Journal of Hydrology 374 (2009) 1–15

Contents lists available at ScienceDirect

Journal of Hydrology

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

Detecting hydroclimatic change using spatio-temporal analysis of time series in Colorado River Basin

Mukesh Kumar, Christopher J. Duffy *

Pennsylvania State University, Department of Civil and Environmental Engineering, University Park, PA 16802, United States

ARTICLE INFO

Article history:
Received 4 November 2008
Accepted 26 March 2009

This manuscript was handled by
K. Georgakakos, Editor-in-Chief

Keywords:
Singular spectrum analysis
Principal component analysis
Pattern classification
Change detection
Colorado River Basin

SUMMARY

It is generally accepted that the seasonal cycle of precipitation and temperature in cordillera of the western US exhibits a north–south pattern for annual, interannual and decadal time scales related to large-scale climate patterns. In this paper we explore these relationships, with special attention to the role of local and regional physiographic, hydrogeologic and anthropogenic conditions on low-frequency climate and terrestrial response modes. The goal is to try to understand the spatio-temporal structure in historical precipitation, temperature and streamflow records (P–T–Q) in terms of climate, physiography, hydrogeology, and human impacts. Spatial coherence in time series is examined by classification of factor loadings from principal component analysis. Classification pattern of P–T–Q stations indicate that local physiography, the hydrogeology, and anthropogenic factors transform atmospheric forcing and terrestrial response into unique clusters. To study the temporal structure, dominant low-frequency oscillatory modes are identified for a region from historical P–T–Q records using singular spectrum analysis. Noise-free time trajectories are reconstructed from the extracted low-frequency modes (seasonal–decadal) for each contributing watershed area corresponding to streamflow observation stations, and the phase–plane plots are obtained. Together, the spatial classification and phase plane provides a means of detecting how large-scale hydroclimatic patterns relate to major landforms and anthropogenic impacts across the CRB. The main result of this paper is that resolving the relative impact of basin-wide patterns of climate, physiography and anthropogenic factors (irrigation, dams, etc.) on runoff response can be a useful tool for detection and attribution for each source of variability.

Introduction

Detection and attribution of hydroclimatic variability and change in runoff has been a major concern in the Colorado River Basin (CRB), where the US basin States and Mexico rely on runoff from the late summer monsoon and the annual snowmelt runoff from higher elevations (Stewart et al., 2004). The consequences of these climate patterns on the sustainability of water resources have been well documented in Gleick and Adams (2000). Dettinger et al. (1998) provide convincing arguments of the importance of large-scale features of the climate system such as sea-surface temperature and sea-level pressure on interannual and decadal climate variability in the western US. Beyond climatic effects, we expect intrabasin variability in streamflow response to be related to soil and hydrogeologic properties, topographic features (elevation, aspect and location), landuse and landcover impacts. Dettinger and Diaz (2000) highlighted the effect of geographical differences on flow of world's largest rivers, while Potter et al. (2004) elucidated the critical importance of anthropogenic effects which in some cases is stronger than climate effects on runoff response of large rivers. Nonetheless, methods for detecting changes in a physiographically and climatically diverse terrestrial system such as CRB, remains a challenge. Models of the hydroclimatic system provide insight on the terrestrial response, but are also limited by inadequate understanding of the coupling mechanisms, processes, limited data, the spatio-temporal scale of land atmosphere exchange and computational needs. For example Shun and Duffy (1998) have shown that in the Wasatch front, slow responding groundwater flow system can amplify relatively weak climate forcing modes e.g. interannual or decadal effects of precipitation and temperature.

In this paper, we use a data-driven strategy from dynamical systems theory that provides a direct measure of hydroclimatic change within a river basin or a watershed. The first step in our analysis involves identifying the spatial coherences in P, T and Q at locations across different physiographic regions using monthly data. The approach uses spatial principal component analysis (PCA) of respective time series followed by *k-means* clustering of the factor loading (obtained from PCA) to bracket observed variables into various classes. Classifications are then analyzed to identify the commonalities between observation locations (that are

* Corresponding author. Fax: +814 863 7304.
E-mail address: cxd11@psu.edu (C.J. Duffy).

0022-1694/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2009.03.039

clustered in the same group) in terms of their physiography and hydroclimatology. The analysis helps in identification of landform and climate descriptors that leads to distinctive P–T–Q patterns in the CRB. The similarity in dynamics at various observation stations is found to be closely related not only to physiography, topography and local relief but also to human impacts such as diversions, dams, irrigation, drains, etc. In the second step, dominant oscillatory modes from the observed noisy P–T–Q data are extracted using singular spectrum analysis (SSA, Hanson et al., 2004). The dominant modes (semi-annual, annual, interannual, decadal time scales) are observed as peaks in the eigenspectrum. These modes carry a sizeable fraction of the total variance of the record. Distinctions between P, T and Q observation stations are identified based on the spatial trends of the dominant modes. Next, in order to study the simultaneous change of P, T and Q over different watersheds in the basin, noise-free $\widehat{P}-\widehat{T}-\widehat{Q}$ trajectory plots (time implicit) are obtained. We note that since the objective of this step is to study temporal changes in the integrated watershed response at different locations, reconstructed time series $(\widehat{P}_i, \widehat{T}_i, \widehat{Q}_i)$ for $i = 1, 2, \ldots, T_R$ (where T_R is the time series length) obtained from the dominant modes of Q and spatially averaged P (*avgP*) and T (*avgT*) are used. Spatial averaging for P and T time series are performed over the contributing area to Q observation location. In our interpretation, temperature reconstruction is a proxy for summer evapotranspiration and winter snow accumulation and melt. The $\widehat{P}-\widehat{T}-\widehat{Q}$ time implicit trajectory plot represents a qualitative measure of the local dynamics of the water and energy balance. The atmospheric, physiographic, and anthropogenic factors that may lead to changes in the $\widehat{P}-\widehat{T}-\widehat{Q}$ phase plane are also discussed. By comparison of the phase plane plot for different periods in the historical record, we are able to get some indication of how the local system evolves in time.

Study area

The Colorado River Basin extends from Rocky Mountains of Colorado in the north and Wyoming to the Gulf of California in Mexico in south, covering parts of seven states: Arizona, Colorado, California, Nevada, New Mexico, Utah, and Wyoming. The aerial extent of the watershed is 246,000 mile^2 (637,000 km^2). The Colorado River itself is approximately 1250 mile (2330 km) long and has many tributaries, the most significant for this analysis being the Green, the San Juan, and the Gila rivers. The size and relief of the CRB introduces long range and local climatic gradients as well as complex hydrologic, geomorphic, and anthropogenic conditions which control basin storage and runoff. The complex changes in the geotopographic and climatic framework in space and time in CRB is discussed in the following subsections.

Geo-topographic framework

The physiographic regions within the CRB are the Basin and Range on the eastern and southern portion of the CRB, the Colorado Plateau in the central part, the Wyoming Basin in the north, and the Rocky Mountains in the east and Middle Rocky Mountains in the north- and north-central region (Fig. 1). The range of relief in the Basin/Range is from around 24 m–1200 m, in the central plateau from 1200 to 2300 m and the eastern mountains reach a height of 4300 m (Fig. 2). Hydrologically, the Rocky mountain region has highly permeable surficial metamorphic rocks while Basin and Range region surficial bedrock is predominantly low to moderately permeable Ogallala formation. The pattern and scale of the topography and near surface geology of each physiographic region along with the landcover/landuse and climate, influence the spatial scales of change in the terrestrial hydrologic response (Winter, 2001).

Hydro-climatic framework

Hydrology of the CRB is critically dependent on the amount of snow accumulation, and the air mass temperature conditions that control the onset of snowmelt over the basin (Piechota and Dracup, 1996; Cayan, 1996; Stewart et al., 2004; Dettinger et al., 2004). In fact, Gleick and Chaleki (1999) and McCable and Wolock (1999) reported that even a small increase in temperature in snowmelt-dominated basins can cause considerable shifts in timing and volume of runoff. Stewart et al., 2004 observed an earlier onset of melting in the southwest during recent decades. These papers indicate that in recent years the snow pack melts around 1–3 weeks earlier than it did during the early part of the 20th century.

Spatial and temporal variability in precipitation also have significant effect in the amount of runoff. Overall, precipitation in the CRB is bi-seasonal with moisture contributions from frontal systems, tropical cyclones, and the summer monsoon. Frontal systems contribute to the maximum amount of annual moisture and lead to orographic precipitation in the high elevation plateaus and mountains (above 1500 m) during winter and spring. Runoff generated from winter and spring precipitation is the main source for irrigation in summer. Sometimes, warm winter storms that originate in the tropical Pacific Ocean cause rainfall on snow conditions, resulting in early generation of large runoff leading to floods in major rivers. Flooding in Basin/Range is also caused by tropical cyclones but is a relatively rare phenomenon. The summer monsoon originating in the Gulf of Mexico and/or the eastern Pacific Ocean, causes high intensity rainfall leading to flash floods locally (Webb et al., 2004).

Temporally, there has been considerable change in the pattern and amount of precipitation (Cayan et al., 1998) and runoff over the last hundred years. Merideth, 2000 observed that the 20th century was initially wetter than average, followed by a mid-century dry period that was in turn followed by a wetter period at the end of century. Subsequently the early part of 21st century faced persistent drought which is still continuing (Webb et al., 2004). These decadal-scale oscillations (Merideth, 2000) have been suggested to be due to the reversal of the tropical Pacific trade winds and warming of tropical oceans due to the El Nino/Southern Oscillation (ENSO). Mantua and Steven, 2002 have reported interdecadal variations on the order of 10–25 years related to Pacific Decadal Oscillation (PDO). Extended droughts during that period were likely the results of larger scale coordination in sea-surface temperature of the North Atlantic and the Pacific Ocean basins. In general we can say that precipitation or lack of it in the CRB is affected by circulation associated with ocean–atmosphere interactions in the Tropical and North Pacific Ocean basin, with wet or dry years depending on the position of the annual storm track. Significant variability in streamflow was documented by Christensen et al., 2004 with minimum and maximum annual flow being 6.5 and 29.6 billion m^3, respectively, in Colorado River in the last century.

In this analysis we use 50–90 years of monthly average precipitation (cm), temperature (degrees F) and streamflow (m^3/s) data collected at National Weather Service (NWS) cooperative station network (NCDC Climate Online, 2007) and at USGS observation stations (USGS, 2007). Missing data values are estimated using correlation with previous year's data for the same month and from the data observed at adjacent stations for the same months using a modified normal ratio method by Young, 1992. Spatially averaged data (of P and T) for time implicit plots are obtained from PRISM (Daly et al., 1994). Out of the numerous streamflow gauging stations operated by USGS, the stations selected here have the longest contiguous records of Q within a river-reach or a watershed and are also distributed over all the different physiographic regions

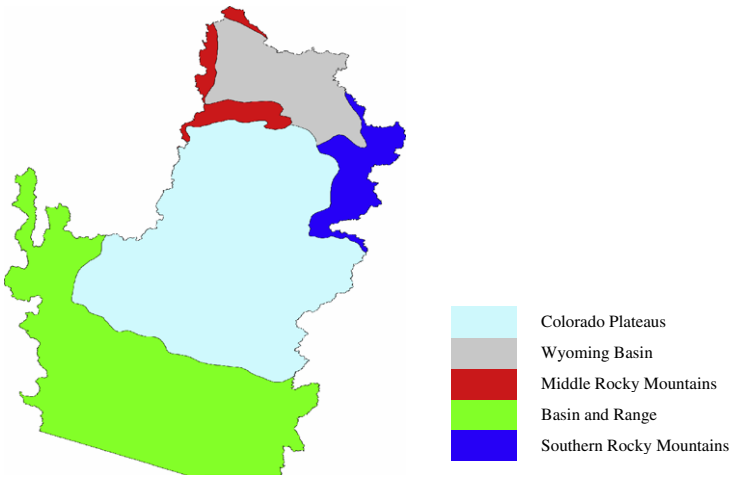

Fig. 1. Physiographic regions in Colorado River Basin (http://tapestry.usgs.gov/).

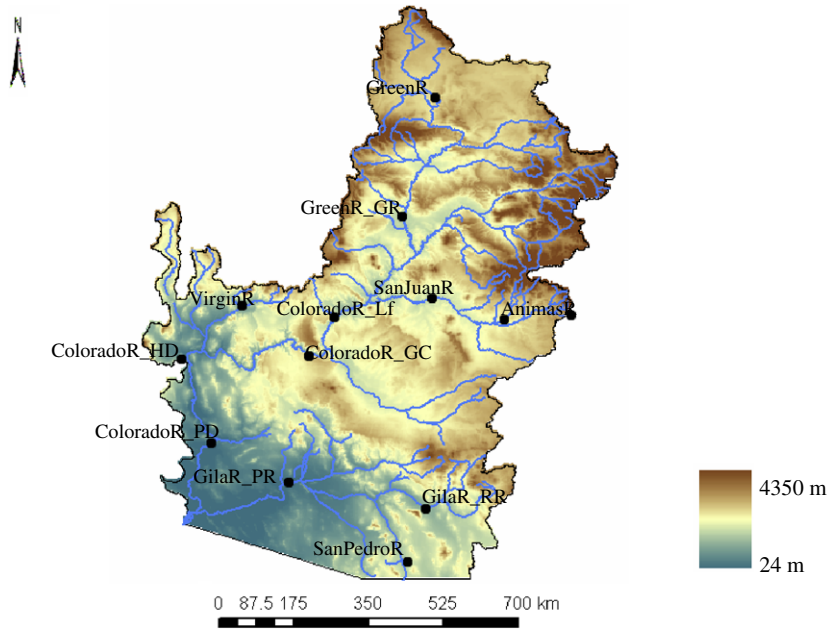

Fig. 2. Topographic map of Colorado River Basin. Black dots represent the streamflow gauging stations with nearby long term concurrent temperature and precipitation data (~50–90 years). Spatio-temporal structure of forcings and response in the basin is obtained by analysis of these point data sets. $\widehat{P}-\widehat{T}-\widehat{Q}$ phase plane analysis uses averaged data for contributing area upland of each streamflow gauging stations. More information about observation stations (shown as black dots) is given in Table 1.

of CRB. Temperature and precipitation stations from different physiographic regions are shown in Table 1 with the NCDC COOP ID (precipitation and temperature station) and the associated USGS Station ID (streamflow station). Length of the record, geographical location and station identifier of each Q station corresponding to which implicit P–T–Q plots are obtained are also listed in the table.

Spatial structure of P–T–Q time series

Spatial analysis of a high dimensional P–T–Q dataset (dimension = $N \times L$, where N is the number of observation stations and L is the length of dataset) is discussed next.

Spatial PCA

Spatial principal component analysis (PCA, also related to the Karhunen–Loève transform, Karhunen, 1947; Loève, 1948) is performed on P, T and Q time series at N observation stations, and the dominant eigenvalues (variance fraction) and corresponding eigenvectors are determined. The analysis involves spectral decomposition of covariance/correlation matrix, S of size $N \times N$ obtained by

$$S = C \cdot C^T \tag{1}$$

where $C = (X - \mu)$, X is the time series at a observation station, μ is the mean of X. Matrix S in Eq. (1) is symmetric and real and so can be decomposed as

Table 1
Hydroclimatic observation stations used in the assessment of temporal structure of P–T–Q time series. COOP ID and Station ID are for NCDC and USGS data repository for climate and streamflow data, respectively.

Precipitation–temperature station (COOP ID)	Stream flow station (station ID #)	Time duration (temp.)	Time duration (precep.)	Time duration (streamflow)	Time duration (concurrent)	P–T–Q triplet identifier	Lat./long. coordinates
Aztec (Az) [290692]	AnimasR. at Farmington [09364500]	1909/1–2004/11	1910/1–2004/11	1930/1–2002/9	1930/1–2002/9	AnimasR	36.75/−108.2
Bluff (Bl) [420778]	San Juan R. Nr. Bluff [09379500]	1911/1–2004/11	1911/6–2004/11	1914/10–2003/9	1914/10–2003/9	SanJuanR	37.12/−109.84
Boulder City (BC) [261071]	ColoradoR. below Hoover Dam [09421500]	1931/1–2004/11	1931/9–1975/12	1934/4–2003/9	1934/4–1975/12	ColoradoR_HD	36.0/−114.75
Buckeye (B) [21026]	Gila R. at Painted Rock Dam [09519800]	1900/1–2003/12	1900/1–1996/4	1959/10–2003/9	1959/10–1996/4	GilaR_PR	33.07/−113
Grand Canyon National Pk. (GcNP) [23596]	ColoradoR. Nr. Grand Canyon [09402500]	1903/1–2004/11	1903/1–2004/11	1922/10–2003/9	1922/10–2003/9	ColoradoR_GC	36.1/−112.09
Green River Aviation (GRA)[423418]	Green R. at Green River [09315000]	1907/12–2004/11	1900/1–2004/11	1905/3–2003/9	1907/12–2003/9	GreenR_GR	38.95/−110.15
Green River (GR) [484065]	Green R. Nr. Green River [09217000]	1900/1–2004/11	1904/12–1995/7	1951/10–2003/9	1951/10–1995/7	GreenR	41.5/−109.35
Lees ferry (Lf) [24849]	ColoradoR. at Lees Ferry [09380000]	1928/1–2004/11	1943/6–2004/11	1921/10–2003/9	1943/6–2003/9	ColoradoR_Lf	36.9/−111.5
Parker (Pk) [26250]	ColoradoR. below Parker Dam [09427520]	1900/1–2004/11	1900/1–2004/11	1934/11–2003/9	1934/11–2003/9	ColorradoR_PD	34.3/−114.05
Safford Agriculture Center (SAC) [27390]	Gila R. at Red Rock [09431500]	1900/1–2004/11	1903/9–2004/11	1962/10–2002/9	1962/10–2002/9	GilaR_RR	32.72/−108.67
St. George (StG) [427516]	Virgin R. at Littlefield [09415000]	1900/1–2004/11	1900/1–2004/11	1929/10–2003/9	1929/10–2003/9	VirginR	36.9/−113.9
Tombstone (T) [28619]	San Pedro R. at Charleston [09471000]	1900/1–2004/11	1900/1–2004/11	1935/5–2003/9	1935/5–2003/9	SanPedroR	31.6/−110.16

$$S = E\Lambda E^T \tag{2}$$

where columns of the diagonalizing matrix E are orthonormal eigenvectors and the diagonal entries in the matrix Λ are eigenvalues of S. Principal components vector A, is obtained from the eigenvectors by

$$A = EX \tag{3}$$

Eigenvalues obtained using Eq. (2) represents the variances of corresponding principal components. A reduced dimension, D ($D < N$) is determined which corresponds to the principal components (PCs) contributing most of the information or variance to the original data. Factor loadings or Correlation coefficients of time series (corresponding to station i) are calculated with PC j ($j \leqslant D$) by

$$\rho_{X_i A_j} = \frac{\lambda_j e_{ji}}{\sqrt{\sigma_{ii}}\sqrt{\lambda_j}} = \frac{e_{ji}\sqrt{\lambda_j}}{\sqrt{\sigma_{ii}}} \tag{4}$$

where σ_i is the ith column of Λ in Eq. (2), e_j is the jth column of matrix E. Factor loading quantifies how time series at each observation station loads or is associated with each principal component. Essentially a factor loading calculation based on PCA reduces the data dimension from $N \times L$ to $N \times D$ where $D \ll L$. We note that all the D principal components are mutually independent and orthogonal to each other and so any observation station can be identified by its $D \times$ one-dimensional coordinate vector of factor loadings in an orthogonal D-dimensional space. Pattern classification of the factor loadings is then used to cluster observation stations into groups that can ideally allow interpretation for climatology, geology, topography and landuse of the region.

Classification of P–T–Q stations

Spatial classification of the respective time series is performed using k-means clustering methodology (MacQueen, 1967). The methodology has inherent advantages over other clustering algorithms (Mantegna, 1999) in terms of its computational efficiency and modification of false classifications over subsequent iterations thus making it relatively robust.

The k-means clustering algorithm is implemented in a sequence of the following five steps: (a) identify the value of k *a priori* where k is the number of clusters (b) initialize k cluster centers randomly (c) determine the class membership of each station by assigning them to nearest cluster center (d) when all the stations have been assigned, recalculate k cluster centers (e) if none of the stations changed membership in last iteration, the implementation stops. Else go to step (c). Mathematically, the algorithm aims to minimize the cost function given by

$$\text{Mean square error (MSE)} = \sum_{i=1}^{N}\sum_{j=1, f_i \in C_j}^{k} |f_i - m_j|^2 \tag{5}$$

where C_j is class j, f_i is factor loading coordinate vector of station i and m_j is the mean coordinate vector of class j. Here the number of classes is obtained in an automated way by minimizing the ratio of intra-cluster to inter-cluster distance as proposed by Ray and Turi (2001).

Spatial PCA (i.e. PCA applied to time series distributed in space) on the historical data series followed by *k-means* clustering of factor loading (obtained from PCA) produces three separate classification sets, one each for precipitation, temperature and streamflow time series.

Change detection using P–T–Q clusters

Classification was first performed on the precipitation data set. Twenty-one precipitation stations (listed in Table 2) cluster into five distinct classes (see Fig. 3a). Saint George, Boulder City, Yuma Citrus, Parker, Grand Canyon National Park and Buckeye, all located in the south-western part of CRB, form the first cluster. The region is generally characterized as Basin and Range south of the Colorado plateau. The higher elevation plateau region to the north appears to serve as an effective orographic barrier creating a distinct precipitation regime. In the south-east, Luna RS, Tucson, Tombstone and

Table 2
Classification of temperature/precipitation (in Table 2a) and streamflow (in Table 2b) stations based on clustering of time series in derived factor loading based low dimensional space. The five obtained classes viz. R (red), B (black), P (purple), G (green) and Bl (blue) are also plotted in Figs. 3a–3c, with each class shown in the same color as their name. Only the stations in bold are used in concurrent $\widehat{P}-\widehat{T}-\widehat{Q}$ phase plane analysis.

Id	Station name (P and T) [Coop ID]	Class (P)	Class (T)	Physiography
1	**Buckeye** (B) [21026]	R	R	B&R
2	Canyon-de-chelly [21248]	B	B	CP
3	**Grand Canyon National Pk.** (GcNP) [23596]	R	R	CP
4	Holbrook [24089]	B	B	CP
5	**Lees ferry** (Lf) [24849]	B	B	CP
6	**Parker** (P) [26250]	R	R	B&R
7	Prescott [26796]	P	R	B&R
8	**Safford Agriculture Center** (SAC) [27390]	G	B	B&R
9	**Tombstone** (T) [28619]	G	R	B&R
10	Tucson [28815]	G	R	B&R
11	Yuma Citrus [29652]	R	R	B&R
12	**Aztec** (Az) [290692]	B	B	CP
13	Luna RS [295273]	G	R	B&R
14	**Boulder City** (BC) [261071]	R	R	B&R
15	**Bluff** (Bl) [420778]	B	G	CP
16	Fort Duschene [422996]	Bl	G	CP
17	**Green River Aviation** (GRA) [423418]	Bl	G	CP
18	Loa	B	B	CP
19	Moab	Bl	G	CP
20	**St. George** (StG) [427516]	R	R	B&R
21	**Green River** (GR) [484065]	Bl	G	WB/ M

Id	Station name (Streamflow) [USGS stn. ID]	Class (Q)	Physiography
1	**GilaR. at Painted Rock Dam** [09519800]	R	B&R
2	San Pedro R. Nr. Redington [09472000]	R	B&R
3	**ColoradoR. Nr. Grand Canyon** [09402500]	Bl	CP
4	GilaR. Below Coolidge Dam [09469500]	R	B&R
5	**ColoradoR. at Lees Ferry** [09380000]	Bl	CP
6	**ColoradoR. below Parker Dam** [09427520]	G	B&R
7	San Juan R. at Shiprock [09368000]	B	CP
8	**Gila R. at Red Rock** [09431500]	R	B&R
9	**San Pedro R. at Charleston** [09471000]	R	B&R
10	ColoradoR. Nr. Cameo [09095500]	B	CP
11	ColoradoR below Baker Gulch [09010500]	B	SRM
12	**AnimasR. at Farmington** [09634500]	B	CP
13	**Los Pinos R. at La Boca** [09354500]	B	CP and SRM
14	Yampa R. Nr. Maybell [09251000]	B	WB/CP
15	**ColoradoR. below Hoover Dam** [09421500]	G	B&R
16	**SanJuanR. Nr. Bluff** [09379500]	B	CP
17	ColoradoR. Nr. Cisco [09180500]	B	CP
18	**GreenR. at Green River** [09315000]	B	CP
19	**VirginR. at Littlefield** [09415000]	R	B&R
20	**GreenR. Nr. Green River**[09217000]	B	WB/ M

Safford Agriculture center, lying on the Basin/Range and Colorado Plateau transition are found to fall in the same cluster, although the south-western stations cluster separately from south-eastern stations. This partitioning seems to be indicative of the importance of atmospheric moisture sources in terms of their timing and magnitude, in addition to local terrain effects. The south-western part of Basin/Range gets precipitation from summer monsoons and tropical cyclones that originate from the Gulf of Mexico that spawn high intensity thunderstorms. The south-eastern region gets its share of moisture from the hurricanes and the tropical depressions of the northern pacific (Webb et al., 2004). These hurricanes are rare and only make the landfall occasionally. The next cluster group is for stations belonging to the central Colorado plateau. In general this region is also referred to as the high plateau region, with dominant orographic influence on precipitation patterns, with snow dominating in winter. As in the first cluster, there can be within-cluster effects due to atmospheric sources and the storm direction. The third significant cluster is the northern Colorado plateau. Generally these stations are at higher elevation than the first cluster, but with a large range in altitude compared to the second.

Fig. 3b represents the factor loading plot of streamflow time series data. Observation stations in the Basin/Range provinces of Gila and San Pedro River, respectively, fall in the same group. These rivers are characterized by low streamflow conditions with

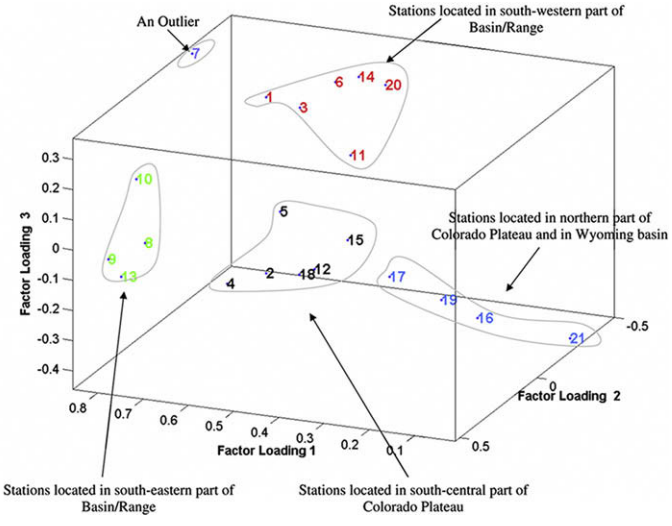

Fig. 3a. Stations IDs (see Table 2 for detail) corresponding to precipitation time series at respective observation stations, plotted in three-dimensional factor loading space. Twenty one stations are classified into five clusters using *k-means* classification of factor loadings. Note that stations with different precipitation sources are classified into different groups. Envelopes around the groups drawn here are merely schematic.

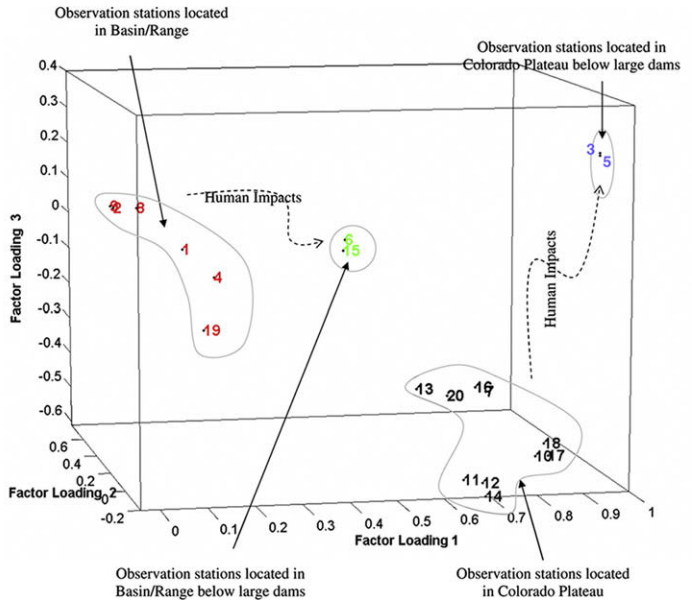

Fig. 3b. Stations IDs (see Table 2 for detail) corresponding to streamflow time series at respective observation stations, plotted in three-dimensional factor loading space. Twenty stations are classified into four clusters using *k-means* classification of factor loadings. Note that stations in different physiographic regions group together unless there is a significant anthropogenic impact due to dams. Envelopes around the groups drawn here are merely schematic.

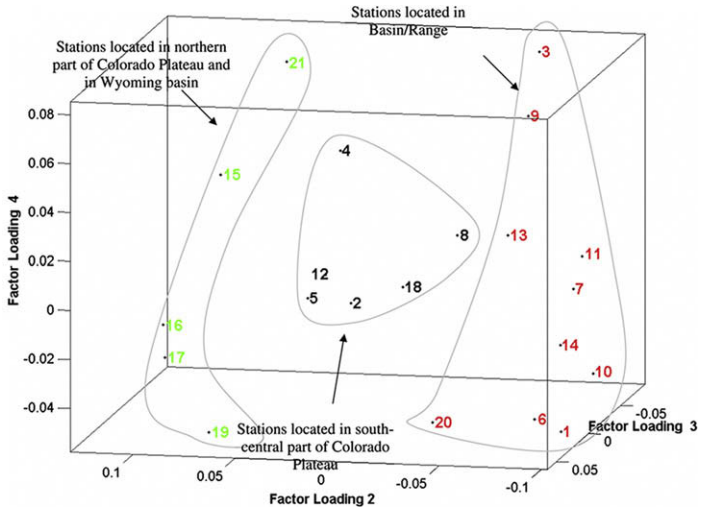

Fig. 3c. Stations IDs (see Table 2 for detail) corresponding to temperature time series at respective observation stations, plotted in three-dimensional factor loading space. Twenty stations are classified into three clusters using *k-means* classification of factor loadings. Note that stations in different physiographic regions group together. Envelopes around the groups drawn here are merely schematic.

intermittent large flash floods from monsoon and tropical cyclonic thunderstorms. Similarly observation stations located in the upper Colorado and San Juan River along with its tributaries fall in a unique group. Notably, anthropogenic effects dramatically change the stream characteristics in the basin. Stations in Basin/Range and Colorado Plateau which lie below large dams, stand out and form separate groups.

In the case of temperature time series, the first principal component carried ~98% of the variance. Because of the dominant annual variation in temperature at all locations, factor loading of temperature time series to first principal component is found to be quite high (close to 0.97). In order to account for other low-frequency modes, we performed classification of the temperature stations based on the next three largest factor loadings (i.e. not including the factor loading that is largely influenced by the annual cycle) which are shown in Fig. 3c. Three primary groups were obtained (Table 2) distinguished by the relative difference in relief among the Basin/Range, and the northern and southern Colorado plateau.

Temporal structure of P–T–Q time series

The goal here is to estimate the dominant low-frequency periodic, nearly periodic, or oscillatory variance components in the P–T–Q records, and to examine these components with respect to physiographic and hydrologic conditions across the CRB. First, spectral decomposition of the time series using Singular Spectral Analysis (SSA) is performed to reveal the oscillatory modes inherent in the time series. In the second step, dominant oscillations (spectral peaks) are recovered from the original series and the eigenspectrum of the remaining stochastic components or noise spectrum is estimated and examined.

Singular spectrum analysis (SSA)

SSA is closely related to Empirical Orthogonal Function (EOF) analysis or principal component analysis (Elsner and Tsonis, 1996) and has been used to identify low-frequency components in the data (Lall and Mann, 1995; Vautard et al., 1992). It is based on the idea of sliding a window over a particular time series and looking for patterns which account for a high proportion of the variance. SSA is performed on a lagged cross-covariance matrix which is computed by the product of time series trajectory matrix and its transpose. The basis functions are data adaptive, empirical and orthogonal. The method can be shown to be optimal in the sense of capturing the maximum variance with fewest independent components. Defining a times series vector X of length N, a Toeplitz matrix S that captures the *lagged* cross-covariance structure of the data series can be calculated as (Elsner and Tsonis, 1996)

$$S_{i,j} = \frac{1}{L_t - |i-j|} \sum_{t=1}^{L_t-|i-j|} x_{|i-j|+t} x_t \tag{6}$$

where $1 \leqslant (i,j) \leqslant M$, M being the window length and L_t is the length of time series. Matrix S calculated in Eq. (6) is similar to the one calculated in Eq. (1) except that S here is a *lagged* cross-covariance matrix (or Toeplitz matrix) obtained using only one time series while the later is a cross-covariance matrix between time series at different locations. Similar to Eq. (2), singular value decomposition is again performed on the matrix S to obtain eigenvalues and the corresponding eigenvectors. Each column in the eigenvector matrix represents the amplitude of oscillations at different frequencies. Since the data time series length varies from station to station, the window length considered while performing SSA is chosen to preserve the climate modes with time scales ranging from annual to decadal. We note that variations in the window length in a sufficiently large range have the effect of stretching or compressing the spectrum of eigenvalues only, leaving their relative magnitudes unchanged (Yiou et al., 1994). In the present analysis, window length is chosen such that interannual climate variability due to reversal of the tropical pacific trade winds (of typical time scale ranging from 2 to 6 years) and one related to cycles in Monsoon moisture flow (of typical time scale ranging from 6 to 10 years) can be captured. Interdecadal variations on the order of 10–25 years which are related to PDO are also captured in cases where the data length is sufficiently long. A rule of thumb for selection of window length is – it should be less than one-third of the time series length (Shun and Duffy, 1998).

Eigenvectors from SSA represent the basis of the time series while eigenvalues quantify the variance carried by each mode. The estimated eigenvalues from SSA are plotted against the estimated frequency of each mode, and the resulting eigenspectrum plot is used to separate signal (e.g. dominant oscillatory modes or spectral peaks) from the noise.

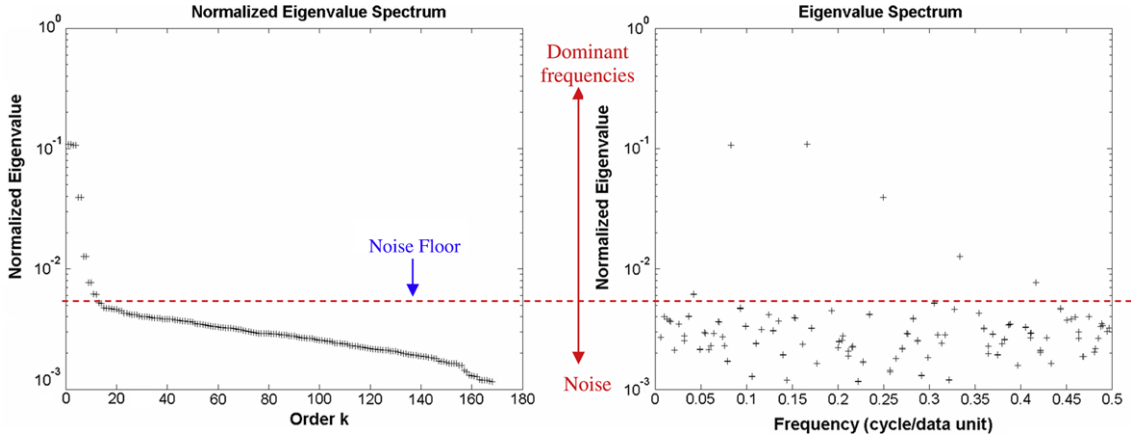

Fig. 4a. Eigenspectrum of precipitation time series at Tombstone. In the case of white noise, eigenvalues lying above the noise floor can be identified with relative ease using either rank–order or frequency ordered plots.

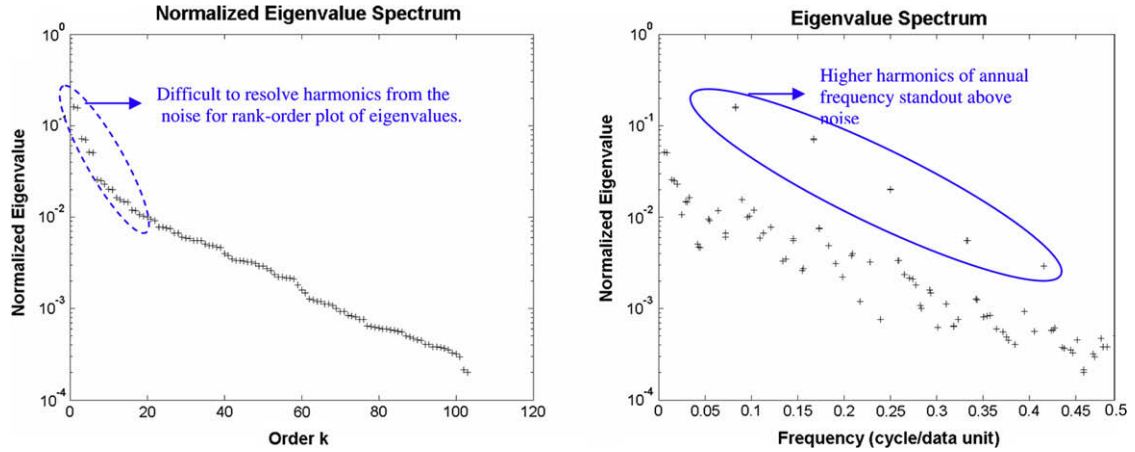

Fig. 4b. Eigenspectrum of discharge time series at Lees Ferry. For signals with a red noise eigenspectrum, re-ordering of the eigenvalues by frequency helps to identifying harmonics of the annual cycle.

Eigenspectra and dominant oscillatory modes

The Eigenspectrum is a rank–order plot of eigenvalues λ_i in descending order of magnitude. Although each eigenvalue represents the fraction of variance corresponding to a particular eigenmode, in the presence of correlated noise rank–order plotting of the eigenvalues may cause statistically significant modes to be ignored because of their plotting position being adjacent to higher frequency eigenvalues of the same magnitude. This problem is resolved by re-ordering of eigenvalues with respect to the dominant frequency of each mode as suggested by Allen and Smith, 1996 (Figs. 4a and 4b). The dominant frequency of each eigenmode is determined by applying a Fast Fourier transform and estimating the dominant frequency of each mode. The eigenspectrum estimator is then presented as a plot of eigenvalues (λ_i) versus frequency (f_i) corresponding to each eigenmode. Although Fourier-covariance analysis can also be used for signal decompositions, it can be shown SSA will represent the original data with the fewest possible independent modes (Priestley, 1981). SSA can also provide a statistical dimension or a first estimate of the degrees of freedom required to describe the dynamical system represented by the record (Elsner and Tsonis, 1996) which is roughly the number of eigenvalues above the noise floor (Vautard and Ghil, 1989). Adjacent eigenvalues with nearly the same magnitude in an eigenspectrum plot and with their respective eigenvectors being in quadrature are assumed to be an oscillation pair (periodic or nearly periodic). Plotted eigenmodes allow visual inspection of the period and the phase of the oscillation pairs. Fig. 5 shows the eigenspectrum plot for P, T and Q at four different stations viz. GreenR (in Rocky Mountains), ColoradoR_GC (in Plateau), SanJuanR (in Plateau) and SanPedroR (in B/R). The "low-pass" streamflow eigenspectrum is fitted with the "red noise" theoretical spectrum (Priestley, 1981) using

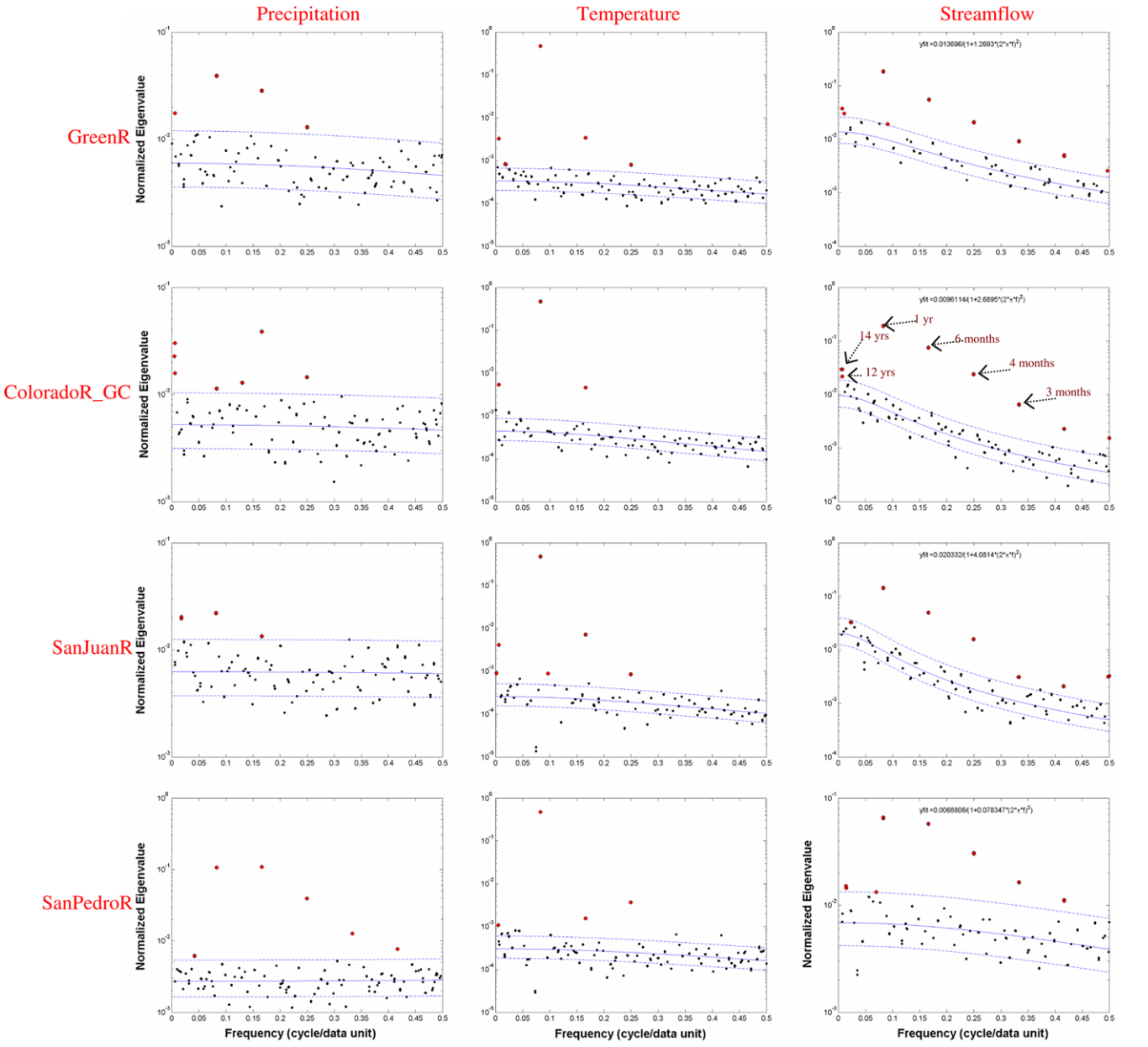

Fig. 5. Eigenspectrum of precipitation, temperature and streamflow data at GreenR, ColoradoR_GC, SanJuanR, SanPedroR stations in the CRB (More information on data at these stations is in Table 1).

$$\phi_r(f) = \frac{A}{B + (2\pi f T_c)^2} \quad (7)$$

where T_c is the characteristic relaxation time for physical phenomena or the threshold frequency for low-pass filtering (Jin and Duffy, 1994). For generating the fit, eigen-pairs are removed till no other pairs can be removed without removing a single eigenvalue and also without violating the chi-square test for the data ($\alpha = 0.025$ and $\beta = 0.975$ level, Priestley, 1981). Dominant frequencies (cycle/month) are shown as circles in Fig. 5. In Fig. 5, a frequency of 0.166 (=6/12) corresponds to semi-annual oscillation, 0.0833 (=1/12) corresponds to an annual oscillation, 0.00833 (=1/120) to decadal oscillation and so on.

Change detection using P–T–Q eigenmodes

Table 3 lists the dominant time periods of P, T and Q at 12 different stations which have long concurrent records (listed in Table 1). The dominant low-frequency oscillation for precipitation for stations in plateau region is the annual cycle, while the semi-annual dominates in the Basin/Range (see Table 3 and Fig. 5). The total variance explained by all low-frequency modes is on the order of 15–20% of the total variance of the time series with annual, interannual and semi-annual components contributing around 4–6% each. As we move northwards from the southern Basin and Range in Arizona towards the Colorado Plateau, the dominant mode shifts from semi-annual to annual: Tombstone (semi-annual) → Buckeye (semi-annual) → Parker (semi-annual & annual) → Boulder City (annual) → Lees Ferry (annual). This change in precipitation pattern from south to north was also corroborated by Higgins et al. (1997). At a higher latitude and elevation, annual snowfall in winter is the major precipitation source; while the bi-seasonal summer monsoon and winter precipitation dominates the Basin/Range. For Basin/Range precipitation is mainly in the form of rainfall, with distinct summer and winter maxima. Winter precipitation comes from incursions of Pacific air, while summer precipitation is monsoonal. The bi-seasonal precipitation regime in the Basin/Range was also noticed by Hereford et al., 2002. Precipitations at Bluff and Green River also have significant decadal oscillations. On the other hand Tombstone has seasonal (intra-annual) dominant time modes as well. We note that for precipitation the interannual and decadal modes tend to be relatively weak, with most of the spectrum variance being associated with the annual cycle and higher frequency components.

In the case of temperature, as expected the dominant low-frequency mode is the annual oscillation for all stations in the CRB. The annual and other low-frequency modes explain a cumulative variance of ~90%. By examining each mode individually, we see that the annual, semi-annual, interannual and decadal components contribute 60%, 8–16%, 0–4% and 0–14%, respectively. Thus significant variance is found in frequencies other than the annual oscillation. The interannual component (2–6 years) likely relates to the El Nino/Southern Oscillation (ENSO) and decadal oscillations (12–15 year) to the Pacific Decadal Oscillations (PDO) (Hanson et al., 2002).

Finally we perform SSA on the streamflow time series, with particular attention to how relatively weak atmospheric modes observed in P and T are expressed in the streamflow response of the terrestrial system. It is expected that dominant modes in streamflow will reflect both atmospheric forcing and human activities with smoothing, sharpening, amplification, and phase effects of particular modes depending on the time scale (memory), and filtering effect operating in the watershed or river-reach.

Fig. 5 also shows streamflow eigenspectra at four locations located in three of the physiographic regions. The annual and intra-annual harmonics are found to be strengthened in all these cases as witnessed by larger eigenvalue magnitudes relative to that of precipitation and temperature eigenvalues. This is due to the filtering of higher frequency components in streamflow from soil moisture, groundwater storage and even dams and irrigation which act as a low-pass filter. The interannual modes for streamflow remain just above or within the noise floor (confidence interval). The cumulative variance for annual oscillatory harmonics explains approximately 60% of the variance of streamflow, with annual, semi-annual, third-annual (4 months) and fourth-annual (3 months) harmonics contributing ~40%, 16%, 4% and 2%, respectively.

In Fig. 5, the variance of dominant low-frequency components in runoff is proportionally greater than for precipitation and temperature at the same frequencies. For a watershed or a stream reach we expect "storage" to act as a "low-pass" filter. So relatively higher frequency components existing in temperature (and thus evapotranspiration and snowmelt) and precipitation time series are reduced or filtered. We note that runoff at stations with large dams show significant decadal components (e.g. ColoradoR_HD and ColoradoR_PD). It would appear that the eigenspectrum for these records documents the longer term memory associated with storage and diversion on downstream runoff. Also from Fig. 5, the fitted time scale of low-pass noise spectrum (Eq. (7)) is shown to be larger in the plateau region (ColoradoR_GC and SanJuanR) than in the Basin/Range (SanPedroR) streamflow time series. This may be related to geologic conditions, with bedrock controls on groundwater and springflow in the plateau region having a lower filtering time scale, than the alluvial basin region where shallow water table conditions and rapid communication with groundwater is expected. Within the plateau, we also expect that the characteristic

Table 3
Dominant time periods present in temperature, precipitation and streamflow record. Periods are written in decreasing order of corresponding eigenvalue magnitudes. P, B/R and M correspond to Colorado Plateau, Basin/Range and Mountain, respectively.

Station name (P–T–Q triplet ID)	Physiographic region	Dominant oscillations with time periods (in years) for precipitation data	Dominant oscillations with time periods (in years) for temperature data	Dominant oscillations with time periods (in years) for streamflow data	Streamflow (USGS stn. ID)
AnimasR	P	1, 0.5, 2, 12, 4, 27, 0.33, 0.25	1, 0.5, 14, 0.33, 20	1,0,5, 0.33, 0.25, 0.2	09364500
SanJuanR	P	1, 4.5, 0.5	1,0.5, 14, 0.85, 0.33	1, 0.5, 3.6, 0.33, 0.25	09379500
ColoradoR_HD	B/R	0.5, 0.75, 0.33, 1	1, 0.5, 4	10, 14, 1, 0.33, 0.2	09421500
GilaR_PR	B/R	0.5, 1, 0.33, 2, 0.75, 8	1, 0.5, 14, 3.6	Not resolvable	09519800
ColoradoR_GC	P	0.5, 14, 0.33, 0.6, 1	1, 14, 0.5,	1, 0.5, 14, 12, 0.33, 0.2	09402500
GreenR_GR	P	1, 0.25, 2	1, 0.5, 0.33, 14	1, 0.5, 12, 0.33, 0.25, 0.2	09315000
GreenR	M	1, 0.5, 14, 0.33	1, 0.5, 14, 4, 0.33	1, 0.5, 12, 0.33, 0.25, 0.2	09217000
ColoradoR_Lf	P	1, 0.5, 0.33,	1, 0.5, 0.33,	1, 0.5, 14, 9, 0.33, 0.25	09380000
ColorradoR_PD	B/R	0.5, 1, 14, 0.6, 5	1, 0.5, 14, 4, 8, 0.33	12, 14, 10	09427520
GilaR_RR	B/R	0.5, 1, 0.33, 14, 5	1, 0.5, 0.33, 14	1, 0.5, 0.25	09431500
VirginR	B/R	1, 0.5, 0.2, 14	1, 0.5, 12, 0.33	1, 20, 0.5	09415000
SanPedroR	B/R	0.5, 1, 0.33, 0.25, 0.2, 2	1, 0.33, 0.5, 15	1, 0.5, 0.33, 0.25, 6, 1.25	09471000

time scale of the noise floor to be large where large dams increase the retention time and overall time scale of storage (Vorosmarty and Sahagian, 2000) while inducing a large bank exchange with groundwater.

Extracting the hydroclimatic signal

The process of separating the signal or dominant oscillatory modes in a time series from the residual noise, allows 'optimal' reconstruction of the time series. Optimal in this case implies that we wish to reconstruct only those modes which exhibit a statistically significant spectral peak. Hidalgo et al., 2000 discusses optimal reconstruction using a minimum number of significant components, which we adopt here.

Reconstructing $\hat{P}-\hat{T}-\hat{Q}$ *modes*

In this study, optimal reconstruction of the data is performed by using the dominant frequency components (again using the modes above the noise floor as discussed earlier in "Temporal structure of P–T–Q time series") obtained by performing SSA on each series in Table 1. The reconstruction is performed by using (Plaut and Vautard, 1994)

$$\hat{x}_{i+j-1} = \sum_{j=1}^{D} a_i^k e_j^k \tag{8}$$

where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, D$, e_j^k's and a^k's are the kth eigenvector and principal component (of the corresponding matrices as calculated in Eq. (2)). We note that the reconstruction also includes the estimated mean of the record.

Fig. 6 shows the noisy observed data series and the corresponding reconstructed signal using the all dominant eigenmodes for P–T–Q data on Colorado River at Grand Canyon National Park (ColoradoR_GC). Note that the reconstructed precipitation signal carried only a small fraction of the original noisy precipitation series. However, it captures the low-frequency modal rise and fall in the historical record as is witnessed by a wet period in the initial 20th century followed by a mid century drought and thereafter a slight increase in precipitation in the later part of the century. The reconstructed signal nicely captures the gradual and the sharp dramatic shifts in mean magnitude over time as is seen in the reconstruction of streamflow time series in Fig. 6.

The $\hat{P}-\hat{T}-\hat{Q}$ *phase plane*

Next we define a hydroclimatic phase plane in terms of a 3D parametric plot (time implicit) of the trajectories of normalized reconstructed precipitation, temperature and streamflow. We reiterate that the P and T time series used in time implicit plot are low-frequency reconstructions of spatially averaged precipitation (*avgP*) and temperature (*avgT*) over the contributing watershed area, upland of the streamflow gauge station. Spatially averaged

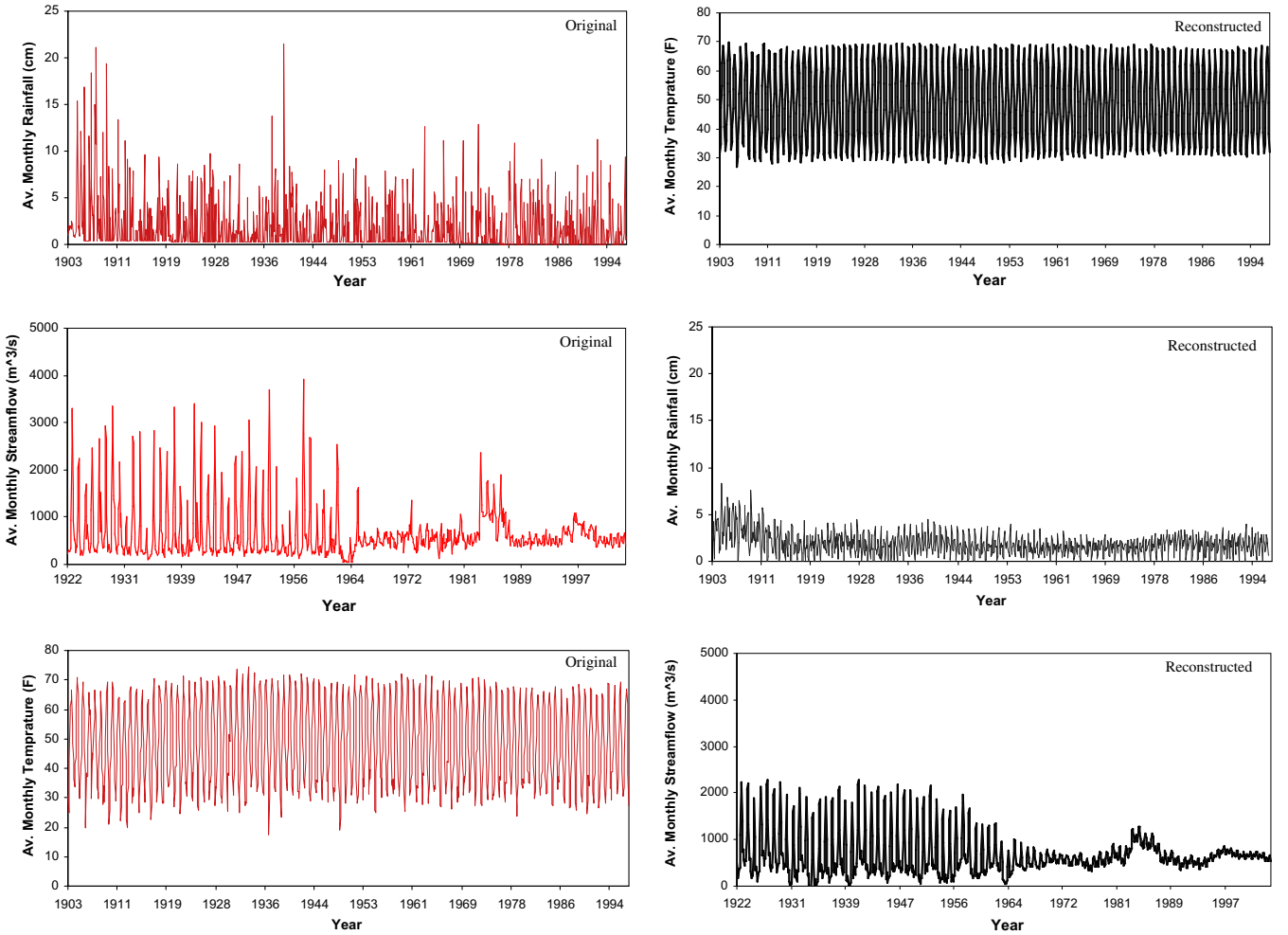

Fig. 6. Original noisy time series and noise-removed reconstructed time series using dominant frequency modes for gauging station at Grand Canyon National Park (ColoradoR_GC).

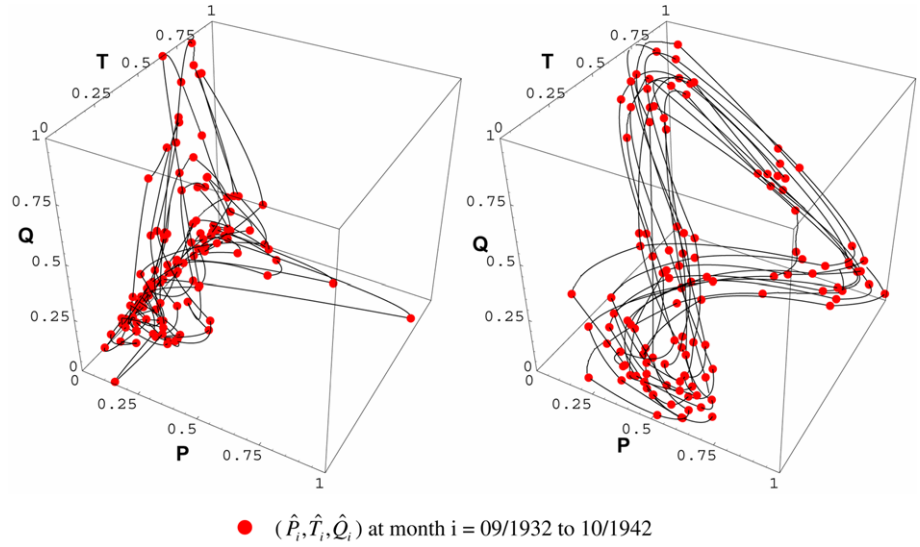

Fig. 7. Phase-plane trajectories for normalized $\hat{P}-\hat{T}-\hat{Q}$ for (a) Original noisy time series (b) Noise removed time series at Grand Canyon National Park for a ten year period of 09/1932 to 10/1942.

Table 4
Time ranges for which $\hat{P}-\hat{T}-\hat{Q}$ trajectory plot is shown in Fig. 8.

Time interval	AnimasR	SanJuanR	ColoradoR_Lf	VirginR	GreenR_GR	SanPedroR
Pre 1960	1/(1943–63)	1/(1943–63)	1/(1943–63)	1/(1943–63)	10/(1951–63)	1/(1943–63)
Post 1976	8/(1977–97)	7/(1975–95)	5/(1976–96)	6/(1976–96)	10/(1976–96)	5/(1976–96)
Post 1997	8/(1997–2002)	8/(1997–2003)	8/(1997–2003)	8/(1997–2003)	8/1997–4/2002	8/(1997–2003)

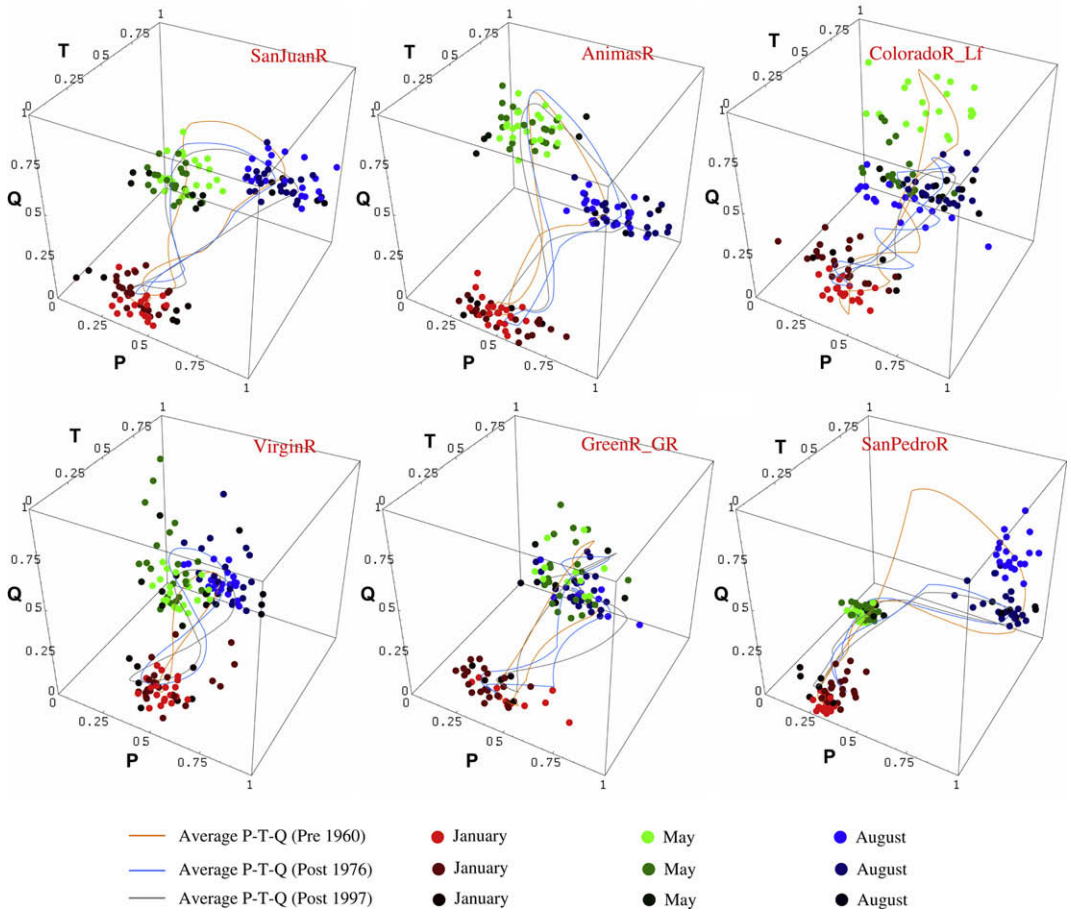

Fig. 8. Time-averaged Normalized $\hat{P}-\hat{T}-\hat{Q}$ trajectory plot of low-frequency reconstructed time records for three time intervals at SanJuanR, AnimasR, ColoradoR_Lf, VirginR, GreenR_GR and SanPedroR. More information about observation stations is in Table 1 and the precise time averaging interval is listed in Table 4.

P and T data for time implicit plots are obtained from PRISM (Daly et al., 1994). The spatial resolution of the data is 2.5′. Clearly this could be done on a local watershed with climate station and runoff records as well, however the goal here is to examine large-scale and low-frequency response and the spatial average data are most appropriate. Normalization of each record is performed on each reconstructed time series by rescaling it by its respective maximum and minimum values using

$$\hat{X}(t) = \frac{\hat{x}(t) - \hat{x}_{\min}}{\hat{x}_{\max} - \hat{x}_{\min}} \tag{9}$$

where $\hat{X}(t)$ is normalized $\hat{P}, \hat{T}$ or $\hat{Q}$'s magnitude at time t, $\hat{x}_{\max}$ and $\hat{x}_{\min}$ are the maximum and minimum of the variable over the concurrent record length (as listed in Table 1 for respective stations) under consideration. The path traced by the vector, Norm $|\hat{P}, \hat{T}, \hat{Q}|$ can be thought of as a qualitative measure of degree of interaction between modes which exist in the atmosphere and land. Note that we interpret temperature as a proxy for evapotranspiration rate in the warm season, and snowmelt in the cold season. Since streamflow Q contains the response from the terrestrial system, Norm $|\hat{P}, \hat{T}, \hat{Q}|$ provides a measure of dynamic behavior of the atmosphere-terrestrial hydroclimatic system. Changes along either the P or T axes during periods of large changes in streamflow indicate the influence of strong atmospheric forcing. Anthropogenic impacts are marked by significant changes along the Q axis. By averaging the pattern of changes over different parts of the historical record we also get some idea of the effect of longer term change from climate and human activity on watershed dynamics.

Fig. 7 shows a $\hat{P}$–$\hat{T}$–$\hat{Q}$ plot for a 10 year reconstruction for the normalized unfiltered P–T–Q time series, and the normalized and filtered (high frequency noise removed) $\hat{P}$–$\hat{T}$–$\hat{Q}$ time series near Grand Canyon on Colorado River. Note that the filtered $\hat{P}$–$\hat{T}$–$\hat{Q}$ phase plane plot is relatively smooth representing the lower frequency oscillators in the data and serves as a tool for detecting change in the dynamic hydrologic response across the CRB. A reconstructed $\hat{P}$–$\hat{T}$–$\hat{Q}$ time series set used for each phase plane plot are shown in Fig. 6.

Change detection using $\hat{P}$–$\hat{T}$–$\hat{Q}$ phase plane

Next we study the hydroclimatic change over the historical records at locations identified in Fig. 2. To account for secular change, the reconstructed time series is partitioned into three periods: the mid 20th century (1943–1963) drought, mid-to-late 20th century (1976–1997) relatively wet period, and the most recent drought period (1997–2004). Noise-removed reconstructed low-frequency

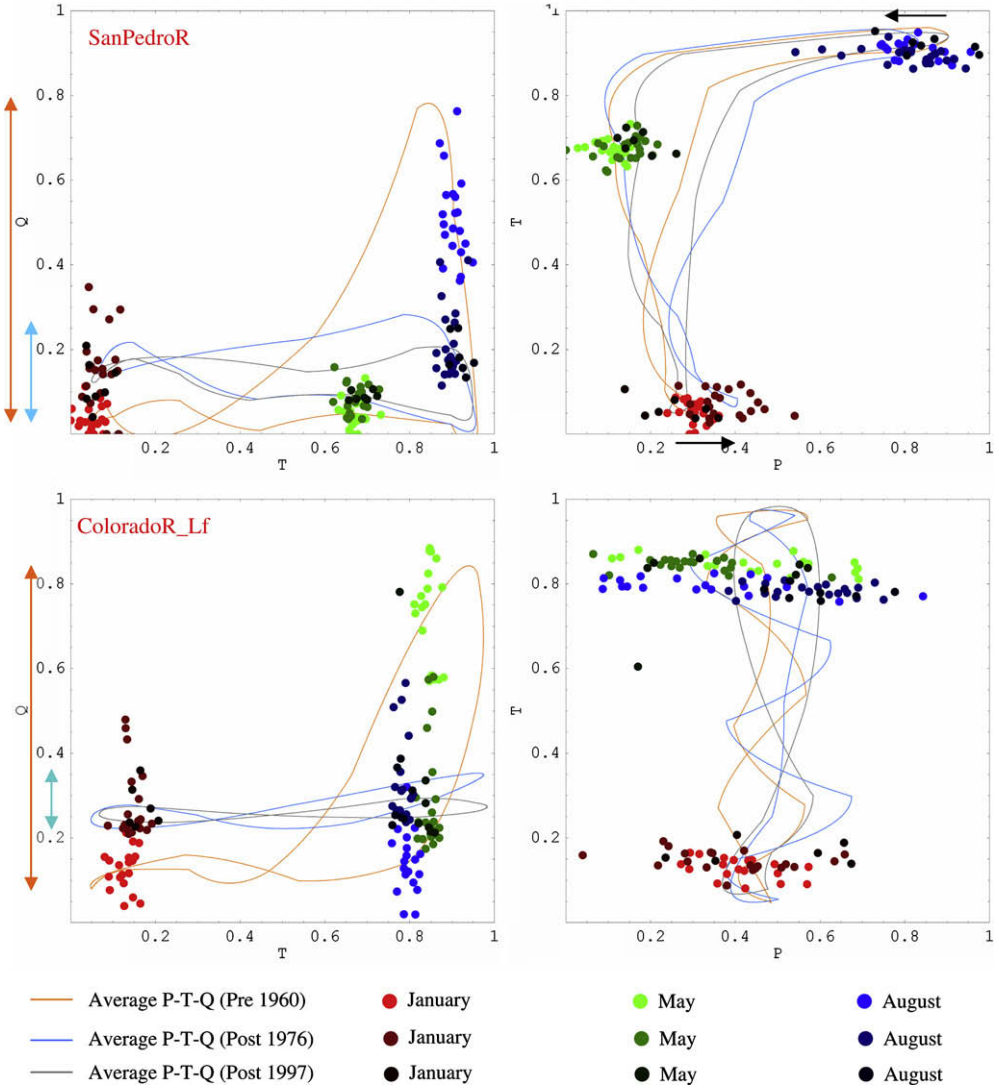

Fig. 9. Q–T and P–T plots at SanPedroR and ColoradoR_Lf. Note the contraction of the phase space (highlighted by vertical lines along Q-axis) in Q–T plane in the later part of century. Also, a distinct increase in winter precipitation and decrease in summer precipitation (highlighted by horizontal arrows along P-axis) is observed at SanPedroR.

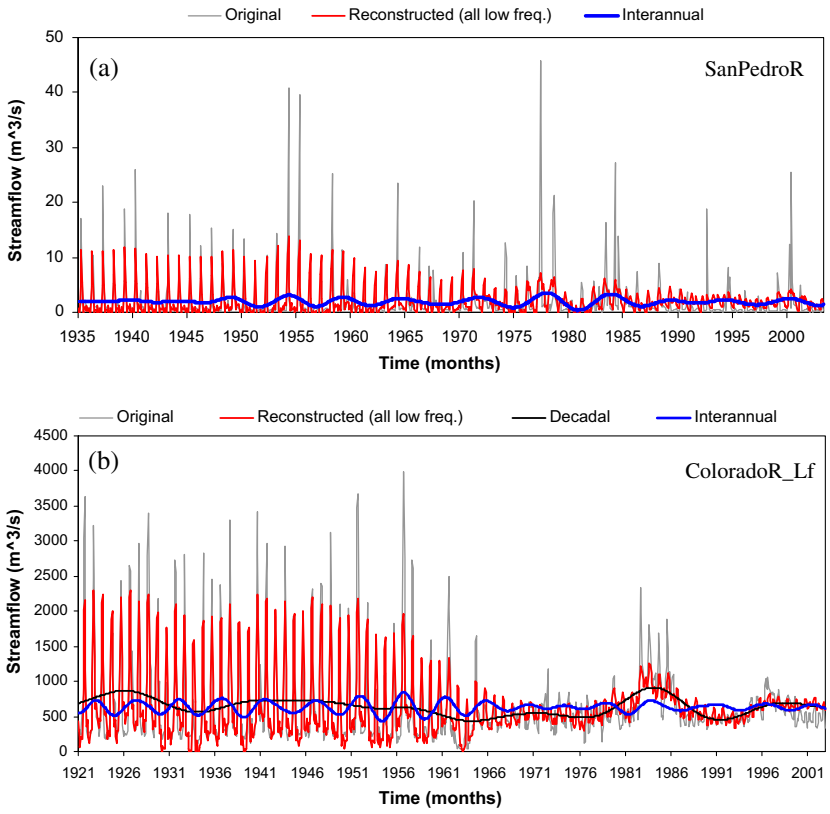

Fig. 10. Streamflow time series (a) on San Pedro River at Tombstone and (b) on Colorado River at Lee Ferry. Note the decrease in the streamflow in later part of the century at San Pedro. The interannual and decadal low-frequency components in both the time series capture the longer time scale changes in streamflow response.

modes for $\hat{P}-\hat{T}-\hat{Q}$ are estimated from monthly averages. Because of the differences in available record lengths, the precise averaging intervals are slightly different for different sets of $\hat{P}-\hat{T}-\hat{Q}$ (see Table 4).

Figs. 8 and 9 show the time averaged $\hat{P}-\hat{T}-\hat{Q}$ as well as planar projections in P–T and Q–T planes for the three time intervals listed in Table 4. Note that trajectories are color coded: 1943–63 (orange), 1976–97 (blue)[1], recent 1997–2004 (grey), indicating the changes in the $\hat{P}-\hat{T}-\hat{Q}$ phase plane pattern over time at each location. Also shown in Figs. 8 and 9 are large dots which correspond to the reconstructed unaveraged-value of P, T and Q data for the months of January, May and August, respectively. These three months are indices for assessing the variation about the mean $\hat{P}-\hat{T}-\hat{Q}$ trajectory, during the winter low-flow (January), early spring runoff (May), and either the late tail of the runoff response or the early runoff from the summer monsoon (August).

We see in Fig. 8 that the minimum streamflow and temperature occur in January at most stations, while maximum annual streamflow is observed in May for all the snowmelt dominated streamflow stations. In the south-east, the San Pedro River at Tombstone peaks in August with significant runoff from the Monsoon. A bi-lobed structure in the $\hat{P}-\hat{T}-\hat{Q}$ plot suggests the presence of important intra-annual frequencies. Multi-lobe plots generally reflect harmonics of the annual cycle in $\hat{P}-\hat{T}-\hat{Q}$ or other higher frequency components.

We note that in all the watersheds the basic pattern (shape) of the low-frequency $\hat{P}-\hat{T}-\hat{Q}$ trajectory remains the same for the three averaging periods for all the stations in Figs. 8 and 9, though the magnitude and range of the trajectories is significantly different. The consistency in shapes (bimodal or multimodal dumbells) of the trajectory plots, for all periods, is maintained because of the existence of same oscillatory modes. Though we observe that atleast at two locations (ColoradoR_Lf and SanPedroR) the phase volumes have shrunk over the century due to changes in the amplitude and/or phase of dominant oscillations. At Lee Ferry (ColoradoR_Lf) the $\hat{P}-\hat{T}-\hat{Q}$ phase volume changed due to the construction of a major dam. Large dams introduce low-frequency modes and trends in the downstream streamflow. The contracted phase volume in Colorado River at Lee Ferry for the period ranging from 1977 to 1997 is indicative of the reduction in the extremes (variance) of seasonal to interannual modes of streamflow below the dam compared to the period before 1960. This is illustrated by the low-frequency reconstructed time series in Fig. 10. The suppression of variability in later part of century is easily discernible in the interannual reconstructions. Notably, for the period 1997–2004, there is further contraction in the phase volume which is likely an indicator of the recent drought (Webb et al., 2004). Contraction of the phase volume is also observed in the San Pedro River at Tombstone. The streamflow plot in Fig. 10 illustrates that the flow magnitude has been reduced significantly since early 1970's. Low-frequency reconstruction based on interannual modes highlight the extended low streamflow period from 1976 to 1984. This might be the result of decreasing precipitation, but the more likely reason is increase in ground-water pumping that is lowering the water table and baseflow, and possibly because of the changes in landcover and riparian vegetation that have occurred (Thomas and Pool, 2006). Low-frequency reconstructions exhibit a gradual decrease in peak streamflow as well. The explanation of increased pumping and aquifer interaction has been suggested by several authors (Corell et al., 1996; Vionnet and Maddock, 1992). Besides decreasing baseflow, pumping also lowers the water table height over time which can induce bank storage. It is interesting to note

[1] For interpretation of color in Figs. 1–11, the reader is referred to the web version of this article.

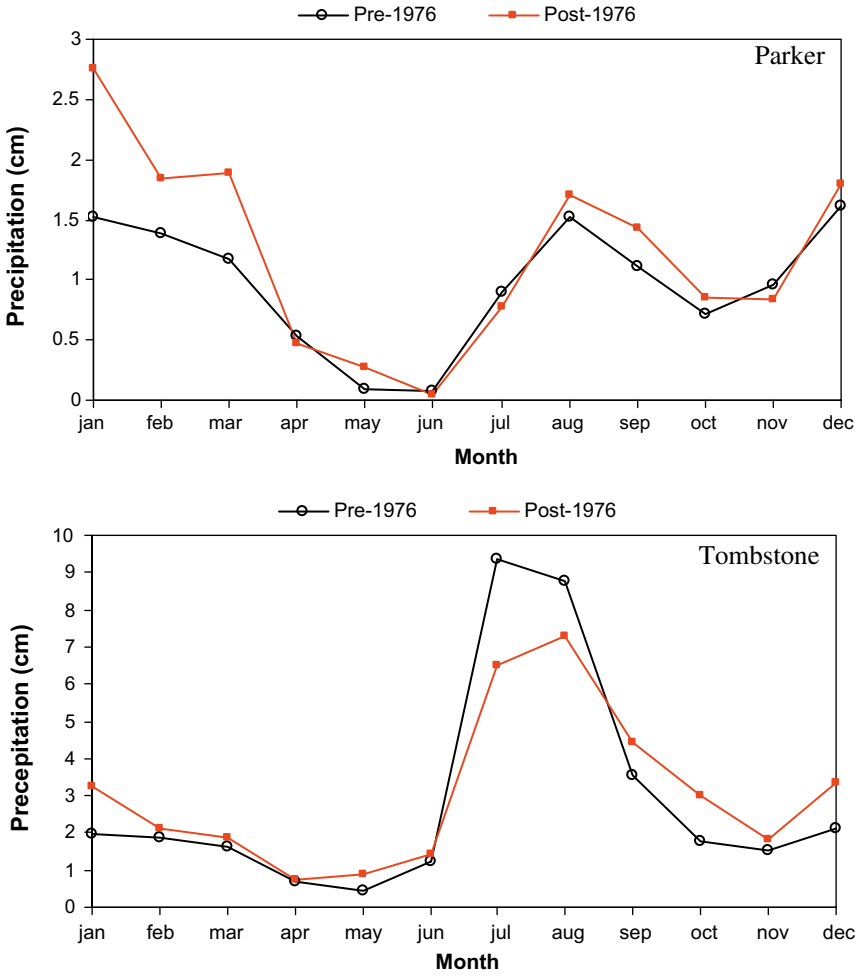

Fig. 11. Change in summer and winter precipitation pattern at Parker and Tombstone. Note the increase in winter precipitation at both the stations after 1976 and decrease in summer precipitation at Tombstone. The change in summer precipitation (June–August) at Parker is negligible.

that the winter streamflow magnitude has increased which can be attributed to the increase in winter precipitation and/or deep recharge under irrigated agricultural areas.

Figs. 8 and 9 show that the stations in basin and range (e.g. SanPedroR), show a distinct increase in winter precipitation and a slight decrease in summer precipitation (Webb and Betancourt, 1990). We see that winter precipitation after 1976 has increased while the summer precipitation has decreased possibly due to negative Southern Oscillation phase (Swetnam and Betancourt, 1997). Average monthly precipitation plot (Fig. 11) for a 20 year period at both locations supports the observation from the phase plane plot analysis. The minimum streamflow magnitude (January) is also found to have increased in the later part of the 20th century, while the maximum streamflow magnitude has decreased. This can be related to factors such as the summer–winter precipitation reversal in the later part of century (Swetnam and Betancourt, 1997), earlier snowmelt (Stewart et al., 2004) and possible irrigation effects that have altered the evapotranspiration/recharge/baseflow balance over this time frame. The snowmelt dominated portion of the basin shows a reduced May runoff peak.

Conclusions

The spatial pattern of precipitation in the CRB seems to be more complex than a simple north–south variation (Dettinger et al., 1998), which also happens to be the alignment of major landforms or physiographic regions in the basin. Classification of factor loadings from principal component analysis reveals regional patterns in precipitation including variations due to: latitude, physiographic features, and the sources of precipitation. Stations belonging to the south-eastern Basin/Range such as Tucson and Tombstone fall in a cluster different from the stations from the south-western Colorado plateau such as Parker and Buckeye. Notably the sources of precipitation, and hence precipitation intensity and timing, for the two clusters are also different with one being dominated by monsoonal precipitation while the other is affected by hurricanes that originate as tropical depressions in the northern pacific (Webb et al., 2004). Stations located in the Colorado Plateau cluster in two groups – one in south-central plateau region that includes Lees Ferry and Holbrooke and the other in northern plateau region with stations like Green River and Moab – reflecting a generic orographic precipitation trend produced as a result of high altitude frontal systems. Classification of loading factors for air temperature, which is a proxy for evapotranspiration and snowmelt, is almost entirely explained as a function of altitude.

Classification of streamflow stations indicates that local physiography, the hydrogeology, and anthropogenic factors may transform atmospheric forcing into unique clusters. A Basin and Range cluster of streamflow stations is identified which includes the Gila at Painted Rock, and the San Pedro at Reddington and is characterized by low streamflow conditions with intermittent large flash floods. A second cluster is found for plateau streamflow sites including the San Juan at Bluff, and the Green River at Green River Wyoming. Streamflow observation stations just below large dams

in each physiographic regions form a distinct cluster. Within the plateau region this cluster includes the Colorado at Lees Ferry and the Colorado at Grand Canyon, while the Basin and Range cluster includes discharge sites below Hoover dam and Parker dam on the Colorado. Clearly anthropogenic impacts of large dams stand out as unique spatial clusters for runoff from the CRB but differences are also found from dams within different physiographic regions.

Singular spectrum analysis further explores the spatial pattern of low-frequency temporal modes associated with each P–T–Q time series by estimating the variance of low-frequency periodic and nearly periodic modes inherent in the time series. This exercise also reinforces and refines the interpretations made through the spatial analysis. The dominant frequency in precipitation time series is observed to gradually change from a semi-annual to annual mode as we move from Tombstone and Buckeye in the south to Boulder city and Lees Ferry in the north. This is because at higher latitude and elevations, annual winter snowfall is the primary source of precipitation while in Basin/Range, precipitation is bi-seasonal. The annual cycle represents the dominant mode in temperature across the basin, however significant interannual modes also exist. Observation stations lying in the same physiographic regions, with significant elevation and latitudinal differences, have similar interannual modes. For streamflow, locations gauges located just below a dam, such as the Hoover dam and Parker dams on the Colorado, introduce a shift in the response to interannual, decadal and longer time scales of runoff. It seems that dams act as a low-pass filter to higher frequency modes in runoff.

The $\widehat{P}-\widehat{T}-\widehat{Q}$ phase plane is used to explore the temporal change in regional hydroclimatology, either due to changes in forcing pattern/magnitude or due to anthropogenic impacts. A reversal in precipitation pattern due to a negative southern oscillation is immediately identified in Basin/Range region including at Lee Ferry and Tombstone (Swetnam and Betancourt, 1997). A dramatic contraction of the phase plane shape (reduction in phase volume) at Lee Ferry on the Colorado and the San Pedro at Tombstone is indicative of anthropogenic impacts in form of dam construction and excessive ground-water pumping, respectively. We recognize the qualitative nature of the change analysis largely because of the sparseness of long term distributed data of all the characteristic hydroclimatic descriptors at same locations. Nonetheless, the $\widehat{P}-\widehat{T}-\widehat{Q}$ phase plane plot analysis along with spatio-temporal classification is shown to be a useful diagnostic tool for detecting and attributing change. Together, spatial principal components, spectral analysis of time series, and the reconstructed phase plane through time, appear to be useful for physical interpretation and detecting hydrologic changes from physiographic, climatic, and human forcing.

References

Allen, M.R., Smith, L.A., 1996. Monte Carlo SSA: detecting irregular oscillations in the presence of colored noise. J. Clim. 9 (12), 3373–3404.

Cayan, D.R., 1996. Interannual climate variability and snow pack in the western United States. J. Clim. 9 (5), 928–948.

Cayan, D.R., Dettinger, M.D., Diaz, H.F., Graham, N.E., 1998. Decadal climate variability of precipitation over western North America. J. Clim. 11 (12), 3148–3166.

Christensen, N.S., Wood, A.W., Voisin, N., Lettenmaier, D.P., Palmer, R.N., 2004. The effect of climate change on the hydrology and water resources of the Colorado River Basin. Climatic Change 62, 337–363.

Corell, S.W., Corkhill, F., Lovvik, D., and Putnam, F., 1996. A groundwater flow model of the Sierra Vista subwatershed of the Upper San Pedro Basin – Southeaster Arizona. Modeling report no. 10, ADWR, 107 pp.

Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteorol. 33, 140–158.

Dettinger, M.D., Diaz, H.F., 2000. Global characteristics of streamflow seasonality and variability. J. Hydromet. 1, 289–310.

Dettinger, M.D., Cayan, D.R., Diaz, H.F., Meko, D., 1998. North–south precipitation patterns in western North America on interannual-to-decadal time scales. J. Clim. 11, 3095–3111.

Dettinger, M.D., Cayan, D.R., Meyer, M.K., Jeton, A.E., 2004. Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American River basins, Sierra Nevada, California, 1900–2099. Climatic Change 62, 283–317.

Elsner, J.B., Tsonis, A.A., 1996. Singular Spectrum Analysis: A New Tool in Time Series Analysis. Plenum, New York.

Gleick, P.H., Adams, D.B., 2000. Water: the potential consequences of climate variability and change for the water resources of the United States. The Report of the Water Sector Team of the National Assessment of the Potential Consequences of Climate Variability and Change for the US Global Research Program. Pacific Institute for studies in Development, Environment, and Security, 151 p.

Gleick, P.H., Chaleki, E.L., 1999. The impacts of climate change for the water resources of the Colorado and Sacramento-San Joaquin River Basins. J. Amer. W. Res. Assn. 35 (6), 1429–1441.

Hanson, R.T., Martin, P., Koczot, K.M., 2002. Simulation of ground-water/surface-water flow in the Santa Clara-Calleguas Basin, Ventura County, California. US Geological Survey Water-Resources Investigation, WRIR 02-4136, 214 p.

Hanson, R.T., Newhouse, M.W., Dettinger, M.D., 2004. A methodology to assess relations between climatic variability and variations in hydrologic time series in south western United States. J. Hydrol. 287, 252–269.

Hereford, R., Webb, R.H., Graham, S., 2002. Precipitation History of the Colorado Plateau Region, 1900–2000. U.S. Geol. Surv. Fact Sheet 119-02, 4 p.

Hidalgo, H.G., Piechota, T.C., Dracup, J.A., 2000. Alternative principal component regression procedures for dendrohydrologic reconstructions. Water Resour. Res. 36 (11), 3241–3249.

Higgins, R.W., Yao, Y., Wang, X.L., 1997. Influence of the North American monsoon system on the United States summer precipitation regime. J. Climatol. 10, 2600–2622.

Jin, M., Duffy, C.J., 1994. Spectral and bispectral analysis for single and multiple input nonlinear phreatic aquifer systems. Water Resour. Res. 30 (7), 2073–2095.

Karhunen, K., 1947. Ueber lineare Methoden in der Wahrscheinlichtskeitsrechnung. Ann. Acad. Sci. Fennicae Ser., 37.

Loève, M.M., 1948. Fonctions Aleatories de Seconde Ordre. In: Levt, P. (Ed.), Process Stochastiques et Movement Brownien. Hermann, Paris.

MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, vol. 1, pp. 281–297.

Mantegna, R.N., 1999. Hierarchical structure in financial markets. Eur. Phys. J. 11, 193–197.

Mantua, N., Steven, H., 2002. Review: the Pacific decadal oscillation. J. Oceanogr. 58 (1), 35–44.

Merideth, R., 2000. A primer on climatic variability and change in the southwest. Udall Center for Studies in Public Policy and the Institute for the Study of Planet Earth, University of Arizona, Tucson, AZ. pp. 28.

NCDC Climate Online. (accessed 2007) <http://cdo.ncdc.noaa.gov/CDO/cdo>.

Piechota, T., Dracup, J.A., 1996. Drought and regional hydrologic variation in the United States: associations with El Nino/southern oscillation. Water Resour. Res. 32, 1359–1373.

Plaut, G., Vautard, R., 1994. Spells of low frequency oscillations and weather regimes in the northern hemisphere. J. Atmos. Sci. 51, 210–236.

Priestley, M.B., 1981. Spectral Analysis and Time series, vol. 1. Academic, San Diego, CA. pp. 890.

Ray, S., Turi, R.H., 2001. Determination of number of clusters in *K*-means clustering and application in colour image segmentation (invited paper). In: De, A.K., Pal, N.R., Das, J. (Eds.), ICAPRDT'99. Narosa Publishing House, pp. 137–143.

Shun, T., Duffy, C.J., 1998. Low-frequency oscillations in precipitation, temperature, and runoff on a west facing mountain front: a hydrogeologic interpretation. Water Resour. Res. 35, 191–201.

Stewart, I.T., Cayan, D.R., Dettinger, M.D., 2004. Changes in snowmelt runoff timing in western North America under a 'business as usual' climate change scenario. Climatic Change 62, 217–232.

Swetnam, T.W., Betancourt, J.L., 1997. Mesoscale disturbance and ecological response to decadal climatic variability in the American southwest. J. Clim. 11, 3128–3146.

Thomas, B., Pool, D.R. 2006. Seasonal Precipitation and Streamflow Trends in Southeastern New Mexico. USGS Professional Paper, 2005-1712, 79 pp.

USGS. (accessed 2007) <http://nwis.waterdata.usgs.gov/usa/nwis/discharge>.

Vautard, R., Ghil, M., 1989. Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series. Physica D 35, 395–424.

Vionnet, L.B., Maddock, T., 1992. Modeling of Groundwater Flow and Surface Water/Ground Water Interactions in the San Pedro River Basin. Part 1 – Cananea, Mexico to Fairbank, Arizona. HWR No. 92-010. University of Arizona.

Webb, R., Betancourt, J.L., 1990. Climatic Variability and Flood Frequency of the Santa Cruz River, Pima County, Arizona. USGS Open-File Report 90-553.

Webb, R.H., McCabe, G.J., Hereford, R. Wilkowske, C., 2004. Climatic Fluctuations, Drought, and Flow on the Colorado River. U.S. Geological Survey Fact Sheet, 3062-04. <http://pubs.usgs.gov/fs/2004/3062/>.

Winter, T.C., 2001. The concept of hydrologic landscapes. J. Am. Water Resour. Assoc. 37, 335–349.

Yiou, P., Ghil, M., Jouzel, J., Paillard, D., Vautard, R., 1994. Nonlinear variability of climate system from singular and power spectral of quaternary records. Clim. Dynam. 9, 371.

Young, K.C., 1992. A three-way model for interpolating monthly precipitation values. Mon. Wea. Rev. 120, 2561–2569.