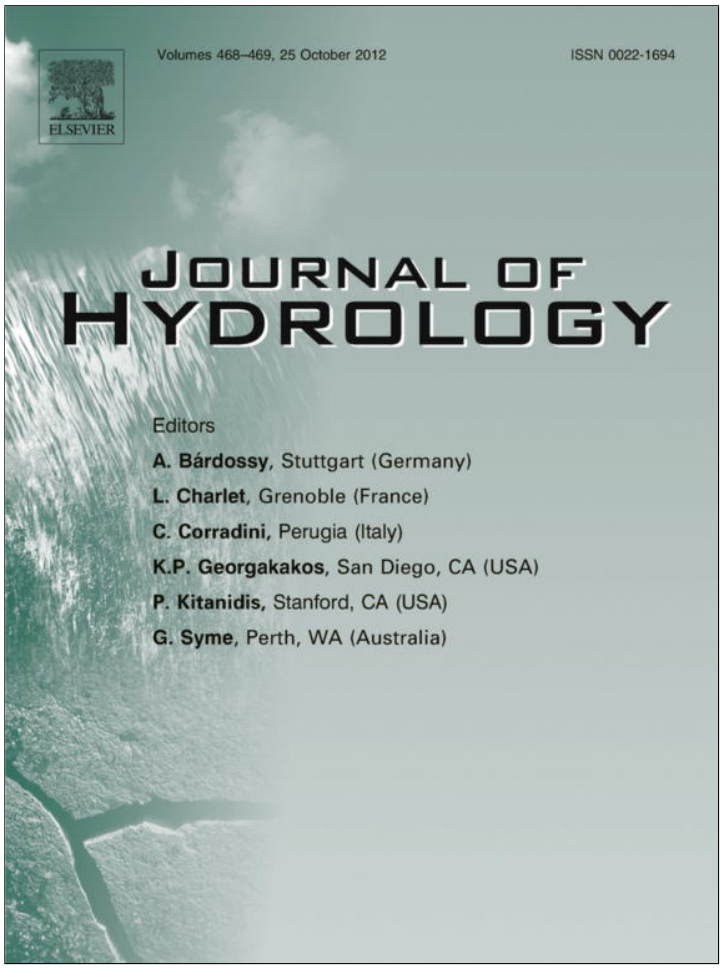

Journal of Hydrology 468–469 (2012) 139–150

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Seasonality of monthly runoff over the continental United States: Causality and relations to mean annual and mean monthly distributions of moisture and energy

Thomas Petersen [a], Naresh Devineni [b], A. Sankarasubramanian [a,*]

[a] Department of Civil, Construction and Environmental Engineering, North Carolina State University, Raleigh, NC 27695-7908, United States
[b] Columbia Water Center, The Earth Institute, Columbia University, New York, NY 10027, United States

ARTICLE INFO

Article history:
Received 18 February 2012
Received in revised form 24 July 2012
Accepted 16 August 2012
Available online 27 August 2012

This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Ashish Sharma, Associate Editor

Keywords:
Streamflow seasonality
Runoff climatology
Aridity index
Intra-annual distributions of moisture and energy
Precipitation seasonality
Soil moisture holding capacity
Principal component analysis

SUMMARY

Observed monthly climatology of streamflow over the continental United States showed significant differences from the monthly precipitation climatology. The objective of this study is to provide an overview of the seasonality of streamflow over the continental US and also to understand the processes that control the streamflow seasonality. For this purpose, we employ principal component regression on five predictors – both climatic and land-surface characteristics – that explain the spatial variability in the streamflow seasonality. While the distribution of mean monthly precipitation is uniform throughout the year over most of the eastern United States (except peninsular Florida), mean monthly streamflow exhibits pronounced seasonality with peak runoff occurring during the winter (early spring) over the Southeast (Mid-Atlantic and Northeast) regions. The spatial variability in the seasonality index – the ratio of peak mean monthly value to the annual total – of runoff over the eastern US primarily depends on the covariability between monthly moisture and energy cycles. As the coherence between these two cycles change from negative to positive over the eastern US, increased moisture availability during the summer results in decreased seasonality index. In contrast, over the western US, both precipitation and streamflow exhibit strong seasonality with respective monthly peaks occurring in early and late winter months. Given that the moisture and energy cycles over the west exhibit significant negative correlation, limited energy availability during peak months of precipitation results in peak monthly runoff occurring in the same season as that of precipitation. Thus, the spatial variability in runoff seasonality over the western United States is strongly dependent on the basin aridity and the seasonality index in precipitation. For catchments over the Midwest and peninsular Florida, given the significant positive correlations in moisture and energy cycles, mean monthly runoff peaks occur in the spring and early summer season with the magnitude of streamflow seasonality being dependent on the aridity index and soil moisture holding capacity of the basin.

1. Introduction

The partitioning of mean annual precipitation into mean annual streamflow and mean annual evapotranspiration is a key aspect in understanding the long-term water balance of a region. It has been shown that the long-term water balance of a region primarily depends on the aridity index – the ratio between mean annual potential evapotranspiration and mean annual precipitation (Budyko, 1974; Milly, 1994a). Milly (1994b) utilized a conceptual water balance model to investigate the role of climate seasonality and soil water holding capacity in determining the long-term water balance over the eastern United States. Recently, Potter et al. (2005) and Hickel and Zhang (2006) analyzed the impact of rainfall seasonality on the mean annual water balance. Similarly, Pui et al. (2011) showed the importance of the role of the antecedent soil moisture conditions in understanding the relationship between rainfall intensity and annual maximum floods. However, limited efforts have been made to understand the effect of seasonality of climate on mean monthly runoff. Farmer et al. (2003) explored the role of climate, soil, and vegetation in controlling the annual, monthly, and daily water balance of several arid basins. Syed et al. (2004) analyzed the 2 years of land surface data from the North American Land Data Assimilation System (NLDAS) to identify the process controls in the land surface hydrological cycle over the conterminous US. They found that precipitation, potential evapotranspiration and soil moisture being the dominant variables that drives the annual and seasonal hydrological cycle. Yokoo et al. (2008) explored the effects of seasonal variability of climate on mean annual and mean monthly water balances theoretically using a hill-slope model.

* Corresponding author.
E-mail addresses: tapeters@ncsu.edu (T. Petersen), nd2339@columbia.edu (N. Devineni), sankar_arumugam@ncsu.edu (A. Sankarasubramanian).

0022-1694/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2012.08.028

While the above studies have focused on understanding the process controls of monthly streamflow theoretically using conceptual models, there has been no systematic analysis in understanding the process controls of seasonality of streamflow based on the observed data over large spatial scales. The primary intent of this study is to provide a broad overview of streamflow seasonality over the continental United States and also to identify the climate and basin controls that influence the streamflow seasonality over 18 water resources regions.

The manuscript is organized as follows: Section 2 introduces the national databases and basin characteristics employed for the study. Section 3 presents the overview of the seasonality in streamflow (Q) and precipitation (P) over the continental United States and also provides the basis for the regional categorization of watersheds into Moisture (Precipitation) and Energy (Potential Evapotranspiration) Availability Scenarios (MEAS) over the continental United States. Section 4 identifies the process controls on seasonality of streamflow for the entire continental United States as well as for each MEAS. Finally, in Section 5, we discuss and summarize the findings from our study.

2. Data description

In analyzing the role of climatic influences on streamflow seasonality over the conterminous United States, we employed monthly time series of precipitation (P), potential evapotranspiration (PET), streamflow (Q) and soil moisture holding capacity available for 1373 Hydroclimatic Data Network (HCDN) watersheds developed by Vogel and Sankarasubramanian (2005). HCDN watersheds, identified by Slack et al. (1993), are basins whose streamflow records are minimally affected by anthropogenic influences such as upstream storages or groundwater pumping. Recently, Martinez and Gupta (2010) emphasized the importance of HCDN watersheds in estimating the various water balance components over the continental United States. A detailed description of the development of the monthly precipitation and potential evapotranspiration time series for the HCDN watersheds can be found in Vogel et al. (1999), Vogel and Sankarasubramanian, 2000 and Sankarasubramanian and Vogel (2005).

2.1. Streamflow database

The HCDN database contains records of average daily streamflow for 1659 locations in the United States and its territories. Streamflow records range from 1874 to 1988 with an average record length of 44 years. Streamflow of basins from the HCDN database is minimally affected by anthropogenic influences such as land use changes and ground water retrieval. Data specialists at the US Geological Survey (USGS) district offices, as well as data specialists at the national headquarters, reviewed records on the

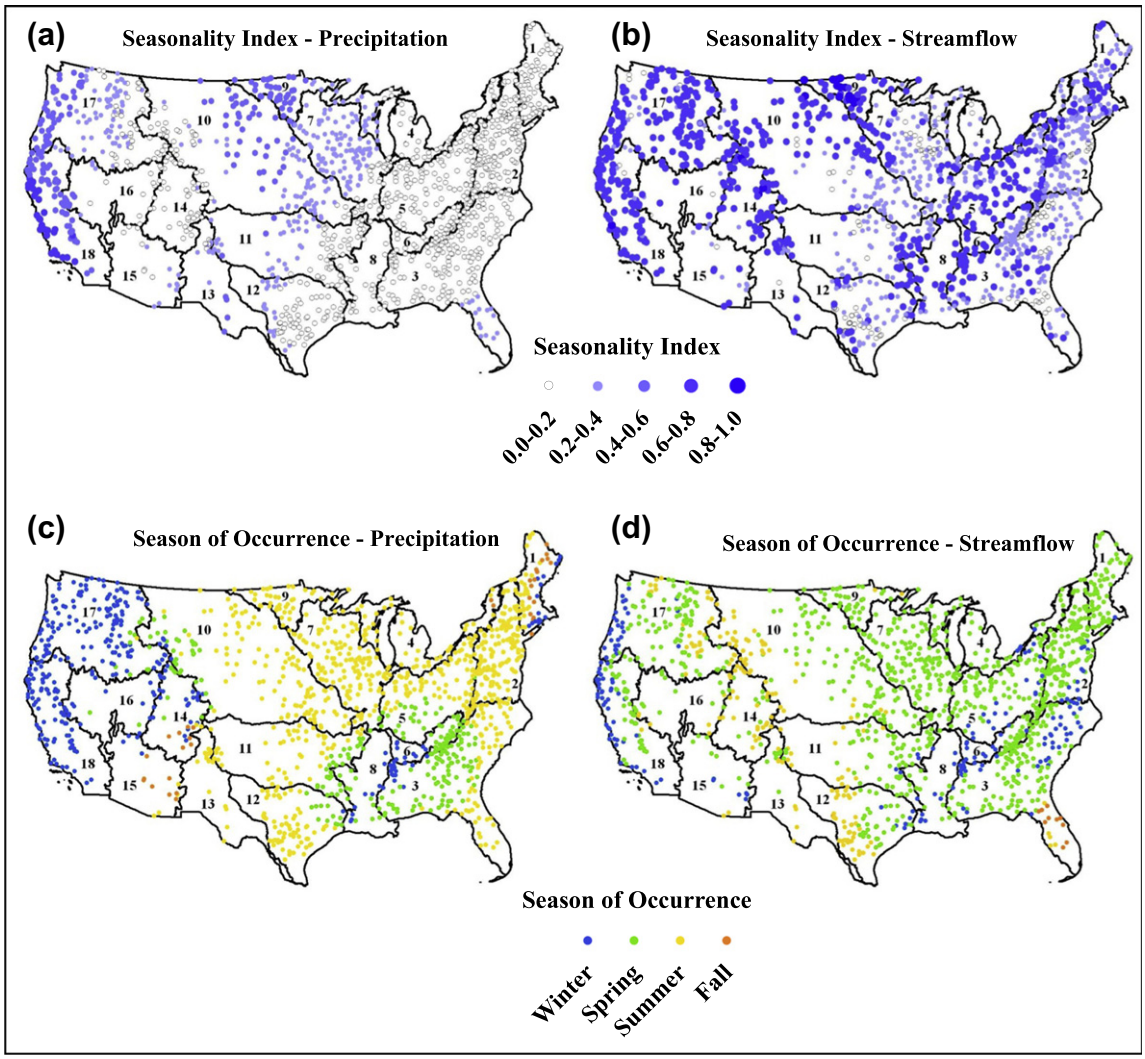

Fig. 1. Seasonality index (a and b) and peak season of occurrence (c and d) for precipitation and streamflow over 1373 HCDN basins across the continental United States.

Table 1
Description of Moisture and Energy Availability Scenarios.

Scenario	# of Basins	Aridity index	Primary locations	Seasonality in P	Covariability between P and PET
MEAS-1	36	$0 \leqslant R \leqslant 1$	Pacific Northwest (Washington, Western Oregon, Northern Idaho)	Strong	$\rho(P,PET) < -0.40$
MEAS-2	36	$R \geqslant 1$	California	Strong	$\rho(P,PET) < -0.40$
MEAS-3	250	$0 \leqslant R \leqslant 1$	New England and Upper Mid-Atlantic	Weak/no seasonality	$-0.40 < \rho(P,PET) < 0.40$
MEAS-4	432	$R \geqslant 1$	Mid-Atlantic, South Atlantic, Great Lakes, Ohio, Tennessee, Upper and Lower Mississippi	Weak/no seasonality	$-0.40 < \rho(P,PET) < 0.40$
MEAS-5	250	$R \geqslant 1$	Florida, Texas, Western Great Lakes, Great Plains	Strong	$\rho(P,PET) > 0.40$

basis of the following criteria: (1) Availability of data in electronic form; (2) record lengths in excess of 20 years unless site location is underrepresented; (3) accuracy ratings of records are at least 'good' as defined by USGS standards; (4) no overt adjustment of 'natural' monthly streamflows by flow diversion, groundwater pumping, or other forms of regulation; and (5) only measured discharge values are tabulated; reconstructed, or estimated records are not used. For this study, we considered streamflow data from 1952 to 1988 with catchments from the 48 contiguous states in the continental United States. This resulted in a total of 1373 basins with 37 years of continuous daily streamflow observations.

2.2. Precipitation and potential evapotranspiration database

Thirty-seven year time series of monthly precipitation and average minimum and average maximum daily temperature for the continental United States were obtained for the 1373 HCDN watersheds using 0.5 time series grids based on the precipitation–elevation regressions on independent slopes model (PRISM) climate analysis system (Daly et al., 1994). PRISM uses a precipitation–elevation regression relationship to distribute point measurements to evenly spaced grid cells. PRISM is considered an improvement over other spatial interpolation methods such as inverse distance weighting or kriging because it attempts to account for orographic effects by using precipitation–elevation regression functions. PRISM also employs adiabatic lapse rate corrections in its temperature interpolations. The monthly climate time series grids were spatially averaged over each HCDN basin using a geographic information system (GIS) and a digital elevation map (DEM) of the United States. A DEM of the United States was used to delineate the watershed boundaries for each of the HCDN river sites. A relatively coarse DEM (1 km resolution raster grid) was employed in this analysis owing to the computational challenge of delineating 1373 watersheds. The end result is a unique national time series data set of monthly precipitation and temperature measurements over the period 1951–1988 corresponding to each of the 1373 watersheds. Using the monthly time series of average minimum and average maximum temperature data along with extraterrestrial solar radiation, estimates of monthly potential evapotranspiration were obtained using a method introduced by Hargreaves and Samani (1982). Extraterrestrial solar radiation was estimated for each HCDN basin by computing the solar radiation over 0.1_grids using the method introduced by Duffie and Beckman (1980) and then summing those estimates over the entire basin. The Hargreaves method was the highest ranked temperature-based method for computing potential evapotranspiration reported in the American Society of Civil Engineers (ASCE) Manual 70 analysis (Jensen et al., 1990).

2.3. Basin characteristics

Information on the soil moisture holding capacity for each of the 1373 HCDN basins was retrieved from the latest database for the United States, developed by Miller and White (1998). It is available at http://www.soilinfo.psu.edu/index.cgi?soil_data&conus&data_cov&awc&datasets. For a given basin, soil characteristics dataset provides the gridded mean available water capacities at 30 arcsec expressed as volumetric percentages of three column lengths: 100 cm, 150 cm, and 250 cm. Our study considered 100 cm column length to estimate the available water capacity, which indicates the approximate depth of water available to plants if the soil, inclusive of rock fragments, were at field capacity. This available water capacity is quantified by the soil moisture holding capacity (b) for each grid and spatially averaged over each HCDN basin in Arc-GIS. We normalized each storage capacity by relating it to the basin's mean annual precipitation $\left(\gamma = \frac{b}{P}\right)$. Additional details regarding this storage index could be found in Sankarasubramanian and Vogel (2002), who used the index to explain mean annual water balance and interannual variability in streamflow over the continental United States. This ratio, termed the storage index, expresses the relative magnitudes of soil moisture holding capacity and moisture availability. Apart from this, we also used basin characteristics of the HCDN sites, developed by Chuck Kroll (http://www.esf.edu/ere/kroll/), in our diagnostic analysis.

3. Seasonality index – description

We used the seasonality index (SI), devised by Markham (1970), to understand the seasonality of the selected hydroclimatic variables. The simplest way to obtain the seasonality of a variable for a given site is to plot the mean monthly values and visually identify the peak month of occurrence. The amount contributed by the peak month to the annual total provides the SI of the variable. However, this approach is infeasible for analyzing the seasonality at continental scales. Hence, Markham (1970) provided an analytical approach to calculate the SI and peak month of occurrence of a variable. Under this approach, mean monthly values are represented as vectors whose magnitude and direction are denoted, respectively, by the mean monthly values and their time of occurrence over the calendar year. Thus, 360° of the polar coordinate system are apportioned to the twelve months of the calendar year. The resultant of these 12 vectors, specified with an angle and magnitude, represents the peak month and the amount of seasonality of a variable. The magnitude of the resultant is then divided by the annual total to calculate the SI of the variable, which can range from 0 to 1. For explicit expressions on how to calculate SI from the mean monthly data, see Markham (1970) or Dingman (2002).

Hydroclimatic attributes having a seasonality index between 0 and 0.2 are defined as non-seasonal and are assumed to exhibit uniform monthly distributions throughout the year; Values close to 1 suggest that the entire annual total occurs in a single month. The SI prescribed by Markham (1970) has shortcomings particularly for attributes that exhibit bimodality with two distinct seasons of high monthly averages. In such instances, the estimated SI may neglect to indicate seasons of pronounced climatic occurrences. Bimodality in precipitation and streamflow is typically

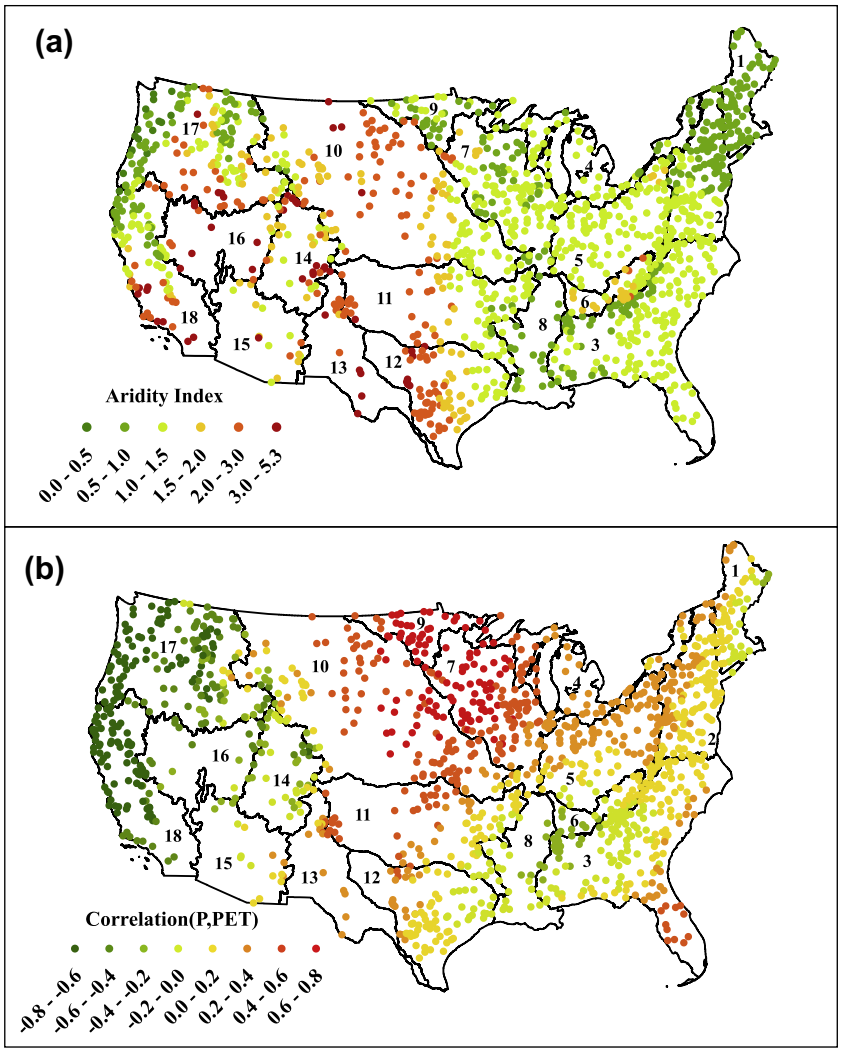

Fig. 2. Aridity index (a) and correlation coefficient between monthly *P* and *PET* (b) for 1373 HCDN basins over the continental United States.

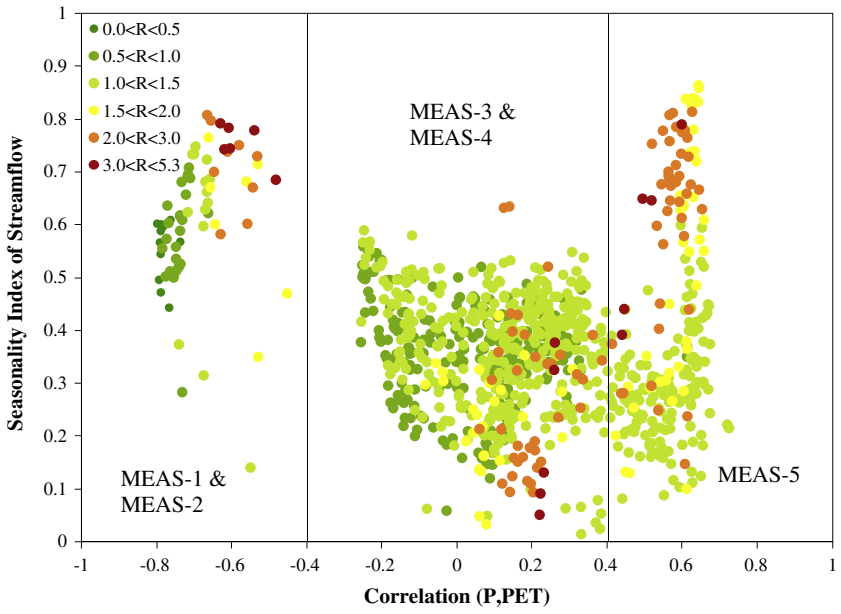

Fig. 3. Observed relationship between the seasonality index of streamflow and the correlation coefficient between monthly precipitation and monthly potential evapotranspiration as a function of aridity index. This figure includes 1005 continental HCDN basins whose average elevation falls below 1000 m.

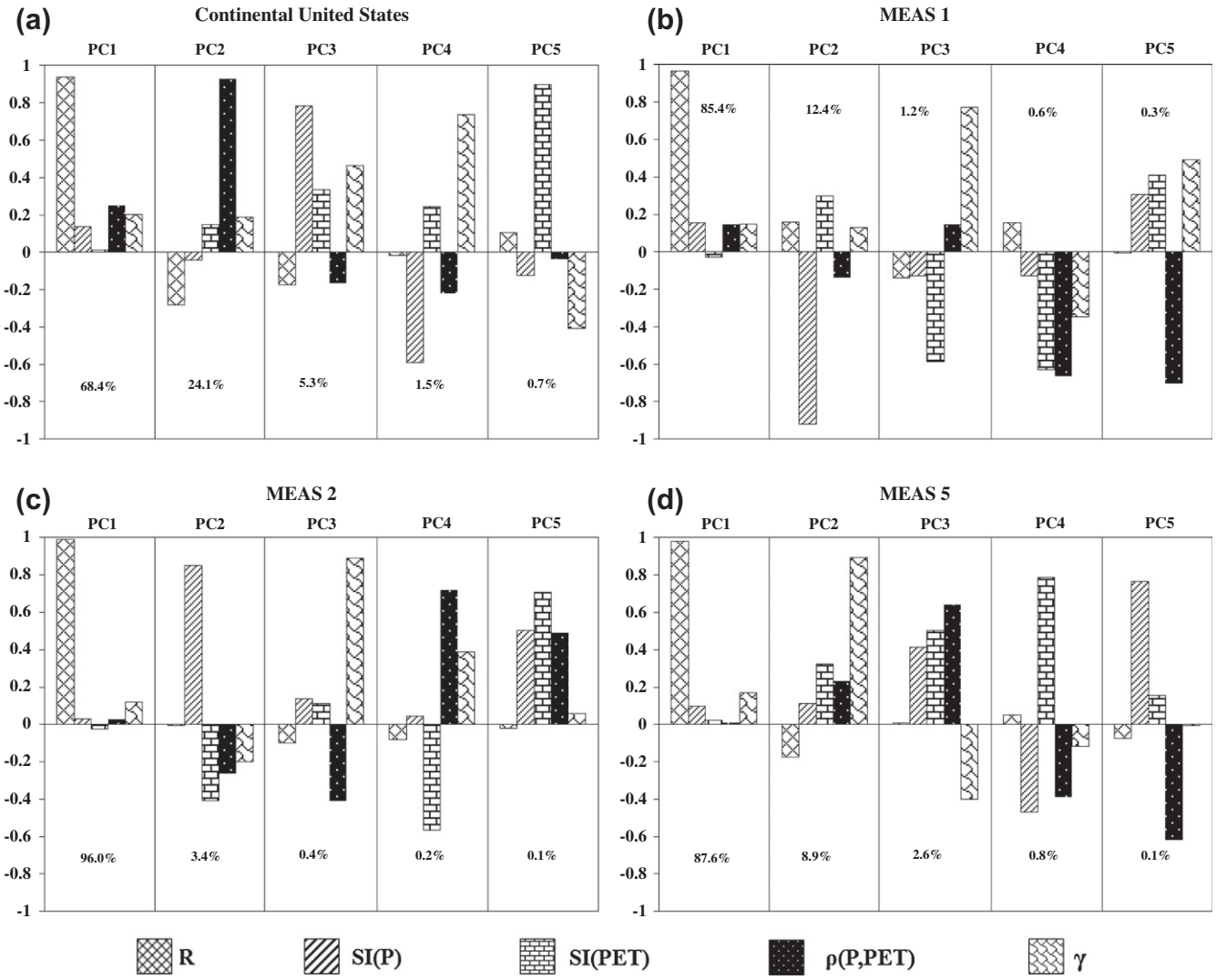

Fig. 4. Eigenvector plots from the principal component analysis performed on various groups of basins: (a) Continental United States, (b) MEAS-1, (c) MEAS-2 and (d) MEAS-5. The percentage under each component indicates the percentage variance explained by that component.

observed in tropical settings (Dingman, 2002). Thus, using *SI* to estimate the seasonality of precipitation and streamflow over the continental United States can be expected to give accurate representations of the intra-annual distribution of hydroclimatic attributes. The peak of mean monthly *PET* is expected to be in the summer, which corresponds to the time of peak energy availability over the study area.

3.1. Overview of the streamflow seasonality over the continental United States

The seasonality index (Fig. 1a and b) and the peak month of occurrence (Fig. 1c and d) of precipitation and streamflow are summarized for the continental United States. We have grouped the peak month of occurrence based on the four seasons: winter (December–February), spring (March–May), summer (June–August), and fall (September–November). The seasonality information presented in Fig. 1 for precipitation is consistent with Markham's findings. Delineating the seasonality of precipitation and streamflow over the continental US elicits several remarkable relations between patterns of rainfall seasonality and runoff seasonality. We observe that California and the Northwestern United States show strong seasonality indices in both precipitation ($SI(P)$) and streamflow ($SI(Q)$). While the peak month of occurrence for precipitation (Fig. 1a) in these basins occurs largely during January, peak month of occurrence for streamflow occurs during late winter and early spring. This indicates that the seasonality in streamflow is influenced by processes of snowmelt and basin storage resulting as a delayed response. Nonetheless, a high $SI(Q)$ over the west coast primarily results from the high $SI(P)$. Winter dominated (December–January; see Fig. 1c) precipitation induces winter and spring (February–March; see Fig. 1d) dominated streamflow. Mean monthly precipitation over the Colorado region and eastern portions of United States Geological Survey (USGS) water-resources region 17 (primarily the Rocky Mountains) indicates no pronounced seasonality. The region's $SI(Q)$, on the other hand, is around 0.6–0.8 with the time of occurrence in the spring and summer months. The elevations of these basins range from 1000 m to 3000 m with snowmelt controlling the $SI(Q)$. As we move towards the Great Plains and Midwest, the $SI(P)$ in region 07 (Upper Mississippi), region 09 (Souris-Red-Rainy), region 10 (Missouri), and region 11 (Arkansas-White-Red) is found to be around 0.2–0.6 with the peak period of occurrence during the summer. Streamflow over the same four regions also shows pronounced seasonality, primarily due to snowmelt, with a period of occurrence in late spring.

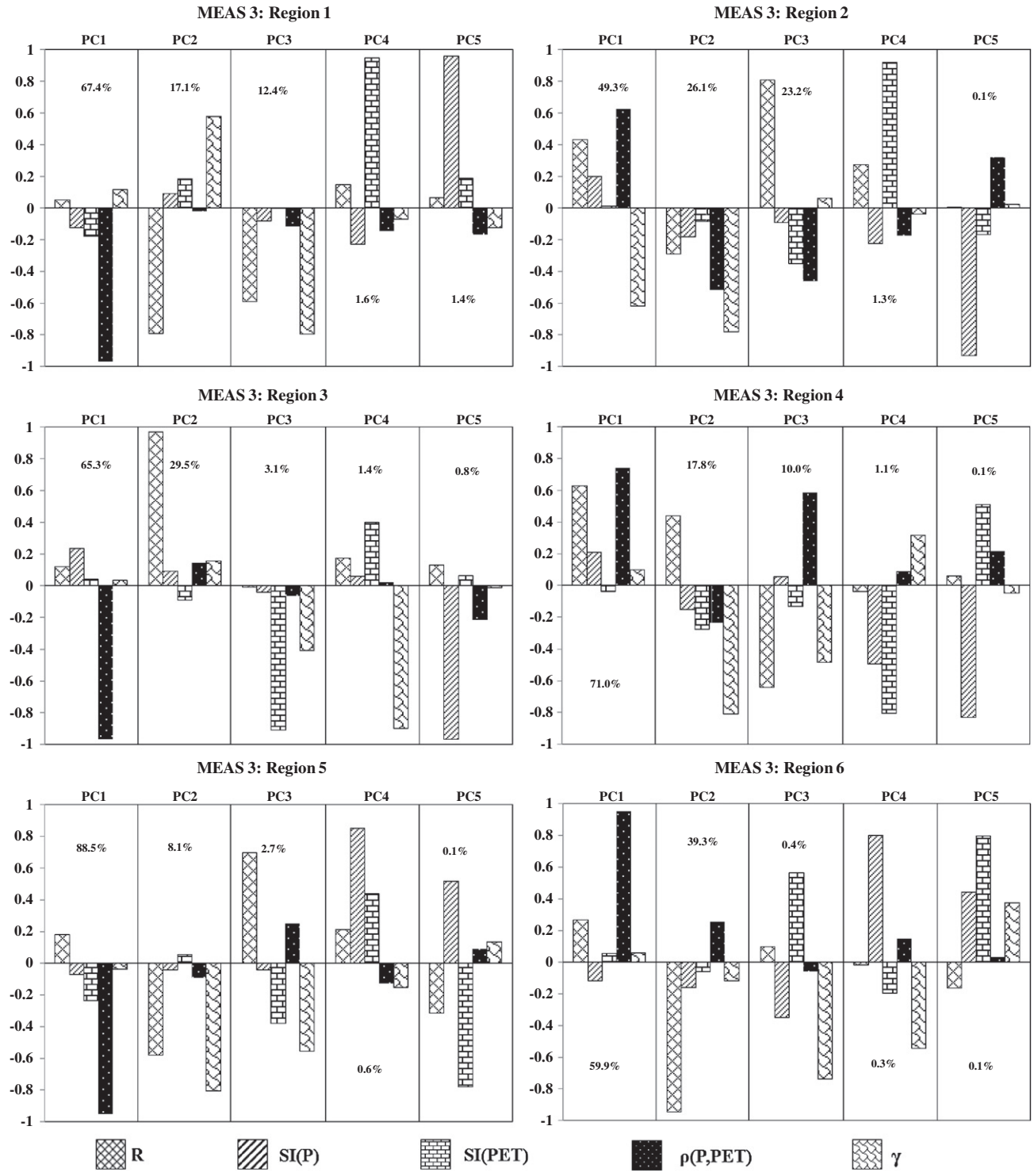

Fig. 5. Eigenvector plots from the principal component analysis performed on the humid basins from MEAS-3 that exhibits no seasonal behavior in moisture availability. The percentage under each component indicates the percentage variance explained by that component.

Over the eastern United States, we infer that peninsular Florida exhibits moderate seasonality in both precipitation and streamflow with *SI* around 0.2–0.4 and 0–0.6 respectively. The peak months of occurrence are experienced during summer and early fall, which primarily arise from hurricane and tropical storm events. The mean monthly precipitation over the eastern United States, including Ohio (region 05) and Tennessee (region 06), Lower Mississippi (region 08), New England (region 01), Mid-Atlantic (region 02), and Southeast (region 03), exhibits no pronounced seasonal behavior as the *SI*(*P*) falls below 0.20. Though precipitation is uniformly dispersed throughout the year, streamflow in these regions exhibits significant seasonality with *SI*(*Q*) values around 0.2–0.6. Winter and spring dominated streamflow events indicate a large portion of mean monthly runoff occurring during periods of minimal evapotranspiration, arising from low energy availability. This overview clearly presents the marked differences in the

Table 2
Adjusted R^2 for the PCR performed on basins across the Continental United States, MEAS-1, MEAS-2, and MEAS-5.

Number of PCs used in regression	Adjusted R^2			
	Continental US (1005 basins)	MEAS-1 (36 basins)	MEAS-2 (36 basins)	MEAS-5(250 basins)
1	0.03	0.01	0.16	0.44
2	0.12	0.60	0.32	0.58
3	0.35	0.71	0.45	0.62
4	0.38	0.74	0.53	0.63
5	0.40	0.74	0.52	0.71

$SI(P)$ and $SI(Q)$ over various regions and thus provides motivation to identify the process controls that explain the spatial variability in SI (Q) over the continental US.

3.2. Moisture and Energy Availability Scenarios

An investigation into the hydroclimatic attributes of the 1373 chosen HCDN basins shows that the behavior of most catchments can be categorized into one of five Moisture and Energy Availability Scenarios (MEAS). Thus, each basin was grouped into one of five MEAS based on aridity index (R) and intra-annual covariability between moisture and energy availability. Details on the grouping of HCDN basins as well as the number of basins under each MEAS are listed in Table 1. Aridity index, $R = \overline{PET}/\bar{P}$, is defined as the ratio of mean annual potential evapotranspiration ($\overline{PET}$) to mean annual precipitation ($\bar{P}$). Basins with R less than 1 are considered humid since, on average, these basins are moisture unlimited. Conversely, basins with R values greater than 1 are classified as temperate ($1 \leqslant R \leqslant 2$) and arid basins ($R > 2$) (Budyko, 1974). The covariability between the seasonal cycles of moisture and energy availability is estimated by the correlation ($\rho(P, PET)$) between monthly precipitation (P) and monthly potential evapotranspiration (PET). $\rho(P, PET)$ ranges from -1 to 1 with large positive (negative) values, $\rho(P, PET) > 0.4$ ($\rho(P, PET) < -0.40$), denoting basins whose precipitation and temperature cycles are (not) coherent/out-of-phase. The regional distributions of R and $\rho(P, PET)$ are indicated in Fig. 2a and b, respectively. Fig. 3 shows the ranges for MEAS-1 to MEAS-5 as a function of R and $\rho(P, PET)$. As illustrated by Fig. 2b, the entire western United States (Regions 14–18) shows a strong negative correlation between P and PET indicating a significant difference in cycles between the moisture and energy availability, while much of the Midwest and peninsular Florida exhibits strong positive correlation, witnessed by coincidental seasonal occurrences of P and PET during the summer. Given the uniformly distributed mean monthly rainfall across the rest of the eastern United States, the covariability between moisture and energy cycles ($\rho(P, PET)$) results ranging from -0.4 to 0.4.

MEAS-1 (MEAS-2) corresponds to humid (temperate to arid) basins in the Pacific Northwest (California and Southwest) where precipitation exhibits strong seasonality and does not coincide with energy availability, resulting in $\rho(P, PET)$ being lesser than -0.40. MEAS-3 (MEAS-4) corresponds to humid (temperate to arid) basins in the Northeast (Mid-Atlantic States and Southeast except Peninsular Florida) where precipitation is uniformly distributed throughout the year and $\rho(P, PET)$ is between -0.40 and 0.40. Finally, MEAS-5 is comprised of both semi-arid and arid basins with moisture and energy cycles coinciding together ($\rho(P, PET) > 0.40$). These basins experience dry winters followed by wet, hot summers and are typically located in the Midwestern United States and peninsular Florida. This overview on the seasonality of streamflow clearly shows that peak mean monthly runoff differs substantially over the United States and depends both on climatic and land surface controls that vary over different MEAS. Our subsequent analyses identifying process controls that explain spatial variability in the seasonality of streamflow will be based on these MEAS groupings as well as for the entire conterminous United States.

4. Diagnostic analysis using Principal Components Regression (PCR)

Given the overview of the seasonality of streamflow over the continental United States, we focus on explaining the spatial variability in $SI(Q)$ and its process controls using PCR. Since basins located at high elevations exhibit significant snow accumulation during the winter, peak periods of runoff occur during the spring and summer as the snow melts. More than 35% of the HCDN basins within the continental United States lie above 1000 m in elevation. These basins are located primarily over Sierra Nevada, Cascades and Rocky Mountain ranges and exhibit very high SI in streamflow ($SI(Q)$) ranging between 0.5 and 0.6. Arid basins exhibit high $SI(Q)$ in these higher elevation basins (figure not shown), since the increased energy availability results in rapid runoff during late spring and early summer. Thus, the seasonal runoff dynamics differ from the rest of the basins over MEAS-1 and MEAS-2 with most of the annual runoff arising from spring and summer melt. Given the distinct seasonal runoff dynamics and limited variability in $SI(Q)$ (0.5–0.6) over these basins, our PCR analysis only considers basins whose average elevation is lower than 1000 m. This resulted in a total of 1005 basins throughout the country (Table 1). We performed PCR to identify the process controls on $SI(Q)$ over the entire continental US and individually for each MEAS.

As previously discussed, both R and $\rho(P, PET)$ can be considered as potential attributes for explaining the seasonality index of streamflow. Next, it is intuitive to consider $SI(P)$ as an additional attribute, since months of substantial rainfall are likely to result in periods of substantial runoff. Similarly, high potential evapotranspiration occurs during the summer over the continental United States with $SI(PET)$ ranging between 0.4 and 0.6. Thus, basins with periods of $SI(Q)$ during the summer (e.g., MEAS-5) are more likely to depend on $SI(PET)$. Additionally, it is important to consider basin characteristics that could influence the seasonality of streamflow. For instance, basins with significant soil moisture holding capacity could reduce the seasonality of streamflow, since increased soil storage, acting as a buffer, could hold the moisture for a longer period of time resulting in reduced mean monthly streamflow. Hence, we considered the normalized soil moisture holding capacity, $\gamma = \frac{b}{\bar{P}}$ as suggested by Sankarasubramanian and Vogel (2002), as a potential attribute in influencing the $SI(Q)$. We also considered hydraulic conductivity, catchment slope and drainage area as additional basin characteristics that could potentially influence the $SI(Q)$. However, these basin characteristics were found to be uncorrelated with $SI(Q)$. Hence, our analyses considered five variables, namely the seasonality indices of precipitation ($SI(P)$) and PET ($SI(PET)$)), the aridity index (R), the correlation between monthly P and PET ($\rho(P, PET)$) and the soil moisture storage index (γ), for explaining the spatial variability in $SI(Q)$.

Since significant spatial correlation exists between the five hydroclimatic and basin attributes (i.e., predictors), we employ Principal Component Regression (PCR), which utilizes the orthogonal scores or the principal components (PCs), for developing a regression relationship with $SI(Q)$. Syed et al. (2004) considered loadings from the principal component analysis (PCA) to identify dominant process controls over various regions in the conterminous United States based on NLDAS data. Similar to Syed et al. (2004), we intend to identify the dominant predictors or process controls by analyzing the eigenvectors of the principal components

Table 3
Adjusted R^2 for the PCR performed on MEAS-3.

Number of PCs used in regression	Adjusted R^2						
	MEAS 3						
	Region 1 (45 basins)	Region 2 (73 basins)	Region 3 (48 basins)	Region 4 (14 basins)	Region 5 (17 basins)	Region 6 (25 basins)	Region 8 (20 basins)
1	0.01	0.09	0.72	0.04	0.62	0.01	0.36
2	0.04	0.27	0.80	0.52	0.64	0.65	0.38
3	0.14	0.55	0.80	0.58	0.67	0.72	0.40
4	0.28	0.62	0.80	0.77	0.55	0.71	0.52
5	0.29	0.62	0.81	0.75	0.52	0.70	0.58

(PCs) that explain the spatial variability in $SI(Q)$ based on PCR. The primary advantage in using PCR is in identifying dominant predictors among correlated predictor variables. The variances of the PCs denote the eigenvalues of the variance–covariance matrix of the predictors. Each principal component is obtained by transforming the original predictors using the respective eigenvectors. The larger the absolute value of an eigenvector, the greater the influence of that predictor in determining that principal component. By identifying the dominant variable in a given PC, based on eigenvectors, we do not claim that dominant variable alone is a good predictor for explaining the spatial variability in $SI(Q)$. It is possible for principal components that explain significant variability in the predictors' space may not be significantly correlated with the predictand. Selecting principal components purely based on their ability to explain the variability of the predictand in PCR could potentially result in poor regression relationships particularly if the selected PCs explain very low percentage of variance on the predictors' total variance (Mason and Gunst, 1985). The aim in using PCR is to identify the dominant predictors that explain the spatial variability in the $SI(Q)$ using the low-dimensional components of the predictors. Hence, we develop regression relationships by first adding principal component with largest variance (PC-1) followed by components explaining lower variances (PC-2 to PC-5). But, we infer dominant variables from the PCR only if the included PCs (i.e., from PC1 to PC5) provide a statistically significant adjusted R^2 in explaining the PCR. The eigenvector plots for each MEAS (Figs. 4 and 5) were created to identify the dominant variables in each PCs. For more information on using PCs to identify dominant hydrologic process controls, see Syed et al. (2004) or Wilks (1995).

PCR results for the entire continental United States (1005 basins) and for each MEAS are summarized in Tables 2–4 and include the adjusted R^2 computed between the PCR estimated $SI(Q)$ and the observed $SI(Q)$by adding principal components (PC) 1–5. The adjusted R^2 helps us to identify the minimum number of PCs required to explain the spatial variability in the observed $SI(Q)$. We show adjusted R^2 from the regression instead of the R^2 since the R^2 in linear regression analysis increases by chance as the number of predictors increase (Helsel and Hirsch, 2002). On the other hand, adjusted R^2 increases with increasing number of predictors, only if the new predictor improves the overall model (Helsel and Hirsch, 2002). Here afterwards, all the discussion on the explained variance on $SI(Q)$ is based on adjusted R^2.

4.1. Continental US

PCR results for the entire continental United States show that the first three PCs account for 98% of the variance in the predictors (Fig. 4a) and explain 35% (Table 2) of the variance in $SI(Q)$. The correlation between the observed $SI(Q)$ and the PCR predicted $SI(Q)$ using these three components is 0.59, which is statistically significant with the 1005 basins. Adding the fourth and fifth components did not significantly improve the adjusted R^2 of the PCR. Analyzing the eigenvectors (Fig. 4a) clearly shows that aridity index, $\rho(P,PET)$ and $SI(P)$ are the dominant processes that determine the first three PCs, thereby indicating their importance in explaining the spatial variability in $SI(Q)$. Since $SI(Q)$ varies substantially over the entire continental US, it is natural to expect a low adjusted R^2 using the selected predictors. Hence, we focus our analyses on each MEAS to understand how different processes control $SI(Q)$ over each region. Our analyses will use eigenvectors (i.e., loadings), eigenvalues (i.e., variance explained on the predictors) and the scores (i.e., PCs) from the PCR to identify the dominant variables in the selected PCs that provide statistically significant regression relationship in explaining the spatial variability in $SI(Q)$.

4.2. Western US (MEAS-1 and MEAS-2)

For MEAS-1, the first principal component did not explain much variability on $SI(Q)$ (Table 2). However, adding the second principal component, the ability to explain the spatial variability in $SI(Q)$ improved substantially. From Fig. 4b, for PC-2, the loadings for $SI(P)$ is the highest indicating the seasonality of precipitation in controlling the $SI(Q)$. As seen in Fig. 4b, the eigenvectors of PC-1 for MEAS-1 are composed almost entirely of the aridity index and explain 85% of the variance in the selected five predictors. Under PC-2 and PC-3, $SI(P)$ and γ result as the dominant predictors. From Table 2, under MEAS-1, the first three PCs are able to explain 71% of the spatial variability in $SI(Q)$. We also infer that adding beyond three PCs did not result in any substantial improvements in adjusted R^2. Under MEAS-2 (Fig. 4c), which are primarily arid basins over the Western United States, aridity index and $SI(P)$ seem to be the dominant controls of $SI(Q)$ based on the eigenvectors of the first two PCs. Under MEAS-2, the first three PCs are able to explain 45% of the spatial variability in $SI(Q)$.

Given that the moisture and energy cycles are negatively correlated under MEAS-1 and MEAS-2 scenarios, peak runoff occurs during the winter. Thus, $SI(Q)$ is primarily controlled by the ratio between mean annual moisture and energy availability (R) and by the seasonality of the precipitation ($SI(P)$). As the basin aridity increases from Northwest to Southern CA, increased energy availability reduces the summer flows resulting in proportionally higher peak flows during the winter months and increased $SI(Q)$. Thus, the attribute that plays a critical role next to the aridity index in explaining the selected PCs is $SI(P)$which specifies the fraction of annual moisture supply available during the winter. Thus, under MEAS-1 and MEAS-2, the critical dynamics between the climate forcings and runoff are observed during the winter season with $SI(P)$and R being the two dominant processes that explain the variability in the selected three PCs for explaining $SI(Q)$.

4.3. Mid-Western US and Peninsular Florida (MEAS-5)

The adjusted R^2 (Table 2) from the PCR using the first three components explained about 62% of the observed variance in $SI(Q)$.Based on the eigenvectors from the first three PCs (Fig. 4d), the dominant variables that control $SI(Q)$ under MEAS-5 are R, γ

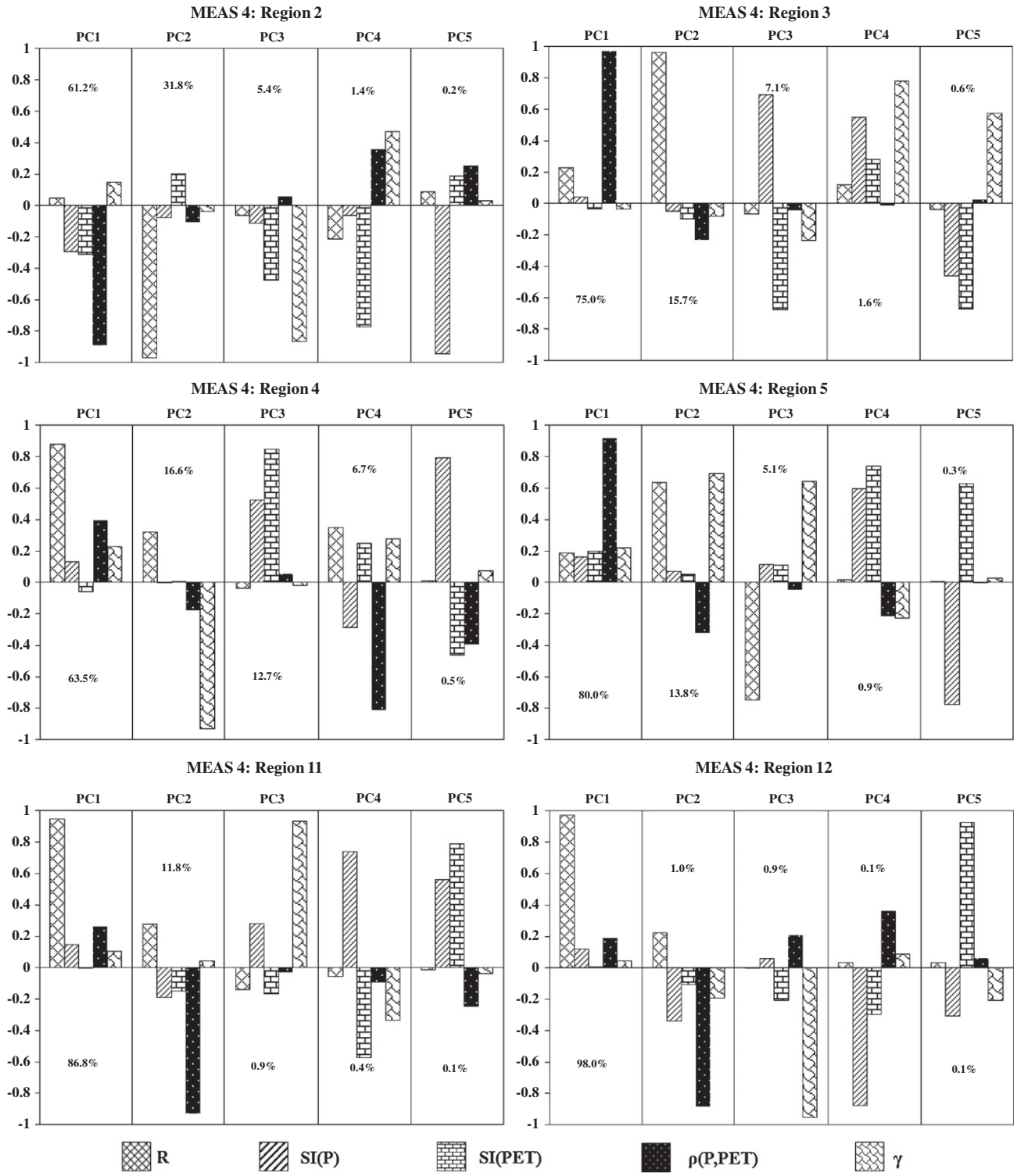

Fig. 6. Eigenvector plots from the principal component analysis performed on the humid basins from MEAS-4 that exhibits no seasonal behavior in moisture availability. The percentage under each component indicates the percentage variance explained by that component.

and $\rho(P,PET)$. All MEAS-5 basins over the Mid-west, as we go from west to east, experience increased moisture availability and reduced $SI(P)$ (Fig. 1a). This increased moisture reduces aridity (Fig. 2a) and positive $\rho(P,PET)$ provides more evapotranspiration opportunity, which results in reduced $SI(Q)$ in the west-east direction over the Mid-west. Since $\rho(P,PET)$ is positive, sufficient moisture is available for evaporation during the summer, thereby the evapotranspiration is also close to PET. Owing to this high evapotranspiration in the summer, peak month of runoff mostly occurs in the winter.

The dominant eigenvector of the 1st component (Fig. 4d) is positive, indicating a positive relation between aridity and $SI(Q)$. The role of soil moisture is reflected under the second component (Fig. 4d) with a negatively signed eigenvector. This indicates that

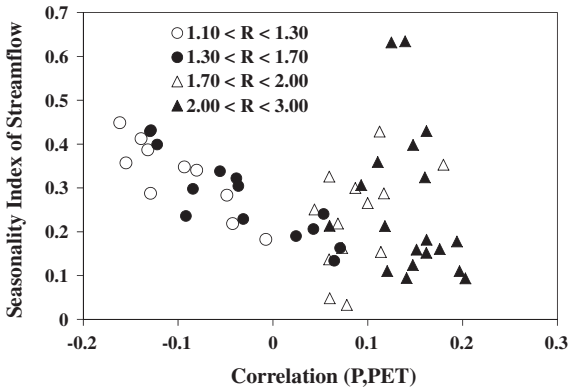

Fig. 7. Observed relationship between the seasonality index of streamflow and the correlation coefficient between monthly precipitation and monthly potential evapotranspiration as a function of aridity index for water-resources region 12.

increased γ provides more opportunity for evapotranspiration, thereby reducing $SI(Q)$. This pattern is particularly evident by the low streamflow seasonality in basins over the sand hills of Nebraska. Peninsular Florida, which also falls under MEAS-5, has $SI(Q)$ from 0.2 to 0.6 with peaks occurring in the summer. Thus, for Florida, streamflow seasonality is observed in late summer to early fall as a response to intense tropical moisture transport.

4.4. Eastern US (MEAS-3 and MEAS-4)

MEAS-3 and MEAS-4 categorize basins with limited/no covariability between moisture and energy ($-0.4 < \rho(P,PET) < 0.4$). The primary reason for the limited covariability between moisture and energy is due to insignificant seasonality in precipitation. From Fig. 2, we can see that MEAS-3 and MEAS-4 cover basins from water resources regions 1–3 (except peninsular Florida), 4–6, 8 and part of the basins from regions 7, 10, 11 and 12. One distinct feature of the seasonality of streamflow in MEAS-3 and MEAS-4 (Fig. 1b) is that basins west of the Appalachians exhibit increased $SI(Q)$ compared to basins on the east. This is primarily due to increased snow from the winter precipitation resulting in relatively concentrated snow-melt season over the western Appalachian basins (Dingman, 2002; Graybeal and Leathers, 2006). Most of the basins in MEAS-3 and MEAS-4 have an aridity index ranging between 0.5 and 1.5 (Fig. 2a) with only few basins from region 12 having aridity index greater than 1.7.

A preliminary PCR analysis (not shown) was performed on all the 626 basins from MEAS-3 and MEAS-4 with the selected five predictors. This resulted in adjusted R^2 values of 0.18 and 0.33 for MEAS-3 and MEAS-4 respectively using all the five PCs. It is important to note that both adjusted R^2 values are statistically significant by comparing the correlation between the observed and the predicted $SI(Q)$ using PCR. These low adjusted coefficients of determination are primarily due to the increased spatial variability in $SI(Q)$ across 646 basins (Fig. 3). Thus, to gain a better understanding of the process controls on $SI(Q)$, PCR was performed on each water resources region separately. The eigenvectors and the adjusted R^2 from PCR are presented for MEAS-3 (Fig. 5 and Table 3) and MEAS-4 (Fig. 6 and Table 4) for each region within the MEAS.

From Fig. 5, we infer that the first component explained more than 50% of the spatial variability among the five predictors in regions 1–6 and 8. Except for region 1, the adjusted R^2 of the developed PCR varied between 0.52 and 0.81 (Table 3). Seasonality of streamflow in New England (region 1) is very homogenous in $SI(Q)$with values ranging between 0.30 and 0.42 over 45 basins, thereby resulting in very low adjusted R^2. Thus, the developed PCR relationship did not explain the observed variability in $SI(Q)$.

For other regions, eigenvector plots (Fig. 5) clearly show that $\rho(P,PET)$ is the most dominant predictor in determining the first principal component (PC-1). Following $\rho(P,PET)$, aridity index and normalized soil moisture storage capacity also play an important role in explaining the spatial variability in $SI(Q)$. Eigenvectors for region 8 (not shown in Fig. 5 for better layout) also showed that $\rho(P,PET)$ followed by aridity index are the dominant predictors in determining the first two PCs.

MEAS-4, arid basins with $\rho(P,PET)$ between -0.4 and 0.4, exhibits similar behavior with $\rho(P,PET)$ being the dominant predictor that explains the variability in the first PC (Fig. 6) for regions 2, 3 and 5. Following $\rho(P,PET)$, the predictors that are dominant under PC-2 and PC-3 are R and γ respectively. The explained variance by the first three PCs in PCR for these regions is between 0.41 and 0.45. For regions 4, 7 and 11, the first three PCs explain 0.72, 0.40 and 0.45 of the observed variance in $SI(Q)$ respectively. As is inferable from Fig. 6, the most dominant variable that explain the first PC is aridity index followed by $\rho(P,PET)$ (region 7 and region 11) or γ (region 4) under PC-2. Given regions 4, 7 and 11 are semi-arid within MEAS-4, we understand that most of the basins in these regions are more arid compared to the basins in the regions 2, 3 and 5 (Fig. 2a). This could be one reason for the increased role of aridity index in controlling the seasonality index of streamflow.

For region 12 (Fig. 7), the selected five predictors explained only 19% of the observed variance in $SI(Q)$. This is primarily due to high aridity index with more than 50% of the basins having R within 1.7–3.0. Further, $SI(Q)$ also varies over a large range (0.05–0.65). Thus, in these basins, apart from the selected five predictors, we expect that other predictors such as storm arrival rate could be playing a critical role. For the rest of the basins (i.e., basins with aridity index lesser than 1.7) in region 12, adjusted R^2 of the PCR explained 80% of the variance using the selected five predictors with $\rho(P,PET)$, R and γ being the dominant predictors. Due to these two distinct groups, the PCR resulted in very low adjusted R^2.

Thus, for the eastern US (except region 1 and region 12 under MEAS-3 and MEAS-4), the most dominant variables in the first three PCs are the seasonal covariability between moisture and energy, followed by the aridity index and the normalized soil moisture holding capacity (Table 5). MEAS-3 and MEAS-4 are predominantly humid and semi-arid basins with R in the range

Table 4
Adjusted R^2 for the PCR performed on MEAS-4.

Number of PCs used in regression	Adjusted R^2						
	MEAS-4						
	Region 2 (60 basins)	Region 3 (118 basins)	Region 4 (22 basins)	Region 5 (61 basins)	Region 7 (18 basins)	Region 11 (48 basins)	Region 12 (76 basins)
1	0.40	0.32	0.01	0.24	0.27	0.18	0.01
2	0.41	0.41	0.01	0.25	0.29	0.44	0.01
3	0.44	0.41	0.72	0.41	0.40	0.45	0.13
4	0.48	0.42	0.72	0.44	0.41	0.59	0.19
5	0.57	0.43	0.70	0.44	0.53	0.58	0.18

Table 5
Processes exhibiting dominant role under each principal component for explaining the spatial variability in $SI(Q)$ under different groupings of the HCDN basins.

PCs	Dominant influences on $SI(Q)$							
	Continental US	MEAS-1	MEAS-2	East of Appalachian		West of Appalachian		MEAS-5
				MEAS-3	MEAS-4	MEAS-3	MEAS-4	
PC-1	R	R	R	$\rho(P,PET)$	$\rho(P,PET)$	$\rho(P,PET)$	R	R
PC-2	$\rho(P,PET)$	$SI(P)$	$SI(P)$	R	R	R	$\rho(P,PET)$	γ
PC-3	$SI(P)$	γ	γ	γ	$SI(PET)$	$SI(PET)$	γ	$\rho(P,PET)$

from 0.5 to 1.5. Given that these basins have no seasonality in precipitation, any seasonal shift in moisture towards the summer results in increased evaporation. Fig. 3 clearly captures this physical control. That is, as $\rho(P,PET)$ increases from −0.4 to 0.4, indicating increased shift in the rainfall over the summer, $SI(Q)$ clearly decreases due to the simultaneous availability of moisture and energy. However, beyond a certain threshold ($\rho(P,PET) > 0.6$), further increase in summer precipitation results in increased $SI(Q)$, since the ability to retain moisture by the soil is bounded by its holding capacity. Thus, the normalized soil moisture holding capacity, γ, plays a dominant role next to the aridity index, R, in explaining the variations in seasonal streamflow concentrations over the Mid-western US (MEAS-5). Aridity index plays a more important role than γ in MEAS-5, since these are basins moisture limited ($R > 1$). Hence, the amount of runoff is first controlled by the amount of the precipitation and then by the variability in the soil moisture holding capacity over the region. In contrast, over the Western US (MEAS-1 and MEAS-2), where $\rho(P,PET) < -0.4$, seasonality in precipitation and aridity index play a critical role in explaining the spatial variability in $SI(Q)$.

5. Discussion

Observed monthly climatology of streamflow over the continental United States showed significant differences from the monthly precipitation climatology. To summarize, $SI(Q)$ depends primarily on two key climatic controls over the continental United States: (a) the aridity index and $\rho(P,PET)$. Fig. 3 plots the $SI(Q)$ over the continental United States as a function of R and $\rho(P,PET)$. The summary of the identified dominant process controls for each MEAS is given in Table 4. PCR analysis shows that $SI(P)$ and R are the primary controls over MEAS-1 and MEAS-2, whereas R and γ are the primary controls over MEAS-5. Over MEAS-3 and MEAS-4, the variations in $\rho(P,PET)$ showed significant influence on $SI(Q)$ for basins with limited seasonality in rainfall. From Fig. 3, we observe that arid basins with $\rho(P,PET) < -0.4$ generally exhibit the greatest $SI(Q)$ across the United States. Since moisture and energy peaks do not occur simultaneously, we can expect substantial runoff to occur in winter during which precipitation peaks. For basins under MEAS-3 and MEAS-4 with no seasonality in precipitation, $SI(Q)$ generally decreases as $\rho(P,PET)$ varies from −0.40 to 0.40. Given uniform monthly precipitation throughout the year, the dominant process control on $SI(Q)$ shifts from R to $\rho(P,PET)$. This implies that increased $\rho(P,PET)$ results in decreased $SI(Q)$ as moisture is shifted from the winter to the summer months. However, once $\rho(P,PET)$ reaches beyond 0.4, which correspond to basins having strong seasonality in precipitation under MEAS-5, $SI(Q)$ shows little relation to $\rho(P,PET)$. As mean monthly distributions of P and PET closely coincide ($\rho(P,PET) \sim 0.60$), $SI(Q)$ variations primarily depend on the spatial variability in the aridity index. Thus, in assessing the spatial variations in streamflow seasonality for basins across the continental United States, we can identify two critical climatic controls – R and $\rho(P,PET)$ – which quantify the mean annual and seasonal distributions of moisture and energy respectively. Following these two variables, the ratio of soil moisture holding capacity in relation to the mean annual precipitation also influences the $SI(Q)$, particularly over the mid-western United States. Increased γ reduces $SI(Q)$, as trapped moisture in the soil increases evapotranspiration opportunity resulting in flattened monthly distributions of runoff. These process controls – aridity index, covariability between precipitation and potential evapotranspiration and soil moisture holding capacity – provide the basis for developing a physically based model that could estimate the mean monthly streamflow purely based on the climate forcings and soil water holding capacity. Our future effort will focus on developing these models based on the understanding from this study.

Acknowledgements

This study was partially supported by the US NSF under CAREER Grant CBET-0954405. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not reflect the views of the NSF. We also thank the three anonymous reviewers and the Associate Editors whose comments led to significant improvements in the manuscript.

References

Budyko, M.I., 1974. Climate and Life, Translated from Russian by D.H. Miller. Academic, San Diego, Calif.
Daly, C., Neilson, R.P., Philips, D.L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteorol. 33 (2), 140–158.
Dingman, S.L., 2002. Physical Hydrology, second ed. Prentice Hall, NJ.
Duffie, J.A., Beckman, W.A., 1980. Solar Engineering of Thermal Processes, John Wiley, New York, 109 pp.
Farmer, D., Sivapalan, M., Jothityangkoon, C., 2003. Climate, soil and vegetation controls upon the variability of water balance in temperate and semi-arid landscapes: downward approach to hydrological prediction. Water Resour. Res. 39 (2), 1035. http://dx.doi.org/10.1029/2001WR000328.
Graybeal, D.Y., Leathers, D.J., 2006. Snowmelt-related flood risk in Appalachia: first estimates from a historical snow climatology. J. Appl.Climatol. Meteorol. 45 (1), 178–193.
Hargreaves, G.H., Samani, Z.A., 1982. Estimating potential evapotranspiration. J. Irrig. Drainage Eng. 108 (3), 225–230.
Helsel, D.R., Hirsch, R.M., 2002, Statistical Methods in Water Resources, Techniques of Water Resources Investigations, US Geological Survey, 522pp.
Hickel, K., Zhang, L., 2006. Estimating the impact of rainfall seasonality on mean annual water balance using a top-down approach. J. Hydrol. 331 (3–4), 409–424.
Jensen, M.E., Burman, R.D., Allen, R.G., 1990. Evapotranspiration and irrigation water requirements. Manuals and Rep. on Eng. Practice 70, 350 pp. (Am. Soc. of Civ. Eng., New York).
Markham, C.G., 1970. Seasonality of precipitation in United States. Ann. Assoc. Am. Geogr. 60 (3), 593–597.
Martinez, G.F., Gupta, H.V., 2010. Toward improved identification of hydrological models: a diagnostic evaluation of the "*abcd*" monthly water balance model for the conterminous United States. Water Resour. Res. 46, W08507. http://dx.doi.org/10.1029/2009WR008294.
Mason, R.L., Gunst, R.F., 1985. Selecting principal components in regression. Stat. Prob. Lett. 3, 299–301.
Miller, D.A., White, R.A., 1998. A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling. Earth Int. <http://EarthInteractions.org>.
Milly, P.C.D., 1994a. Climate, soil water storage, and the average annual water balance. Water Resour. Res. 30 (7), 2143–2156.
Milly, P.C.D., 1994b. Climate, interseasonal storage of soil water and the annual water balance. Adv. Water Resour. 17, 19–24.
Potter, N.J., Zhang, L., Milly, P.C.D., McMahon, T.A., Jakeman, A.J., 2005. Effects of rainfall seasonality and soil moisture capacity on mean annual water balance

for Australian catchments. Water Resour. Res. 41 (6). http://dx.doi.org/10.1029/2004WR003697.
Pui, A., Lal, A., Sharma, A., 2011. How does the interdecadal Pacific oscillation affect design floods in Australia? Water Resour. Res. 47, W05554. http://dx.doi.org/10.1029/2010WR009420.
Sankarasubramanian, A., Vogel, R.M., 2002. Annual hydroclimatology of the United States. Water Resour. Res. 38 (6) (art. no.1083).
Sankarasubramanian, A., Vogel, R.M., 2005. Comment on the paper: "Basin hydrologic response relations to distributed physiographic descriptors and climate" by Karen Plaut Berger, Dara Entekhabi, 2001. Journal of Hydrology 247, 169–182. J. Hydrol. 263 (1), 257–261.
Slack, J.R., Lumb, A.M., Landwehr, J.M., 1993. Hydroclimatic data network (HCDN): a US Geological Survey streamflow data set for the United States for the study of climate variation, 1874–1988. Water Resour. Invest. Rep. 93-4076.
Syed, T.H., Lakshmi, V., Paleologos, E., Lohmann, D., Mitchell, K., Famiglietti, J.S., 2004. Analysis of process controls in land surface hydrological cycle over the continental United States. J. Geophys. Res. 109, D22105. http://dx.doi.org/10.1029/2004JD004640.
Vogel, R.M., Sankarasubramanian, A., 2000. Scaling properties of annual streamflow in the continental United States. J. Hydrol. Sci. 45, 465–476.
Vogel, R.M., Sankarasubramanian, A., 2005. USGS Hydro-Climatic Data Network (HCDN): Monthly Climate Database, 1951–1990, Data Set Available On-line from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. doi:http://dx.doi.org/10.3334/ORNLDAAC/810.
Vogel, R.M., Wilson, I., Daly, C., 1999. Regional regression models of annual streamflow for the United States. J. Irrig. Drainage Eng., ASCE 125 (3), 148–157.
Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences. Academic Press, New York.
Yokoo, Y., Sivapalan, M., Oki, T., 2008. Investigating the roles of climate seasonality and landscape characteristics on mean annual and monthly water balances. J. Hydrol. 357, 255–269.