AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2017JD027018

Key Points:

- Clustering analysis of the oscillated anomaly correlation for CFSv2 forecasts of summer precipitation in China
- Illustration of two robust patterns regarding the relationship between anomaly correlation and forecast lead time
- Revelation of the grids where CFSv2 forecasts exhibit strong correlations with observations at a long lead time

Correspondence to:
P. Liu,
liupan@whu.edu.cn

Citation:
Zhao, T., P. Liu, Y. Zhang, and C. Ruan (2017), Relating anomaly correlation to lead time: Clustering analysis of CFSv2 forecasts of summer precipitation in China, *J. Geophys. Res. Atmos.*, *122*, doi:10.1002/2017JD027018.

Received 24 APR 2017
Accepted 18 AUG 2017
Accepted article online 24 AUG 2017

Relating anomaly correlation to lead time: Clustering analysis of CFSv2 forecasts of summer precipitation in China

Tongtiegang Zhao[1,2], Pan Liu[1], Yongyong Zhang[3], and Chengqing Ruan[4]

[1]State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China, [2]Department of Infrastructure Engineering, University of Melbourne, Melbourne, Victoria, Australia, [3]Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, [4]North China Sea Marine Forecasting Center, State Oceanic Administration of China, Qingdao, China

Abstract Global climate model (GCM) forecasts are an integral part of long-range hydroclimatic forecasting. We propose to use clustering to explore anomaly correlation, which indicates the performance of raw GCM forecasts, in the three-dimensional space of latitude, longitude, and initialization time. Focusing on a certain period of the year, correlations for forecasts initialized at different preceding periods form a vector. The vectors of anomaly correlation across different GCM grid cells are clustered to reveal how GCM forecasts perform as time progresses. Through the case study of Climate Forecast System Version 2 (CFSv2) forecasts of summer precipitation in China, we observe that the correlation at a certain cell oscillates with lead time and can become negative. The use of clustering reveals two meaningful patterns that characterize the relationship between anomaly correlation and lead time. For some grid cells in Central and Southwest China, CFSv2 forecasts exhibit positive correlations with observations and they tend to improve as time progresses. This result suggests that CFSv2 forecasts tend to capture the summer precipitation induced by the East Asian monsoon and the South Asian monsoon. It also indicates that CFSv2 forecasts can potentially be applied to improving hydrological forecasts in these regions. For some other cells, the correlations are generally close to zero at different lead times. This outcome implies that CFSv2 forecasts still have plenty of room for further improvement. The robustness of the patterns has been tested using both hierarchical clustering and k-means clustering and examined with the Silhouette score.

1. Introduction

Coupled ocean-atmosphere global climate models (GCMs) have been steadily improved over the past years due to accumulations of scientific and technological advances, such as supercomputing and global data observation/assimilation systems [*Barnston et al.*, 2012; *DelSole et al.*, 2014; *Bauer et al.*, 2015; *Li et al.*, 2015; *Infanti and Kirtman*, 2016; *Tian et al.*, 2016]. One key strength of GCMs is that they formulate the interactions among the atmosphere-ocean-land processes and therefore take advantage of the slowly varying components, e.g., sea surface temperature and soil moisture, to predict global climate at subseasonal, seasonal, and even interannual time scales [*Koster et al.*, 2010; *Yuan et al.*, 2011; *Kirtman et al.*, 2014]. Nowadays, GCMs have been developed and adopted by major climate agencies around the world to provide climate outlooks, for example, the Climate Forecast System Version 2 (CFSv2) at the U.S. National Centers for Environmental Prediction [*Saha et al.*, 2014] and the European Centre for Medium-Range Weather Forecasts's System 4 model [*Molteni et al.*, 2011]. In the North American Multimodel Ensemble (NMME) project, the operational predictive capabilities and the strengths/weaknesses of more than 10 GCMs have been investigated [*Kirtman et al.*, 2014].

GCMs provide forecast information at a long lead time, and their forecasts have a great potential to improve environmental management [*Leung and Qian*, 2005; *Maurer and Lettenmaier*, 2004; *Cloke and Pappenberger*, 2009]. The applications include flood warning [e.g., *Alfieri et al.*, 2013; *Siegmund et al.*, 2015], drought preparation and recovery [e.g., *Pan et al.*, 2013; *Sheffield et al.*, 2014], and agricultural planning [e.g., *Ines and Hansen*, 2006]. However, raw GCM forecasts are largely unusable, though they contain an ensemble of scenarios regarding future climate. This is because raw ensemble mean is, in general, biased; raw ensemble spread is typically overconfident; and further raw ensemble forecasts are usually not as skillful as reference climatology forecasts [*Gneiting et al.*, 2005; *Wilks and Hamill*, 2007; *Zhao et al.*, 2017]. As a result, postprocessing is a necessary step before GCM forecasts can be readily used. Various postprocessing methods are available. They range from simple regression and ensemble dressing models [*Wilks and Hamill*, 2007] to more complicated nonhomogeneous Gaussian regression [*Gneiting et al*., 2005; *Thorarinsdottir and Gneiting*, 2010] and Bayesian joint probability models [*Robertson et al*., 2013; *Shrestha et al*., 2015; *Schepen et al*., 2016].

One common characteristic of postprocessing methods is that they explicitly account for "how well raw ensemble forecasts are correlated with observations" [*Gneiting et al*., 2005; *Wilks and Hamill*, 2007; *Zhao et al*., 2017]. The spatial-temporal GCM forecasts have overall five dimensions, namely, latitude, longitude, initialization time, lead time, and ensemble members [*Molteni et al*., 2011; *Saha et al*., 2014; *Kirtman et al*., 2014]. To characterize the performance of raw GCM forecasts, the simplest and most popular measure is probably the anomaly correlation between the mean values of raw ensemble forecasts and the corresponding observations [e.g., *Yuan et al*., 2011; *Luo et al*., 2013; *Liu et al*., 2014; *Ma et al*., 2016; *Infanti and Kirtman*, 2016; *Dirmeyer and Halder*, 2017]. Although to take the ensemble mean help to eliminate the fifth dimension, the anomaly correlation still has up to four dimensions. The dimensionality complicates the analysis of GCM forecasts. In forecast evaluation, people tend to additionally fix one or two dimensions and further simplify the problem. For example, in analyzing the CFSv2 forecasts, *Regonda et al*. [2016] investigated the forecasts at 1 month lead time considering forecasts at longer lead times would be less skillful; *Lang et al*. [2014] and *Ma et al*. [2016] paid attention to case study river basins, instead of GCM grid cells, and evaluated the forecast skill of spatial averaging precipitation.

This paper aims to derive the patterns that relate the anomaly correlation of GCM forecasts to lead time in the three-dimensional space of latitude, longitude, and forecast initialization time. Focusing on a certain period of the year and a certain grid cell, the correlations for forecasts initialized at different preceding periods form a vector. We propose to link the vectors and pool the correlations across different cells using clustering. In this way, the performance of GCM forecasts, as is indicated by anomaly correlation, across many cells can be illustrated. While clustering is a popular method in the data mining area, there are few studies using clustering to explore the anomaly correlation for GCM forecasts. This paper provides a novel application of clustering. As will be demonstrated through the case study of CFSv2 forecasts of summer precipitation in China, the clustering analysis efficiently reveals the relationship between anomaly correlation and lead time for all the CFSv2 grid cells across China. One remarkable outcome is that it tells where the forecasts exhibit a positive correlation with observations even at a long lead time and where the correlation is close to zero across different lead times.

The remainder of the paper is organized as follows. Section 2 introduces the data sets of forecasts and observations. Section 3 describes the methods, including the clustering algorithm and the selection of clusters. Section 4 presents the results and illustrates the spatial and temporal patterns of anomaly correlation. Section 5 discusses the results, and section 6 concludes the paper.

2. Data

This paper investigates monthly precipitation forecasts from CFSv2, the current operational forecasting model at the National Centers for Environmental Prediction (NCEP). CFSv2 was implemented in 2011. It is built upon CFSv1, the operational model at the NCEP between 2004 and 2011, by a number of new packages for atmosphere-ocean-land processes and also a new data assimilation system [*Jiang et al*., 2013; *Saha et al*., 2014; *Dirmeyer and Halder*, 2017]. CFSv2 is known for its promising performance in the NMME project that have compared multiple GCMs [*Kirtman et al*., 2014]. The global hindcast data of CFSv2 are downloaded from the data library of the International Research Institute for Climate and Society (https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.NCEP-CFSv2/.HINDCAST/). The forecasts are at a spatial resolution of 1.0° × 1.0°. They cover the period from 1982 to 2010 and are initialized in each month. It provides monthly ensemble precipitation forecasts for the next 10 months, each ensemble containing 24 members.

The observed precipitation is obtained from the Global Precipitation Climatology Centre (GPCC, https://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html). The GPCC data set is established on the basis of quality-controlled observation data from more than 85,000 precipitation gauges, which are selected from between 150,000 and 250,000 stations worldwide [*Schneider et al*., 2014]. GPCC overcomes the data format issues and provides integrated gridded global monthly precipitation at three spatial resolutions, which are 0.5° × 0.5°, 1.0° × 1.0°, and 2.5° × 2.5°. This study uses the 1.0° × 1.0° full V7 data. This data set lasts from 1901 to 2013 and covers the whole period of CFSv2 hindcast.

It is necessary to note that the grid latitude and longitude of CFSv2 and GPCC data sets vary by 0.5°. Specifically, the centers of CFSv2 grid cells are at (longitude, latitude) (longitude = 0, 1, …, 359; latitude = −90, 89, …, 89, 90), while the centers of GPCC grid cells are at (longitude, latitude) (longitude = 0.5, 1.5, …, 359.5; latitude = −89.5, −88.5, …, 88.5, 89.5). To facilitate the cell-by-cell correlation analysis, the GPCC data set is regridded to match the CFSv2 data set using bilinear interpolation [*Alfieri et al.*, 2013; *Jiang et al.*, 2013; *Tian et al.*, 2016]. In this paper, the analysis is concentrated on the forecasts of summer (June, July, and August) precipitation in China. In summer, heavy precipitation caused by monsoon leads to widespread flood inundation [*Ding and Chan*, 2005]. The analysis of the performance of CFSv2 precipitation forecasts is expected to provide some information for seasonal flood forecasting [e.g., *Luo et al.*, 2013; *Lang et al.*, 2014; *Ma et al.*, 2016]. We investigate the seasonal forecasts initialized in January, February, March, April, May, and June. These forecasts are respectively at lead times of 5, 4, 3, 2, 1, and 0 months.

3. Methods

There are two steps in data processing and analysis. First, the cell-by-cell anomaly correlation is derived for the CFSv2 forecasts of summer precipitation in China. The correlation vector for forecasts initialized in January, February, March, April, May, and June depicts the performance of forecasts with lead time. Then, the clustering is employed to group the vectors and deal with the oscillation of anomaly correlation as time progresses from January to June. The Silhouette score is used to elicit the clusters that best characterize the relationship between anomaly correlation and lead time.

3.1. Spearman's Rank Correlation

The ensemble mean of GCM forecasts can exhibit nonlinear relationships with observations [e.g., *Yuan et al.*, 2011; *Luo et al.*, 2013; *Liu et al.*, 2014]. Considering this, the Spearman's rank correlation, instead of the Pearson's linear correlation, is used in this study. On the other hand, it is noted that the Spearman's correlation treats the variables as ranked and is otherwise similar to the Pearson's correlation. Let's denote y_i as the summer precipitation at the grid cell under investigation in year k (k = 1982, 1983, …, 2010) and $x_{m,k}$ as the corresponding ensemble mean of forecasts made in month m (m = 1, 2, …, 6). The Pearson's correlation ρ_m between x_m and y is calculated as follows:

$$\rho_m = \frac{\sum_k \left(x_{m,k} - \bar{x}_m\right)\left(y_k - \bar{y}\right)}{\sqrt{\sum_k \left(x_{m,k} - \bar{x}_m\right)^2}\sqrt{\sum_k \left(y_k - \bar{y}\right)^2}} \tag{1}$$

As shown in equation (1), ρ_m measures how the anomaly of $x_{m,k}$—the departure from its mean—is linearly correlated with the anomaly of y_k. In calculating the Spearman's correlation r_m, the formula is similar to equation (1), but the rank of data values is used:

$$r_m = \frac{\sum_k \left(rx_{m,k} - \overline{rx}_m\right)\left(ry_k - \overline{ry}\right)}{\sqrt{\sum_k \left(rx_{m,k} - \overline{rx}_m\right)^2}\sqrt{\sum_k \left(ry_k - \overline{ry}\right)^2}} \tag{2}$$

In equation (2), the ranks of $x_{m,k}$ and y_k, i.e., $rx_{m,k}$ and ry_k, are used. In this way, r_m is able to capture nonlinear relationships between $x_{m,k}$ and y_k.

The Spearman's correlation is calculated for forecasts initialized in the months from January to June. These six correlations form a vector:

$$R = [r_1 \;\; r_2 \;\; r_3 \;\; r_4 \;\; r_5 \;\; r_6] \tag{3}$$

In equation (3), R tells how the correlation between raw ensemble mean and observations changes as time progresses from January to June.

3.2. Agglomerative Hierarchical Clustering

In exploratory analysis of the anomaly correlation, we find that in most cases it does not increase as time progresses but oscillates instead. In other words, for a certain grid cell, the forecasts can exhibit a positive correlation with observations in May but a negative correlation in June. The noisy oscillations are attributable to the chaotic nature of the climate system and relate to the issues of GCM setting, grid resolution, and ensemble size [*Barnston et al*., 2012; *DelSole et al*., 2014; *Saha et al*., 2014]. In the field of data mining, the clustering has been shown to be a robust method that filters noise and exploits useful information [*Cheng and Wallace*, 1993; *Xu and Wunsch*, 2005; *Rau et al*., 2017]. In this study, we use the agglomerative hierarchical clustering to investigate the correlation vectors across CFSv2 grid cells. For hierarchical clustering, there are, in general, three steps [*Xu and Wunsch*, 2005; *Zhang et al*., 2015]. First, each vector is treated as an independent cluster. Then, small clusters are gradually merged into large ones. And finally, a hierarchy of clusters is created. At the top of the hierarchy is one single cluster comprised by all the vectors; at the bottom are the smallest clusters, each containing one individual vectors; and the hierarchy illustrates how the clusters are organized.

To facilitate clustering, one basic issue is the distance metric. It not only measures the similarity among the vectors but also determines merging vectors into clusters. The popular Euclidean distance is used in this study. That is, for cell *i* and cell *j*, the distance between R_i and R_j is calculated as follows:

$$d(i,j) = \sqrt{\left|R_i - R_j\right|^2} = \sqrt{\sum\nolimits_{m=1}^{6} \left(r_{m,i} - r_{m,j}\right)^2} \tag{4}$$

Based on equation (4), the total within-cluster variance, which tells how similar the vectors within a cluster are to each other, can be calculated:

$$\mathrm{var}(d(i,j)) \quad (i,j\in C) \tag{5}$$

In equation (5), var() is the operator of the variance of pair-wise distances among all the vectors in the cluster *C*. To create a hierarchy, small clusters are merged using the Ward's method [*Xu and Wunsch*, 2005]. Specifically, two clusters are merged into a new cluster if the merging leads to the minimum increase in total within-cluster variance. From the perspective of distance, it means that the vectors would be the most similar within the new cluster.

3.3. Silhouette Score

After obtaining the hierarchy of clusters, the next task is to pick out the clusters that most efficiently represent all the vectors. The clusters at the bottom of the hierarchy represent individual correlation vectors; they are step by step merged until finally becoming one cluster at the top of the hierarchy. To elicit the best clusters, we use the classical Silhouette score [*Rousseeuw*, 1987]:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{6}$$

In equation (6), $a(i)$ is the average distance of vector *i* in cluster *C* to the other vectors in *C*, whereas $b(i)$ represents the shortest average distance of *i* to the clusters other than *C*. When *i* is very similar to the other vectors in *C* and quite distinct from vectors in other clusters, we have $a(i) \ll b(i)$. In this case, $s(i)$ is close to 1. On the other hand, when *i* is neither similar to the other vectors in *C* nor distinct from vectors in other clusters, $s(i)$ would be close to 0 since $a(i) \approx b(i)$. In an extreme case where *i* is distinct from the other vectors in *C* but similar to vectors in other clusters, $s(i)$ would approach -1 as $a(i) \gg b(i)$. By pooling the Silhouette score for individual vectors, the average Silhouette score indicates whether similar vectors have been merged into one cluster.

In the hierarchy, the number of selected clusters is gradually increased from 2 to 10 and the Silhouette scores are recorded. The optimal number of clusters is the one that corresponds to a peak average Silhouette score; i.e., the corresponding clusters best characterize the anomaly correlation vectors [*Rousseeuw*, 1987; *Xu and Wunsch*, 2005; *Rau et al*., 2017].

4. Results

The anomaly correlation between CFSv2 forecasts and summer precipitation in China is calculated. The six correlations, which are for forecasts initialized in January, February, March, April, May, and June, form a vector; the vectors across different CFSv2 grid cells are grouped into clusters through hierarchical clustering. The clusters show where CFSv2 forecasts perform well and where the forecasts are not skillful.

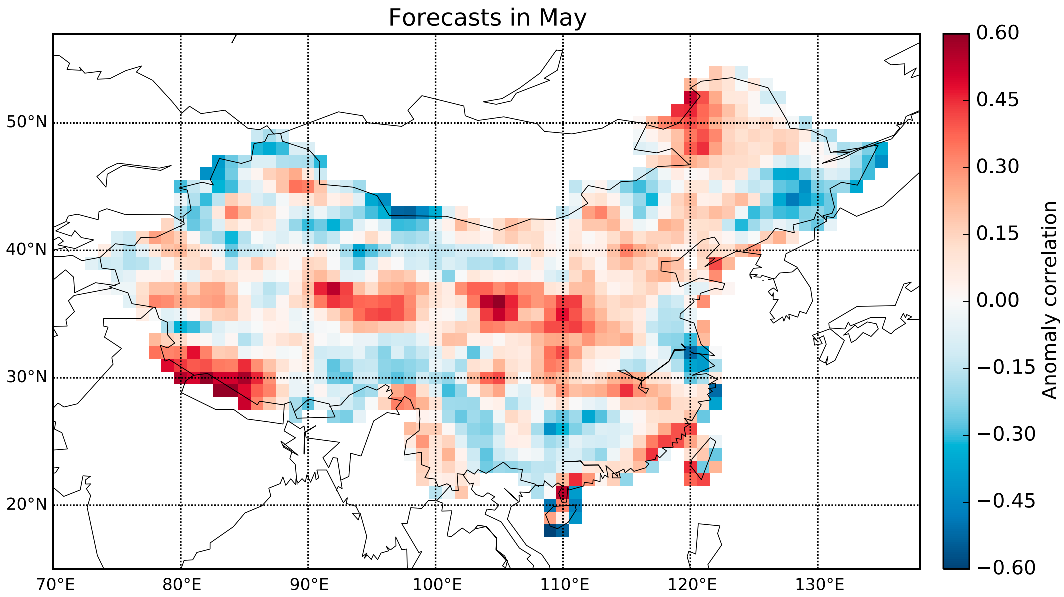

Figure 1. Spatial distribution of anomaly correlation between CFSv2 forecasts initialized in May and GPCC precipitation.

4.1. Spatial and Cumulative Distributions of Anomaly Correlation

For the purpose of data exploration, the anomaly correlation for CFSv2 forecasts in May and June are illustrated in Figures 1 and 2, respectively. In the plots, each CFSv2 grid cell is represented by a square. The color of the squares indicates the value of correlation: redder colors indicate higher positive correlations, while bluer colors represent lower negative correlations. Both figures show that CFSv2 forecasts exhibit positive correlations with observations in many cases. In the meantime, there also exist negative correlations. Comparing Figure 1 to Figure 2, it can be observed that the spatial distributions of anomaly correlations are different in May and June. For a certain cell, a high anomaly correlation in May, in general, does not correspond to a similarly high or even higher correlation in June. The indication is that measured by anomaly correlation, the forecast skill of summer precipitation does not steadily improve as time progresses. Or in other words, CFSv2 forecasts at a short lead time are not necessarily more skillful than CFSv2 forecasts at a

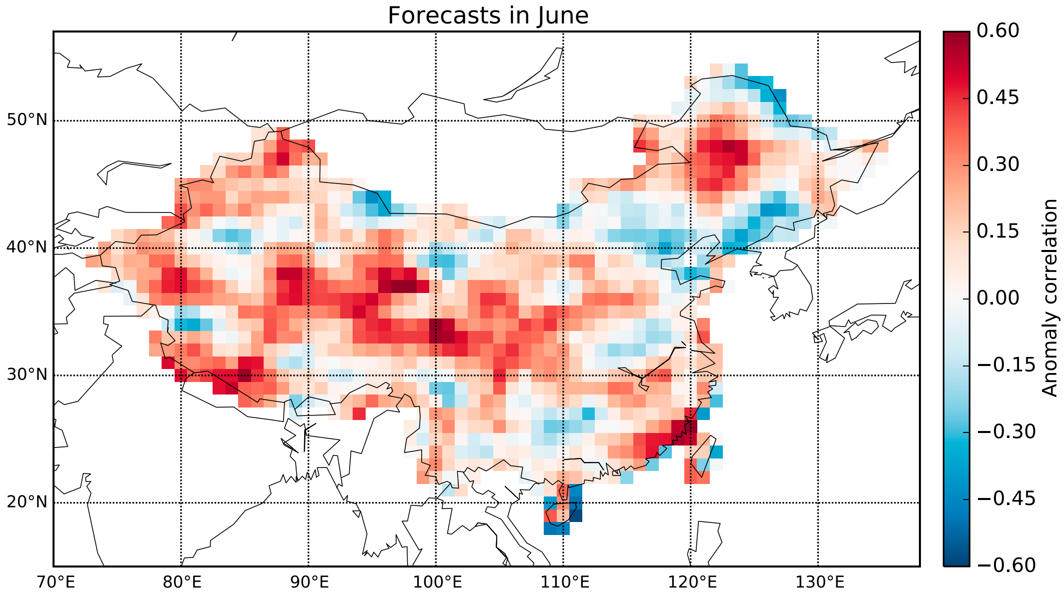

Figure 2. As for Figure 1 but for CFSv2 forecasts initialized in June.

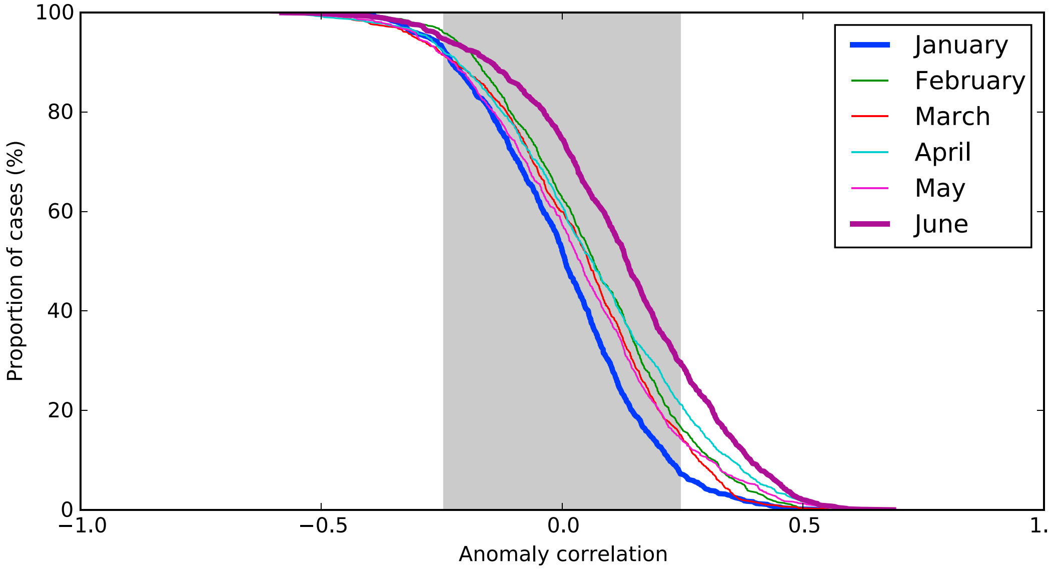

Figure 3. Cumulative distribution of anomaly correlation for CFSv2 forecasts initialized in the months from January to June; the shaded region represents the interval (−0.245, 0.245) within which the anomaly correlation is nonsignificant and with *P* value larger than 20%.

long lead time. This characteristic of precipitation forecasts is quite different from that of streamflow forecasts, for which a shorter lead time in most cases corresponds to a higher forecast skill [*Alfieri et al.*, 2013; *Bennett et al.*, 2016; *Zhao et al.*, 2016a].

The cumulative distribution of anomaly correlation across the grid cells is presented in Figure 3. While the cell-based analysis does not show patterns regarding the relationship between anomaly correlation and lead time (Figures 1 and 2), the results across the cells collectively illustrate that the anomaly correlation tends to improve as time progresses from January to June. In Figure 3, the *x* axis indicates the value of correlation from negative to positive and the *y* axis illustrates the proportion of grid cells. The shaded region represents the interval (−0.245, 0.245), within which the anomaly correlation is not significant. The distribution at different months are marked by different colors. As is shown by the thick purple line, the distribution for forecasts in June is at the top. Since anomaly correlation is a positively oriented metric, this result means that for a certain correlation value, the proportion of cells, at which the anomaly correlation between forecasts and observations is equal to or higher than that value, is the highest in June. That is, forecasts in June are overall the most skillful. They exhibit significant anomaly correlation with observations in nearly 30% of the cells. By contrast, the cumulative distribution of correlation in January is at the bottom (the thick blue line). This outcome suggests that forecasts in January tend to be the least skillful.

4.2. Silhouette Score-Based Hierarchical Clustering

In the previous section, the comparison between Figure 3 and Figures 1 and 2 suggests that some patterns of the anomaly correlation can be observed by investigating the cells collectively. In this section, the hierarchical clustering pools the correlation vectors across the cells in the analysis and explores the relationship between anomaly correlation and lead time.

The average Silhouette score is employed to determine the number of clusters under which the clustering is the most efficient. With the number of clusters increasing from 2 to 10, the average Silhouette score is presented in Figure 4 (left). It can be observed that this score exhibits a peak value when the number of clusters is 2. According to its definition (equation (6)), the peak value means that when the anomaly correlation vectors are partitioned into two clusters, anomaly correlation vectors tend to be the most similar within one cluster and the most different from the other cluster. For the case of two clusters, we denote the clusters as A and B. For individual anomaly correlation vectors in A and B, we rank their Silhouette scores from lowest to highest and illustrate them in Figure 4 (right). It can be observed that the Silhouette scores for individual vectors in the two clusters are mostly positive. There are only a few, less than 10%, anomaly correlation

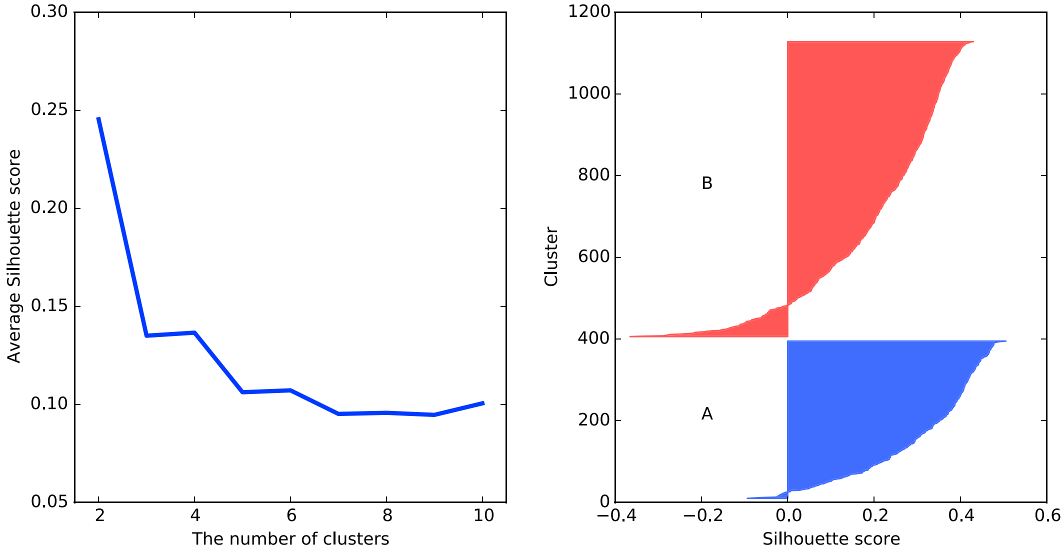

Figure 4. (left) The average Silhouette score as the number of clusters increases from 2 to 10 and (right) the ranked silhouette score for individual anomaly correlation vectors when the number of clusters is 2.

vectors with negative Silhouette scores. This result confirms the peak average Sihouette score at Figure 4 (left). Meanwhile, it is pointed out that the maximum Silhouette score is around 0.5 as opposed to 1.0 (the theoretical maximum value). This is mainly because of the oscillation of anomaly correlation with lead time (Figure 5), which makes it difficult to obtain distinct clusters to classify the correlation vectors. In this study, in addition to the hierarchical clustering, the k-means clustering is later on applied to examine the robustness of the results.

For clusters A and B, the corresponding anomaly vectors are illustrated in Figure 5. It can be observed that the two clusters represent two distinct patterns. In cluster A, the anomaly correlations at different lead times are overall positive. In addition, they tend to increase as time progresses from January to June. This result suggests that the CFSv2 forecasts corresponding to the vectors in cluster A tend to be skillful in capturing the observed summer precipitation and that they gradually improve as time progresses. On the other hand,

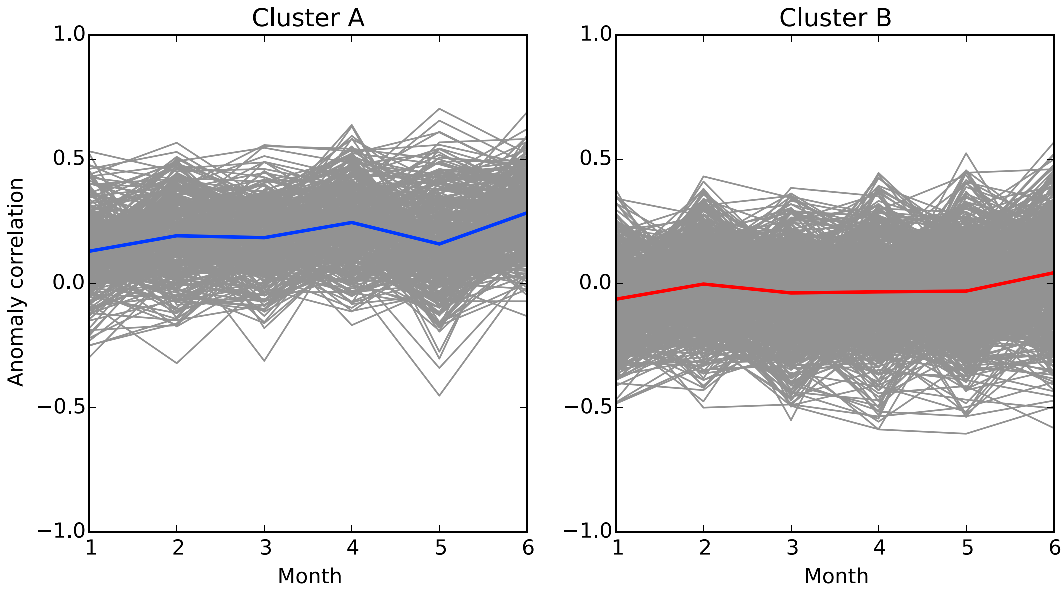

Figure 5. Clustering of anomaly correlation vectors when the number of clusters is 2 (the blue and red lines indicate the average anomaly correlation for clusters A and B, respectively).

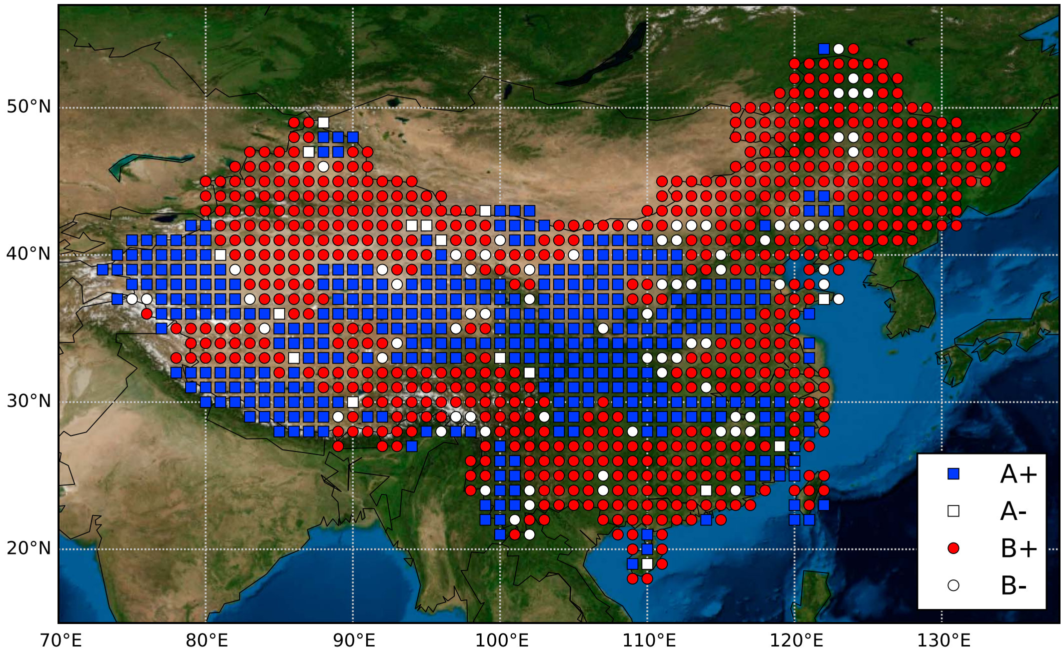

Figure 6. CFSv2 cells in cluster A, where the anomaly correlation tends to be positive at different lead times, and cluster B, where anomaly correlation is generally neutral at different lead times (the "plus" and "minus" signs represent positive and negative Silhouette scores, respectively).

the anomaly correlation vectors in cluster B tend to be close to zero. This result implies that the corresponding CFSv2 forecasts are not skillful and that the anomaly correlation does not seem to improve as time progresses.

It is pointed out that the correlation vectors for individual cells are shown to be noisy in Figure 5. That is, the correlation oscillates as time progresses, which implies that the skill of raw ensemble does not steadily improve. In addition, there exist a considerable number of instances with negative correlations, which mean that a high forecast ensemble mean may correspond to a low observation. Nevertheless, clustering analysis efficiently filters the noises and reveals two meaningful patterns for the relationship between anomaly correlation and lead time.

4.3. Clusters A and B by Hierarchical Clustering

The grid cells corresponding to the clusters A and B are illustrated in Figure 6. Overall, there are 386 cells in cluster A: 370 cells are with positive Silhouette scores (blue squares) and 16 cells are with negative scores (white squares). In the meantime, there are 723 cells in cluster B: 646 cells are with positive scores (red circles) and 77 cells are with negative scores (white circles). According to the patterns of anomaly correlation in Figure 5, CFSv2, in general, provides potentially skillful forecasts for about 35% of the cells covering China. These forecasts even exhibit positive correlations with observations at a lead time of 5 months. The forecasts can be further postprocessed to generate reliable and sharp forecasts for environmental management [*Gneiting et al*., 2005; *Wilks and Hamill*, 2007; *Zhao et al*., 2017]. For the remainder 65% of the cells, the correlation is overall close to zero at different lead times, although the correlation can be positive at certain cells and lead times. In these cases, postprocessing would be of limited use since raw CFSv2 forecasts are not informative, i.e., poorly correlated with observations.

The spatial map of CFSv2 grid cells helps to identify for which parts of China CFSv2 forecasts tend to be skillful. From Figure 6, it is observed that a considerable number of cells in cluster A are in Central China, in particular the Yellow River basin. This river is known for its water scarcity problem. While its catchment area is as large as 752,546 km^2, its flow has partially ceased in 19 years between 1972 and 1996 mainly due to a dry climate and a huge agricultural/industrial water use. The blue cells in Figure 6 suggest that the CFSv2 forecasts can provide useful information regarding summer precipitation for the Yellow River, which can be

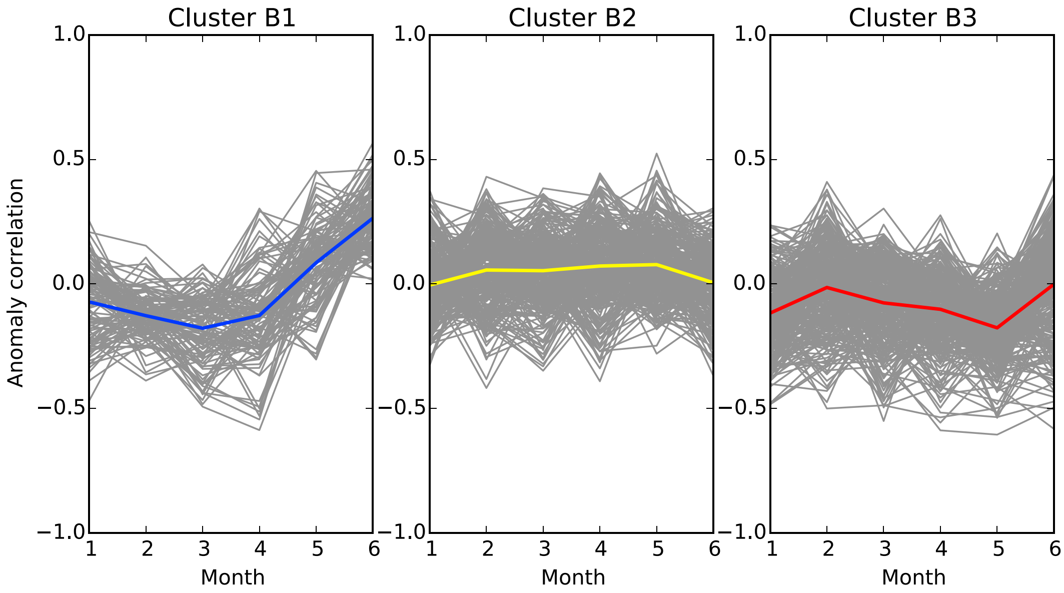

Figure 7. Anomaly correlation vectors in clusters B1, B2, and B3 (the blue, yellow, and red lines indicate the mean of the vectors in clusters B1, B2, and B3, respectively).

useful for long-range water resources scheduling. In the meantime, it is noted that the cells in cluster A also cover part of Southwest China. In this region are the catchments of the Brahmaputra River (the upstream of Ganges River) and the Lancang River (the upstream of the Mekong River). This suggests that CFSv2 forecasts can also be helpful for international water management.

4.4. Clusters B1, B2, and B3 by Hierarchical Clustering

One feature of hierarchical clustering is that it illustrates how a cluster at a higher hierarchy is composed of clusters at a lower level [*Xu and Wunsch*, 2005]. By setting the number of clusters as 4, the anomaly correlation vectors are partitioned into four clusters—A, B1, B2, and B3. The clusters B1, B2, and B3 comprise the cluster B in Figures 5. For B1, B2, and B3, the anomaly correlation vectors are illustrated in Figure 7 and the mean of the vectors are highlighted. It can be observed that in cluster B1, the anomaly correlations are close to zero or negative in January, February, and March, but they tend to improve from April to June and turn to be positive in June. For clusters B2 and B3, the anomaly correlations are overall close to zero or negative, even in June.

The spatial distribution of grid cells in clusters B1, B2, and B3 are further illustrated in Figure 8. Cells marked by B1+, B2+, and B3+ are with positive Silhouette scores, while these by B123− are with negative Silhouette scores. It is observed that more than half of the cells for cluster B1+ (blue circles) with positive Silhouette scores are in Northeast China. This pattern can be confirmed from the spatial distribution of anomaly correlation in Figures 1 and 2: some cells in Northeast China tend to exhibit a positive correlation in June, but the correlation is low in May (and also in other months). Again, it is illustrated that a high correlation at one lead time does not necessarily correspond to a high correlation at other lead times. For cells marked by B1+, skillful forecasts of summer precipitation can only be achieved at a short lead time, i.e., in June. For cells in B2+ and B3+, they spread in South, East, and North China. The anomaly correlations in these two clusters cannot be deemed satisfactory across different lead times, which implies that CFSv2 forecasts still have plenty of room for further improvement.

Further, it is pointed out that there are more instances of negative Silhouette scores in Figure 8 compared to Figure 6. The indication is that the anomaly correlation vectors are not quite distinct across the clusters B1, B2, and B3. The similarity leads to low and even negative Silhouette scores. It also supports Figure 4 in that with two clusters, the patterns of anomaly correlation are the most evident.

4.5. Analysis by k-Means Clustering

There is also a popular k-means clustering algorithm [*Xu and Wunsch*, 2005]. Compared to hierarchical clustering that derives a hierarchy of clusters, k-mean is more straightforward. It directly classifies the anomaly

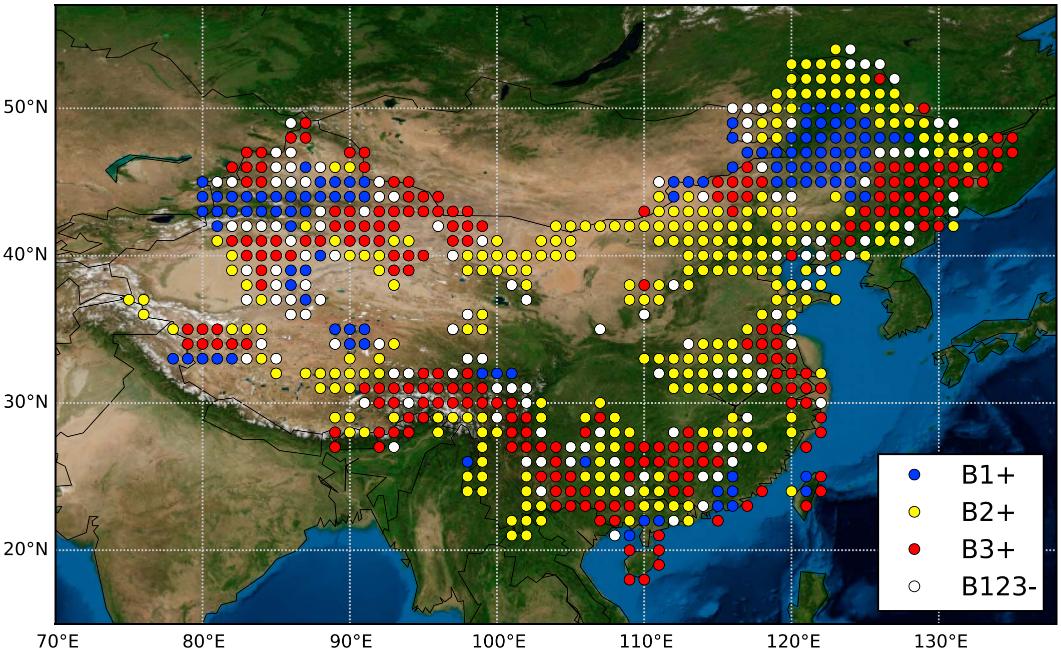

Figure 8. CFSv2 cells with positive Silhouette scores (marked by plus sign) in clusters B1, B2, and B3; the cells with negative Silhouette scores are pooled and marked as B123− for the sake of simplicity.

correlation vectors into k clusters (k is a predefined number). The application of k-means clustering yields similar results to that of hierarchical clustering. The optimal number of clusters is 2. That is, k = 2 corresponds to the peak average Silhouette score as the value of k increases from 2 to 10. The anomaly correlation vectors in the two clusters are shown in Figure 9. It can be observed that one cluster contains anomaly correlation vectors that tend to be positive across different lead times and the other cluster includes vectors that are close to zero. The patterns are similar to those in Figure 5. Further, the cells in the two clusters are illustrated in Figure 10. It can be seen that the cells, which exhibit consistently high anomaly correlation and are with positive Silhouette scores, are situated mostly in Central and Southwest China. Therefore, the patterns of

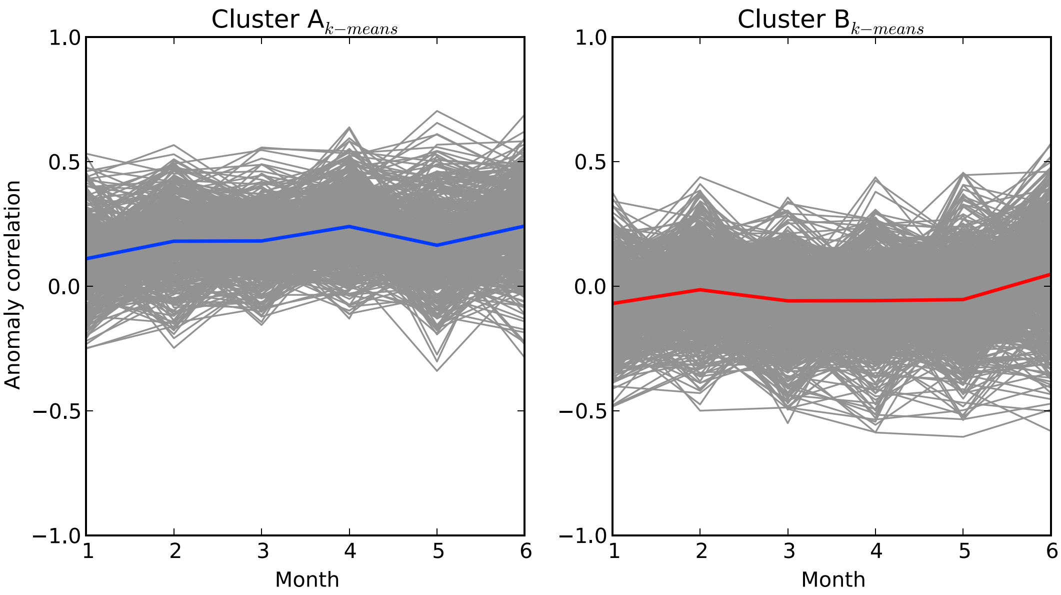

Figure 9. k-means clustering of the anomaly correlation vectors when the number of clusters is set as 2 (k = 2).

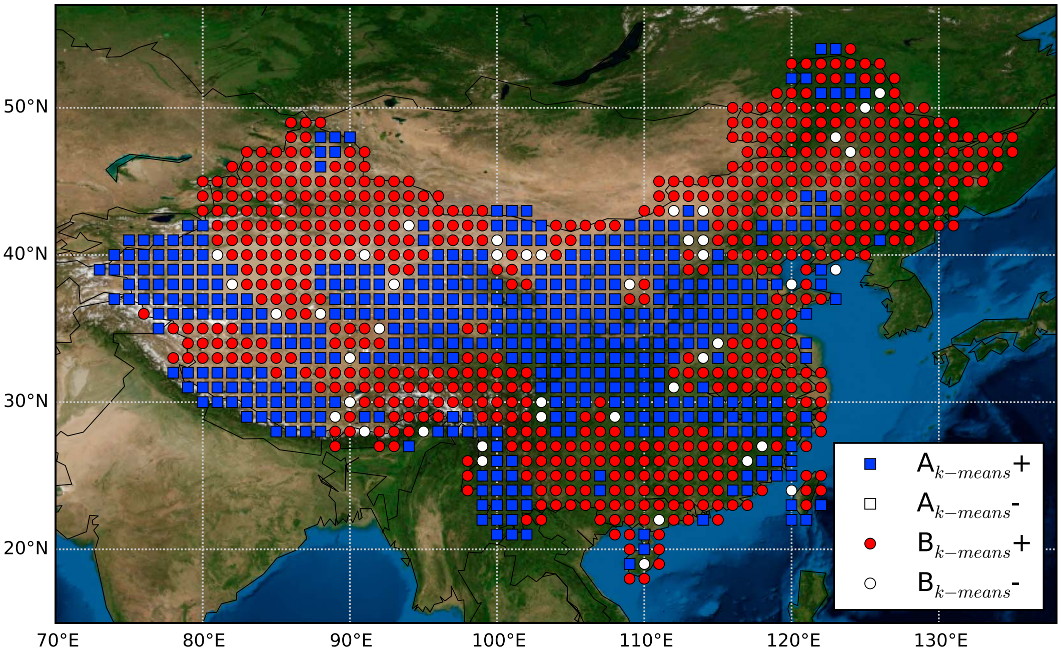

Figure 10. CFSv2 cells in clusters A and B by k-means clustering (the plus and minus signs represent positive and negative Silhouette scores, respectively).

anomaly correlation are not only illustrated by the hierarchical clustering. The similar outputs by the two clustering methods suggest that the patterns are robust and that the clustering is efficient in exploring the patterns.

5. Discussion

The clustering analysis reveals useful patterns from the anomaly correlation of CFSv2 forecasts. We note that the summer precipitation in China is largely dominated by the Asia-Pacific monsoon, which is generally due to the land-sea thermal contrast between the Asian continent and the adjacent Pacific and Indian Oceans [*Ding and Chan*, 2005; *Zhou and Zou*, 2010]. Specifically, it is subject to the East Asian Monsoon (EAM) and the South Asian Monsoon (SAM). The EAM is a subtropical monsoon featured by the low-level winds reversing from winter northerlies to summer southerlies; by contrast, the SAM is a tropical monsoon characterized by the low-level easterlies in winter and westerlies in summer [*Ding and Chan*, 2005]. The patterns illustrated by clustering reflect the capacity of CFSv2 in predicting the monsoon climate in China.

The EAM has a prevailing effect on precipitation in Central, South, East, and North China [*Ding et al.*, 2009; *Qian et al.*, 2009; *Zhao et al.*, 2016b]. The grid cells in Cluster A cover part of Central China. This result implies that CFSv2 tends to capture the effect of EAM in this region. On the other hand, the cells in Cluster B cover large parts of South, East, and North China, which means the performance of CFSv2 is not satisfactory. We note that in these regions, the effect of EAM is subject to other climate processes. In South China, there are interactions between EAM and SAM [*Ding and Chan*, 2005]; in East China, EAM is additionally affected by the zonal thermal contrast between East Asian and the North Pacific Ocean [*Qi et al.*, 2008]; and in North China, EAM is impacted by the midlatitude westerly wind belt in the northern hemisphere [*Qian et al.*, 2009; *Zhao et al.*, 2016b]. In these regions, one cause of the unsatisfactory CFSv2 forecasts can relate to the setting of GCM. For example, *Gao et al.* [2008] observed that GCM with a low resolution leads to the displacement of monsoon front in China and that the displacement can be mitigated by regional climate models at a high resolution; *Chen et al.* [2010] found that cumulus convection affects large-scale circulation and that modifications of the convection scheme improve the simulation of EAM. Another cause is about the predictability. In an investigation of EAM, *Zhou and Zou* [2010] identified that the meridional thermal contrast between East Asia and the tropical western Pacific is reasonably predictable, but the zonal thermal contrast across the East Asian continent and the North Pacific is largely unpredictable. In the future, more efforts are in demand to associate the output by clustering with the physical influencing factors and to yield understandings of the EAM.

The SAM mainly determines the precipitation in the Indian subcontinent, but it can extend to the southern part of China [*Ding et al.*, 2009; *Qian et al.*, 2009; *Zhou and Zou*, 2010]. The cells in Cluster A cover part of Southwest China (Figures 6 and 10). This result suggests that CFSv2 tends to capture the effect of SAM in this region. In this meantime, it implies that CFSv2 may to some extent capture the effect of SAM in South Asia. On the other hand, it is noted that the onset of SAM is, in general, earlier than the onset of EAM and that SAM usually starts in May [*Ding and Chan*, 2005]. For future studies, it is worthwhile to investigate the performance of CFSv2 in predicting the SAM-induced precipitation.

6. Conclusions

This paper has investigated the performance of raw CFSv2 forecasts, as measured by anomaly correlation between raw ensemble mean and observations, in predicting summer precipitation in China. By forming the correlations for forecasts initialized in different months as a vector, we use clustering and reveal two meaningful patterns. For CFSv2 forecasts, they exhibit high correlations with summer precipitation in Central China and Southwest China; the corresponding anomaly correlations tend to be positive even in January. Meanwhile, the forecasts do not show positive anomaly correlations at different lead times in South, East, and North China.

For GCM forecasts, it is well known that the forecast skill exhibits spatial and temporal variations. In this paper, we highlight that the clustering represents an efficient approach to exploring how the forecast skill varies spatially and temporally. There are two prominent advantages for the clustering analysis. First, it can be implemented conveniently, and second, the investigation can be scaled up easily. In future studies, the analysis can be extended to forecasts from GCMs other than CFSv2. To know the comparative performance of multiple GCM forecasts can lead to more efficient combined use of GCM outputs for subseasonal to seasonal hydroclimatic forecasting.

Acknowledgments
We thank the Editors and reviewers for the constructive comments that have led to substantial improvements to the paper. The observed precipitation data are available at the Global Precipitation Climatology Centre (GPCC, https://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html). The CFSv2 forecast data are available at the International Research Institute for Climate and Society (https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.NCEP-CFSv2/.HINDCAST/). This paper is supported by the National Natural Science Foundation of China (51422907 and 41671024) and the Ministry of Science and Technology of China (2016YFC0402208 and 2016YFC0401903). The first author is in particular grateful to the State Key Laboratory of Water Resources and Hydro-power Engineering at Wuhan University for the support (2016SWK01).

References

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger (2013), GloFAS - global ensemble streamflow forecasting and flood early warning, *Hydrol. Earth Syst. Sci.*, *17*(3), 1161–1175.

Barnston, A. G., M. K. Tippett, M. L. L'Heureux, S. H. Li, and D. G. DeWitt (2012), SKill of real-time seasonal ENSO model predictions during 2002-11: Is our capability increasing?, *Bull. Am. Meteorol. Soc.*, *93*(5), 631–651.

Bauer, P., A. Thorpe, and G. Brunet (2015), The quiet revolution of numerical weather prediction, *Nature*, *525*(7567), 47–55.

Bennett, J. C., Q. J. Wang, M. Li, D. E. Robertson, and A. Schepen (2016), Reliable long-range ensemble streamflow forecasts: Combining calibrated climate forecasts with a conceptual runoff model and a staged error model, *Water Resour. Res.*, *52*, 8238–8259, doi:10.1002/2016WR019193.

Chen, H. M., T. J. Zhou, R. B. Neale, X. Q. Wu, and G. J. Zhang (2010), Performance of the new NCAR CAM3.5 in East Asian summer monsoon simulations: Sensitivity to modifications of the convection scheme, *J. Clim.*, *23*, 3657–3675.

Cheng, X. H., and J. M. Wallace (1993), Cluster-analysis of the northern-hemisphere wintertime 500-hpa height field - spatial patterns, *J. Atmos. Sci.*, *50*(16), 2674–2696.

Cloke, H. L., and F. Pappenberger (2009), Ensemble flood forecasting: A review, *J. Hydrol.*, *375*(3–4), 613–626.

DelSole, T., J. Nattala, and M. K. Tippett (2014), Skill improvement from increased ensemble size and model diversity, *Geophys. Res. Lett.*, *41*, 7331–7342, doi:10.1002/2014GL060133.

Ding, Y. H., and J. C. L. Chan (2005), The East Asian summer monsoon: An overview, *Meteorog. Atmos. Phys.*, *89*(1–4), 117–142.

Ding, Y. H., Y. Sun, Z. Y. Wang, Y. X. Zhu, and Y. F. Song (2009), Inter-decadal variation of the summer precipitation in China and its association with decreasing Asian summer monsoon. Part II: Possible causes, *Int. J. Climatol.*, *29*(13), 1926–1944.

Dirmeyer, P. A., and S. Halder (2017), Application of the land-atmosphere coupling paradigm to the operational Coupled Forecast System, Version 2 (CFSv2), *J. Hydrometeorol.*, *18*(1), 85–108.

Gao, X., Y. Shi, R. Song, F. Giorgi, Y. Wang, and D. Zhang (2008), Reduction of future monsoon precipitation over China: Comparison between a high resolution RCM simulation and the driving GCM, *Meteorog. Atmos. Phys.*, *100*(1–4), 73–86.

Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman (2005), Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation, *Mon. Weather Rev.*, *133*(5), 1098–1118.

Ines, A. V. M., and J. W. Hansen (2006), Bias correction of daily GCM rainfall for crop simulation studies, *Agric. For. Meteorol.*, *138*(1–4), 44–53.

Infanti, J. M., and B. P. Kirtman (2016), Prediction and predictability of land and atmosphere initialized CCSM4 climate forecasts over North America, *J. Geophys. Res. Atmos.*, *121*, 12,690–12,701, doi:10.1002/2016JD024932.

Jiang, X. W., S. Yang, Y. Q. Li, A. Kumar, X. W. Liu, Z. Y. Zuo, and B. Jha (2013), Seasonal-to-interannual prediction of the Asian Summer Monsoon in the NCEP Climate Forecast System Version 2, *J. Clim.*, *26*(11), 3708–3727.

Kirtman, B. P., et al. (2014), The north american multimodel ensemble phase-1 seasonal-to-interannual prediction, phase-2 toward developing intraseasonal prediction, *Bull. Am. Meteorol. Soc.*, *95*(4), 585–601.

Koster, R. D., S. P. P. Mahanama, B. Livneh, D. P. Lettenmaier, and R. H. Reichle (2010), Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow, *Nat. Geosci.*, *3*(9), 613–616.

Lang, Y., A. Z. Ye, W. Gong, C. Y. Miao, Z. H. Di, J. Xu, Y. Liu, L. F. Luo, and Q. Y. Duan (2014), Evaluating skill of seasonal precipitation and temperature predictions of NCEP CFSv2 forecasts over 17 hydroclimatic regions in China, *J. Hydrometeorol.*, *15*(4), 1546–1559.

Leung, L. R., and Y. Qian (2005), Downscaling extended weather forecasts for hydrologic prediction, *Bull. Am. Meteorol. Soc.*, *86*(3), 332–333.

Li, H. Y., L. R. Leung, T. Tesfa, N. Voisin, M. Hejazi, L. Liu, Y. Liu, J. Rice, H. Wu, and X. F. Yang (2015), Modeling stream temperature in the Anthropocene: An earth system modeling approach, *J. Adv. Model. Earth Syst.*, *7*(4), 1661–1679.

Liu, X. W., S. Yang, Q. P. Li, A. Kumar, S. Weaver, and S. Liu (2014), Subseasonal forecast skills and biases of global summer monsoons in the NCEP Climate Forecast System version 2, *Clim. Dyn.*, *42*(5–6), 1487–1508.

Luo, L. F., W. Tang, Z. H. Lin, and E. F. Wood (2013), Evaluation of summer temperature and precipitation predictions from NCEP CFSv2 retrospective forecast over China, *Clim. Dyn.*, *41*(7–8), 2213–2230.

Ma, F., A. Z. Ye, X. X. Deng, Z. Zhou, X. J. Liu, Q. Y. Duan, J. Xu, C. Y. Miao, Z. H. Di, and W. Gong (2016), Evaluating the skill of NMME seasonal precipitation ensemble predictions for 17 hydroclimatic regions in continental China, *Int. J. Climatol.*, *36*(1), 132–144.

Maurer, E. P., and D. P. Lettenmaier (2004), Potential effects of long-lead hydrologic predictability on Missouri River main-stem reservoirs, *J. Clim.*, *17*(1), 174–186.

Molteni, F., et al. 2011, The new ECMWF Seasonal Forecast System (System 4). ECMWF Tech. Memo. 656, 49 pp.

Pan, M., X. Yuan, and E. F. Wood (2013), A probabilistic framework for assessing drought recovery, *Geophys. Res. Lett.*, *40*, 3637–3642, doi:10.1002/grl.50728.

Qi, L., J. H. He, Z. Q. Zhang, and J. N. Song (2008), Seasonal cycle of the zonal land-sea thermal contrast and East Asian subtropical monsoon circulation, *Chin. Sci. Bull.*, *53*(1), 131–136.

Qian, W. H., T. Ding, H. R. Hu, X. Lin, and A. M. Qin (2009), An overview of dry-wet climate variability among monsoon-westerly regions and the monsoon northernmost marginal active zone in China, *Adv. Atmos. Sci.*, *26*(4), 630–641.

Rau, P., L. Bourrel, D. Labat, P. Melo, B. Dewitte, F. Frappart, W. Lavado, and O. Felipe (2017), Regionalization of rainfall over the Peruvian Pacific slope and coast, *Int. J. Climatol.*, *37*(1), 143–158.

Regonda, S. K., B. F. Zaitchik, H. S. Badr, and M. Rodell (2016), Using climate regionalization to understand Climate Forecast System Version 2 (CFSv2) precipitation performance for the Conterminous United States (CONUS), *Geophys. Res. Lett.*, *43*, 6485–6492, doi:10.1002/2016GL069150.

Robertson, D. E., D. L. Shrestha, and Q. J. Wang (2013), Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting, *Hydrol. Earth Syst. Sci.*, *17*(9), 3587–3603.

Rousseeuw, P. J. (1987), Silhouettes - a graphical aid to the interpretation and validation of cluster-analysis, *J. Comput. Appl. Math.*, *20*, 53–65.

Saha, S., et al. (2014), The NCEP Climate Forecast System Version 2, *J. Clim.*, *27*(6), 2185–2208.

Schepen, A., Q. J. Wang, and Y. Everingham (2016), Calibration, bridging, and merging to improve GCM seasonal temperature forecasts in Australia, *Mon. Weather Rev.*, *144*(6), 2421–2441.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf (2014), GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, *Theor. Appl. Climatol.*, *115*(1–2), 15–40.

Sheffield, J., et al. (2014), A drought monitoring and forecasting system for sub-sahara african water resources and food security, *Bull. Am. Meteorol. Soc.*, *95*(6), 861–882.

Shrestha, D. L., D. E. Robertson, J. C. Bennett, and Q. J. Wang (2015), Improving precipitation forecasts by generating ensembles through postprocessing, *Mon. Weather Rev.*, *143*(9), 3642–3663.

Siegmund, J., J. Bliefernicht, P. Laux, and H. Kunstmann (2015), Toward a seasonal precipitation prediction system for West Africa: Performance of CFSv2 and high-resolution dynamical downscaling, *J. Geophys. Res. Atmos.*, *120*, 7316–7339, doi:10.1002/2014JD022692.

Thorarinsdottir, T. L., and T. Gneiting (2010), Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression, *J. R. Stat. Soc. Ser. A-Stat. Soc.*, *173*, 371–388.

Tian, D., M. Pan, L. W. Jia, G. Vecchi, and E. F. Wood (2016), Assessing GFDL high-resolution climate model water and energy budgets from AMIP simulations over Africa, *J. Geophys. Res. Atmos.*, *121*, 8444–8459, doi:10.1002/2016JD025068.

Wilks, D. S., and T. M. Hamill (2007), Comparison of ensemble-MOS methods using GFS reforecasts, *Mon. Weather Rev.*, *135*(6), 2379–2390.

Xu, R., and D. Wunsch (2005), Survey of clustering algorithms, *IEEE Trans. Neural Netw.*, *16*(3), 645–678.

Yuan, X., E. F. Wood, L. F. Luo, and M. Pan (2011), A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction, *Geophys. Res. Lett.*, *38*, L13402, doi:10.1029/2011GL047792.

Zhang, Y. Y., G. B. Fu, B. Y. Sun, S. F. Zhang, and B. H. Men (2015), Simulation and classification of the impacts of projected climate change on flow regimes in the arid Hexi Corridor of Northwest China, *J. Geophys. Res. Atmos.*, *120*, 7429–7453, doi:10.1002/2015JD023294.

Zhao, T., A. Schepen, and Q. J. Wang (2016a), Ensemble forecasting of sub-seasonal to seasonal streamflow by a Bayesian joint probability modelling approach, *J. Hydrol.*, *541*, 839–849.

Zhao, T. T. G., J. S. Zhao, H. C. Hu, and G. H. Ni (2016b), Source of atmospheric moisture and precipitation over China's major river basins, *Front. Earth Sci.*, *10*(1), 159–170.

Zhao, T. T. G., J. C. Bennett, Q. J. Wang, A. Schepen, A. W. Wood, D. E. Robertson, and M. H. Ramos (2017), How suitable is quantile mapping for postprocessing gcm precipitation forecasts?, *J. Clim.*, *30*, 3185–3196.

Zhou, T. J., and L. W. Zou (2010), Understanding the predictability of East Asian summer monsoon from the reproduction of land-sea thermal contrast change in AMIP-type simulation, *J. Clim.*, *23*(22), 6009–6026.