INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (2014)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.4066

Identification and trend analysis of homogeneous rainfall zones over the East Asia monsoon region

Jehangir Ashraf Awan,[a] Deg-Hyo Bae[a*] and Kyung-Joong Kim[b]
[a] *Department of Civil & Environmental Engineering, Sejong University, Seoul, Republic of Korea*
[b] *Department of Computer Engineering, Sejong University, Seoul, Republic of Korea*

ABSTRACT: Rainfall is a complex phenomenon with high spatiotemporal variability. Identification of homogeneous rainfall zones to better analyse the rainfall intensity and extent is of vital significance for water resources management and mitigation of potential hazards from extreme events, i.e. droughts and floods. Appropriate zoning of homogeneous rainfall regions may give better understanding of rainfall patterns by resolving small scale variations. Although homogeneous rainfall zones have been established at country scale based on climatological mean behaviour, there has been little attempt to identify zones over broader scale with consistently homogeneous rainfall variability. This study employed K-means and Hierarchical clustering methods to establish homogeneous rainfall zones in the East Asia monsoon region (20°N–50°N, 103°E–149°E) using 30 years (1978–2007) monthly rainfall data at 0.5° grid resolution. Various cluster validation indices were used to assess the optimal number of homogeneous rainfall zones. The comparison of K-means and Hierarchical clustering showed that although both methods were able to define the homogeneous rainfall zones well with spatial contiguity, the K-means clustering outperformed the Hierarchical clustering in identifying more distinct zones with diverse rainfall characteristics. Mann-Kendall and linear regression tests were used for seasonal and annual rainfall trend analysis in the homogeneous rainfall zones. The study revealed that the region experiences distinct rainfall regimes over different zones. Furthermore, significant increasing and decreasing trends were observed over different zones with strong seasonal variation that indicate the aggravated stress of climate induced disasters, i.e. droughts and floods over the East Asia monsoon region.

KEY WORDS K-means clustering; Hierarchical clustering; cluster validation indices; homogeneous rainfall zones; East Asia monsoon; trends; Mann-Kendall.

Received 9 December 2013; Revised 7 May 2014; Accepted 9 May 2014

1. Introduction

Rainfall is a major source of water that is essential for life, industry and agriculture. The abnormalities in rainfall, i.e. excess or lack of rainfall often leads to flooding or droughts that exert severe impacts on human lives and economy of a region. For optimal management of water resources, rainfall characteristics of a specific area are necessary. The zoning of land based on homogeneous rainfall characteristics may give a better comprehension of rainfall distribution and intensity for a specific region. There are several studies for identification of climate divisions and homogeneous rainfall zones using cluster analysis methods in different countries. Türkeş and Tatlı (2011) used spectral clustering to determine homogeneous rainfall regions in Turkey. They used annual precipitation totals of 96 stations and obtained seven coherent rainfall regions in Turkey. Dikbas *et al.* (2012) used fuzzy cluster method to identify hydrologically homogeneous zones in a Turkish basin. They determined the optimal number of groups as six based on several performance evaluation indices. Bieniek *et al.* (2012) established 13 climate divisions for Alaska by employing three cluster analysis methods, i.e. Average Linkage, K-means and Ward's method on monthly average temperature data for the period of 1977–2010. Park *et al.* (2009) used empirical orthogonal function (EOF) and Ward's method to obtain climate zones in South Korea using air temperature and rainfall data. They showed that Ward's method gave better results as compared to EOF for classification of climate zones. Dambul and Jones (2007) used K-means and principal component analysis (PCA) for regional and temporal classification in Borneo, respectively. They used precipitation data of 20 stations for regional classification and used both precipitation and temperature data of six stations for temporal classification. Wolter and Allured (2007) employed Average Linkage and Ward's method for climate classification of the United States. They used precipitation and temperature data of 4324 stations for the period of 1979–2006. Sahin and Cigizoglu (2012) established sub-climate regions and sub-precipitation regime regions in Turkey using Ward's method and neuro-fuzzy method. They obtained 7 main precipitation and 16 sub-precipitation regime regions, and 7 main climate and 15 sub-climate regions for Turkey. Most of the studies for identification of homogeneous rainfall zones were limited to a country and use of station based

* Correspondence to: D.-H. Bae, Department of Civil & Environmental Engineering, Sejong University, 98 Gunja-Dong, Gwangjin-Gu, Seoul 143-747, Republic of Korea. E-mail: dhbae@sejong.ac.kr

data. However, major rainfall systems are not confined to any political boundaries. Therefore, to understand the large scale rainfall phenomenon and its regional variability, a broader scale zoning is essential.

In this study, we focused on the East Asia monsoon region (as shown in Figure 1) that encompasses several countries including eastern China, Korea, Japan and the adjacent marginal seas (Wu *et al.*, 2013). The East Asia monsoon has the largest meridional extent among all the regional monsoons on globe (Liu *et al.*, 2009). It is a dominant atmospheric phenomenon that affects about one-third of the global population, influencing hydrologic conditions and climate of East Asia. The East Asia monsoon is well distinguished with two subsystems: summer and winter monsoons (Lau, 1992). The East Asia summer monsoon (EASM), known as 'Meiyu' in China, 'Baiu' in Japan and 'Changma' in Korea, accounts for a large percentage of the annual rainfall. It has a prominent feature of rainfall concentration in the east-west elongated rain belt that is associated with a quasi-stationary subtropical front and stretches over thousands of kilometres. Seasonal march of the EASM displays distinct step-wise advances towards north and northeast with two abrupt jumps and three stationary periods. The stationary periods correspond to the eminent East Asia monsoon rainy seasons, including pre-summer rainy season over South China, Baiu/Meiyu over Japan and eastern China, and Changma over Korea (Chang, 2004). The East Asia winter monsoon (EAWM) accompanied by strong Siberian high and cold surges is also a distinct and active component of the global climate system. The strong EAWM generally leads to heavy snowstorms or low temperatures over East Asia. It not only dominates wintertime weather and climate of East Asia but also carries cold air from high northern latitudes to the Southern Hemisphere. The East Asia monsoon region also experiences significant amount of rainfall during spring and autumn seasons over some parts. The spring rainy season also known as pre-rainy season or pre-summer rainy season is generally more pronounced over South China and the East China Sea, while the autumn rainy season is more obvious in Japan (Qian *et al.*, 2002). The autumn rainy season is known as 'Kaul Changma' in Korea, 'Qiu-yu' in China and 'Shurin' or 'Akisame' in Japan.

Owing to the complex spatiotemporal structure of the East Asia monsoon rainfall, it is extremely valuable to define the broad-scale homogeneous rainfall zones with distinct rainfall features. In this study, we established homogeneous rainfall zones over the East Asia monsoon region using K-means and Hierarchical clustering. Seasonal and annual rainfall features and trends were also investigated to demonstrate distinct spatiotemporal characteristics and long-term variability of rainfall in each of the identified homogeneous rainfall zones. Rainfall trends have been examined by many researchers in various regions (Bae *et al.*, 2008; Jung *et al.*, 2011; Casimiro *et al.*, 2013; Chang *et al.*, 2012; Iwasaki, 2012; Wang *et al.*, 2013; Yang *et al.*, 2013; Duan *et al.*, 2013; Huang *et al.*, 2013; Sang *et al.*, 2013). Most of the research on rainfall trend analysis in East Asia has been focused on the

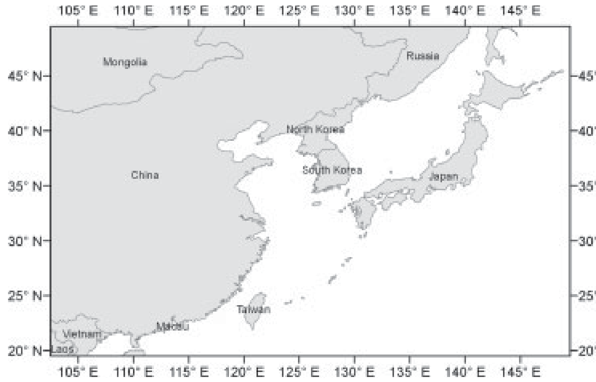

Figure 1. Study area for the identification of homogeneous rainfall zones in the East Asia monsoon region.

EASM with little attention paid to other rainy seasons, i.e. spring, autumn and winter. In this study, we investigated rainfall trends for all four seasons over different homogeneous rainfall zones of the East Asia monsoon region using Mann-Kendall and linear regression tests.

The objectives of this study were to (1) obtain homogeneous rainfall zones in the East Asia monsoon region using cluster analysis methods, (2) assess the viability of K-means and Hierarchical cluster analysis for clustering the gridded rainfall data, (3) evaluate the applicability of cluster validation indices, i.e. Calinski–Harabasz (CH), Krzanowski-Lai (KL) and Davies-Bouldin (DB) to obtain the optimum number of homogeneous rainfall zones, and (4) analyse the rainfall features and trends in each of the identified homogeneous rainfall zones.

2. Study area and dataset

The study area is shown in Figure 1. It spans over the domain size of 20–50° latitude and 103–149° longitude. The domain was selected to cover the East Asia monsoon region in which the dominant monsoon phenomenon contributes a major portion of annual rainfall. The selected domain includes eastern China, Korea, Japan and Taiwan.

To identify the homogeneous rainfall zones over the East Asia monsoon region, daily gridded rainfall data was obtained at 0.5° horizontal grid resolution from APHRODITE (Asian Precipitation – Highly Resolved Observational Data Integration Towards Evaluation of Water Resources) (Yatagai *et al.*, 2012). Monthly rainfall data was computed from the daily gridded rainfall dataset for a period of 30 years (1978–2007) and arranged in time series format for each grid point. We used the monthly rainfall time series data for cluster analysis to establish the homogeneous rainfall zones with grid points having similar rainfall intensity and consistent temporal variation during the 30-year period.

3. Methods and model configuration

In this study, we employed the most commonly used cluster analysis methods, i.e. K-means clustering and

Hierarchical clustering, to identify the homogeneous rainfall zones in the East Asia monsoon region. To obtain the optimal number of homogeneous rainfall zones, several cluster validation indices were used. The following sections describe the cluster analysis methods, cluster validation indices, trend analysis methods and model configuration used in this study.

3.1. K-means and Hierarchical clustering

K-means (MacQueen, 1967) is an unsupervised clustering algorithm that partitions the input data into K mutually exclusive clusters, where K is a positive integer that specifies the total number of clusters. To obtain clusters using K-means, these steps are followed: (1) define the K initial cluster centroids at random, (2) compute distances of all samples to the cluster centroids and assign each sample to the cluster with smallest distance, and (3) recalculate the cluster centroids and repeat steps ii and iii until there is no change in the cluster centroids. At the stage when there is no change in the cluster centroids, the clusters are stable and the clustering process ends.

Another method used in this study is the Hierarchical clustering that creates hierarchy of cluster tree by grouping data over different scales. The Hierarchical clustering method only uses similarities of samples and attempts to cluster samples which are more similar to each other within a same cluster than the samples in different clusters (Alpaydin, 2004). The hierarchical tree can be pruned either by specifying the arbitrary number of clusters or by finding natural division in the data to partition the samples into K clusters. For further details on K-means and Hierarchical clustering, the reader may refer to (Jones, 1993).

3.2. Cluster validation indices

The key task in cluster analysis is the identification of appropriate number of clusters. To determine the optimal number of clusters ‘K’, there are several cluster validation indices. In this study, we used three well-known cluster validation indices that include Calinski–Harabasz (CH), Krzanowski–Lai (KL) and Davies–Bouldin (DB). The CH index proposed by Caliński and Harabasz (1974) is known as variance ratio criterion. Given a set of N samples and partition of these samples in K mutually disjoint clusters, the CH index can be computed as:

$$\mathrm{CH} = \frac{\mathrm{SS_B}/(K-1)}{\mathrm{SS_W}/(N-K)} \quad (1)$$

where $\mathrm{SS_W}$ is the overall within-cluster variation and $\mathrm{SS_B}$ is the overall between-clusters variation. The large value of CH indicates more distinct clustering. The KL index proposed by Krzanowski and Lai (1988) is defined as:

$$\mathrm{KL} = \left| \frac{\mathrm{DIFF}(K)}{\mathrm{DIFF}(K+1)} \right| \quad (2)$$

$$\mathrm{DIFF}(K) = (K-1)^{2/P}\, W(K-1) - (K)^{2/P}\, W(K) \quad (3)$$

where the term P denotes the number of features in dataset and $W(K)$ represents the within-cluster sum of squares for any given partition of the samples given K. The number of clusters that maximizes KL is taken as optimal. The DB index introduced by Davies and Bouldin (1979) is defined as:

$$\mathrm{DB} = \frac{1}{n} \sum_{i=1, i\neq j}^{n} \max \left[\frac{\sigma_i + \sigma_j}{d\left(c_i, c_j\right)} \right] \quad (4)$$

where n denotes the number of clusters and σ_i represents the average distance of all samples in cluster i to their cluster center c_i. The term σ_j denotes the average distance of all samples in cluster j to their cluster center c_j, and $d(c_i, c_j)$ represents the distance of cluster centers $\mathrm{c_i}$ and c_j. A small value of DB index indicates better clustering.

3.3. Trend analysis methods

In this study, parametric linear regression and non-parametric Mann-Kendall tests were applied for detecting seasonal and annual rainfall trends. Linear regression and Mann-Kendall tests have been widely used for trend analysis in climate and hydrology data (Burn and Hag Elnur, 2002; Schmidli & Frei, 2005; Cheung *et al.*, 2008; Longobardi & Villani, 2010; Mondal *et al.*, 2012). Both methods have their own strengths and limitations. Linear regression method is more powerful for normally distributed data, while Mann-Kendall test does not require the data to be normally distributed. Mann-Kendall test is more resistant to outliers and also has the ability to deal with missing data values and values below a detection limit.

3.4. Model configuration for the identification of homogeneous rainfall zones

The MATLAB Statistics toolbox (Jones, 1993) was used for the implementation of K-means and Hierarchical clustering methods. The K-means clustering method was employed with random initialization and squared Euclidean distance. To avoid inconsistency in the results which may occur due to random initialization, we ran the K-means for 50 replicates, each with a new set of initial cluster centroids, and selected the one that generated clusters with a minimum total within-cluster sum of distances. Each K-means run was simulated for a maximum of 150 iterations. The Hierarchical clustering method was employed with the squared Euclidean distance and Ward’s method. We attempted different methods, i.e. single, complete and average, and found Ward’s method superior to others.

4. Homogeneous rainfall zones

To obtain the optimal number of clusters (homogeneous rainfall zones), the CH, KL and DB indices were computed for each cluster configuration obtained by applying the K-means and Hierarchical clustering for a different number of clusters (2–10). The results of CH, KL and DB indices for the K-means clustering are shown in Figure 2. A large value of CH and KL indices indicates a better clustering. Figure 2 shows that seven clusters are optimal

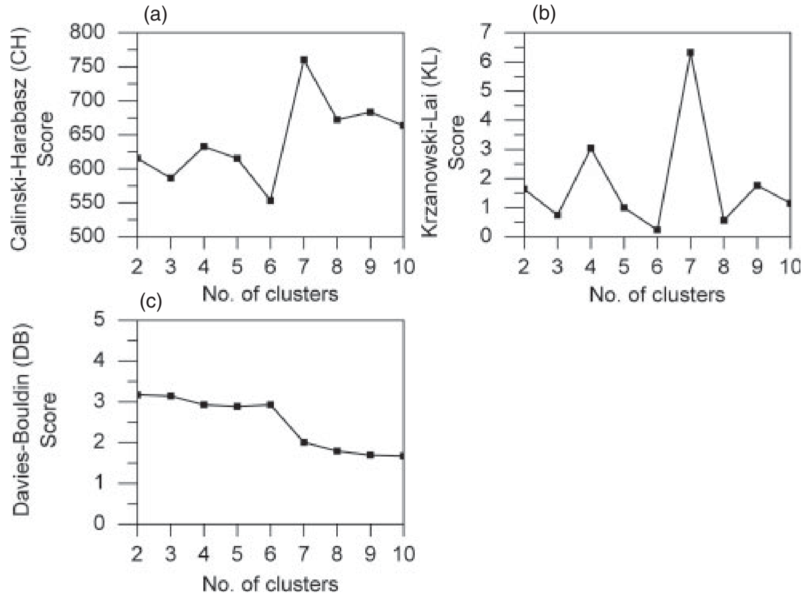

Figure 2. Identification of the optimal number of clusters for K-means clustering using cluster validity indices: (a) Calinski-Harabasz (CH); (b) Krzanowski-Lai (KL); and (c) Davies-Bouldin (DB).

with a consistent large value of both CH and KL indices for the K-means clustering in identifying the homogeneous rainfall zones over the East Asia monsoon region. For DB index, the number of clusters that minimizes the index value is considered as optimal. The index does not show a clear distinctive score; rather, it is observed to decrease consistently with increase in the number of clusters that may result in a big number of small zones rather than clear distinctive major zones. Similarly, we computed the CH, KL and DB indices for the Hierarchical clustering using a different number of clusters (2–10). The results of all three indices are shown in Figure 3. The values of both CH and KL indices reveal that six clusters are more appropriate for clustering rainfall data of the East Asia monsoon region using Hierarchical clustering. On the other hand, the DB index is again observed to decrease consistently with increase in the number of clusters resulting in no distinctive score. It reveals that the CH and KL cluster validation indices are more appropriate indicators to obtain the optimal number of clusters for identification of homogeneous rainfall zones.

The homogeneous rainfall zones over the East Asia monsoon region were defined using the optimal number of clusters determined by CH and KL indices. Figure 4(a) and (b) shows the identified homogeneous rainfall zones using the K-means (seven zones) and the Hierarchical clustering (six zones), respectively. The zones are indexed in ascending order using integers, i.e. 1–7 for K-means clustering and 1–6 for Hierarchical clustering, based on their average annual rainfall values. For instance, in Figure 4(a), Zones 1 and 7 represent the regions of the lowest and highest average annual rainfall, respectively. However, the highest rainfall zones vary for different seasons due to high seasonal rainfall variation, which is explained in the next section. Figure 4(a) and (b) illustrates that both clustering methods clearly defined the distinct homogeneous rainfall zones with spatial contiguity. The Gobi desert that covers parts of north and northwest China, and south Mongolia was identified as the lowest rainfall zone, whereas the Pacific coastal region of Japan was identified as the highest rainfall zone. However, the rainfall zones defined by the K-means and Hierarchical clustering significantly differ over Japan and Taiwan. The Hierarchical clustering method failed to capture the distinct Seto inland Sea region of Japan which is surrounded by the Chugoku and Shikoku mountains, and receives relatively small amount of rainfall. The K-means, on the other hand, successfully separated Seto inland Sea region from high rainfall coastal regions of Japan. The Hierarchical clustering also did not capture the distinct rainfall characteristics of Taiwan which differ significantly from north to south. The K-means, on the other hand, defined four rainfall zones over Taiwan. Interestingly, the K-means identified the small region of Taiwan with lowest rainfall as well, i.e. coast of the northern Chianan plain and the Penghu archipelago. Both clustering methods showed similar results for Korea and defined it as part of one zone. To further compare the performance of K-means and Hierarchical clustering, we generated seven homogeneous rainfall zones using Hierarchical clustering as well (as shown in Figure 4(c)). However, it did not show any improvement for rainfall zones in Japan and Taiwan. The results reveal that the K-means is superior to Hierarchical clustering, and CH and KL are more suitable cluster validation indices for the identification of homogeneous rainfall zones over selected domain of the East Asia monsoon region.

5. Rainfall features and trend analysis

To obtain the distinct spatiotemporal characteristics and long-term variability of the East Asia monsoon rainfall, the rainfall features and trends were investigated in each of the

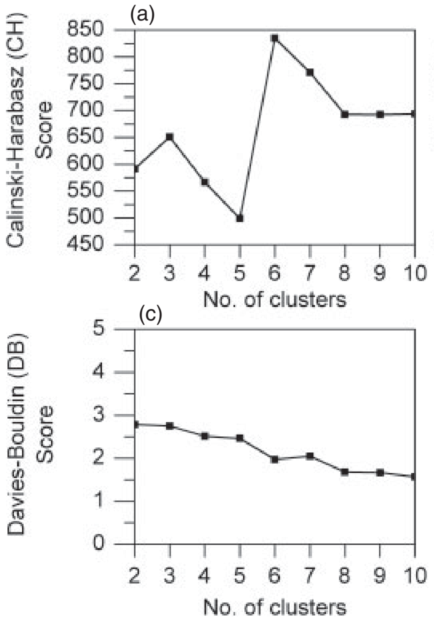

Figure 3. Identification of the optimal number of clusters for Hierarchical clustering using cluster validity indices: (a) Calinski-Harabasz (CH); (b) Krzanowski-Lai (KL); and (c) Davies-Bouldin (DB).

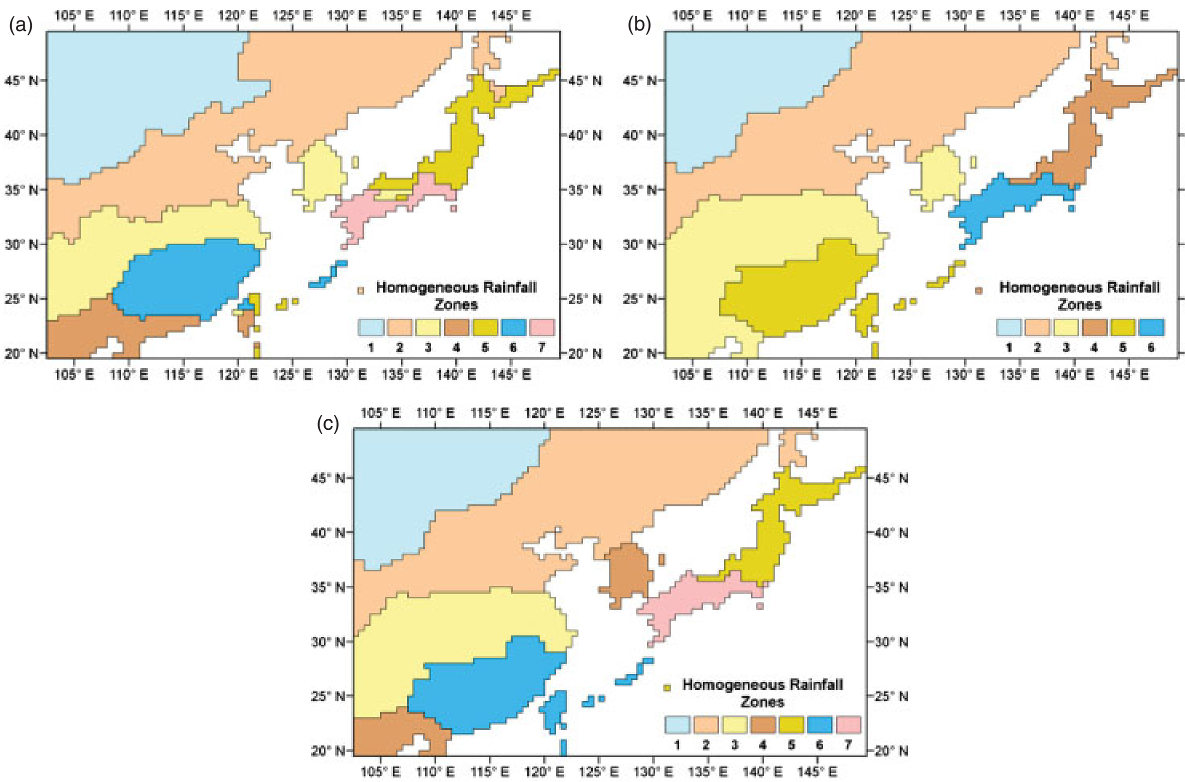

Figure 4. Homogeneous rainfall zones in the East Asia monsoon region: (a) Seven zones using K-means clustering; (b) Six zones using Hierarchical clustering; and (c) Seven zones using Hierarchical clustering.

seven homogeneous rainfall zones, which are presented in the following sections.

5.1. Rainfall features

The mean monthly and seasonal rainfall distributions were obtained using spatial rainfall averages of each zone for a period of 30 years from 1978 to 2007 (as shown in Figure 5). The seasonal rainfall distributions show that although the summer rainy season is dominant over most of the zones, the spring and autumn seasons are also well pronounced over some parts of the region. For instance, Zones 5 and 6 receive significant amount of rainfall in autumn and spring seasons, respectively, and they even exceed the summer rainfall. The detailed seasonal and

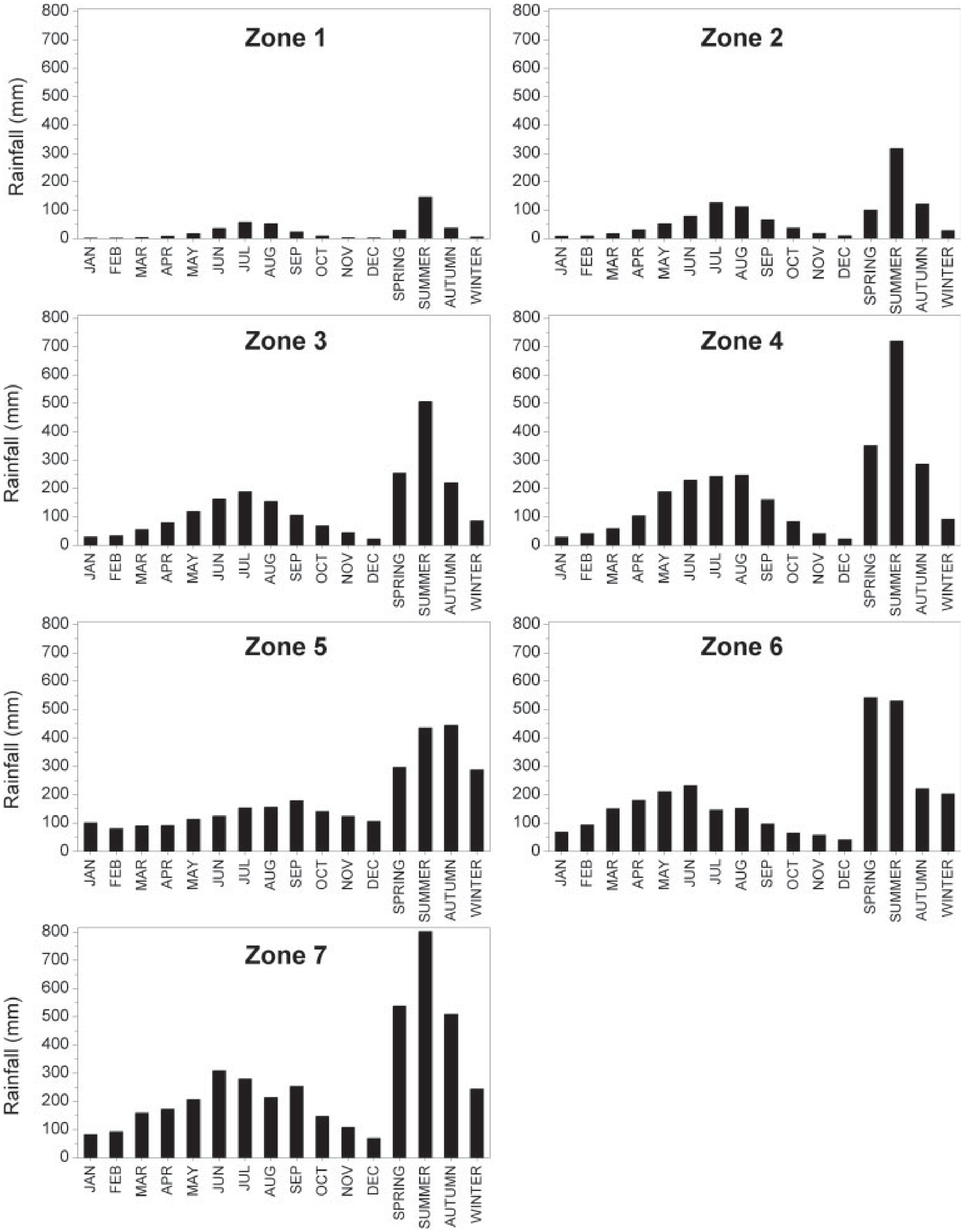

Figure 5. Temporal distribution of mean monthly and seasonal rainfall during the period from 1978 to 2007.

annual rainfall characteristics of each zone are presented in Table 1. In the East Asia monsoon region, Zone 1 is the lowest rainfall region with 219.0 mm mean annual rainfall. It experiences a rainfall regime in the months from June to September with a peak in July (57.8 mm), and 77.6% of the mean annual rainfall occurs in this period. On the other hand, it receives very little or no rainfall during the months of November to April. Examining the seasonal rainfall, it is evident that summer (June-July-August) rainfall is dominant with 66.6% of the mean annual rainfall, while only 2.7% of the mean annual rainfall occurs in the winter (December-January-February)

Table 1. Seasonal and annual rainfall (mm) characteristics of the homogeneous rainfall zones in the East Asia monsoon region, 1978–2007.

	Mean seasonal rainfall					Rainfall range[a]			
	Spring	Summer	Autumn	Winter	Mean annual rainfall	Spring	Summer	Autumn	Winter
Zone 1	30.1	145.9	37.0	6.0	219.0	33.8	120.4	27.4	6.5
Zone 2	101.2	317.4	122.0	28.5	568.9	76.4	137.2	77.3	24.3
Zone 3	254.0	506.7	221.1	87.3	1068.6	167.2	249.0	163.4	82.0
Zone 4	351.8	719.1	285.8	93.3	1449.5	304.5	477.7	257.1	171.4
Zone 5	297.1	435.2	444.6	288.3	1464.9	145.0	298.5	271.2	127.6
Zone 6	542.4	530.7	220.9	204.2	1497.1	287.9	325.8	219.6	296.5
Zone 7	538.1	801.9	509.0	244.3	2093.1	365.5	987.8	597.7	253.7

[a]Difference between maximum and minimum rainfall values.

season. The spring (March-April-May) and autumn (September-October-November) seasons receive 13.7 and 16.9% of the mean annual rainfall, respectively.

Zone 2 has the second lowest mean annual rainfall (568.9 mm) of all zones in the East Asia monsoon region. The rainfall distribution of Zone 2 is similar to that of Zone 1. It receives about 55.8% of the mean annual rainfall during summer season with peak rainfall in the month of July (126.3 mm), while only 5.0% of the mean annual rainfall occurs in the winter season. About 17.8 and 21.4% of the mean annual rainfall occur during the spring and autumn seasons, respectively.

The mean annual rainfall of Zone 3 is about 1068.6 mm with peak rainfall in the month of July (188.2 mm). The summer rainfall is dominant in Zone 3 with 47.4% of the mean annual rainfall followed by spring and autumn seasons that receive 23.8 and 20.7% of the mean annual rainfall, respectively. The winter season receives the lowest rainfall with a small share of only 8.2% of the mean annual rainfall.

Zone 4 receives a mean annual rainfall of about 1449.5 mm. It has the second highest summer rainfall (719.1 mm) of all zones in the East Asia monsoon region. About 49.6% of the mean annual rainfall occurs in the summer season with peak rainfall in the month of August (246.7 mm) that slightly exceeds the July average (243.0 mm). The spring and autumn seasons receive 24.3% and 19.7% of the mean annual rainfall, respectively, while only 6.4% of the mean annual rainfall occurs in the winter season.

Zone 5 receives a mean annual rainfall of about 1464.9 mm and has the highest winter rainfall (288. 3 mm) of all zones in the East Asia monsoon region. The rainfall regime is relatively uniform throughout the year with a slightly higher share during summer (29.7%) and autumn (30.4%) seasons. Peak rainfall occurs in the month of September (178.9 mm). The spring and winter seasons receive about 20.3 and 19.7% of the mean annual rainfall, respectively.

Zone 6 has the second highest mean annual rainfall (1497.1 mm) and the highest spring rainfall (542.4 mm) of all zones in the East Asia monsoon region. The spring and summer rainfall seasons are dominant in Zone 6 with 36.2% and 35.4% of the mean annual rainfall, respectively. Peak rainfall occurs in the month of June (232.9 mm). The autumn and winter seasons receive relatively less rainfall with shares of about 14.8 and 13.6% of the mean annual rainfall, respectively.

Zone 7 has the highest mean annual rainfall (2093.1 mm) with the highest summer (801.9 mm) and autumn (509.0 mm) rainfall, and the second highest spring (538.1 mm) and winter (244.3 mm) rainfall of all zones in the East Asia monsoon region. Peak rainfall occurs in the months of June (309.1 mm) and July (279.7 mm) followed by a second peak in the month of September (253.0 mm). The summer season is dominant in Zone 7 with 38.3% of the mean annual rainfall followed by spring and autumn seasons having 25.7% and 24.3% of the mean annual rainfall, respectively. The winter season receives relatively less rainfall with only 11.7% share of the mean annual rainfall.

5.2. Trend analysis

Figure 6 shows the annual rainfall time series for each of the seven homogeneous rainfall zones with a linear regression trend line and total change in the annual rainfall over the past 30 years (1978–2007). The significant downward trends are evident in Zones 1 and 2 with a decrease of 11.8% (27.5 mm) and 10.9% (65.7 mm) in annual rainfall, respectively. The long-term rainfall variability in Zone 3 is almost flat with a little increase of about 1.2% (13.0 mm). The trend line for Zone 4 shows slight downward trend with a decrease of 4.4% (65.0 mm) in the annual rainfall. The significant increasing trend is evident in Zone 5 with 11.0% (153.1 mm) increase in the annual rainfall. Zones 6 and 7 also show a small increase in annual rainfall of about 4.4% (63.9 mm) and 4.1% (84.2 mm), respectively.

The detailed seasonal and annual rainfall trends statistics were estimated using Mann-Kendall and linear regression tests. Table 2 lists the Mann-Kendall tests statistics for each of the seven homogeneous rainfall zones. It shows decreasing annual rainfall trends for Zones 1–4 and increasing trends for Zones 5–7. The annual rainfall trends for Zones 2 and 5 are statistically significant at the 0.10 level. The spring season shows decreasing trend for all the zones except Zones 1, 3 and 5. Particularly, the decreasing spring rainfall trends for Zones 4 and 6 are statistically significant at the 0.10 level. During summer season, the Mann-Kendall test shows an increasing trend for Zones 4–6 and decreasing trend for all other zones. Specifically, the decreasing trend for Zone 1 and increasing trend for Zone 6 during the summer season

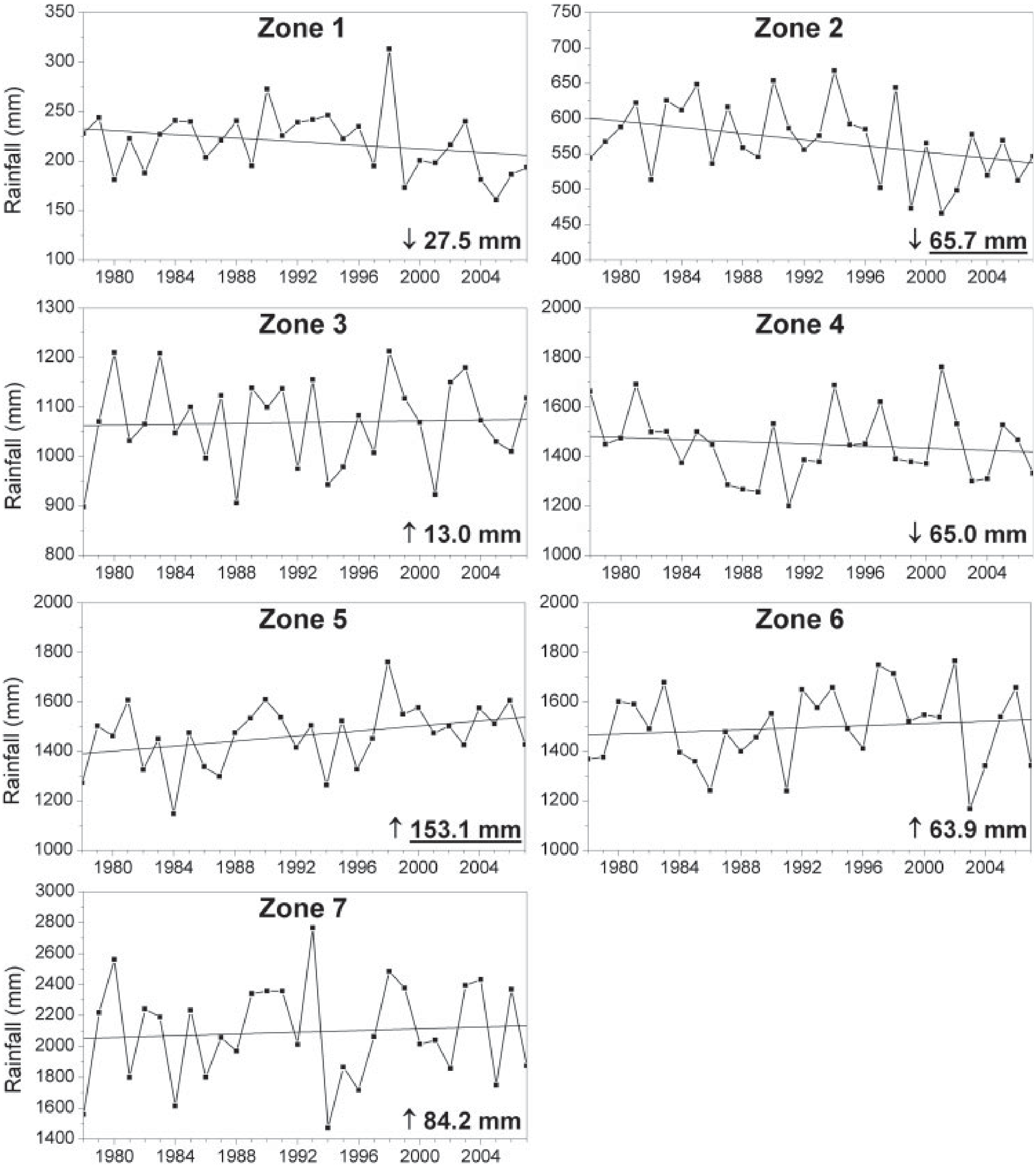

Figure 6. Time series of the annual rainfall with linear regression trend line and total change (↑ increase; ↓ decrease) during the period from 1978 to 2007. Note that the *y*-axis has different scales, and the underlined values indicate statistically significant trend at the 0.10 level.

are statistically significant at the 0.05 level. The autumn season has decreasing rainfall trend for all the zones except Zones 5 and 7, which exhibit increasing rainfall trends. Particularly, Zone 5 shows statistically significant increasing trend at the 0.10 level. The trend statistics show increasing trend in winter rainfall for all the zones; specifically, the increasing trends for Zones 3 and 5 are statistically significant at the 0.10 level.

The linear regression test statistics of seasonal and annual rainfall trends are presented in Table 3. The annual and spring rainfall trends obtained from the linear regression test are similar to that obtained from the Mann-Kendall test for all the zones except Zone 3 (annual) and Zone 2 (spring). Contrary to the Mann-Kendall test, the linear regression test exhibits increasing annual and spring rainfall trends for Zones 3 and 2, respectively. During the summer season, the linear regression test statistics show increasing rainfall trend in all the zones except Zones 1 and 2. Particularly, the decreasing summer rainfall trend for Zone 2 and increasing trend for Zone 6 are statistically significant at the 0.10 and 0.05 levels, respectively. Like the Mann-Kendall test, the linear regression test shows decreasing rainfall trend for the autumn season in all the zones except Zones 5 and 7.

Table 2. Seasonal and annual rainfall trends using Mann-Kendall test, 1978–2007.

	Spring		Summer		Autumn		Winter		Annual	
	β	p	β	p	β	p	β	p	β	p
Zone 1	0.098△	0.643	−1.198$\underline{\blacktriangledown}$	0.030	−0.101▽	0.335	0.066△	0.216	−0.814▽	0.101
Zone 2	−0.252▽	0.695	−1.459▽	0.108	−1.013▽	0.134	0.024△	0.694	−2.486▼	0.094
Zone 3	0.293△	0.593	−0.714▽	0.858	−1.626▽	0.154	0.993▲	0.063	−1.356▽	0.775
Zone 4	−1.056▼	0.094	3.889△	0.318	−2.330▽	0.116	0.405△	0.955	−0.299▽	0.354
Zone 5	0.817△	0.748	1.764△	0.144	2.095▲	0.087	1.579▲	0.078	4.721▲	0.080
Zone 6	−1.972▼	0.080	4.985$\underline{\blacktriangle}$	0.020	−1.206▽	0.301	1.325△	0.488	4.565△	0.412
Zone 7	−2.169▽	0.544	−3.500▽	1.000	2.710△	0.643	2.209△	0.464	3.127△	0.475

△ increasing (not significant); ▲ increasing ($\alpha = 0.10$); $\underline{\blacktriangle}$ increasing ($\alpha = 0.05$); ▽ decreasing (not significant); ▼ decreasing ($\alpha = 0.10$); $\underline{\blacktriangledown}$ decreasing ($\alpha = 0.05$); β = Mann-Kendall trend slope; p = probability of obtaining test statistic value.

Table 3. Seasonal and annual rainfall trends using linear regression test, 1978–2007.

	Spring		Summer		Autumn		Winter		Annual	
	b	p	b	p	b	p	b	p	b	p
Zone 1	0.174△	0.358	−0.305▽	0.107	−0.195▽	0.301	0.248△	0.197	−0.253▽	0.182
Zone 2	0.024△	0.901	−0.359▼	0.058	−0.320▼	0.090	−0.074▽	0.702	−0.363▼	0.055
Zone 3	0.085△	0.654	0.049△	0.797	−0.298▽	0.114	0.353▲	0.067	0.043△	0.821
Zone 4	−0.314▼	0.097	0.292△	0.122	−0.439$\underline{\blacktriangledown}$	0.020	−0.060▽	0.754	−0.137▽	0.469
Zone 5	0.046△	0.810	0.208△	0.271	0.282△	0.136	0.403$\underline{\blacktriangle}$	0.037	0.354▲	0.061
Zone 6	−0.361▼	0.056	0.446$\underline{\blacktriangle}$	0.018	−0.136▽	0.472	0.122△	0.527	0.121△	0.522
Zone 7	−0.129▽	0.494	0.011△	0.954	0.119△	0.529	0.209△	0.277	0.077△	0.683

△ increasing (not significant); ▲ increasing ($\alpha = 0.10$); $\underline{\blacktriangle}$ increasing ($\alpha = 0.05$); ▽ decreasing (not significant); ▼ decreasing ($\alpha = 0.10$); $\underline{\blacktriangledown}$ decreasing ($\alpha = 0.05$); b = standardized coefficient of slope; p = probability of obtaining test statistic value.

However, the decreasing autumn rainfall trends for Zones 2 and 4 using the linear regression test are statistically significant at the 0.10 and 0.05 levels, respectively. In the winter season, the linear regression test statistics show increasing rainfall trends for all the zones except Zones 2 and 4. Specifically, the increasing winter rainfall trends for Zones 3 and 5 are statistically significant at the 0.10 and 0.05 levels, respectively.

The spatial distribution of seasonal and annual rainfall trends over the East Asia monsoon region using the Mann-Kendall test is shown in Figure 7. The trends are statistically significant at the 0.10 level. The bar-chart in each panel shows the percentage of grid points with statistically significant ($\alpha = 0.10$) positive or negative trends in each zone. The trend analysis for annual rainfall shows decreasing trend over Far East Russia, parts of Mongolia and northeast, central and south China. The decreasing annual rainfall trend is mainly confined in Zone 2 with 40% of its grid points having statistically significant decreasing trend followed by Zone 4 (20%) and Zone 1 (18%). On the other hand, the increasing annual rainfall trend is prominent over east Japan, Korea and south Taiwan. The small regions in east China and central Mongolia also show an increasing annual rainfall trend. Zone 5 exhibits high percentage (18%) of grid points with statistically significant increasing annual rainfall trend.

The spatial trend analysis for the spring season reveals a dominant decreasing trend over southeast China. Some parts of central China and Far East Russia also experience decreasing rainfall trend. The decreasing spring rainfall trend is mainly focused over Zones 4 and 6 having 32 and 35% of their grid points with statistically significant decreasing trend, respectively. The increasing spring rainfall trend is apparent only in some parts of northeast China and North Korea with maximum of only 8% grid points in Zone 2.

In the summer season, the decreasing rainfall trend is evident over Far East Russia, north Mongolia and parts of northeast and central China with a large spatial coverage of 21 and 35% in Zones 1 and 2, respectively. On the other hand, the increasing summer rainfall trend is well pronounced over southeast China and Taiwan, being well confined in Zones 6 and 4 with spatial extent of 45 and 19%, respectively.

In the autumn season, the decreasing rainfall trend is prominent over central Mongolia, Far East Russia and parts of northeast, central and south China with spatial extent of 11, 43, 30 and 59% in Zones 1–4, respectively. On the other hand, the increasing rainfall trend is shown only in some parts of Mongolia, Korea and eastern Japan.

In the winter season, the increasing rainfall trend is well pronounced over Mongolia, eastern Japan and Yangtze River valley in China with a large spatial extent of 31% in Zone 3 followed by 20 and 23% in Zones 1 and 5, respectively. The decreasing trend for winter rainfall is mainly confined in Far East Russia with spatial extent of 15% in Zone 2.

6. Discussion

In this study, we identified the homogeneous rainfall zones to simplify the complex spatiotemporal structure of rainfall

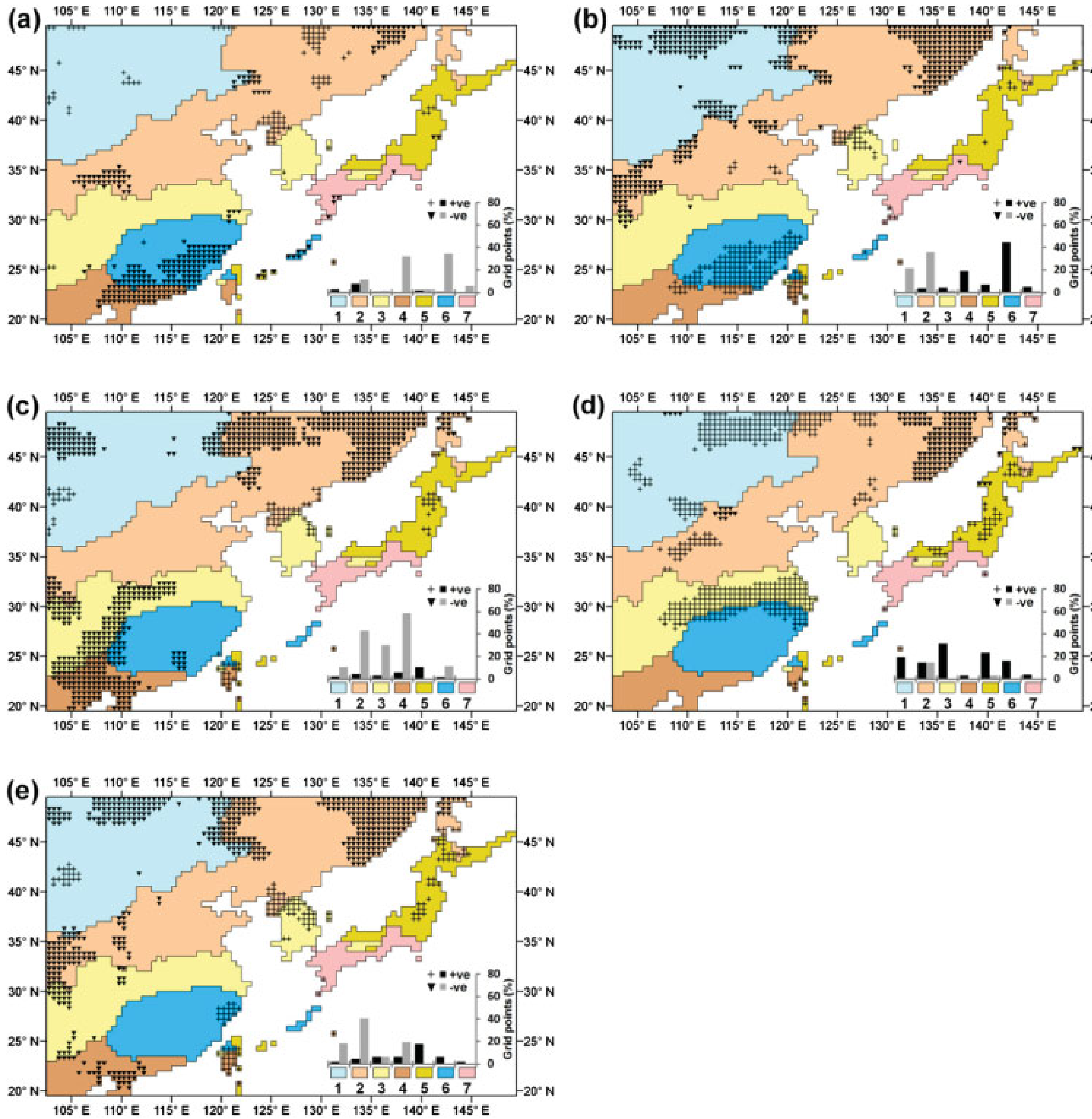

Figure 7. Spatial distribution of positive (+) and negative (▼) trends, and percentage of grids points (bar chart) with positive and negative trends in each of the seven homogeneous rainfall zones using Mann-Kendall test ($\alpha = 0.1$) for (a) spring, (b) summer, (c) autumn, (d) winter, and (e) annual rainfall during the period from 1978 to 2007.

over the East Asia monsoon region. The rainfall characteristics of the homogeneous rainfall zones show that although EASM is dominant over the region, most of the zones have unique rainfall regime under the influence of distinct rainfall share from different rainy seasons. Moreover, the rainfall trends are not uniform over the East Asia monsoon region. In general, most of the zones exhibit unique long-term rainfall variability during different seasons.

The decreasing summer rainfall trend observed in low rainfall zones covering Mongolia, north China and Far East Russia, exhibits weakening of EASM. This drying trend in northern China concurs with the increasing trend of summer rainfall over southeast China and parts of Korea, Taiwan and Japan that seems to correspond with declined northward and northeastward penetration of the rain belt. Similar trends often referred to as 'northern drought and southern flooding in China' have been reported by several studies (Hu *et al.*, 2003; Zhou *et al.*, 2009; Li *et al.*, 2010). In seasonal perspective, the rainfall trend analysis exhibits negative trend in spring and autumn seasons, while exhibiting positive trend in summer and winter seasons over south China. Specifically, the decreasing trend in spring and increasing trend in summer over southeast China suggests that the spring-drought and summer-flooding is a more apt description of long-term rainfall variations over southeast China. Xin *et al.* (2006) also reported significant decreasing trend in spring rainfall over south China and discussed associated atmospheric circulation changes. In addition to spring and summer trends, the decreasing trend in autumn and increasing trend in winter over south China make this region more susceptible to climate induced disasters. On the other hand, the northern parts also exhibit decreasing tendency in autumn rainfall. The decline in autumn rainfall along with decreasing summer rainfall trend signifies the aggravated drought stress over northern parts of the East Asia monsoon region. Here, the proposed homogeneous rainfall zones are helpful in confining the boundaries of these long-term rainfall variations and offering unique insight into the spatial extent of rainfall trends over the region. For instance, the statistically significant increasing summer rainfall trend over south China is mainly confined in Zone 6. On the other hand, the zone has minor influence of long-term rainfall variations during autumn and winter seasons which are well pronounced over surrounding zones alongside its boundary, such as the decreasing autumn rainfall trend over Zones 3 and 4, and the increasing winter rainfall trend over Zone 3. Similarly, the decreasing spring rainfall trend over the East Asia monsoon region is mainly confined in Zones 4 and 6. In Japan, Zone 7 that covers the west and southwest Pacific coastal region has not shown any significant trend throughout the year, while Zone 5 that covers east and Japan Sea coastal region exhibits significant increasing rainfall trends. Particularly, the increasing trend in Zone 5 is well pronounced for the winter and annual rainfall. Further, this study suggests that the seasonal rainfall trend analysis is of vital importance over the region with high seasonal variation as mere annual rainfall trend analysis may often be misleading. For instance, Zone 6 has not shown any significant trend in annual rainfall, while it experiences a significant decreasing trend in the spring and increasing trend in the summer season.

We believe that the proposed zoning will assist in effectively planning and implementing climate change mitigation and adaptation strategies by determining more precise climate change impacts in the homogeneous rainfall zones over the East Asia monsoon region.

7. Conclusions

In this article, the K-means and Hierarchical clustering methods were used to obtain the homogeneous rainfall zones in the East Asia monsoon region. Three cluster validation indices, i.e. CH, KL and DB were employed to assess the optimal number of clusters (homogeneous rainfall zones) for both the K-means and Hierarchical clustering. The CH and KL indices were found more suitable with the consistent results. The comparison of the K-means and Hierarchical clustering revealed the superiority of K-means in defining the homogeneous rainfall zones with more distinct rainfall features, especially over Japan and Taiwan.

The rainfall trend analysis was conducted using the Mann-Kendall and linear regression tests. Annual rainfall showed decreasing tendency in low rainfall zones and increasing tendency in high rainfall zones. Specifically, the decreasing trend in Zone 2 and the increasing trend in Zone 5 were statistically significant at the 0.10 level. The spatial distribution of statistically significant ($\alpha = 0.10$) trends, in general, showed the dominant decreasing trend in the East Asia monsoon region during all seasons except winter which exhibited well pronounced increasing trend. However, the rainfall trends were not uniform over the region, and most of zones exhibited unique long-term rainfall variability in different seasons. The seasonal trend analysis revealed prominent spring-drought (Zones 4 and 6) and summer-flooding (Zone 6) patterns over southeast China. Other significant trends observed over the region include: decreasing rainfall trend for all seasons in parts of Far East Russia, decreasing trends for summer and autumn rainfall and increasing trend for winter rainfall in the northern parts of the East Asia monsoon region, decreasing trend for autumn rainfall and increasing trend for winter rainfall in southern China, and increasing trend for winter rainfall in eastern Japan. These trends with strong seasonal variation indicate the increasing vulnerability to climate induced natural disasters (i.e. droughts and floods) in the East Asia monsoon region.

In the future, we will attempt to analyse the trends and frequency of droughts in the homogeneous rainfall zones using drought indices. Further research will also be carried out to analyse the influence of various driving factors on the unique rainfall regimes of different zones.

Acknowledgements

This work was funded by the Korea Meteorological Administration Research and Development Program under Grant CATER 2012–3100 and also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2011-0030040).

References

Alpaydin E. 2004. Introduction to Machine Learning. MIT Press: Cambridge, MA.

Bae DH, Jung IW, Chang H. 2008. Long-term trend of precipitation and runoff in Korean river basins. *Hydrol. Processes* **22**: 2644–2656.

Bieniek PA, Bhatt US, Thoman RL, Angeloff H, Partain J, Papineau J, Fritsch F, Holloway E, Walsh JE, Daly C, Shulski M, Hufford G, Hill DF, Calos S, Gens R. 2012. Climate divisions for Alaska based on objective methods. *J. Appl. Meteorol. Climatol.* **51**: 1276–1289.

Burn DH, Hag Elnur MA. 2002. Detection of hydrologic trends and variability. *J. Hydrol.* **255**: 107–122.

Caliński T, Harabasz J. 1974. A dendrite method for cluster analysis. *Commun. Stat. Theory Methods* **3**: 1–27.

Casimiro L, Sven W, Labat D, Ronchail J, Espinoza JC, Guyot JL. 2013. Trends in rainfall and temperature in the Peruvian Amazon–Andes basin over the last 40 years (1965–2007). *Hydrol. Processes* **27**: 2944–2957.

Chang C-P. 2004. *East Asian Monsoon*. World Scientific: Singapore.

Chang CP, Lei Y, Sui CH, Lin X, Ren F. 2012. Tropical cyclone and extreme rainfall trends in East Asian summer monsoon since mid-20th century. *Geophys. Res. Lett.* **39**: L18702.

Cheung WH, Senay GB, Singh A. 2008. Trends and spatial distribution of annual and seasonal rainfall in Ethiopia. *Int. J. Climatol.* **28**: 1723–1734.

Dambul R, Jones P. 2007. Regional and temporal climatic classification for Borneo. *Geografia* **3**: 84–105.

Davies DL, Bouldin DW. 1979. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-1**: 224–227.

Dikbas F, Firat M, Koc AC, Gungor M. 2012. Classification of precipitation series using fuzzy cluster method. *Int. J. Climatol.* **32**: 1596–1603.

Duan A, Wang M, Lei Y, Cui Y. 2013. Trends in summer rainfall over China associated with the Tibetan Plateau sensible heat source during 1980–2008. *J. Clim.* **26**: 261–275.

Hu ZZ, Yang S, Wu R. 2003. Long-term climate variations in China and global warming signals. *J. Geophys. Res. Atmos. (1984–2012)* **108**: 4614.

Huang J, Sun S, Zhang J. 2013. Detection of trends in precipitation during 1960–2008 in Jiangxi province, southeast China. *Theor. Appl. Climatol.* **114**: 237–251.

Iwasaki H. 2012. Recent positive trend in heavy rainfall in eastern Japan and its relation with variations in atmospheric moisture. *Int. J. Climatol.* **32**: 364–374.

Jones B. 1993. *MATLAB: Statistics Toolbox User's Guide*. MathWorks: Natick, MA.

Jung IW, Bae DH, Kim G. 2011. Recent trends of mean and extreme precipitation in Korea. *Int. J. Climatol.* **31**: 359–370.

Krzanowski WJ, Lai Y. 1988. A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics* **44**: 23–34.

Lau K-M. 1992. East Asian summer monsoon rainfall variability and climate teleconnection. *J. Meteorol. Soc. Jpn.* **70**: 211–242.

Li H, Dai A, Zhou T, Lu J. 2010. Responses of East Asian summer monsoon to historical SST and atmospheric forcing during 1950–2000. *Clim. Dyn.* **34**: 501–514.

Liu J, Wang B, Wang H, Kuang X, Ti R. 2009. Forced response of the East Asian summer rainfall over the past millennium: results from a coupled model simulation. *Clim. Dyn.* **1**: 177.

Longobardi A, Villani P. 2010. Trend analysis of annual and seasonal rainfall time series in the Mediterranean area. *Int. J. Climatol.* **30**: 1538–1546.

MacQueen J. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Berkeley, CA, 281–297.

Mondal A, Kundu S, Mukhopadhyay A. 2012. Rainfall trend analysis by Mann-Kendall test: a case study of North-Eastern part of Cuttack district, Orissa. *Int. J. Geol. Earth Environ. Sci.* **2**: 70–78.

Park C, Choi Y, Moon J, Yun W. 2009. Classification of climate zones in South Korea considering both air temperature and rainfall. *J. Korean Geogr. Soc.* **44**: 1–16.

Qian W, Kang H-S, Lee D-K. 2002. Distribution of seasonal rainfall in the East Asian monsoon region. *Theor. Appl. Climatol.* **73**: 151–168.

Sahin S, Cigizoglu HK. 2012. The sub-climate regions and the sub-precipitation regime regions in Turkey. *J. Hydrol.* **450**: 180–189.

Sang YF, Wang Z, Li Z, Liu C, Liu X. 2013. Investigation into the daily precipitation variability in the Yangtze River Delta, China. *Hydrol. Processes* **27**: 175–185.

Schmidli J, Frei C. 2005. Trends of heavy precipitation and wet and dry spells in Switzerland during the 20th century. *Int. J. Climatol.* **25**: 753–771.

Türkeş M, Tatlı H. 2011. Use of the spectral clustering to determine coherent precipitation regions in Turkey for the period 1929–2007. *Int. J. Climatol.* **31**: 2055–2067.

Wang H, Chen Y, Chen Z. 2013. Spatial distribution and temporal trends of mean precipitation and extremes in the arid region, northwest of China, during 1960–2010. *Hydrol. Processes* **27**: 1807–1818.

Wolter K, Allured D. 2007. New climate divisions for monitoring and predicting climate in the US. *Intermountain West Clim. Summ.* **3**: 2–6.

Wu Q, Yan Y, Chen D. 2013. A linear Markov Model for East Asian monsoon seasonal forecast. *J. Clim.* **26**: 5183–5195.

Xin X, Yu R, Zhou T, Wang B. 2006. Drought in late spring of South China in recent decades. *J. Clim.* **19**: 3197–3206.

Yang L, Villarini G, Smith JA, Tian F, Hu H. 2013. Changes in seasonal maximum daily precipitation in China over the period 1961–2006. *Int. J. Climatol.* **33**: 1646–1657.

Yatagai A, Kamiguchi K, Arakawa O, Hamada A, Yasutomi N, Kitoh A. 2012. APHRODITE: constructing a long-term daily gridded precipitation dataset for Asia based on a Dense Network of Rain Gauges. *Bull. Am. Meteorol. Soc.* **93**: 1401–1415.

Zhou T, Gong D, Li J, Li B. 2009. Detecting and understanding the multi-decadal variability of the East Asian Summer Monsoon – recent progress and state of affairs. *Meteorol. Z.* **18**: 455–467.

