Streamflow Variability in the United States: 1931–78

HARRY F. LINS
U.S. Geological Survey, Reston, VA 22092

(Manuscript received 26 June 1984, in final form 11 November 1984)

ABSTRACT

Systematic modes of spatial and temporal variation in a 48-year record of streamflow are defined using principal components. The components were calculated from a matrix of annual streamflow departures for 106 grid cells covering the United States in the years 1931–78. Five statistically significant components are found to account for more than 56% of the total variance. A varimax orthogonal rotation of the original components describes regional anomaly cores located in the middle Mississippi Valley, Pacific Northwest, Far West, Northeast, and northern Great Plains. Each of these patterns is an enhancement of a less well defined spatial form apparent in the unrotated solution. Temporal variations in the scores of all five components closely agree with contemporaneous national and regional departure patterns in several climatic variables.

1. Introduction

Various authors have recognized a tendency for the occurrence of uniform and recurrent annual streamflow patterns over broad regions of the United States and, at times, across nearly the entire nation (Harbeck and Langbein, 1949; Busby, 1963; Nace and Pluhowski, 1965; Langbein and Slack, 1982). Often, such systematic tendencies are attributed to vague or generalized “controlling climatic anomalies.” Detailed relationships and linkages between nationwide streamflow and specific climatic forcing functions, however, have yet to be defined. Moreover, there has heretofore been no quantitative characterization of the long-term covariance structure in annual flows from a nationwide sample of United States streams. As a result, the ability of hydrologists to model the synoptic-scale climate-streamflow system is severely constrained.

In this study, a statistical-descriptive examination of streamflow in the United States is presented. The purpose is 1) to identify the principal patterns of spatial and temporal variability in annual streamflow across the United States and, using these patterns, 2) to illustrate the similarity between the variability in streamflow and that in several climatic parameters. The latter purpose provides a basis for future quantitative characterizations of the relationship between nationwide streamflow and climate. Streamflow variability is defined using principal components analysis.

2. Nationwide streamflow record

Mean annual values of streamflow for the years 1931–78 have been assembled for 106 2.0° latitude by 3.0° longitude grid cells (Fig. 1). These data were excerpted from a recent study of long-term patterns in United States runoff reported by Langbein and Slack (1982). A total of 182 gaging stations, located in basins ranging in area from 40 to 60,000 km², were used. The values computed for each grid cell were based on records from at least one but not more than four stations.

The annual data presented in the Langbein and Slack report were prepared using gaging station records primarily on watercourses where there was no reported regulation or diversion or where diversion amounted to less than 10% of the mean flow and storage capacity was less than 10% of the mean annual runoff. In a few instances, where station records could not be found meeting the above criteria, records including alterations due to regulation or diversion were used. In all such instances, preference was given to streamflow records least so affected. Of the 106 grid cells used in the present study, only 21 were based on data in this last category, while 37 cells were based solely on unaltered gaging records.

As is common with streamflow, the data at each location followed a log-normal distribution. Thus, the annual means were logarithmically transformed and then standardized. The standardization is given by

\[ q_i = \frac{q_{in} - (\mu_{48})}{\sigma_i} \]

1 Throughout this report the terms “year” and “annual” refer to the water year which begins on 1 October of the previous year and ends on 30 September of the given calendar year. Thus water year 1931 began on 1 October 1930, and ended on 30 September 1931.
where $q_{in}^*$ is the standardized value of the log-transformed mean annual streamflow observation $q_{in}$ at the $i$th grid cell in the $n$th year of the 48-year series; $(\mu_{ik})$ is the 48-year mean value of $q$ at the $i$th grid cell, and $\sigma_i$ is the standard deviation of $q$ from the 48-year mean at grid cell $i$.

3. Principal component analysis

The dominant modes of spatial variability in nationwide streamflow are represented in terms of the principal components (PC) of the $48 \times 106$ observation matrix of $q_{in}^*$. Principal components are calculated herein using the correlation matrix. In light of the recent literature documenting the efficacy of component rotation in deriving more physically meaningful patterns of variation (Karl and Koscielny, 1982; Walsh et al., 1982; Richman, 1983a,b), both unrotated and rotated (orthogonal and oblique) solutions are computed. The unrotated component solution serves two purposes. First, it facilitates evaluation of the streamflow component shapes in terms of their correspondence to “Buell” patterns (Buell, 1975, 1979). That is, the question of whether the unrotated component patterns represent actual streamflow anomalies or instead conform to characteristic patterns commonly observed in the principal components of geophysical data is answered. This point is discussed in more detail later. Secondly, it provides derived quantities for which a basis exists to evaluate statistical significance. Current selection criteria for principal components have been designed for application to the results of unrotated solutions (Preisendorfer et al., 1981). Thus, if physically meaningful spatial patterns are sought, rotating only those components that first pass objective significance criteria provides a conservative framework for subsequent physical interpretation.

The percentage of variance and the cumulative percentage of variance explained by the first six unrotated principal components appear in Table 1. The values reveal that the first six components of nationwide streamflow account for nearly 61% of the total variance. These six components were tested for statistical significance using Preisendorfer and Barnett's (1977) dominant variance rule N. The rule-N test is designed to determine if the eigenvalues, or percentages of variance, calculated in a principal component analysis (unrotated solution) of a geophysical data set are distinguishable from those produced from a spatially and temporally uncorrelated random process. This particular test was applied in lieu of several other dominant variance selection rules because it 1) consistently exhibits the lowest variance in rule performance and 2) is a basically conservative

<table>
<thead>
<tr>
<th>Principal component</th>
<th>Percent variance explained</th>
<th>Cumulative percent variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.8301</td>
<td>22.8301</td>
</tr>
<tr>
<td>2</td>
<td>12.8597</td>
<td>35.6898</td>
</tr>
<tr>
<td>3</td>
<td>9.3743</td>
<td>45.0641</td>
</tr>
<tr>
<td>4</td>
<td>6.2264</td>
<td>51.2905</td>
</tr>
<tr>
<td>5</td>
<td>5.1314</td>
<td>56.4219</td>
</tr>
<tr>
<td>6</td>
<td>4.4306</td>
<td>60.8525</td>
</tr>
</tbody>
</table>
test, but less likely than others to omit potentially useful information (Preisendorfer et al., 1981). As the rule is applied, the statistical significance of a component is suspect when its corresponding geophysical eigenvalue is less than that generated from a set of random data. Results of the application of rule N to the components of nationwide streamflow appear in Table 2. As indicated in the table, the variances explained by the first five streamflow components exceed their random data counterparts. The variance explained by the sixth streamflow component (as well as those of all subsequent components), however, falls below its random data counterpart and thus represents the point at which the streamflow signal cannot be separated from noise.

a. Component loadings

The first five unrotated principal components of streamflow are mapped in Fig. 2. The contoured values are the component loadings, or the correlation coefficients between the streamflow variables and the principal components. Component one, representing the dominant mode of variation (22.83% of the total variance), is characterized by a pattern of streamflow anomalies of one sign across most of the contiguous United States. Strongest loadings on the first PC, indicative of the greatest departure from mean flow, occur in two broad regions. One encompasses the northern Great Basin and much of the northern and central Rocky Mountains. The other covers most of the East Central parts of the nation. A proclivity for streamflow to vary as a national unit during some years has long been recognized (Harbeck and Langbein, 1949; Busby, 1963).

The second principal component is characterized by a north–south opposition in the sign of loadings. The core regions of anomalous flows (i.e., areas with loadings ≥ 0.50) are found in the Pacific Northwest and in a broad band across the southwestern quarter of the nation. The basic shape of this component closely agrees with the shape of the distribution of low streamflows characteristic of the south and southwest drought years of the early to mid-1950s (Nace and Pluhowski, 1965) and early 1970s (Langbein and Slack, 1982).

In contrast to this north–south pattern, the third component of streamflow is characterized by an east–west sign opposition. Core areas for this function are located in northern California and western Nevada and in parts of New England. The areal extent and magnitude of loadings associated with the core areas of this PC are notably less than those observed for the first two components.

A continued diminution in the area and loading magnitude of core regions is apparent in the patterns of the fourth and fifth components. The fourth PC of streamflow is characterized by anomaly cores of like sign in central California and in much of New England. In broader terms, this function is described by a pattern of loadings of one sign in the Far West and northeastern regions with those of opposite sign across the northwestern, central, and southeastern regions. The fifth principal component contrasts small opposite-signed core areas of moderate loadings in the northern Great Plains and in southwestern Texas. Notably, the anomaly pattern depicted in PC four is consistent with observed precipitation anomaly patterns (Klein, 1965; Namias, 1966) while the pattern in PC five agrees closely with precipitation contrasts arising from variations in the preferred annual tracks of cyclones and anticyclones (Klein, 1957; Hayden, 1981).

Although numerous empirical associations can be described between the principal component patterns of streamflow and the contemporaneous or antecedent occurrence of various atmospheric conditions, recent research has rendered suspect the assignment of physical meaning to unrotated principal conditions (Buell, 1975, 1979; Richman, 1983a,b). The basis for such suspicion derives from the work of Buell who observed that the anomaly patterns or topography of the principal components of a data set are primarily determined by the geometrical shape of the boundary of the data sampling network and not by the correlation structure of the variables.

Buell defined a sequence of functions for sampling domains with square-, triangular-, and rectangular-shaped boundaries. Considering just the rectangular case, since the sample grid covering the United States used herein (Fig. 1) approximates a rectangle, a comparison is made between Buell's first five component patterns and those depicted in Fig. 2 for streamflow. Comparison results appear in Table 3. Ignoring the anomaly sign, which, as a consequence of the reflectivity property of principal components, is not of significance in this comparison and concentrating thus on the position of the anomalies within the rectangular domain, close agreement between the two sets of functions is apparent. The only differences

---

**Table 2. Comparison of the streamflow percentages of variance explained with those generated by a random process.**

<table>
<thead>
<tr>
<th>Principal component</th>
<th>Runoff data variance explained (T&lt;sub&gt;r&lt;/sub&gt;)</th>
<th>Random data variance explained* (U&lt;sub&gt;r&lt;/sub&gt;)</th>
<th>T&lt;sub&gt;r&lt;/sub&gt;/U&lt;sub&gt;r&lt;/sub&gt;* **</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.83</td>
<td>6.60</td>
<td>3.46</td>
</tr>
<tr>
<td>2</td>
<td>12.86</td>
<td>5.93</td>
<td>2.17</td>
</tr>
<tr>
<td>3</td>
<td>9.37</td>
<td>5.55</td>
<td>1.69</td>
</tr>
<tr>
<td>4</td>
<td>6.23</td>
<td>5.21</td>
<td>1.20</td>
</tr>
<tr>
<td>5</td>
<td>5.13</td>
<td>4.95</td>
<td>1.04</td>
</tr>
<tr>
<td>6</td>
<td>4.43</td>
<td>4.73</td>
<td>0.94</td>
</tr>
</tbody>
</table>

* Computed for a 48 x 106 matrix.
** If T<sub>r</sub>/U<sub>r</sub>* > 1, pass rule N; if < 1, fail.
exist in the assignment of patterns as components four or five and in the asymmetrical position of signs in the fifth streamflow component. The similarity is close enough to suggest that the principal component patterns for streamflow could be attributable to the Buell effect, as has been suggested by Richman (1983a) to be the case for the unrotated principal components of precipitation, temperature, and sea level pressure as reported in numerous studies.

To account for this effect the five streamflow components were rotated, both orthogonally (varimax) and obliquely (promax), following the procedures described by Richman (1983a,b) and previously applied by Walsh and Richman (1981), Horel (1981) and Karl and Koscielny (1982). Results from the orthogonal and oblique solutions are quite similar. Absolute loading values calculated by both methods for each of the 106 variables and 5 PCs never differ by more than 0.15, and the mean difference between methods for the 530 data points (106 × 5) is 0.04. Additional evidence of the similarity between the two approaches is presented in Table 4. Percentages of variance explained by the orthogonally and obliquely rotated streamflow components appear in the table in comparison to their unrotated counterparts. Differences between the two sets of rotated values are very small, although both are notably different than the unrotated results. In essence, the rotation transformations tend to minimize the differences in variance explained between each component while still accounting for most of the total variance explained by the five unrotated components. Indeed, in this case, both rotations account for the same amount of variance (to three decimal places) as explained in the unrotated solution. Given the similarity in results between the rotated solutions and in order to retain components that are uncorrelated with each other, only the results of the simpler orthogonal rotation are discussed here.

Loadings on the five orthogonally rotated components of streamflow are mapped in Fig. 3. Perhaps the most general distinction to be drawn between

Fig. 2. Loadings (×100) on the first five unrotated principal components of annual streamflow.
TABLE 3. Geometrical shape comparison of the first five principal components of streamflow with those produced by Buell (1975) for a rectangular domain.

<table>
<thead>
<tr>
<th>Principal component</th>
<th>Buell pattern*</th>
<th>Streamflow pattern**</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[+ ]</td>
<td>[− ]</td>
<td>Anomaly of one sign over entire domain</td>
</tr>
<tr>
<td>2</td>
<td>[− ]</td>
<td>[+ ]</td>
<td>North-south opposition in anomaly signs</td>
</tr>
<tr>
<td>3</td>
<td>[+ − ]</td>
<td>[+ − ]</td>
<td>East-west opposition in anomaly signs</td>
</tr>
<tr>
<td>4</td>
<td>[− + ]</td>
<td>[− + ]</td>
<td>Buell—positive anomaly in northwest and southeast, negative anomaly in southwest and northeast Streamflow—negative anomaly in west and east, positive anomaly in Central</td>
</tr>
<tr>
<td>5</td>
<td>[− + − ]</td>
<td>[+ − ]</td>
<td>Buell—same as streamflow PC4 Streamflow—positive anomaly in northwest, southwest, and southeast, negative anomaly in northeast</td>
</tr>
</tbody>
</table>

* From Buell (1975; Fig. 9, p. 191).
** See Fig. 2.

these patterns and their unrotated counterparts (Fig. 2) is that there is no diminution in the areal extent and loadings magnitude of the core areas in the third, fourth and fifth components. The effect of the rotation is to isolate into coherent modes and enhance areas of high systematic streamflow variability which, because of inherent axis fitting limitations, are less well defined by the unrotated solution. It is clear in Fig. 3 that the core areas depicted therein are simply refined representations of the patterns in Fig. 2. Nothing new or inconsistent has been introduced by the rotation; no significant existing information has been removed.

In more specific terms, each of the five orthogonal streamflow components emphasize a unique regional mode of variation. The first orthogonally rotated component (PC-1-Or) is characterized by a core region covering much of the southeastern quarter of the United States. Strongest loadings are centered in southern Illinois and in a narrow band extending westward through the Ozarks. This pattern, accounting for 14.38% of the variance, is representative of uniform conditions of either above- or below-mean streamflow across most of the nation, with significant flooding or drought focused in the middle Mississippi Valley. As noted earlier, the tendency for the entire nation to experience either flooding or drought conditions on an annual time scale has been documented for many years.

Loadings on PC-2-Or indicate a focus of strong streamflow anomalies in the Pacific Northwest. The most intense loadings, in excess of 0.90, occur in a narrow band in western Idaho. The PC-2-Or component accounts for 13.15% of the total variance in nationwide streamflow. Components three through five successively describe regional streamflow variations centered in the Great Basin (extending from the Pacific coast of northern and central California eastward to the western Rockies). New England and parts of the Middle Atlantic States, and the northern Great Plains—upper Mississippi Valley. These functions explain 10.42, 9.36 and 9.11% of the total streamflow variance, respectively. Corollaries for each of these regional streamflow functions can be found in the rotated principal components of precipitation (Walsh et al., 1982) and drought (Karl and Koscielny, 1982).

b. Component scores

Principal component scores were calculated according to the formulation of Jöreskog et al. (1976), where

$$F = XA\Lambda^{-1/2}.$$ 

Thus, the $48 \times 5$ matrix of streamflow component scores $F$ is equal to the product of $A$ (the orthogonally rotated $106 \times 5$ PC loadings matrix) premultiplied by $X$ (the original $48 \times 106$ streamflow data matrix), scaled by the inverse square root of the diagonal matrix of eigenvalues $\Lambda$. These scores indicate the relative importance of each principal component for each of the 48 years of record. Time series plots of the scores for the five orthogonally rotated PCs appear in Fig. 4.
The annual scores for the first principal component exhibit a broad quasi-sinusoidal variation. This pattern indicates that during the period 1931–78 two very generalized cycles of below- and above-mean streamflow occurred on a nationwide scale. Specifically, near-normal to much below normal flows characterized most of the United States between the years 1931 and 1941; very deficient flows existed during the 1953–56 period; and intermittent annual occurrences of moderately deficient flows occurred between 1956 and 1968. The 11-year period 1942–52 was one of persistently abundant flows across most of the country, as were eight of the ten years between 1969 and 1978. The time history of deficient and excessive flows, as characterized by this PC function, agrees quite closely with the temporal distribution of positive and negative anomalies of several climatic variables. For example, in an analysis of major dry and wet periods in the United States, Diaz (1983) investigated the temporal variation in Palmer's Drought Area Index (DAI) and Wet Spell Area Index (WSAI). The years noted by Diaz as having a DAI in excess of 0.50 (i.e., more than 50% of the country undergoing moderate to severe drought) agree exactly with the years for which PC-1-Or had positive scores exceeding one standard deviation from the mean score of 0.00. Similarly, each year with a WSAI in excess of 0.50 corresponded to a year where PC-1-Or had a negative score greater than one standard deviation from its mean.

Scores for the first streamflow component were also compared with annual mean precipitation (i.e., a mean value for each year for the entire United States) for the 1931–78 period. A time series plot of the two variables appears in Fig. 5 and, therein, the close association of both variables through time is apparent. The only discrepancies between the plots are associated with an occasional 1-year phase lag in
the streamflow PC behind precipitation. Cross correlation of the streamflow component scores with precipitation produced coefficients of 0.85 and 0.35 at lags $k = 0$ and $k = 1$ respectively. The high zero-lag correlation provides additional support for the designation of PC-1-Or as a "nationwide" function.

Scores for the second streamflow component (Fig. 4) reflect several distinct periods of streamflow anomalies in the Pacific Northwest along with a high degree of interannual variability. The period 1931–47 was one within which streamflow varied from near-normal to much below normal. This was followed by a 9-year period (1948–56) in which flows varied in the normal to above-normal range. Fourteen consecutive years (1957–70) of approximately normal streamflow then ensued, followed by eight years of widely varying and opposing flow extremes. A recent assessment of annual streamflow extremes in the western United States by Meko and Stockton (1984) describes an almost identical sequence of "normal to below—normal to above—normal—variably above and below normal" in the Pacific Northwest. Meko and Stockton further note that large streamflow anomalies in the Pacific Northwest often occur coincident with anomalies of opposite sign in the southwest. This sign opposition is clearly delineated in the map of PC-2-Or component loadings (Fig. 3). Notably, the transition from high annual variability to approximately mean conditions in this function during the mid-1950s coincides with the timing of a significant reduction in precipitation variability in the Pacific Northwest (Diaz and Quayle, 1980).

The temporal distribution of scores for the third rotated component (western United States function) also exhibits considerable variability through the study period. The only recognizable periods of persistence in the sign of streamflow departures occurred during the decade of the 1940s, when streamflow in the West was in the normal to above-normal range, and between 1965 and 1975 when flows were again in the normal to above category.

Component four (Northeast function) scores appear somewhat less variable annually. This is especially true after 1960, in contrast to components two and three. For example, between 1961 and 1968, mildly to extremely deficient streamflows characterized much of the Northeast region. The strongest departures from normal within this period occurred during the years 1962–66, coincident exactly with the well-documented period of drought in the northeastern United States (Spar, 1967; Namias, 1966). By the early 1970s a significant change in the pattern of streamflow anomalies occurred in the Northeast. For seven of the eight years between 1971 and 1978 excessive flows and flooding persisted. Only in 1977, when much of the nation was undergoing severe drought, did the Northeast depart from elevated annual flow conditions to flows that were in the normal range.

Scores for the fifth rotated component (northern Great Plains function) were also less variable from year to year. In fact, four distinct consecutive periods of persistence characterize the PC-5-Or scores during the study period. First, between 1931 and 1941, normal to much below normal flows dominated the northern Great Plains and upper Mississippi Valley. Then, from 1942 until 1953, normal to above-normal runoff persisted throughout the region. In all but one year between 1954 and 1961 streamflow again exhibited normal to below-normal conditions. Finally, in all but three years, including the record dry year of 1977, did streams in the northern Great Plains between 1962 and 1978 flow in the above-normal range.

One additional characteristic of PC-5-Or deserves mention. The area around the central Rio Grande basin in southwestern Texas contains moderately strong loadings opposite in sign to those in the
northern Great Plains (Fig. 3). This pattern indicates that there is some tendency for streamflow anomalies of one extreme (drought or flood) to occur in southwestern Texas contemporaneously with streamflows of the opposite extreme in the northern Great Plains. Notably, the occurrence of such a pattern has been documented for many years (Busby, 1963; Nace and Pluhowski, 1965) and the timing of these occurrences agrees with the distribution of component five scores.

4. Summary

Five statistically significant modes of variation in annual streamflow across the United States are identified using principal components analysis. An orthogonal rotation of the component axes based on the varimax criterion is applied to emphasize coherent regions of variability. The five modes collectively explain more than 56% of the total variance in the streamflow record. The regions associated with these modes are the middle Mississippi Valley, Pacific Northwest, Far West, Northeast, and northern Great Plains. The temporal occurrence of strong departures from mean streamflow conditions in each region, as characterized by the time series of component scores, is consistent with previously documented regional patterns of precipitation and drought.

It thus appears that principal components analysis provides a reasonable description of the large-scale long-term variance structure in annual United States streamflow. Based on these findings, it is suggested that a framework now exists for establishing more rigorous quantitative relationships between streamflow and climate. Further, time series analysis of the component scores may prove applicable to a number of problems in synthetic hydrology. For these purposes, however, similar analyses using monthly and seasonal streamflow data will probably be necessary.

REFERENCES


Nace, R. L., and E. J. Pluhowski, 1965: Drought of the 1950's


