (Fig. 7) with the latest KGCC (Beck et al. 2018) in deserts for reference. The

comparison showed that two deserts in central Eurasia belonged not to the subtropical climatic zone (SS) with arid characteristics but the temperate climatic zone (PS) with polar frontal precipitation. According to the latest KGCC (Beck et al. 2018), the desert climate can be divided into two categories: hot (BWh) and cold (BWk) with mean annual temperatures above and below 18°C ,

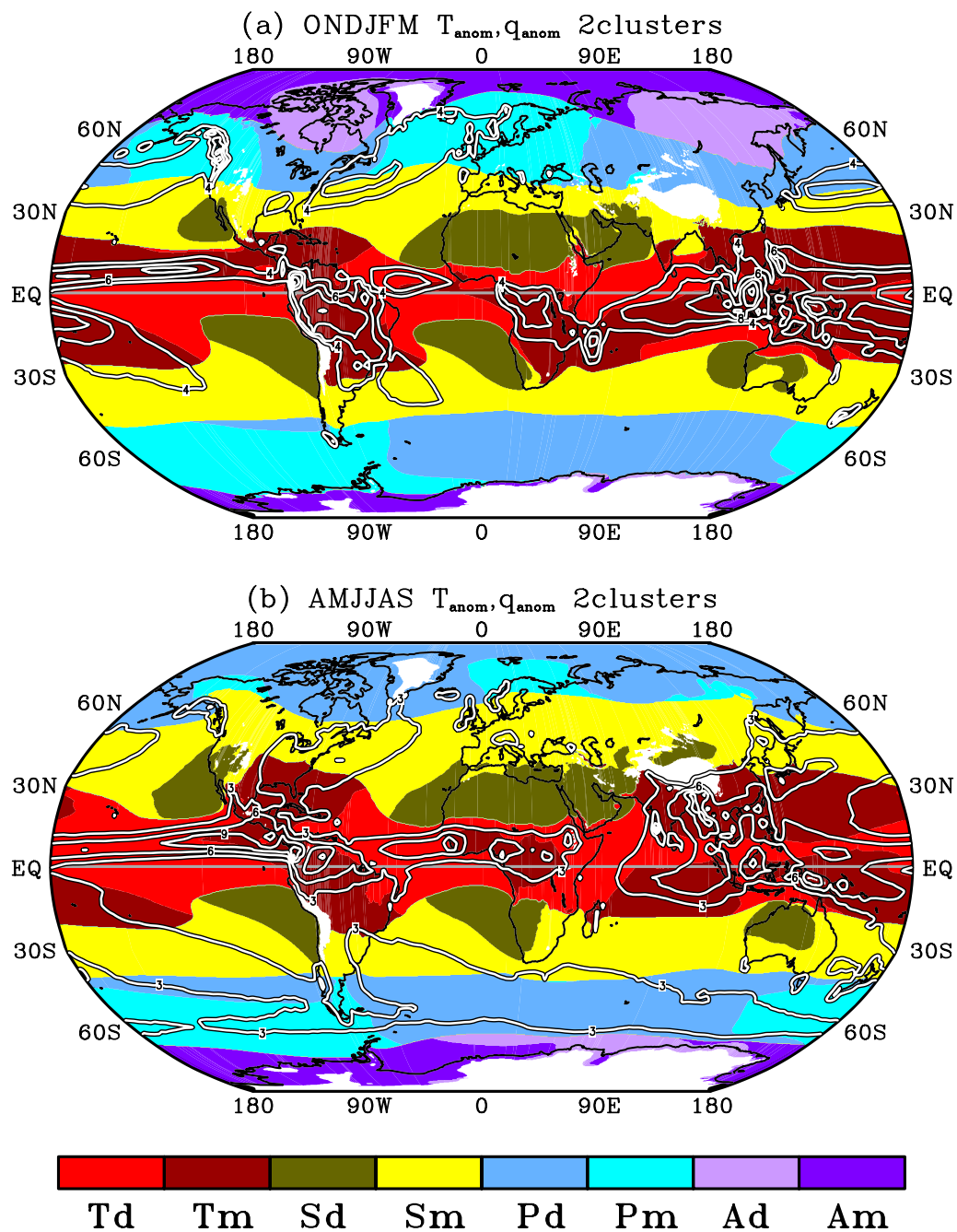


Fig. 9 Regions with dry and moist air masses (shading) obtained by subdividing each air mass zone (Fig. 5) into two clusters, and surface precipitation greater than 3 mm day^{-1} (white contour) in the (a) O–M half-year and (b) A–S half-year. The contour interval is 3 mm day^{-1} . The blank areas mean the same as in Fig. 5. Td: Tropical dry, Tm: Tropical moist, Sd: Subtropical dry, Sm: Subtropical moist, Pd: Polar dry, Pm: Polar moist, Ad: Arctic/Antarctic dry, Am: Arctic/Antarctic moist

respectively. The KGCC classifies the desert climate in central Eurasia as BWk because of the low half-year temperature in the winter (Fig. 2a) and the Somali region of Africa as BWh. However, the renewed ACC classifies the later region as belonging to the tropical zone (TT) with humid climatic features (Fig. 7). It is difficult for the

renewed ACC to explain the desert formation mechanism of this region as the subtropical air mass zone with dry atmospheric characteristics. Thus, we subdivided the climatic zones in Fig. 7 into climatic regions to discuss this issue in more detail.

Table 2 Characteristics of each dry and moist air mass

Symbol	Types of air mass	\bar{T}_{ONDJFM}	\bar{q}_{ONDJFM}	\bar{T}_{AMJJAS}	\bar{q}_{AMJJAS}
Td	Tropical dry	290.79	9.68	291.01	9.96
Tm	Tropical moist	290.59	10.72	290.82	10.80
Sd	Subtropical dry	288.93	4.64	291.48	4.83
Sm	Subtropical moist	282.52	5.26	282.05	5.31
Pd	Polar dry	268.46	2.27	268.92	2.52
Pm	Polar moist	268.54	2.42	269.35	2.61
Ad	Arctic dry	253.39	0.57	243.37	0.24
Am	Arctic moist	254.81	0.67	250.01	0.48

Mean temperature \bar{T} (K) and specific humidity \bar{q} (g kg^{-1}) in each dry and moist air mass for the O–M and A–S half-years

Figure 9 shows the distributions of the air mass zones in Fig. 5 when further subdivided into dry (subscript

d) and moist (subscript m) air masses. The air masses were classified as dry or moist considering the global anomaly distribution (Fig. 3) and mean values of \bar{T} and \bar{q} . First, four air mass zones were prepared, each consisting only of grid points classified into the same air mass zone (uniform values of \bar{T} and \bar{q}) between two half-years (see Sect. 3.2). Then, each air mass zone was divided into two air masses using anomalies of \bar{T} and \bar{q} (Fig. 3) at those grid points, which resulted in $4 \times 2 = 8$ air masses. Table 2 summarizes the mean values of \bar{T} (K) and \bar{q} (g kg^{-1}) of each air mass. The air masses were labeled as dry and moist, although dry and moist air masses are usually equivalent to continental and maritime ones, respectively. Figure 9 shows that, in the tropics, intense precipitation occurred mostly in the moist air mass (Tm) and at its boundary with the dry air mass (Td). In the

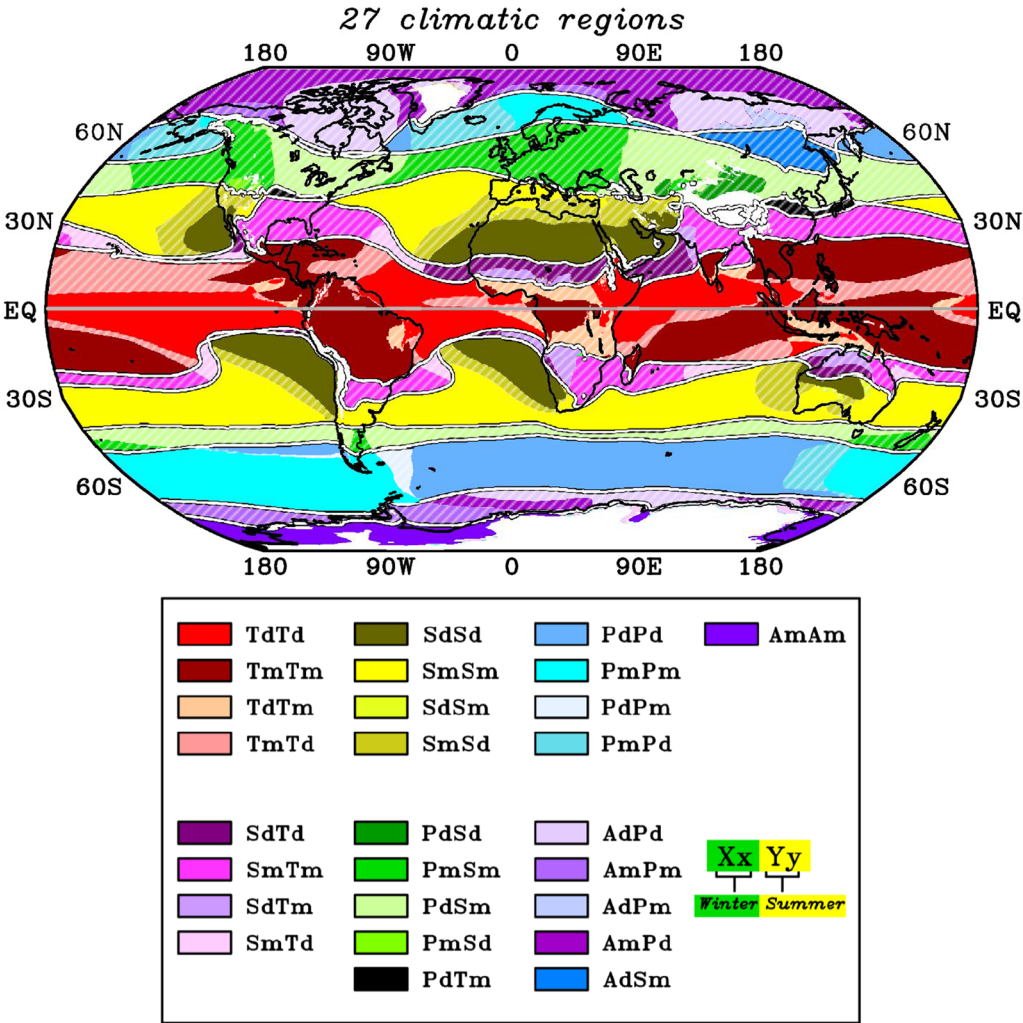


Fig. 10 (a) Global climatic regions obtained by overlapping the air mass distributions of the winter and summer half-years in each hemisphere (Fig. 9; see Table 3 for the symbols). The hatched areas indicate regions where the air masses alternate with the half-year. The white contours indicate the boundaries of climatic zones (Fig. 7). The blank areas mean the same as in Fig. 5

Table 3 Twenty-seven climatic regions and their correspondence to the KGCC

Symbol	Climatic regions (Winter / Summer)	Landscape (cf. Beck et al. 2018)
TdTd	Tropical year-round dry (Td / Td)	Deserts: BWh
TmTm	Tropical year-round moist (Tm / Tm)	Tropical rainforests: Af, (Aw, Am)
TdTm	Tropical summer-moist (Td / Tm)	Savannas: Aw, (BSH)
TmTd	Tropical summer-dry (Tm / Td)	
SdTd	Quasi-tropical year-round dry (Sd / Td)	Steppes: BSh, (BWh)
SmTm	Quasi-tropical year-round moist (Sm / Tm)	Evergreen forest: Cfa, Cwa
SdTm	Quasi-tropical summer-moist (Sd / Tm)	Aw, (BSH, Cwa)
SmTd	Quasi-tropical summer-dry (Sm / Td)	
SdSd	Subtropical year-round dry (Sd / Sd)	Hot-Deserts: BWh
SmSm	Subtropical year-round moist (Sm / Sm)	Mediterranean: Csa, Csb, (BWh, BSh)
SdSm	Subtropical summer-moist (Sd / Sm)	Bwh
SmSd	Subtropical summer-dry (Sm / Sd)	BWh, (BWk, BSk)
PdSd	Temperate year-round dry (Pd / Sd)	Cold-Deserts: BWk
PmSm	Temperate year-round moist (Pm / Sm)	Broadleaf, Mixed forests: Cfb, Dfb, (Dfc)
PdSm	Temperate summer-moist (Pd / Sm)	Coniferous forests: Dfa, Dfb, (Dwa)
PmSd	Temperate summer-dry (Pm / Sd)	BWk
PdTm	Temperate summer-wet (Pd / Tm)	Cfa, Dfa, (Dwa)
PdPd	Polar year-round dry (Pd / Pd)	
PmPm	Polar year-round moist (Pm / Pm)	Dfc
PdPm	Polar summer-moist (Pd / Pm)	
PmPd	Polar summer-dry (Pm / Pd)	Dfc
AdPd	Subarctic year-round dry (Ad / Pd)	Taigas, Tundras: Dfc, ET
AmPm	Subarctic year-round moist (Am / Pm)	Dsb
AdPm	Subarctic summer-moist (Ad / Pm)	Dfd
AmPd	Arctic (Am / Pd)	ET, (Dfc)
AdSm	Subarctic summer-warm (Ad / Sm)	Dwb, Dwc
AmAm	Antarctic (Am / Am)	EF

Names and symbols of each climatic region and the dominant air masses during the winter and summer half-years in each hemisphere (in parentheses in the middle column) and their correspondence to the KGCC (right column; Beck et al. 2018).

mid-latitudes, the subtropical dry air mass (Sd) extended throughout the year over the Sahara and Namib deserts in Africa, the Arabian Peninsula, and dry regions along the western coasts of North and South America. The subtropical dry air mass had a higher temperature and lower humidity of about 9 K and 0.5 g kg^{-1} , respectively, than the subtropical moist air mass during the A–S half-year (see Sd and Sm in Table 2). Thus, the subtropical dry air mass accurately represented the center of subtropical anticyclones with extremely hot and dry characteristics, which are a significant factor contributing to desert formation.

Figure 10 shows the overlapping seasonal changes of the air mass distributions in Fig. 9. The hatched areas indicate regions where air masses alternated between the winter and summer half-years in each hemisphere. Table 3 summarizes the name and symbol of climatic regions, the types of air masses that dominated each half-year for each hemisphere, and their correspondence to the vegetation landscape of the latest KGCC (Beck et al.

2018). A climatic region is expressed by four letters (e.g., XxYy). The first two and last two letters indicate the air masses that dominate the winter and summer half-years, respectively. The capital X and Y can be replaced by T, S, P, or A, and the small x and y can be replaced by d or m.

Each stable zone (i.e., XX in Fig. 7) except for Antarctica was divided into four regions: where a dry or moist air mass dominates all year round (XdXd and XmXm), and where a dry/moist air mass dominates in the winter and a moist/dry air mass dominates in the summer (XdXm and XmXd) (see Table 3). Seasonal one-class change zones (XY) (Fig. 7) were divided into four regions: where dry and moist air masses within the same air mass zone alternated seasonally (XdYd, XmYm), and where dry and moist air masses within adjacent air mass zones alternate seasonally (XdYm, XmYd). The temperate and subarctic climatic zones included seasonal two-class change regions (PdTm and AdSm), which resulted in five types of climatic regions (see Table 3).

We focused on the Somali and central Eurasia regions where the renewed ACC ran into a problem and found that these regions were covered by dry air masses throughout the year (Fig. 9). The Somali region is covered by a tropical dry air mass all year, which forms a dry region (TdTd) in the tropical climatic zone. The central Eurasia region is covered by a subtropical dry air mass centered in the Sahara that expands northeastward during the A–S half-year (Fig. 9b) and by a polar dry air mass in the polar air mass zone (Table 2) during the O–M half-year (Fig. 9a). This explains why two locally arid climatic regions (PdSd) formed in the temperate climatic zone (Fig. 10).

In contrast to the classical ACC, the climatic regions in Fig. 10 better represent east–west climatic differences, especially at the mid- to high latitudes, such as the western and eastern coasts of the continental landmasses in the Northern Hemisphere. Our climatic regions also showed good correspondence with the vegetation landscape of the KGCC (Table 3). This means that the renewed ACC connects, through observations, purely meteorological concept of air masses to the global distribution of vegetation, and thus to the delineation of ecoregions. The renewed ACC can be used to gain insights into the climatic distribution over oceans, which is not classified by the KGCC in principle (Kottek et al. 2006; Peel et al. 2007; Kriticos et al. 2012; Beck et al. 2018).

4 Conclusions

For many years, one of the most critical issues in air mass climatology has been developing an objective method to determine the air mass distribution. In this study, we applied K-means++ clustering to global reanalysis data to quantitatively and objectively determine the boundaries between air masses. Then, we subdivided air mass zones into regions based on continental or maritime climates and east–west climatic differences, and we considered the half-year changes in air mass zones. Thus, we renewed the classical ACC for the first time in almost 70 years.

We began by questioning whether the global climate can be divided into four air mass zones as Alisov did in the 1950s. Our statistical evaluation of four indices (DBI, SC, CHI, and BIC) confirmed that the division of the global climate into four air mass zones is suitable (Fig. 4). We then applied K-means++ clustering to classify the global climate for the O–M and A–S half-years (Fig. 5). The clustering results showed consistency with rain bands along fronts that were probably induced by baroclinic instability in the mid-latitudes. Alisov (1954) devised seven climatic zones (Fig. 1) based on the January–July alternation of air mass zones. By considering the half-year alternation of air mass zones (Fig. 5), we

also obtained nine climatic zones (Fig. 7, Table 1). These zones can be used to diagnose changes in large-scale atmospheric circulations in relation to the seasonal shift of air mass zones.

The seasonal distributions of the air mass zones were used to establish four stable climatic zones (TT, SS, PP, AA), three one-class change climatic zones (ST, PS, AP), and two two-class change climatic zones (AS, PT) (Fig. 7). These climatic zones can explain climatic phenomena in the Northern Hemisphere more realistically than the original ACC because they consider both the temperature and humidity rather than just the temperature.

Finally, we subdivided the climatic zones into climatic regions to reflect differences between the ocean and continent and the east and west coasts. The 27 climatic regions established in this study (Fig. 10, Table 3) accomplish the subdivision of climatic zones left undone by Alisov (1954) and his successors. Our climatic regions can also be used to gain insight into the climatic distribution over oceans, which cannot be done by the KGCC. The renewed ACC improves the correspondence between mid-latitude climatic features and vegetation landscape, which was one of the issues of the original ACC. Thus, the renewed ACC represents a significant advance in air mass climatology. The data-driven machine learning approach can be used to establish a standard for causal CC based on air masses. Note that we did not define the climate at high altitudes higher than 2000 m (e.g., the Tibetan Plateau, Antarctica, and Greenland) in this study.

One of the advantages of using the renewed ACC is to refine the climate change research. The classical ACC cannot diagnose the effects of global warming on air mass distributions. The renewed ACC may be applicable to visualizing climate change projected by global climate models (Netzel and Stepinski 2016). We will be able to trace the meridional shift of climatological fronts associated with climate change, which can assist with carrying out the measures recommended by the Intergovernmental Panel on Climate Change and revive air mass climatology.

Abbreviations

CC	Climate classification
ACC	Alisov's climate classification
KGCC	Köppen-Geiger climate classification
ERA5	Fifth-generation ECMWF atmospheric reanalysis of the global climate
ECMWF	European Centre for Medium-Range Weather Forecasts
GPCP	Global Precipitation Climatology Project
SSE	Sum of squared errors
DBI	Davies–Bouldin index
SC	Silhouette coefficient
CHI	Cailnski–Harabasz index
BIC	Bayesian information criterion

Acknowledgements

We acknowledge Dr. Yamaura of RIKEN Center for Computational Science and two anonymous reviewers for their constructive comments. We would also like to thank Enago (www.enago.jp) for the English language review.

Author contributions

RS and TT proposed the fundamental idea, conceived and designed the study. RS carried out the air mass analysis. KF supported about the interpretation of clustering technique, particularly about the indices for determining the number of clusters. TT collaborated with the corresponding author in the construction of manuscript. All authors read and approved the final manuscript.

Funding

This work was supported in part by "Dynamic Alliance for Open Innovation Bridging Human, Environment and Materials" from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT).

Availability of data and material

The ERA5 reanalysis dataset used in this study are available via Copernicus Climate Change Service Climate Data Store (CDS), <https://cds.climate.copernicus.eu/>. The GPCP Version 2.3 monthly dataset used in this study are available via NOAA National Centers for Environmental Information (NCEI) webserver, <https://www.ncei.noaa.gov/data/global-precipitation-climatology-project-gpcp-monthly/access/>.

Declarations

Competing interests

The authors declare that they have no competing interest.

Received: 15 December 2022 Accepted: 19 March 2023

Published online: 11 April 2023

References

- Adler RF, Sapiano MRP, Huffman GJ, Wang JJ, Gu G, Bolvin D, Chiu L, Schneider U, Becker A, Nelkin E, Xie P, Ferraro R, Shin DB (2018) The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere* 9:138. <https://doi.org/10.3390/atmos9040138>
- Alisov BP (1936) Geographical types of climates. *Meteorol Hydrol* 1:16–25. (in Russian)
- Alisov BP (1954) Die Klimate der Erde (ohne das Gebiet der UdSSR). Deutscher Verlag der Wissenschaften, Berlin, p 277
- Arthur D, Vassilvitskii S (2007) k-means++: The advantages of careful seeding. In: Proceedings symposium discrete algorithms 1027–1035.
- Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF (2018) Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci Data* 5:180214. <https://doi.org/10.1038/sdata.2018.214>
- Bergeron T (1930) Richtlinien einer dynamischen Klimatologie (Guiding principles of a dynamic climatology). *Met Z* 47:246–262 (English translation by Willett HC (1931) Ground Plan of a Dynamic Climatology. *Mo Wea Rev* 59, 219).
- Brunschweiler DH (1957) Die Luftmassen der Nordhemisphäre: Versuch einer genetischen Klimaklassifikation auf aerosomatischer Grundlage. *Geogr Helv* 12:164–195
- Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat - Theory Methods* 3:1–27. <https://doi.org/10.1080/03610927408827101>
- Davies DL, Bouldin DW (1979) A cluster separation measure. In: *IEEE Trans Pattern Anal Machine Intell*, vol PAMI-1. pp 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the 2nd KDD, AAAI Press*, 226–231
- Fukui E (1962) *Climatology. Modern geography series part 1, physical and applied geography*, vol 2. Kokon Shoin, Tokyo, 454 p. (in Japanese)
- Gardner AS, Maclean IM, Gaston KJ (2020) A new system to classify global climate zones based on plant physiology and using high temporal resolution climate data. *J Biogeogr* 47:2091–2101. <https://doi.org/10.1111/jbi.13927>
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, Chiara GD, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S et al (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146:1999–2049. <https://doi.org/10.1002/qj.3803>
- Khlebnikova EI (2009) Classification of the climate of the earth. In: Gruza GV (ed) *Environmental structure and function: climate system*, vol 1. EOLSS, pp 229–245.
- Kriticos DJ, Webber BL, Leriche A, Ota N, Macadam I, Bathols J, Scott JK (2012) CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods Ecol Evol* 3:53–64. <https://doi.org/10.1111/j.2041-210X.2011.00134.x>
- Köppen W (1918) Klassifikation der klima nach temperatur, niederschlag und Jahreslauf. *Pet Mitt* 64:243–248.
- Köppen W, Geiger R (1936) Das geographische system der klimate. In: Köppen W, Geiger R (eds) *Handbuch der Klimatologie*, vol 1. Verlag von Gebrüder Borntraeger, Berlin, pp 1–44.
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. *Meteorol Z* 15:259–263. <https://doi.org/10.1127/0941-2948/2006/0130>
- MacQueen J (1967) Classification and analysis of multivariate observations. In: 5th Berkeley symposium mathematical statistics probability, pp 281–297.
- Mahlstein I, Knutti R (2010) Regional climate change patterns identified by cluster analysis. *Clim Dyn* 35:587–600. <https://doi.org/10.1007/s00382-009-0654-0>
- Metzger MJ, Bunce RGH, Jongman RHG, Sayre R, Trabucco A, Zomer R (2013) A high resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring. *Glob Ecol Biogeogr* 22:630–638. <https://doi.org/10.1111/geb.12022>
- Mizukoshi M, Yamashita S (1985) An introduction to climate. Kokon-Shoin, Tokyo, 144pp. (in Japanese)
- Netzel P, Stepinski T (2016) On using a clustering approach for global climate classification. *J Clim* 29:3387–3401. <https://doi.org/10.1175/JCLI-D-15-0640.1>
- Nishina J (2019) *Fundamental climatology (4th ed.)*, Understanding the world's natural environment through climate. Kokon Shoin, Tokyo, 144pp. (in Japanese)
- Ninomiya K (1984) Characteristics of Baiu front as a predominant subtropical front in the summer Northern Hemisphere. *J Meteor Soc Japan, Ser II* 62:880–894. <https://doi.org/10.2151/jmsj.1965.62.6.880>
- Oliver JE (1970) A genetic approach to climatic classification. *Annals Assoc Amer Geographers* 60:615–637. <https://doi.org/10.1111/j.1467-8306.1970.tb00750.x>
- Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrol Earth Syst Sci* 11:1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duc-hesnay É (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
- Pernin J, Vrac M, Crevoisier C, Chédin A (2016) Mixture model-based atmospheric air mass classification: a probabilistic view of thermodynamic profiles. *Adv Stat Clim Meteorol Oceanogr* 2:115–136. <https://doi.org/10.5194/ascmo-2-115-2016>
- Raschka S, Mirjalili V (2020) *Python machine learning programming: Theory and practice by a master data scientist (3rd Edition)*. Impress, Tokyo, pp 305–311. English edition: Raschka S, Mirjalili V (2019) *Python machine learning: machine learning and deep learning with python, scikit-learn, and TensorFlow 2 (3rd Edition)*. Packt Publishing Ltd, London.
- Rohli RV, Joyner TA, Reynolds SJ, Shaw C, Vázquez JR (2015) Globally extended Köppen-Geiger climate classification and temporal shifts in terrestrial climatic types. *Phys Geogr* 36:142–157. <https://doi.org/10.1080/02723646.2015.1016382>
- Rosenberg A, Hirschberg J (2007) V-measure: A conditional entropy-based external cluster evaluation measure. In: *Proceedings of the 2007 joint*

- conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL), pp 410–420.
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Sathiaraj D, Huang X, Chen J (2019) Predicting climate types for the continental United States using unsupervised clustering techniques. *Environmetrics* 30:e2524. <https://doi.org/10.1002/env.2524>
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–466
- Suzuki H (1961) Problems of climatic classification. *J Geogr (chigaku Zasshi)* 70:215–219. <https://doi.org/10.5026/jgeography.70.215>. (in Japanese)
- Vrac M, Billard L, Diday E, Chédin A (2012) Copula analysis of mixture models. *Comput Stat* 27:427–457. <https://doi.org/10.1007/s00180-011-0266-0>
- Wadachi K (1997) Encyclopedia of meteorology, Latest Edition (3rd ed.). Tokyo-do Shuppan, Tokyo, 607pp. (in Japanese)
- Watanabe T, Takenaka H, Nohara D (2020) Framework of forecast verification of surface solar irradiance from a numerical weather prediction model using classification with a Gaussian mixture model. *Earth Space Sci* 7:e2020EA001260. <https://doi.org/10.1029/2020EA001260>
- Yazawa T (1980) Classification of climates and division into climatic regions – Current of thoughts and problems. *Geogr Rev* 53:357–374. <https://doi.org/10.4157/grj.53.357>. (in Japanese)
- Yoshino M (1978) Climatology, Taimeido, Tokyo, 350 p. (in Japanese)
- Zhang T, Ramakrishnan R, Livny M (1996) BIRCH: an efficient data clustering method for very large databases. *ACM SIGMOD Rec* 25:103–114. <https://doi.org/10.1145/235968.233324>
- Zhang X, Yan X (2014) Spatiotemporal change in geographical distribution of global climate types in the context of climate warming. *Climate Dyn* 43:595–605. <https://doi.org/10.1007/s00382-013-2019-y>
- Zscheischler J, Mahecha MD, Harmeling S (2012) Climate classifications: the value of unsupervised clustering. *Procedia Comput Sci* 9:897–906. <https://doi.org/10.1016/j.procs.2012.04.096>

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)