Vol. 67: 151–163, 2016
doi: 10.3354/cr01365

CLIMATE RESEARCH
Clim Res

Published February 24

Projected variations in the regional clustering of precipitation stations around Chicago

Myoung-Jin Um[1], Momcilo Markus[2], Donald J. Wuebbles[3], Yeonjoo Kim[1,*]

[1]Department of Civil and Environmental Engineering, Yonsei University, Seoul 120-749, Korea
[2]Illinois State Water Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign, 2204 Griffith Drive, Champaign, Illinois 61820, USA
[3]Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, 105 S. Gregory Street, Urbana, Illinois 61801, USA

ABSTRACT: Precipitation frequency analyses are typically performed for regions or groups of neighboring gauges that represent similar topographic and climatic characteristics. For past precipitation frequency analyses, regions are defined using observed datasets. However, for frequency analysis of projected precipitation, climate change can influence the grouping of sites, such as homogeneous regions of extreme precipitation. Therefore, this study investigates the effect of region definition in the regional frequency analysis (RFA) of extreme precipitation in climate change scenarios. Specifically, we use a statistically downscaled climate modeling-based dataset for Chicago, Illinois, USA and 8 climate change cases (4 models with 2 future climate scenarios). The cases were developed using the asynchronous regional regression model, which focuses on accurately resolving the tails of the probability distributions of precipitation data. For the 40 stations around Chicago, the clustering of precipitation stations varies. The precipitation characteristics, such as the averages of the monthly maximum and annual precipitation and the L-moments of the annual maximum daily precipitation, vary significantly over different time periods and regional clusters. Furthermore, the number of stations that exhibit heterogeneity in terms of their clusters is lower when changes in the clustering of the climate regions are considered than when these changes are ignored. The results of this work illustrate the need to consider changes in the regional clustering of precipitation stations in RFA, which is particularly useful for designing water-related infrastructure in response to climate change.

KEY WORDS: Climate change · Regional clustering · Extreme precipitation · Regional frequency analysis

1. INTRODUCTION

The impacts of and adaptation to climate change in sectors such as water resource management and agricultural practices occur at regional or local scales, whereas climate change occurs over the entire globe. Therefore, downscaling techniques are often used to transform global climate model (GCM)-based precipitation and temperature data to impact-relevant spatial and temporal scales. Several different dynamic or statistical downscaling approaches have been developed to downscale monthly GCM-based precipitation and temperature data on grids with resolution of 1° or coarser into daily data at the local scale or grids with higher spatial resolution. The resulting downscaled data can be used to determine the potential magnitude and frequency of climate extremes. It is critical to understand the characteristics of precipitation extremes for designing various hydraulic structures such as dams and bridges and for adapting to climate change.

Site observations of variables such as annual maximum precipitation are frequently utilized to perform frequency analysis at a site; however, accurate quan-

*Corresponding author: yeonjoo.kim@yonsei.ac.kr

tile estimates are often limited due to a scarcity of data. To overcome this problem, regional frequency analysis (RFA), which is based on polling precipitation information from several stations, has been used. RFA utilizes data from other sites to estimate the quantiles and compensates for the temporal limits of precipitation records at a site with spatial randomness by using data from other sites in a homogeneous region (Schaefer 1990). If neighboring stations are correlated, the spatial correlation has to be explicitly taken into account (e.g. Maraun et al. 2010). In general, RFA yields relatively more accurate estimates at a site than does frequency analysis (Cunnane 1989) and may be advantageous in estimating extreme quantiles (Cunnane 1989, Hosking &Wallis 1997).

The traditional stationarity assumption of RFA has been challenged when RFA has been applied to climate change. As a result, non-stationarity RFA has been proposed, but few studies of non-stationarity RFA have been conducted to date. In conducting RFA of climate change scenarios, the grouping of sites or regions could vary among different periods. Several recent studies have assumed that the regional clustering remains the same (e.g. Santos et al. 2011, Gabriele & Chiaravalloti 2013, Basu & Srinvas 2015), although others, such as Cunderlik & Burn (2003) and Leclerc & Ouarda (2007), do not.

In this study, we investigate the influence of region definition in the RFA of extreme precipitation in climate change scenarios. Assuming that the regional grouping of sites that is used for the frequency analysis of extreme precipitation under climate change might change, we evaluate the sensitivity of the region definitions of the climate change scenarios downscaled for sites. Specifically, we use the statistically downscaled at-site data in the Chicago region and the asynchronous regional regression model (ARRM) (Stoner et al. 2013); we consider this downscaled dataset as a pilot case before evaluating the generalizability of our results in future studies. We use the fuzzy c-means clustering method to regionally group stations and preliminarily analyze the RFA results, such as the L-moments and heterogeneity measures. Our results demonstrate that the regional grouping changes when the station numbers in different regions, time periods and future scenarios are varied; this finding illustrates the need for caution when designing water-related infrastructure based on RFA without considering changes in the clustering.

2. MATERIALS AND METHODS

2.1. Study area

Our study focuses on the Chicago (Illinois, USA) area, which is composed of both rural regions and highly urbanized cities (Fig. 1). The 40 stations (Table 1) used in this study are based on the Precipitation-Frequency Atlas of the United States, published by NOAA (Bonnin et al. 2006). In the atlas, regions were subjectively delineated into 3 regions based on climate, season(s) of highest precipitation, type of precipitation (e.g. general storm, convective storm, tropical storm or hurricane, or a combination), topography and the homogeneity of these characteristics within a given geographic area. As shown in Fig. 1, 21 stations are grouped into region 53, 6 stations into region 54 and 13 stations into region 55. Although this grouping of stations is arbitrary in the atlas, the present study uses the fuzzy c-mean clustering method for the climate change scenario data, maintaining the total number of regions at 3. Given the widely variable

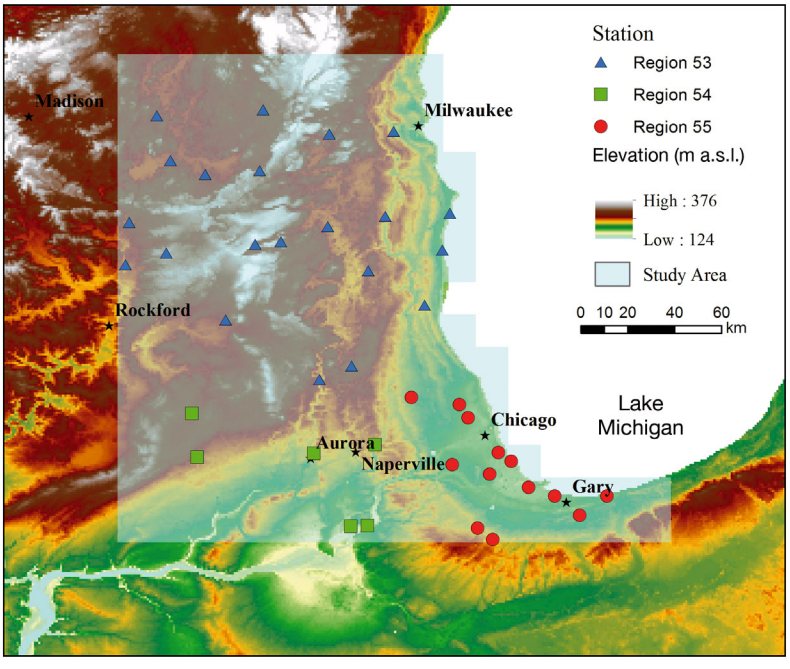

Fig. 1. Study area (light blue overlay) and stations near Chicago from Atlas 14, Precipitation-Frequency Atlas of the United States, NOAA (Bonnin et al. 2006)

Table 1. Descriptions of the precipitation stations from NOAA's Atlas 14 (Bonnin et al. 2006)

Station	Region	Longitude (°W)	Latitude (°N)	Elevation (m)	Data period (mo/yr) Start	End
11-0203	53	88.0994	42.4811	228.6	07/1901	12/2000
11-0442	53	88.1639	42.1153	266.7	01/1963	12/2000
11-2736	53	88.2861	42.0628	232.6	02/1898	12/2000
11-5326	53	88.6469	42.2928	248.4	01/1887	12/1999
11-9029	53	87.8828	42.3492	213.4	01/1923	12/2000
47-0696	53	89.0311	42.5039	237.7	01/1893	12/2000
47-1205	53	88.2544	42.6506	228.9	01/1949	12/2000
47-1667	53	88.8753	42.5492	292.6	01/1950	12/2000
47-2302	53	88.5167	42.8667	274.3	07/1942	12/1993
47-2869	53	88.8589	42.9050	243.8	09/1941	12/2000
47-3979	53	89.0167	42.6667	232.0	01/1945	12/1986
47-4174	53	87.8156	42.5608	182.9	02/1944	12/2000
47-4457	53	88.4347	42.5936	257.9	04/1945	12/2000
47-4482	53	88.9117	43.0761	259.7	01/1892	12/2000
47-6200	53	88.5036	43.1003	260.9	01/1944	12/2000
47-6922	53	87.7861	42.7022	181.4	05/1896	12/2000
47-8723	53	88.0336	42.6903	222.5	06/1941	12/2000
47-8937	53	88.2492	43.0064	253.0	01/1893	12/2000
47-9046	53	88.0017	43.0175	220.4	01/1952	11/2000
47-9190	53	88.7247	42.8508	266.7	06/1941	12/2000
47-9226	53	88.5333	42.5833	271.0	04/1921	12/1957
11-0338	54	88.3092	41.7806	201.2	01/1888	12/2000
11-2223	54	88.7756	41.9342	266.1	01/1893	12/2000
11-4530	54	88.1028	41.5033	165.5	01/1939	12/2000
11-4535	54	88.1667	41.5000	181.1	01/1894	11/1974
11-9010	54	88.7561	41.7650	240.8	01/1943	12/1996
11-9221	54	88.0728	41.8128	207.3	01/1936	12/2000
11-1542	55	87.7500	41.9667	198.1	07/1948	12/1979
11-1549	55	87.9336	41.9950	200.6	07/1940	12/2000
11-1552	55	87.6333	41.7000	200.9	07/1948	12/1979
11-1564	55	87.5500	41.7500	185.9	07/1948	12/1979
11-1567	55	87.7167	41.9167	182.9	07/1948	12/1979
11-1572	55	87.6000	41.7833	181.1	01/1926	10/1994
11-1577	55	87.7775	41.7372	189.0	01/1942	12/2000
11-2011	55	87.6222	41.4492	202.4	01/1940	10/1998
11-6616	55	87.6800	41.4933	216.4	01/1940	12/2000
12-3213	55	87.3833	41.6167	182.9	06/1936	12/1978
12-4008	55	87.2881	41.5422	195.1	07/1919	12/1999
12-6542	55	87.1833	41.6167	185.9	01/1952	12/1988
12-9570	55	87.4833	41.6500	189.0	01/1910	12/1961

lengths of the observation periods at precipitation sites (Table 1), the reliability of clustering analysis with observed precipitation data is limited.

2.2. Downscaled future at-site precipitation data

We use the downscaled precipitation data for precipitation sites in the study area. Specifically, downscaled precipitation data were used with the ARRM for 40 precipitation stations around Chicago (A. M. K. Stoner pers. comm.). Eight future cases are developed with 4 atmosphere–ocean general circulation models (AOGCMs) under 2 IPCC Special Report on Emission Scenarios (SRES) (Table 2). The details of the ARRM and the future scenarios are provided in Sections 2.2.1 and 2.2.2, respectively.

2.2.1. ARRM

Stoner et al. (2013) developed the ARRM, which builds on asynchronous quantile regression to define a quantitative relationship between daily observed and simulated surface climate variables that exhibit a symmetric distribution. The model emphasizes the accurate resolution of the relationship at the tails of the distribution. The downscaling procedure includes several steps: (1) It separates the data into 12 vectors by month such that a separate statistical model can be built for each month. (2) It fits a regression function to the ranked values. Piecewise linear regression is used because it provides consistent fitting while accounting for biases in the model values near the tails of the distribution. (3) It uses the statistical regression models that were constructed from the observed and historical simulated time series to downscale the future projections. The resulting downscaled values are rearranged in the original order to retrieve the final product, which is a continuous chronological time series of the downscaled values.

This technique performs reasonably well for capturing severe precipitation and temperature events. The model has been validated with observed precipitation and temperature data from stations across North America, and has been demonstrated to improve the simulation accuracy at the local scale, especially for extreme conditions (Stoner et al. 2013, Janssen et al. 2014). The ARRM analysis can be performed either at the station level or on a latitude–longitude grid. Thus, the ARRM-based datasets are particularly valuable for hydrologists and water resource engineers who are interested in estimating the return periods of extreme precipitation for the design of water-related infrastructure, such as dams and spillways.

Table 2. Descriptions of the 8 cases used in this study. AOGCM: atmosphere–ocean general circulation model

Case	AOGCM			Scenario
	Model	Atmospheric resolution	Reference	
ccsm.a1fi ccsm.b1	Community climate system model version 3 (CCSM 3)	1.4° × 1.4°	Collins et al. (2006)	A1fi B1
gfdl2.1.a1fi gfdl2.1.b1	GFDL climate model version 2.1 (GFDL-CM2.1)	2.0° × 2.5°	Delworth et al. (2006)	A1fi B1
hadcm3.a1fi hadcm3.b1	Hadley climate model version 3 (HadCM3)	2.5° × 3.75°	Pope et al. (2000)	A1fi B1
pcm.a1fi pcm.b1	Parallel climate model (PCM)	2.81° × 2.81°	Maltrud et al. (1998)	A1fi B1

A number of recent studies (e.g. Eden et al. 2012, Ehret et al. 2012, Maraun 2012, Wong et al. 2014) criticize bias correction methods such as ARRM. For example, Eden et al. (2012) note that bias correction methods only increase the agreement of climate model outputs with observations in hindcasts and do not necessarily reduce the uncertainty in future predictions. However, the use of downscaled data with ARRM across North America has been validated (Stoner et al. 2013, Janssen et al. 2014). We therefore argue that the quality of these data is acceptable for studying the sensitivity of regional groups in RFA under climate change scenarios, while acknowledging that it may be limited to some extent, and that other advanced downscaled approaches could be useful.

2.2.2. Climate change scenarios

The downscaling processes are applied to 8 cases from 4 global models (Table 2) under 2 future emissions scenarios, which were obtained from Stoner et al. (2013). The historical AOGCM simulations correspond to the 20th century climate in coupled models (20C3M), and the SRES higher (A1fi) and lower (B1) emission scenarios are used for the future climate scenarios. These 2 emission scenarios assume that atmospheric CO_2 concentrations reach approximately 550 ppm (B1) and 990 ppm (A1fi) by 2100. These levels are much greater than the preindustrial level of CO_2 of approximately 280 ppm and the current concentration of 400 ppm. The A1fi scenario assumes continued heavy use of fossil fuels through 2100, whereas the B1 scenario assumes a rapid movement away from fossil fuels by mid-century.

As an example, Fig. 2 shows the annual maximum daily precipitations from 2 projections (ccsm.a1fi and ccsm.b1). Although the annual maximum of the daily maximum precipitation increases in most cases, such as for ccsm.a1fi (an increase of 0.465 mm decade^{-1}; Fig. 2), cases such as ccsm.b1 (an increase of 0.265 mm decade^{-1}) and hadcm3.b1 (an increase of 0.317 mm decade^{-1}) do not exhibit clear increases. The increase in the annual precipitation ranges from 0.265 to 0.732 mm decade^{-1}, and the average increase for the 8 projections from 1960 to 2099 is 0.489 mm decade^{-1}. As expected, the annual maximum precipitation based on the A1fi scenarios exhibits clear increases compared with the precipitation based on the B1 scenarios. Using 4 models, the average regression slopes for the A1fi and B1 scenarios are 0.529 and 0.448 mm decade^{-1}, respectively.

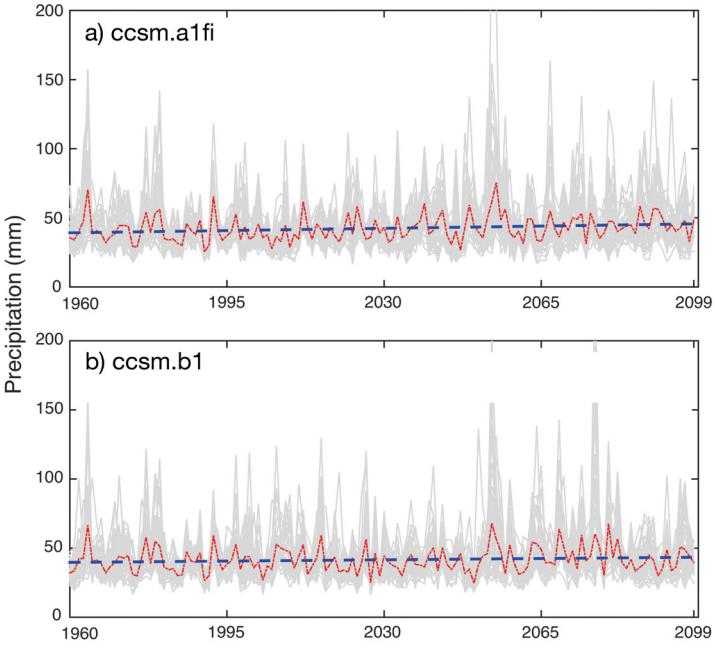

Fig. 2. Projected data from the asynchronous regional regression model (ARRM) downscaling model for scenarios (a) ccsm.a1fi and (b) ccsm.b1. Gray lines: time series of the annual daily maximum precipitation at each site; red line: region-averaged (or 40-station averaged) annual daily maximum precipitation; blue line: trend from 1960 to 2099

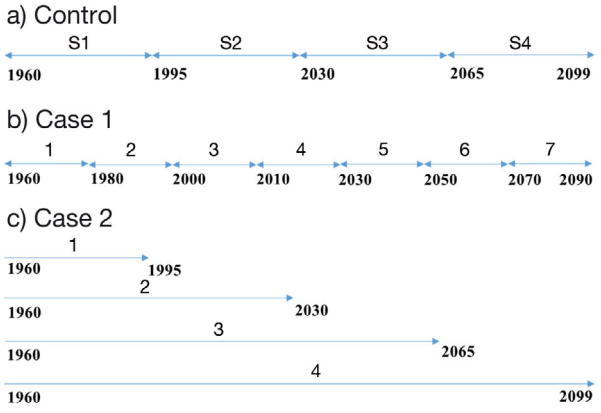

Fig. 3. Division of time periods for (a) control case, (b) Case 1 and (c) Case 2 from 1960 to 2099

To evaluate how the region definitions in RFA would change with climate change scenarios, we arbitrarily divide the entire study period of 1960–2099 into four 35 yr periods: S1 (1960–1994), S2 (1995–2029), S3 (2030–2064) and S4 (2065–2099). In addition, to perform the clustering analysis with this set of periods (Control Case), we also evaluate the effects of other sets of periods: Case 1 has shorter time periods than Control Case, with seven 20 yr periods (1960–1979, 1980–1999, etc.), and Case 2 has longer time periods than Control Case, with periods S1 (1960–1995), S2 (1960–2030), S3 (1960–2065) and S4 (1960–2099) (Fig. 3).

2.3. Fuzzy c-means for clustering

Data clustering is the process of dividing data elements into clusters such that the items of a cluster are as similar as possible and the items in different clusters are as dissimilar as possible. Depending on the nature of the data and the purpose for which the clustering is being used, different measures of similarity may be used to place items into clusters (Elena 2013). Of the many existing clustering methods, such as the k-nearest neighbor method, the Ward method, the hierarchical method and the fuzzy c-means method, fuzzy c-means clustering is used in this study because it has been successfully applied to cluster climate data (DeGaetano 2001, Sarma & Hazarika 2014). The following steps are used in fuzzy c-means clustering to determine the data clusters.

Step 1. Initialize the membership matrix with random values between 0 and 1:

$$\sum_{i}^{C} u_{ij} = 1.0 \tag{1}$$

where u is the membership matrix; C is the number of clusters; $j = 1, \ldots, C$; $i = 1, \ldots, N$; and N is the number of data.

Step 2. Calculate the fuzzy cluster center for the jth cluster:

$$v_j = \sum_{i=1}^{N} u_{ij}^{m} x_j \Big/ \sum_{i=1}^{N} u_{ij}^{m} \tag{2}$$

where $m(\in [1,\infty])$ is the fuzziness index and x is the data.

Step 3. Compute the cost function:

$$J(u,v) = \sum_{i=1}^{N}\sum_{j=1}^{C} (u_{ij})^{m} (d_{ij})^{2} \tag{3}$$

where J is the cost function and $d_{ij} = \|x_i - v_j\|$.

Step 4. Compute a new membership matrix:

$$u_{ij} = \left(\sum_{k=1}^{C} \left(d_{ij} \Big/ d_{kj} \right)^{2/(m-1)} \right)^{-1} \tag{4}$$

where $1 \le k \le C$.

Step 5. Repeat Steps 1 to 4 until the error $(\Delta = u_{ij}^{t+1} - u_{ij}^{t})$, with t as the iteration step, is acceptably small.

2.4. L-moments and heterogeneity measure for RFA

L-moment statistics are employed to evaluate the climate characteristics in the grouped regions. The sampling properties of the L-moment are nearly unbiased even in small samples and are nearly normally distributed. The sample probability-weighted moments estimated from the data ($x_{1:n}$) in increasing rth order from 1 to n are as follows:

$$\beta_0 = n^{-1} \sum_{j=1}^{n} x_j \tag{5}$$

$$\beta_r = n^{-1} \sum_{j=r+1}^{n} (j-1)(j-2)\ldots(j-r) \Big/ (n-1)(n-2)\ldots(n-3)\, x_j \tag{6}$$

where n is the number of data.

L-moments are linear combinations of the probability-weighted moments that provide estimates of the location, dispersion and shape of the data (Hosking & Wallis 1997). The L-moments are defined as follows:

$$l_1 = \beta_0 \tag{7}$$

$$l_2 = 2\beta_1 - \beta_0 \tag{8}$$

$$l_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{9}$$

$$l_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 + \beta_0 \tag{10}$$

The first L-moment (l_1, Eq. 7) is the sample mean, which is a measure of the location, and the second L-moment (Eq. 8) is a measure of the dispersion of the data values. The coefficient of variation of the L-

moment, L-CV, which has a value between 0 and 1, is defined by:

$$t = l_2 / l_1 \quad (11)$$

L-moment ratios are estimated by dividing higher-order L-moments by the dispersion measure. The L-skewness is defined as $t_3 = l_3/l_2$, and the L-kurtosis is defined as $t_4 = l_4/l_2$. These values are dimensionless quantities that range between −1 and +1.

The heterogeneity measure (H) determined from the L-moments (Hosking &Wallis 1997) is used to test whether a region of interest is homogeneous. After fitting a kappa distribution to the regional average L-moment ratios, the estimated kappa distribution is applied to generate 500 homogeneous regions. The heterogeneity measure is estimated by

$$H = (V - \mu_V)/\sigma_V \quad (12)$$

where μ_V is the mean of V, and σ_V is the standard deviation of V. For a sample and a simulated region, V is estimated as

$$V = \left\{ \sum_{i=1}^{N} n_i [t^{(i)} - t^R]^2 / \sum_{i=1}^{N} n_i \right\}^{1/2} \quad (13)$$

where N is the number of sites, n_i is the record length at site i, $t^{(i)}$ is the L-CV at site i and t^R is the regional average L-CV. Hosking & Wallis (1997) suggested that a region can be regarded as acceptably homogeneous if $H < 1$, possibly homogeneous if $1 \leq H < 2$ and definitely heterogeneous if $H \geq 2$.

3. RESULTS

3.1. Clustering of climate regions

The fuzzy c-means clustering method is used to assess the changes in the regional groups of precipitation stations with the 8 climate change cases (4 models and 2 future climate scenarios) divided into 4 periods (S1, S2, S3 and S4). In this analysis, we considered 4 climate-related factors: annual maximum monthly precipitation, annual precipitation and location (longitude and latitude). For simplicity, we assume that the number of clusters of precipitation stations is fixed at 3, which is the same number of clusters as in the NOAA Atlas 14 (Bonnin et al. 2006).

Fig. 4 shows the results of clustering the 40 stations into 3 regions for periods S1 through S4 for the ccsm.a1fi scenario. The 3 clusters are roughly divided by longitude, but such patterns generally become less clear in the future periods. A comparison of the clusters in S1 and S2 clearly demonstrates tempo-

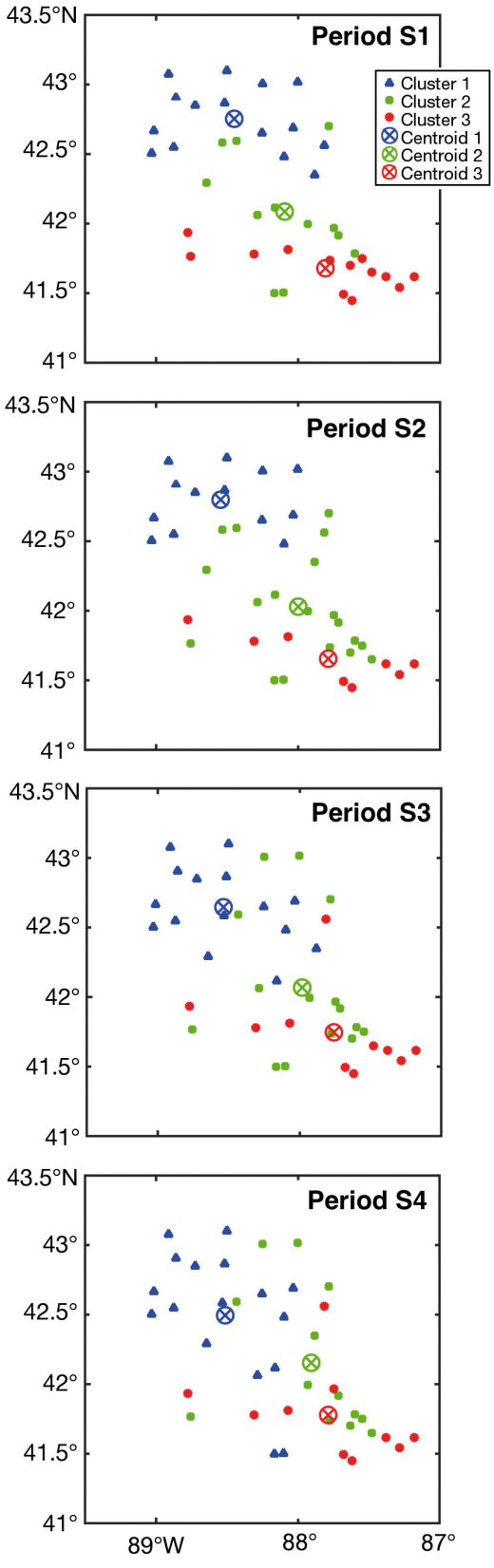

Fig. 4. Fuzzy c-means analysis distinguishing climate regions for scenario ccsm.a1fi

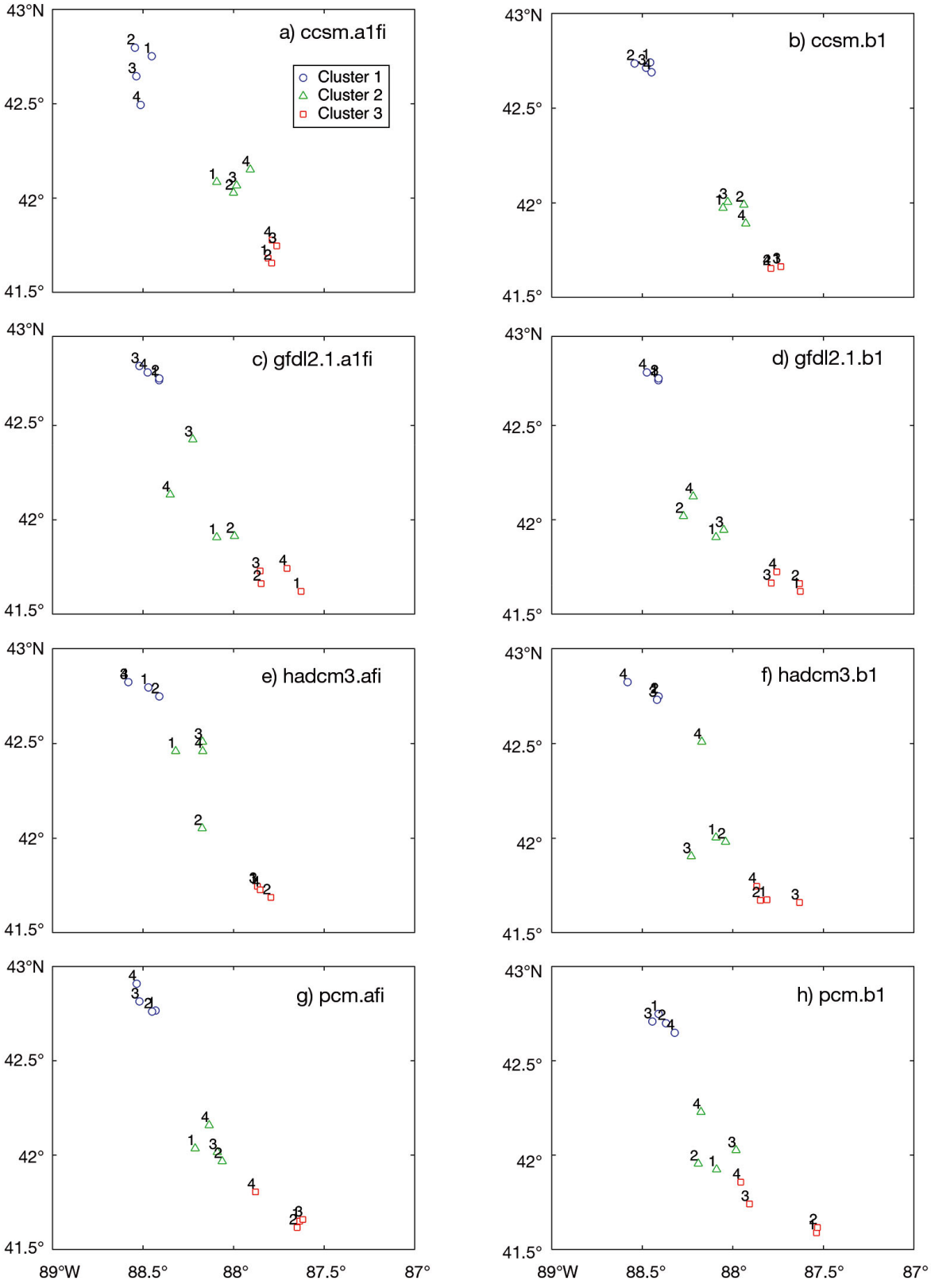

Fig. 5. Movement of cluster centroids by period for (a) ccsm.a1fi, (b) ccsm.b1, (c) gfdl2.1.a1fi, (d) gfdl2.1.b1, (e) hadcm3.a1fi, (f) hadcm3.b1, (g) pcm.a1fi and (h) pcm.b1. Blue: Cluster 1; green: Cluster 2; red: Cluster 3. Periods S1 to S4 (see Fig. 3a) are numbered

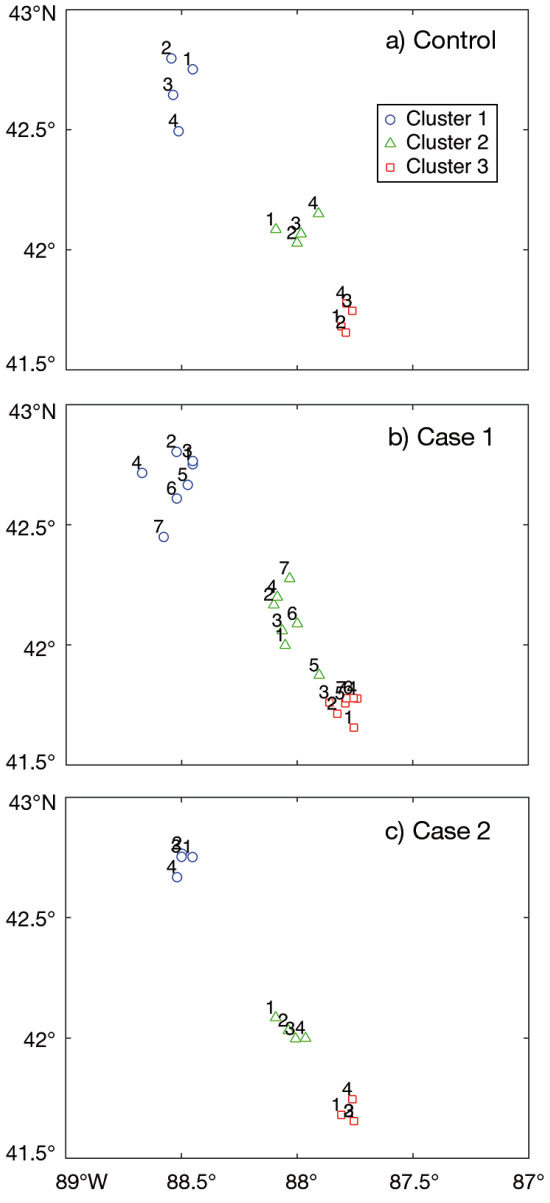

Fig. 6. Movement of cluster centroids for different sets of time periods in ccsm.a1fi for (a) Control Case, (b) Case 1 and (c) Case 2. Periods are numbered as in Fig. 3

ral changes in the clusters. The centroids of the 3 climate regions change from S1 to S4, with different magnitudes in the different scenarios (Fig. 5). Smaller temporal changes in the centroids occur in the B1 scenarios than in the A1fi scenarios. These findings suggest that one cannot assume the same climate regions for past and future precipitation data (Yang et al. 2012, 2013).

Furthermore, to understand the robustness of our findings, we perform additional clustering analyses with different sets of time periods, as described in Section 2.2. Fig. 6 shows the locations of centroids in different clusters for different time periods. Comparing Case 1 with Control Case reveals that the changes in centroid locations in Control Case are not an artifact, suggesting the potential generalizability of our findings. Comparing Control Case with Case 2 reveals that Case 2 shows a relatively small change in cluster locations because the time periods of Case 2 are overlapping (Fig. 6a).

3.2. Monthly maximum and annual precipitation

Based on the results of the clustering analysis, we investigate the averages of the monthly maximum precipitation for the different clusters, periods and scenarios. The averages of the monthly maximum precipitation do not exhibit any discernible patterns in the different clusters (Fig. 7). The averages increase to some extent from periods S1 to S4, particularly for Cluster 3 in most of the A1Fi scenarios (i.e. CCSM, HadCM3 and PCM). In addition, the temporal changes in the maximum precipitation are less in the B1 scenarios than in the A1fi scenarios.

The averages of the annual precipitation for different clusters, periods and scenarios are also examined (Figs. 8 & 9). The annual precipitation increases from S1 to S4, and the increases in the annual precipitation are most evident for Cluster 3 (Fig. 8). These patterns differ across the periods and clusters. The ccsm.a1fi case demonstrates the pattern most clearly, whereas the hadcm3.b1 case exhibits the most ambiguous pattern. Fig. 9 shows the averages of the annual precipitation for each period. The range of the average annual precipitation for each cluster increases from S1 to S4. The average annual precipitation ranges from 820 to 950 mm in all cases in period S1 and from 845 to 1142 mm in period S4.

3.3. Statistics for RFA

For the clustered regions, we evaluate the L-moments and the heterogeneity measure (H) for RFA (Hosking & Wallis 1997). The L-moments of the annual maximum daily precipitation are analyzed for 3 clusters, 4 periods and 8 scenarios. Fig. 10 shows these results for ccsm.a1fi. Whereas the mean values increase from S1 to S4 in all scenarios, no clear temporal trends in L-CV, L-skewness and L-kurtosis are presented. The variations in the L-statistics decrease from Cluster 3 to Clusters 2 and 1.

We also compared the heterogeneity of the region (i.e. H) for the climate regions based on S1 (i.e. a

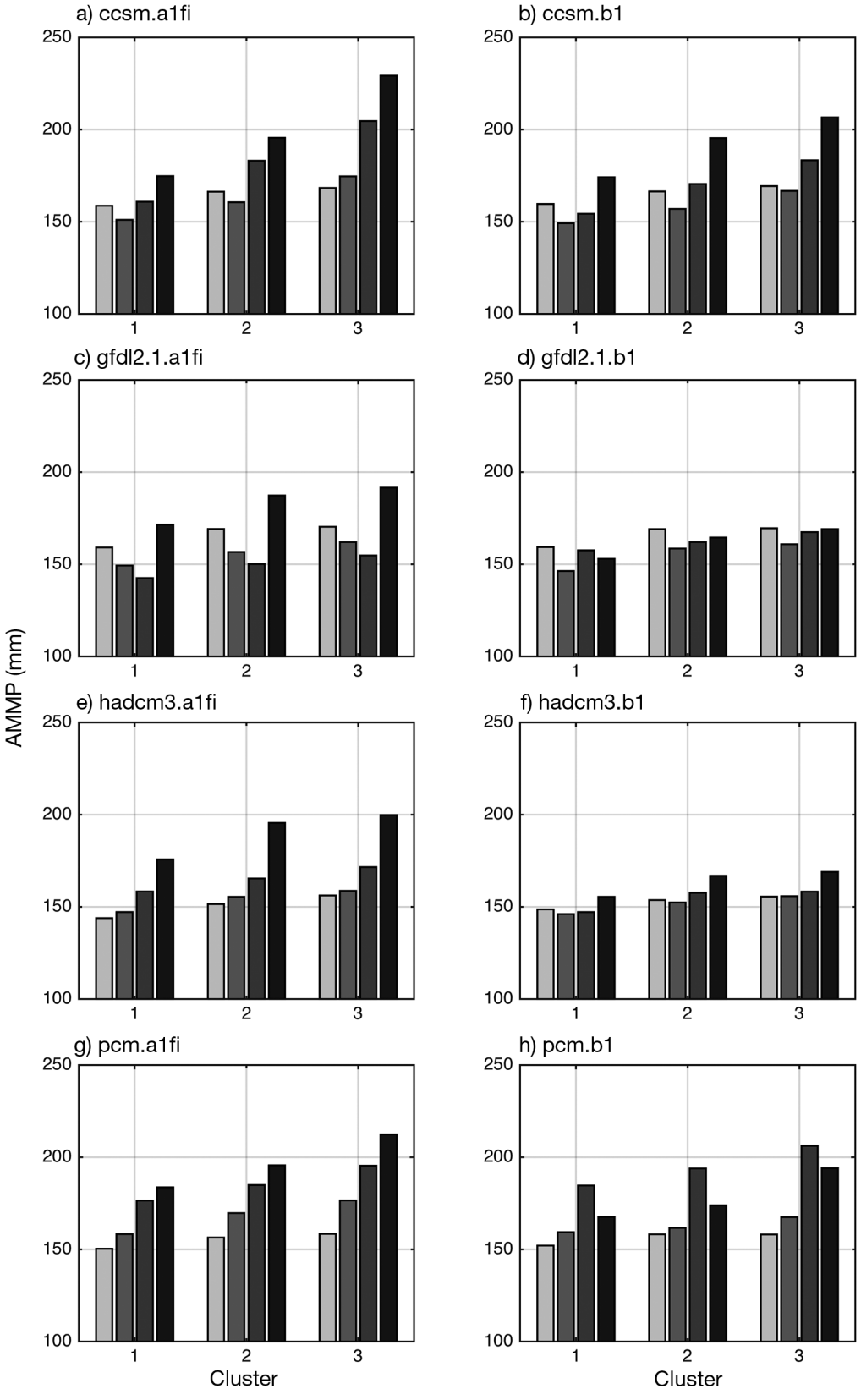

Fig. 7. Variations in the average monthly maximum precipitation (AMMP) in the 3 clusters for (a) ccsm.a1fi, (b) ccsm.b1, (c) gfdl2.1.a1fi, (d) gfdl2.1.b1, (e) hadcm3.a1fi, (f) hadcm3.b1, (g) pcm.a1fi and (h) pcm.b1. Periods S1 to S4 are indicated by light to dark colors, respectively

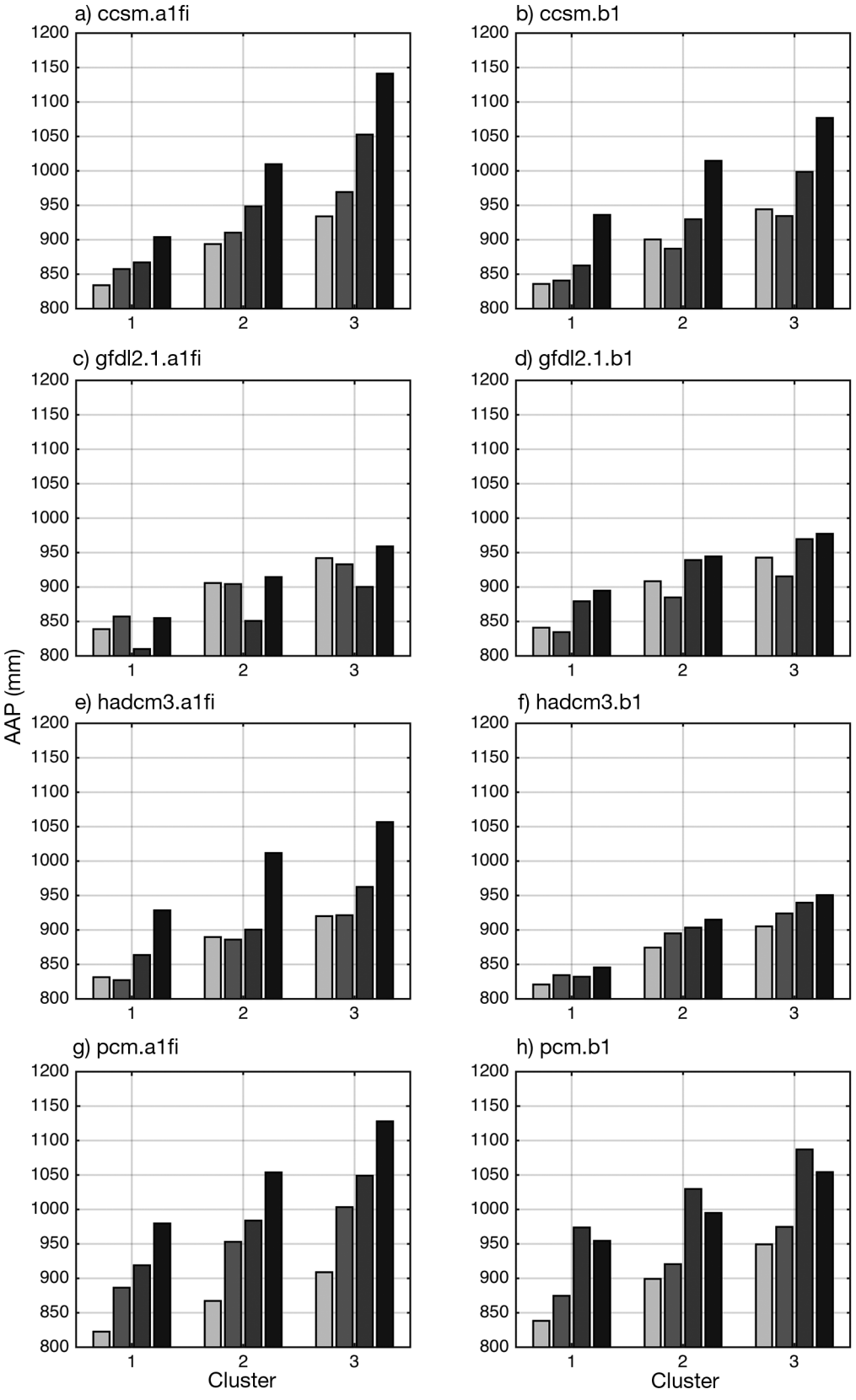

Fig. 8. Variations in the average annual precipitation (AAP) for the 3 clusters for (a) ccsm.a1fi, (b) ccsm.b1, (c) gfdl2.1.a1fi, (d) gfdl2.1.b1, (e) hadcm3.a1fi, (f) hadcm3.b1, (g) pcm.a1fi and (h) pcm.b1. Periods S1 to S4 are indicated by light to dark colors, respectively

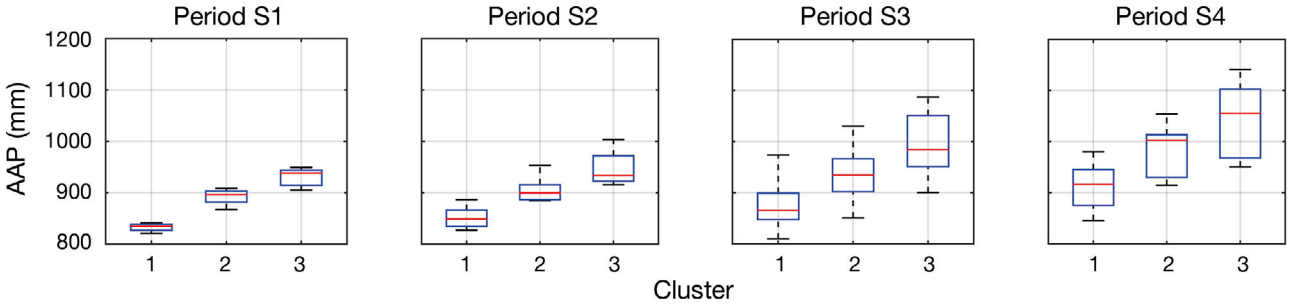

Fig. 9. Variations in the average annual precipitation (AAP) of the 8 projections by period. In each box, the central red line denotes the median, the blue edges of the box represent the 25th and 75th percentiles, and the whiskers extend to the maximum and minimum AAP

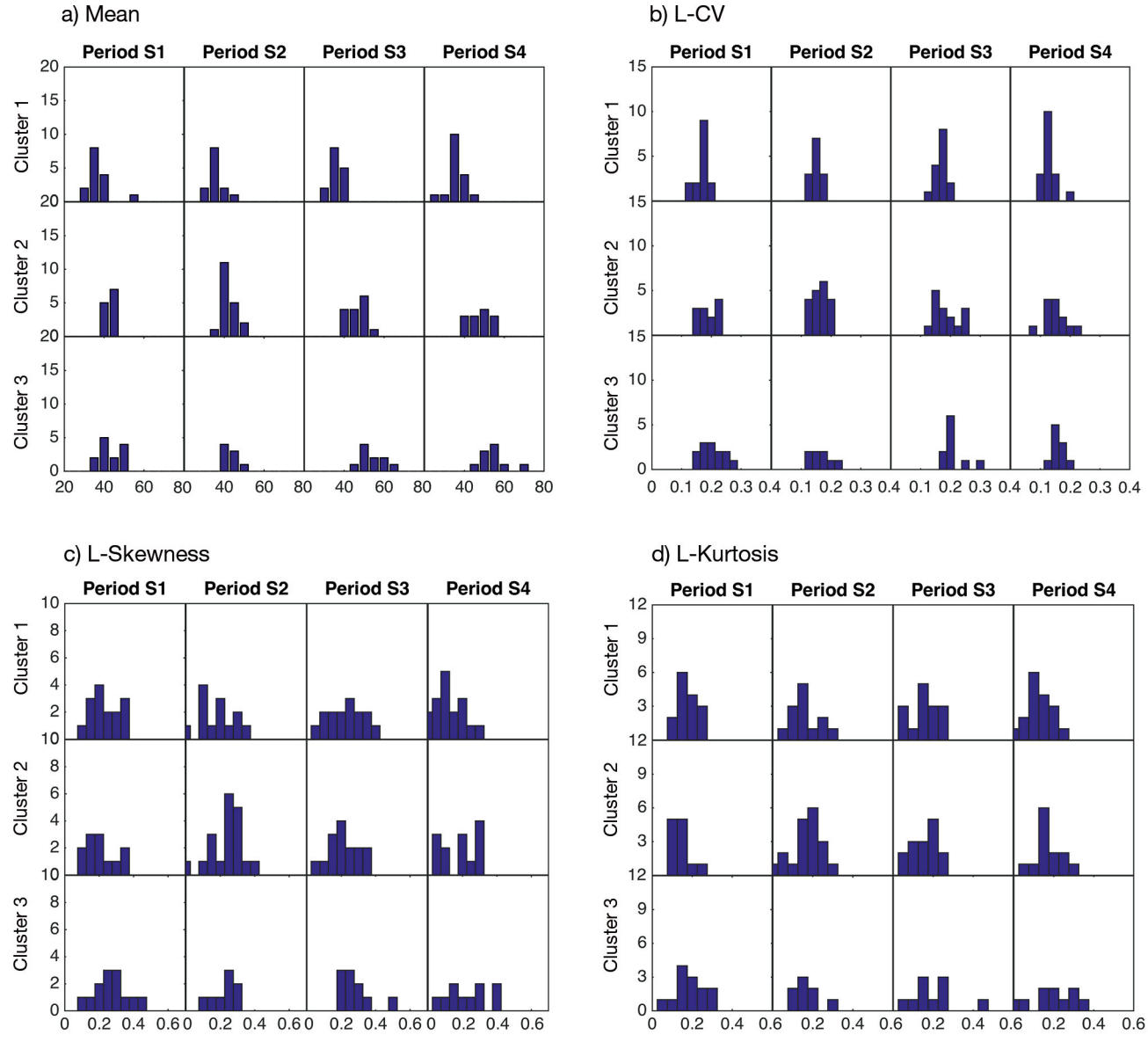

Fig. 10. Histograms of the (a) mean, (b) L-coefficient of variation, (c) L-skewness and (d) L-kurtosis at the sites for different periods and clusters in ccsm.a1fi

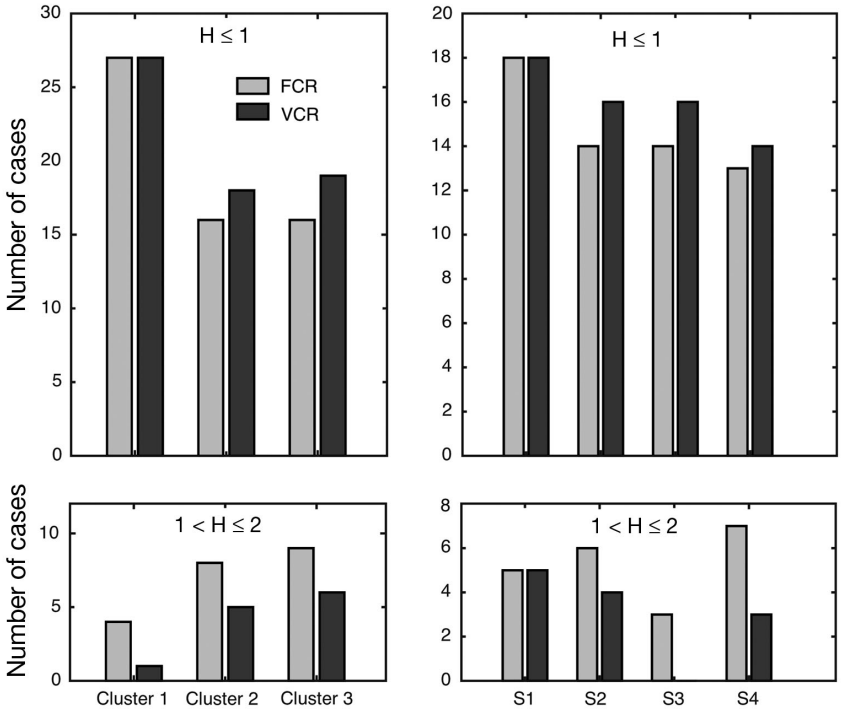

Fig. 11. Statistics of the heterogeneity measure for the 8 scenarios for the clusters (left) and periods (right). H: heterogeneity measure; FCR: fixed climate regions with the S1 cluster; VCR: varying climate regions

fixed climate region [FCR]) with those for the climate regions that vary from S1 to S4 (i.e. a varying climate region [VCR]). As can be observed from Fig. 11, the number of stations with *H* values <1 increases by 2 and 3 in Clusters 2 and 3, respectively, and remains the same for Cluster 1 when the changes in the regional clusters are considered (VCR). The number of stations in the partially heterogeneous region decreases by 3 in Clusters 1 to 3. Similar increases and decreases are found for the different periods (S2, S3 and S4). One can conclude that VCR performs better than FCR based on the number of homogeneous stations in the region.

4. DISCUSSION AND CONCLUSIONS

This study has examined whether the regional clustering of precipitation stations in the Chicago area would change in the future. Based on statistically downscaled climate scenarios yielded by the ARRM, which focuses on accurately resolving the tails of the probability distributions of climate data and thus improves simulations of extreme climates, we analyze 8 climate change cases from GCM simulations (using 4 models) based on the IPCC SRES A1fi and B1 scenarios. We determine how the regional clustering could vary over time and how such variations are related to the characteristics of the maximum precipitation (i.e. its magnitude) and the RFA statistics, such as the L-moments and heterogeneity measure.

In the 40 stations around Chicago, the clustering of the precipitation stations changed by variable amounts. The locations of the centroids of the clusters from the fuzzy c-means analysis changed more in the A1fi scenarios than in the B1 scenarios. The precipitation characteristics, such as the averages of the monthly maximum and annual precipitation and the L-moments of the annual maximum daily precipitation, varied across the time periods and regional clusters. The heterogeneity of the stations decreased when changes in the clustering of the climate regions were considered compared with when the existing clustering based on the historical simulation data was retained.

These results illustrate the need to consider changes in the regional clustering of precipitation stations in RFA. This consideration would be particularly useful for the design of water-related infrastructure in response to climate change. Although this study focused on how clustering would change and influence the statistics relevant to RFA of precipitation, the effect on the RFA performance should be studied in the future.

Our findings are limited to some extent. First, we focus on the frequency analysis of extreme precipitation, for which future projections are uncertain. However, this limitation is inherent to any frequency analysis of extreme precipitation. Second, we use the downscaled data with ARRM. As stated previously, bias correction methods are limited in that they increase the agreement of climate model outputs with observations in hindcasts but obscure rather than reduce the uncertainty for future predictions (Ehret et al. 2012). Given the validation of ARRM data across North America by Stoner et al. (2013) and Janssen et al. (2014), we consider the quality of these data acceptable for studying the sensitivity of regional groups in RFA under climate change scenarios, although some of the cluster changes could be an artifact of the ARRM. Issues that were not within the

focus of the present research but suggested for future studies include the effects of downscaling methods and clustering analysis techniques on the clusters. Finally, although the analyses here are performed with future predictions based on 8 different SRES scenarios for regions around Chicago, future studies could be performed using recent scenarios based on the IPCC representative concentration pathways for larger regions such as the contiguous USA.

Acknowledgements. This study was supported by the Korea Meteorological Administration R&D Program under Grant KMIPA 2015-6180 and also partly by the Climate Change Correspondence R&D Program of the Korea Ministry of Environment (2013001310002).

LITERATURE CITED

Basu B, Srinvas VV (2015) Analytical approach to quantile estimation in regional frequency analysis based on fuzzy framework. J Hydrol (Amst) 524:30–43

Bonnin GM, Martin D, Lin B, Parzybok T, Yekta M, Riley D (2006) NOAA Atlas 14, Precipitation-frequency atlas of the United States, Vol. 2 v. 3.0: Delaware, District of Columbia, Illinois, Indiana, Kentucky, Maryland, New Jersey, North Carolina, Ohio, Pennsylvania, South Carolina, Tennessee, Virginia, West Virginia. US Department of Commerce/NOAA/National Weather Service, Silver Spring, MD

Collins WD, Bitz CM, Blackmon ML, Bonan GB and others (2006) The community climate system model version 3 (CCSM3). J Clim 19:2122–2143

Cunderlik JM, Burn DH (2003) Non-stationary pooled flood frequency analysis. J Hydrol (Amst) 276:210–223

Cunnane C (1989) Statistical distributions for flood frequency analysis. Operational hydrology report No. 33, WMO No. 718. World Meteorological Organization, Geneva

DeGaetano AT (2001) Spatial grouping of United States climate stations using a hybrid clustering approach. Int J Climatol 21:791–807

Delworth TL, Broccoli AJ, Rosati A, Stouffer RJ and others (2006) GFDL's CM2 global coupled climate models. 1. Formulation and simulation characteristics. J Clim 19: 643–674

Eden JM, Widmann M, Grawe D, Rast S (2012) Skill, correction, and downscaling of GCM-simulated precipitation. J Clim 25:3970–3984

Ehret U, Zehe E, Wulfmeyer V, Warrach-Sagi K, Liebert J (2012) HESS opinions: Should we apply bias correction to global and regional climate model data? Hydrol Earth Syst Sci 16:3391–3404

Elena M (2013) Fuzzy c-means clustering in matlab. In: Löster T, Pavelka T (eds) Proc 7th Int Days Stat Econ, Prague, 19–21 September 2013. Melandrium, Slaný, p 905–914

Gabriele S, Chiaravalloti F (2013) Using the meteorological information for the regional rainfall frequency analysis: an application to Sicily. Water Resour Manage 27: 1721–1735

Hosking JRM, Wallis JR (1997) Regional frequency analysis. An approach based on L-moment. Cambridge University Press, Cambridge

Janssen E, Wuebbles DJ, Kunkel KE, Olsen SC, Goodman A (2014) Observational- and model-based trends and projections of extreme precipitation over the contiguous United States. Earth's Future 2:99–113

Leclerc M, Ouarda TBMJ (2007) Non-stationary regional flood frequency analysis at ungauged sites. J Hydrol (Amst) 343:254–265

Maltrud ME, Smith RD, Semtner AJ, Malone RC (1998) Global eddy-resolving ocean simulations driven by 1985–1995 atmospheric winds. J Geophys Res 103: 30825–30853

Maraun D (2012) Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. Geophys Res Lett 39:L06706, doi: 10.1029/2012GL051210

Maraun D, Wetterhall F, Ireson AM, Chandler RE and others (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev Geophys 48:RG3003, doi:10.1029/2009RG000314

Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3. Clim Dyn 16:123–146

Santos JF, Portela MM, Pulido-Calvo I (2011) Regional frequency analysis of droughts in Portugal. Water Resour Manage 25:3537–3558

Sarma AK, Hazarika J (2014) GCM based fuzzy clustering to identify homogeneous climatic regions of north-east India. Int J Env Chem Ecol Geol Geophys Eng 8:727–734

Schaefer MG (1990) Regional analyses of precipitation annual maxima in Washington state. Water Resour Res 26:119–131, doi:10.1029/WR026i001p00119

Stoner AMK, Hayhoe K, Yang X, Wuebbles DJ (2013) An asynchronous regional regression model for statistical downscaling of daily climate variables. Int J Climatol 33: 2473–2494

Wong G, Maraun D, Vrac M, Widmann M, Eden JM, Kent T (2014) Stochastic model output statistics for bias correcting and downscaling precipitation including extremes. J Clim 27:6940–6959

Yang P, Hou W, Feng GL (2012) The characteristics of clusters of weather and extreme climate events in China during the past 50 years. Chin Phys B 21:019201

Yang P, Xiao Z, Yang J, Liu H (2013) Characteristics of clustering extreme drought events in China during 1961–2010. Acta Meteorol Sin 27:186–198

Editorial responsibility: Gouyu Ren, Beijing, China

Submitted: April 13, 2015; Accepted: December 22, 2015
Proofs received from author(s): February 15, 2016