An aggregate drought index: Assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage

John A. Keyantash
Department of Earth Sciences, California State University, Dominguez Hills, Carson, California, USA

John A. Dracup
Department of Civil and Environmental Engineering, University of California, Berkeley, California, USA

Received 20 August 2003; revised 4 March 2004; accepted 4 June 2004; published 11 September 2004.

[1] An aggregate drought index (ADI) has been developed, and evaluated within three diverse climate divisions in California. The ADI comprehensively considers all physical forms of drought (meteorological, hydrological, and agricultural) through selection of variables that are related to each drought type. Water stored in large surface water reservoirs was also included. Hydroclimatic monthly data for each climate division underwent correlation-based principal component analysis (PCA), and the first principal component was deseasonalized to arrive at a single ADI value for each month. ADI time series were compared against the Palmer Drought Severity Index (PDSI) to describe two important droughts in California, the 1976–1977 and 1987–1992 events, from a hydroclimatological perspective. The ADI methodology provides a clear, objective approach for describing the intensity of drought and can be readily adapted to characterize drought on an operational basis. *INDEX TERMS:* 1812 Hydrology: Drought; 1894 Hydrology: Instruments and techniques; 1833 Hydrology: Hydroclimatology; *KEYWORDS:* drought, index, PDSI, principal components

Citation: Keyantash, J. A., and J. A. Dracup (2004), An aggregate drought index: Assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage, *Water Resour. Res.*, *40*, W09304, doi:10.1029/2003WR002610.

1. Introduction

[2] Droughts are the world's costliest natural disasters, causing an average $6–8 billion in global damages annually and collectively affecting more people than any other form of natural disaster [*Wilhite*, 2000]. Of the 46 U.S. weather-related disasters between 1980 and 1999 causing damages in excess of $1 billion, eight of the events were droughts. These included the most costly national disaster, the 1988 drought, which alone was assessed for $40 billion of losses [*Ross and Lott*, 2000]. However, despite the ubiquity and often obvious recognition of drought, precise physical quantification of drought intensity is a complex geophysical endeavor.

[3] Traditionally, drought has been classified according to the hydrologic compartment in which there is a water deficiency. Meteorological drought results from a shortage of precipitation, while hydrological drought describes a deficiency in the volume of water supply, which includes streamflow, reservoir storage, and/or groundwater depths [*Wilhite*, 2000]. Agricultural drought relates to a shortage of available water for plant growth and is assessed as insufficient soil moisture to replace evapotranspirative losses [*World Meteorological Organization* (*WMO*), 1975]. For an extensive listing of the specialized drought indices available for the three physical forms of drought, refer to *WMO* [1975] and *Heim* [2000].

[4] An alternative to specialized drought indices is to use a broad metric to assess the overall availability of water. This was the motivation behind the Palmer Drought Severity Index (PDSI), which has historically been the most widely used index of drought in the United States [*Palmer*, 1965]. However, the PDSI is recognized to have limitations owing to its complex, empirical derivation and the fact that its underlying computation is based on the climates of Midwestern states. For a thorough discussion of these weaknesses, refer to *Guttman et al.* [1992], *Alley* [1984], and *Keyantash and Dracup* [2002].

[5] The Surface Water Supply Index (SWSI), developed by *Shafer and Dezman* [1982], is a drought index that assesses the intensity of hydrological drought by considering snowpack and surface water levels. It was developed for regions such as California and the mountainous West, where spring snowmelt contributes significantly to surface water reserves. Computation of the SWSI requires monthly measurements for snowpack, precipitation, streamflow, and reservoir storage. The hydrological observations are input to a basin-calibrated algorithm that considers the percent contribution of each hydrological component compared with the typical water conditions in the basin [*Garen*, 1992]. However, the SWSI does not directly consider other elements of the hydrological cycle, such as evaporation and soil moisture content.

[6] Introduced here is the Aggregate Drought Index (ADI), a multivariate drought index that considers the bulk quantity of water across the meteorological, hydrological, and agricultural regimes of drought. The domain of the ADI is a region of climatic uniformity, such as a National Climatic Data Center (NCDC) climate division, which has

0043-1397/04/2003WR002610$09.00

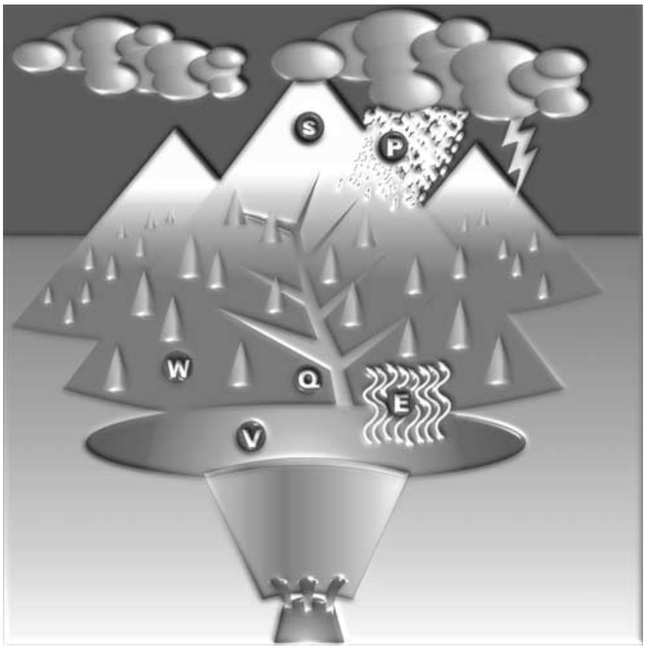

Figure 1. Surface water balance as described by the aggregate drought index (ADI). Precipitation P and snow water content s contribute to streamflow Q, some of which is captured as reservoir storage V. Evapotranspiration E stems from surficial evaporation and vegetative transpiration, both of which diminish the surface soil water content W. See color version of this figure at back of this issue.

its water-related resources aggregately assessed through principal component analysis (PCA).

2. ADI Constituents

[7] The ADI input variables represent volumes of water residing or moving within the climate division. The six parameters adopted for the ADI include precipitation (P), evapotranspiration (E), streamflow (Q), reservoir storage (V), soil moisture content (W), and snow water content (s), as shown in Figure 1. These variables may be used selectively, depending on the characteristics for the region of interest. For instance, if a region does not possess winter snowpack, s should be omitted. Since the approach treats months separately, it is possible to use a different set of variables for each month.

[8] Groundwater can offset surface water shortages during a drought; aquifers may be managed analogously to surface reservoirs for the storage and withdrawal of water [e.g., *Draper et al.*, 2003]. However, groundwater was not included in the suite of ADI variables for three reasons: (1) Historic groundwater heights prior to human disturbances are typically unknown; (2) groundwater flow between distinct, heterogeneous aquifers across sizeable climate divisions is difficult to assess; and (3) groundwater response to drought may be asynchronous with other ADI variables; groundwater recharge/depletion commonly occurs at a timescale of weeks to years, which extends beyond the ADI time step of 1 month.

[9] The hydrologic cycle perspective of the ADI also excludes variables such as temperature and teleconnection indices from the analysis. The rationale is that if these quantities are anticipated to herald drought, and drought represents a shortage of water, then the effects should be observed more directly in water-related variables. The adopted approach thus attempts to preserve a close connection between drought and the basic elements of the hydrologic cycle.

3. Data Sources

[10] Hydrologic variables used in the computation of the ADI were preferentially selected to represent measurements as opposed to modeled results, but this was not possible in the case of soil moisture. Soil moisture measurements are not available on a widespread, national scale. However, modeled soil moisture is available from the Climate Prediction Center (CPC) on a climate divisional basis (see http://www.cpc.ncep.noaa.gov/soilmst). Similarly, monthly divisionally averaged precipitation was available from the National Climatic Data Center (NCDC) in their TD-9640 data set [*NCDC*, 2000a].

[11] Data for E, Q, V, and s exist for discrete points, so multiple observational stations were used to arrive at a single total for each climate division that was studied. The extent of available data differed between climate divisions; the data sources used to construct a regional value for each division are enumerated in Table 1. Geographical locations for the 192 observational stations employed are detailed by *Keyantash* [2001].

[12] Stream discharges and reservoir storages were (separately) summed to produce divisional totals. Snow water content and pan evaporation were spatially averaged using the simple Thiessen polygon technique. The total polygonal area for E was the full climate division, while the domain of s was the approximate areal extent of divisional snow cover as of 1 April for years 1997 through 2001, using imagery produced by the National Operational Hydrologic Remote Sensing Center (available at http://www.nohrsc.nws.gov/snowsurvey.html).

[13] April 1 is the nominal date when alpine snow is considered to be at its greatest accumulation in California [*Department of Water Resources* (*DWR*), 2003].

4. Examination Periods

[14] The objective of the research was to compute a chronological ADI time series from the earliest year in which the full suite of constituent variables were available through water year (WY) 2000. The examination period used for each climate division is listed in Table 2. The periods should be considered reference intervals during

Table 1. Number of Data Stations Used to Construct ADI

Discrete Variable	Division NC	CV	SC	Sources
E	2	5	1	*NCDC* [2001a, 2001b]
Q	15	13	9	CDEC[a] and USGS[b]
V	11	34	28	CDEC
s	17	67	0	CDEC

[a]California Data Exchange Center (http://cdec.water.ca.gov).
[b]U.S. Geological Survey National Water Information System (http://waterdata.usgs.gov/nwis).

which the climatology was established for each hydrologic variable, as well as the ADI itself.

5. Test Regions

[15] The three trial regions selected for application of the ADI methodology lie within California. They are the climate divisions identified as the North Coast Drainage (NC), the San Joaquin Drainage (CV, as the San Joaquin River drains the southern portion of California's Central Valley), and the South Coast Drainage (SC). These three divisions appear in Figure 2 as zones 1, 5, and 6, respectively. The regions are distinguished by the following properties:

[16] 1. The NC (zone 1) is characterized by generally wet conditions and rugged mountainous terrain. It is the least urbanized region and has the highest annual rainfall of any region studied.

[17] 2. The CV (zone 5) contains elevational differences from the alpine crests of the Sierra Nevada to the lowlands of the Great Central Valley, where a warm growing season, fertile soils, and irrigation combine to form the most productive agricultural land in the United States. This region represents the challenges of drought to the farming community.

[18] 3. The SC (zone 6) represents the coastal and canyon landscape of urban Southern California, where precipitation is modest but a large population (16 million people) depends upon the natural and imported water supplies. It is the most urbanized environment in the state.

[19] Given the different climates in each of these regions, application of the ADI to these diverse conditions demonstrates its potential applicability to other regions nationwide.

6. Mathematical Formulation of ADI

[20] Principal component analysis has been used extensively in the atmospheric and hydrologic sciences to describe dominant patterns appearing in observational data [e.g., *Barnston and Livezey*, 1987; *Hidalgo et al.*, 2000; *Lins*, 1997]. In this research, PCA was adopted as the numerical approach to distill the essential hydrologic information from the input data set, which leads to the construction of the ADI. Specifically, we used P-mode PCA [*Cattell*, 1952], where the analysis describes temporal fluctuations of input variables at a fixed location (climate division). Computation of the principal components requires constructing a square ($p \times p$, where p is the number of variables), symmetric, correlation matrix $\mathbf{R}$ to describe the correlations between the original data. Note that correlations were only computed among data representing the same month. Therefore there were twelve separate $\mathbf{R}$ per division to describe the cross correlations between variables. These correlation matrices underwent PCA; for discussion of the routine mathematical steps, refer to *Haan* [1977], *Wilks* [1995], or *Preisendorfer* [1988].

[21] Principal components are a reexpression of the original p-variable data set in terms of uncorrelated components $\mathbf{z}_j$ ($1 < j \leq p$). Eigenvectors derived through PCA are unit vectors (i.e., magnitude of 1) that establish the relationship between the PCs and the original data:

$$\mathbf{Z} = \mathbf{XE}, \tag{1}$$

where $\mathbf{Z}$ is the $n \times p$ matrix of principal components, $\mathbf{X}$ is the $n \times p$ matrix of standardized observational data, and $\mathbf{E}$ is the $p \times p$ matrix of eigenvectors.

[22] The ADI is the first PC (PC1), normalized by its standard deviation:

$$\mathrm{ADI}_{i,k} = \frac{z_{i,1,k}}{\sigma}, \tag{2}$$

where $\mathrm{ADI}_{i,k}$ is the ADI value for month k in year i, $z_{i,1,k}$ is the first principal component during year i, for month k, and σ is the sample standard deviation of $z_{i,1,k}$ over all years i.

[23] The ADI utilizes only the first PC because it explains the largest fraction of the variance described by the full p-member, standardized data set. Considering all months, PC1 described an average of 58%, 63%, and 59% of the data set variance for the NC, CV, and SC climate divisions, respectively. Since PCs are orthogonal vectors, it is not mathematically proper to combine them into a single expression. Therefore only the dominant mode was adopted to describe the bulk of the water anomalies in the observational data.

[24] The first PC is deseasonalized to enable each month's ADI to represent a normalized expression of variability. Without standardization, months that routinely possess a higher degree of hydrologic variability cause a chronological plot of ADI values to predictably jump. It can be shown (Appendix A) that each of the 12 ADI time series has a mean of zero but possesses nonunit standard deviation (averaging 1.8 and ranging between 1.5 and 2.1 across the 36 month-division combinations examined). Monthly standard deviations for the three divisions examined are shown in Figure 3.

[25] The final step in the construction of the ADI is to reorder the terms from the 12 ADI annual time series into a single chronology of $12n$ terms. The steps involved in the calculation of the ADI are summarized in Figure 4.

[26] Computation of the ADI can be readily accomplished using statistical/mathematical software and a personal computer. For this research the ADI was calculated using code written for MATLAB and executed on a Pentium III desktop computer. The computational time for the ADI time series for the three climate divisions was less than 1 s.

[27] For compactness, the remainder of this paper describes the ADI using vector-matrix notation instead of specifying multiple subscripts (as in equation (2)). In doing so, subscript i of equation (2) vanishes, as it is implicitly represented by the element number of the vector. Subscripts 1 and k are also absent; only the first PC is used in the ADI, and it should be understood that any computational expression applies consistently but separately for each month. (Only upon final computation of all 12 monthly ADI series are they combined into a single chronological time series.) Incorporating these modifications into equation (2), the ADI may be compactly expressed as

$$\mathbf{a} = \frac{\mathbf{z}_1}{\sigma}, \tag{3}$$

where $\mathbf{a}$ is the $n-$ member ADI time series, $\mathbf{z}_1$ is the first principal component of the hydrological data, and σ is the sample standard deviation of $\mathbf{z}_1$.

[28] Alternatively, Appendix B demonstrates how $\mathbf{a}$ can be fundamentally expressed as a scalar multiple of a unit

vector $\hat{\mathbf{z}}_1$:

$$\mathbf{a} = \sqrt{n-1}\hat{\mathbf{z}}_1.$$

7. Example Calculation

[29] Consider the computation of the ADI time series for the month of April in the South Coast division. The 25 years of data for P, E, Q, V, and W are arranged columnarly into an 25×5 matrix of observations **O**. A strength of the correlation-based PCA approach used for the ADI is that it is unimpacted by the measurement units of the input data, so the monthly data for P, E, Q, V, and W in **O** are reported in their original units of inches, hundredths of inches, average cubic feet per second, million acre-feet, and millimeters, respectively:

$$\mathbf{O} = \begin{bmatrix} 1.34 & 473 & 61 & 1.15 & 216 \\ 0.07 & 672 & 36 & 0.90 & 176 \\ 2.02 & 464 & 790 & 1.74 & 545 \\ 0.03 & 630 & 567 & 1.73 & 397 \\ 0.91 & 602 & 584 & 1.87 & 476 \\ 0.61 & 549 & 135 & 1.59 & 246 \\ 1.57 & 569 & 426 & 1.63 & 299 \\ 3.6 & 510 & 782 & 1.86 & 541 \\ 0.61 & 723 & 75 & 1.57 & 187 \\ 0.18 & 683 & 103 & 1.44 & 203 \\ 0.71 & 583 & 161 & 1.63 & 316 \\ 0.28 & 725 & 84 & 1.49 & 178 \\ 2.96 & 598 & 90 & 1.48 & 254 \\ 0.12 & 788 & 34 & 1.35 & 166 \\ 0.75 & 646 & 40 & 1.32 & 148 \\ 0.08 & 624 & 273 & 1.33 & 308 \\ 0.26 & 669 & 258 & 1.49 & 348 \\ 0.00 & 652 & 540 & 1.87 & 454 \\ 1.07 & 564 & 119 & 1.60 & 257 \\ 1.09 & 560 & 468 & 1.83 & 459 \\ 0.65 & 723 & 125 & 1.66 & 247 \\ 0.11 & 738 & 83 & 1.53 & 207 \\ 2.1 & 497 & 831 & 1.78 & 502 \\ 2.32 & 552 & 132 & 1.60 & 230 \\ 1.74 & 533 & 184 & 1.51 & 231 \end{bmatrix}$$

Table 2. Examination Periods for Each Climate Division

Division	Water Years
NC	1970–2000
CV	1974–2000
SC	1976–2000

[30] The data in **O** next have their column means subtracted, and each element is divided by the column standard deviation. This expresses the original observations as a series of standardized anomalies, and the new series is referred to as **X**. Correlations between the standardized anomalies are expressed in the symmetric, 5×5 correlation matrix **R**:

$$\mathbf{R} = \frac{1}{24}\mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} 1 & -0.68 & 0.40 & 0.32 & 0.38 \\ -0.68 & 1 & -0.53 & -0.27 & -0.54 \\ 0.40 & -0.53 & 1 & 0.71 & 0.96 \\ 0.32 & -0.27 & 0.71 & 1 & 0.74 \\ 0.38 & -0.54 & 0.96 & 0.74 & 1 \end{bmatrix}$$

[31] Principal component analysis is performed on **R**, and eigenvalues for the April SC time series are determined to be 3.257, 1.067, 0.425, 0.210, and 0.041. We utilize only the first eigenvalue, which single-handedly explains 65 percent ($3.257 \div 5$) of the data set variance. The eigenvector $\mathbf{e}_1$ associated with the first April eigenvalue is

$$\mathbf{e}_1 = [0.36 \;\; -0.40 \;\; 0.51 \;\; 0.43 \;\; 0.51]^{\mathrm{T}}.$$

[32] The first PCs are computed using equation (1), where **E** is simply $\mathbf{e}_1$. The deseasonalized ADI series is constructed

Figure 2. The seven climate zones of California. The North Coast (NC), San Joaquin, and South Coast Drainages are represented by numbers 1, 5, and 6, respectively. Source, *National Climate Data Center* [2000b].

by equation (3). Combining these equations into a single expression, the ADI for April in the SC is

$$\mathbf{a} = \frac{\mathbf{X}\mathbf{e}_1}{\sigma}. \tag{4}$$

Referring to Figure 3, the April SC $\mathbf{z}_1$ series has σ = 1.80. Inserting numbers for all variables,

$$\mathbf{a} = \frac{1}{1.80} \cdot \begin{bmatrix} 0.34 & -1.59 & -0.83 & -1.77 & -0.70 \\ -0.95 & 0.67 & -0.93 & -2.84 & -1.01 \\ 1.03 & -1.69 & 1.95 & 0.77 & 1.92 \\ -1.00 & 0.19 & 1.10 & 0.75 & 0.75 \\ -0.10 & -0.13 & 1.16 & 1.34 & 1.37 \\ -0.40 & -0.73 & -0.55 & 0.12 & -0.46 \\ 0.57 & -0.50 & 0.56 & 0.32 & -0.04 \\ 2.64 & -1.17 & 1.92 & 1.32 & 1.89 \\ -0.40 & 1.25 & -0.78 & 0.07 & -0.93 \\ -0.84 & 0.79 & -0.67 & -0.53 & -0.80 \\ -0.30 & -0.34 & -0.45 & 0.30 & 0.10 \\ -0.74 & 1.27 & -0.74 & -0.30 & -1.00 \\ 1.99 & -0.17 & -0.72 & -0.33 & -0.40 \\ -0.90 & 1.99 & -0.94 & -0.90 & -1.10 \\ -0.26 & 0.37 & -0.91 & -1.04 & -1.24 \\ -0.94 & 0.12 & -0.02 & -0.98 & 0.03 \\ -0.76 & 0.64 & -0.08 & -0.29 & 0.35 \\ -1.03 & 0.44 & 0.99 & 1.35 & 1.20 \\ 0.06 & -0.56 & -0.61 & 0.17 & -0.37 \\ 0.08 & -0.60 & 0.72 & 1.19 & 1.23 \\ -0.36 & 1.25 & -0.59 & 0.46 & -0.45 \\ -0.91 & 1.42 & -0.75 & -0.13 & -0.77 \\ 1.11 & -1.32 & 2.11 & 0.96 & 1.58 \\ 1.34 & -0.69 & -0.56 & 0.17 & -0.58 \\ 0.75 & -0.91 & -0.36 & -0.21 & -0.57 \end{bmatrix} \cdot \begin{bmatrix} 0.36 \\ -0.40 \\ 0.51 \\ 0.43 \\ 0.51 \end{bmatrix} = \begin{bmatrix} -0.43 \\ -1.56 \\ 1.86 \\ 0.46 \\ 1.05 \\ -0.18 \\ 0.45 \\ 2.18 \\ -0.83 \\ -0.89 \\ -0.01 \\ -1.00 \\ 0.04 \\ -1.42 \\ -0.99 \\ -0.44 \\ -0.28 \\ 0.64 \\ -0.10 \\ 0.99 \\ -0.54 \\ -0.96 \\ 1.79 \\ 0.14 \\ 0.03 \end{bmatrix}$$

8. Operational Computation of ADI

[33] The ADI results presented in this paper were obtained using the entire period of record (see Table 2) for each of the parameters. Usage of the full record to compute the parameter means is ultimately responsible for the ADI zero mean property detailed in Appendix A. However, if the ADI were operationally computed on a monthly basis, it would be necessary to assign a baseline period from which to define reference means, reference standard deviations, and reference eigenvectors. The baseline means, standard deviations, and eigenvectors, 12 of each, would be applied to all future data. Defining the new

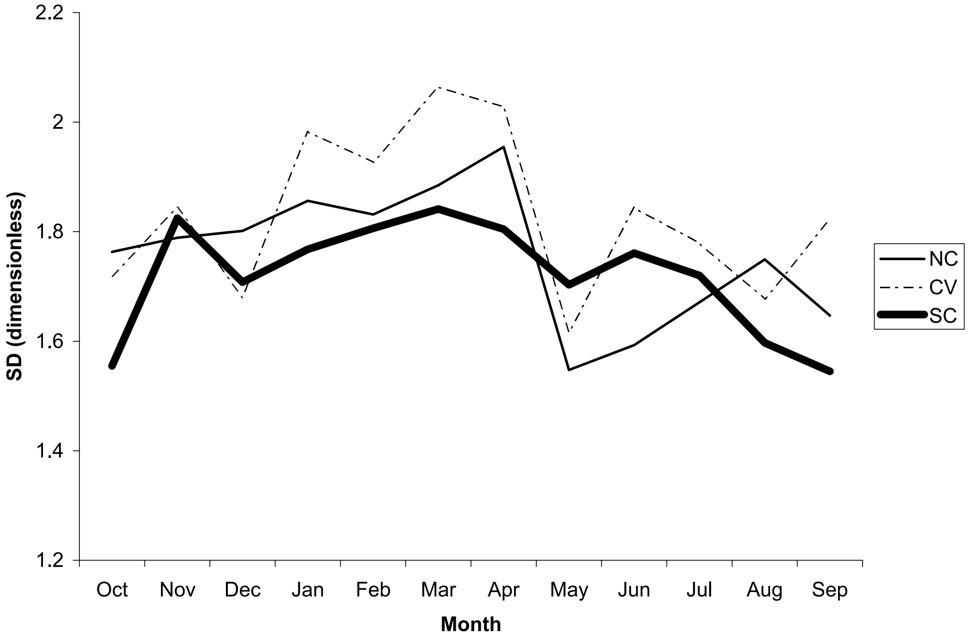

Figure 3. Prior to deseasonalization, the first principal components series from each climate division possesses nonunit standard deviation.

period as an ever-enlarging interval of time following the baseline period, $\mathbf{a_{new}}$ and $\mathbf{X_{new}}$ would be computed

$$\mathbf{a_{new}} = \frac{\mathbf{X_{new}e_1}}{\sigma} \tag{5}$$

where

$$\mathbf{X_{new}} = (\mathbf{O_{new}} - \mathbf{M})\mathbf{D} \tag{6}$$

leading to

$$\mathbf{a_{new}} = \frac{(\mathbf{O_{new}} - \mathbf{M})\mathbf{De_1}}{\sigma} \tag{7}$$

where

- $\mathbf{O_{new}}$ = $n_{new} \times p$ matrix of postreference period observational data;
- $\mathbf{M}$ = $n_{new} \times p$ matrix of data means (every row is identical), based on reference period;
- $\mathbf{D}$ = $p \times p$ diagonal matrix with the inverse of the sample standard deviation for each hydrologic variable x_i (i.e., $\sigma_{x_i}{}^{-1}$, $1 \leq i \leq p$) residing upon each diagonal element $d_{i,i}$;
- $\mathbf{e}_1$ = $p \times 1$ first eigenvector, based on reference period;
- σ = sample standard deviation of $\mathbf{z}_1$ during reference period.

[34] In addition to simplifying computation of the ADI (the numerical steps to redetermine $\mathbf{e}_1$ would be no longer be necessary), use of reference quantities would importantly preserve the historical values of the ADI. In certain instances it might be desirable to reevaluate the severity of wet and dry months by viewing anomalies with respect to the entire historical context, but in routine operations it would be preferable to have the ADI as a function of time that is independent from the date of computation.

[35] There are close relationships between the baseline period and operational formulae for the ADI. Appendix C summarizes expressions for $\mathbf{a}$ and $\mathbf{a_{new}}$ using various data vectors.

9. Comparison With PDSI and Other Drought Indices

[36] The ADI is similar to the PDSI in the fact that both indices' dependent variables are centered upon the physical theme of water balance. Consequently, a comparison of the ADI computations with the PDSI time series provided an important check on the appropriateness of the ADI findings. It should be stressed that a perfect corroboration was neither anticipated or desired. As mentioned in section 1, the PDSI suffers from certain limitations, primarily geographic biases, inadequate treatment of snowfall, and a complex, empirical formulation [*Alley*, 1984; *Guttman et al.*, 1992; *Keyantash and Dracup*, 2002]. Consequently, although the PDSI is a widely used index, it should not be considered as the ultimate standard.

[37] Monthly PDSI data for each climate division were obtained from the NCDC in the TD-9640 data set [*NCDC*, 2000a]. Assessments between the ADI and PDSI time series used the Spearman correlation coefficient, which is the application of the ordinary (Pearson) correlation coefficient performed upon data ranks instead of the data themselves [e.g., *Wilks*, 1995]. Rank correlation is a robust measure that is insensitive to the underlying data distribution and is recognized as a competent tool for determining the best aggregate correlation [*Wilks*, 1995]. Similar to an ordinary correlation coefficient, the range of the Spearman coefficient is constrained between ±1.

[38] Rank correlations between the ADI and PDSI were found to behave similarly across the three climate divisions. For the NC, CV, and SC divisions the correlations were 0.78, 0.72, and 0.65, respectively. However, PDSI values

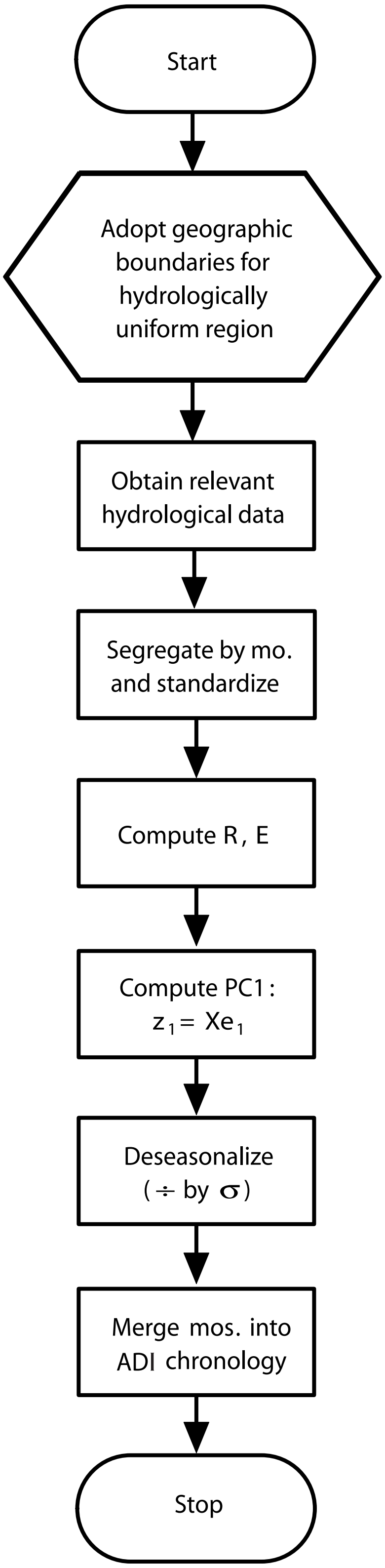

Figure 4. Flow chart of the ADI computational process.

oscillate between larger extrema than the ADI. The chronological comparison between the ADI and the standardized PDSI time series for the three climate divisions is shown in Figure 5.

[39] As an alternative to the PDSI, the National Drought Mitigation Center and the Western Regional Climate Center recommend the Standardized Precipitation Index (SPI) as an improved measure of meteorological drought [*Redmond*, 2000]. The SWSI is also widely utilized for hydrological drought monitoring in numerous Western states. The PDSI, SPI, and SWSI were compared and quantitatively ranked by *Keyantash and Dracup* [2002] using a set of weighted evaluation criteria. A thorough comparison of the ADI with the SPI and the SWSI will be the subject of a future paper.

10. Two Critical Droughts

[40] Time series of the ADI were constructed for each of the three climate divisions to characterize the magnitudes of various droughts. Particular foci are the droughts that occurred in California during water years 1976–1977 and from 1987–1992. Both of these droughts were of critical severity, and their effects reached beyond the borders of California to affect other parts of the nation; the 1977 drought extended over much of the western United States. Therefore these two critical drought periods were used to assess the fidelity of the ADI.

[41] The drought during water year 1976–1977 was a brief but extreme period of dryness, with 1977 having the distinction of being the driest year of record in California's history [*DWR*, 1993]. During the year, annual precipitation was 45 percent of normal, snowpack was a meager 22 percent of normal, and reservoir levels were 35 percent of mean storage. The 1976–1977 drought is best characterized as an extreme event that fortunately only persisted for a brief period.

[42] The relationship between drought severity, magnitude, and duration is displayed in Figure 6. Using this nomenclature, the 1987–1992 drought was a longer duration event, of lesser magnitude, than the 1976–1977 event. However, because of its duration, the severity of the 1987–1992 drought exceeded the 1976–1977 drought, depleting many of California's surface water reservoirs; total storage within 155 California reservoirs averaged two thirds of normal during the 6 years of drought [*DWR*, 1993]. Not surprisingly, the most severe hydrological drought year was near the end, when storage at the start of WY 1993 was 56 percent of normal. Several months later, in February 1993, the drought was declared over by the governor of California upon the rapid refilling of many surface water reservoirs to near-average levels. This occurred as a result of heavy precipitation during the first portion of the 1992–1993 winter, which elevated snowpack levels to 180 percent of normal [*DWR*, 1993].

[43] The chronological ADI series for each climate division, along with the intervals of critical drought, are displayed in Figure 7. Wet and dry spells are punctuated by intermittent months that exhibit anomalies of the opposite sign. The starting and stopping dates are subjective; objective determination of the initiation and termination points for droughts is an unsolved problem in hydrologic research. The adopted months are indicated by the vertical lines in Figure 7 and are specified in Table 3. The computed severities, durations, and magnitudes for the two critical droughts are also shown in Table 3.

[44] In all divisions, the most extreme drought magnitude $M_{\max}$ and mean magnitude M were larger during the 1976–

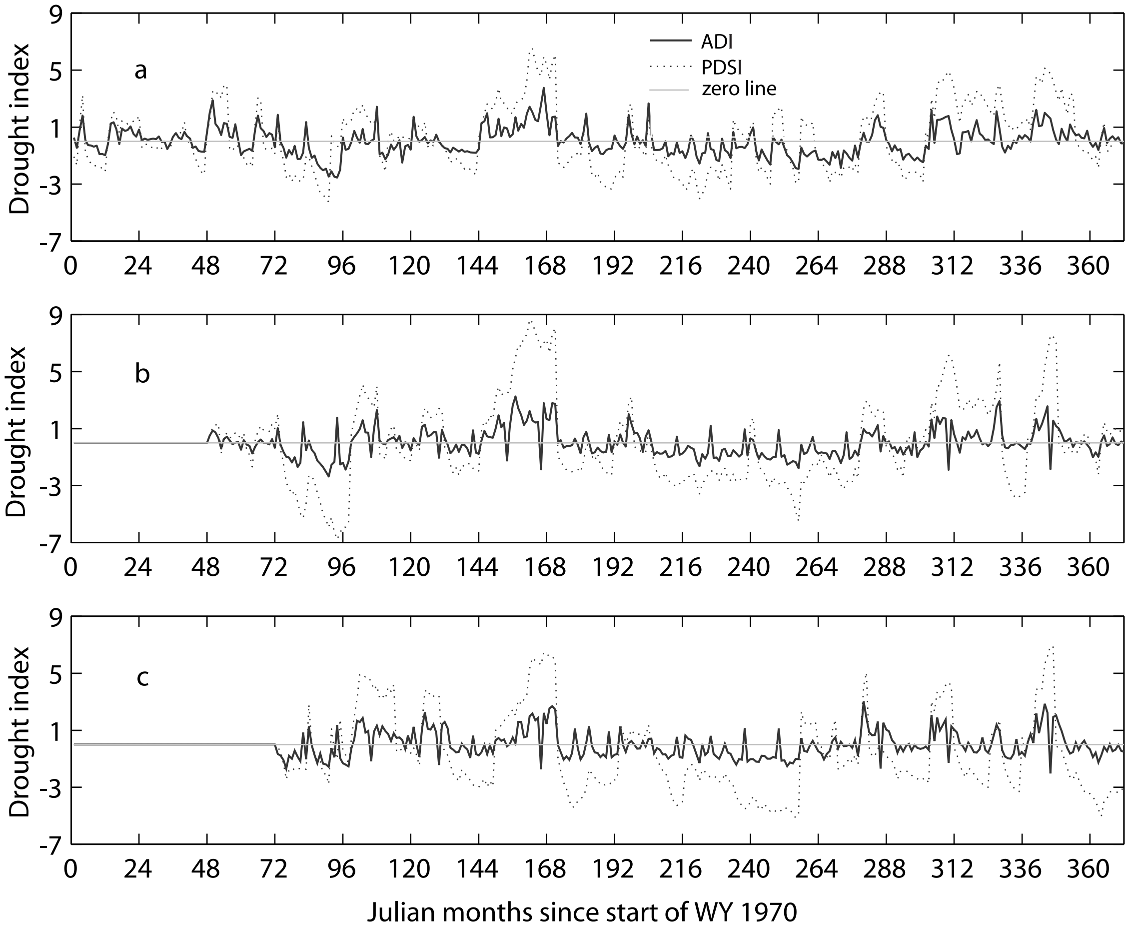

Figure 5. ADI and Palmer Drought Severity Index (PDSI) time series for the (a) North Coast (NC), (b) Central Valley (CV), and (c) South Coast (SC) climate divisions.

1977 drought than the 1987–1992 event. This behavior is consistent with historical hydrological records. Interestingly, Figure 7 indicates that the 1987–1992 drought may have begun as early as WY 1984 in the SC, where the fluctuating ADI during the few years before 1987 appears more noticeably negative than in the NC and CV divisions. Comparison with Figure 5 during this interval shows corroboration by the PDSI.

[45] The drought severity, indicated by the integrated area of the ADI, for each event was qualitatively appropriate in each division (i.e., $S_{1987-1992} > S_{1976-1977}$ for all divisions). The ADI also identifies periods of relative wetness; the most striking example of this is the elevated peak centered near Julian month 167, which represents the end of summer 1983 following the strong 1982–1983 El Niño event. Therefore, while the motivation and emphasis of this research is drought description, it should be noted that the ADI is equally suited to capture wet spells.

[46] The suite of ADI values may be interpreted probabilistically with empirical cumulative distribution functions (EDFs). Selected ADI percentiles may be used as thresholds for drought severity, which is the approach for rainfall deciles [*Keyantash and Dracup*, 2002; *Kinninmonth et al.*, 2000; *Gibbs and Maher*, 1967] and the SPI [*Keyantash and Dracup*, 2002; *Redmond*, 2000; *McKee et al.*, 1993]. Low percentiles are characteristic of drought.

[47] The SPI dryness thresholds used by the National Drought Mitigation Center are Gaussian variates of -2, -1.5, and -1 standard deviations. These correspond to the 2nd, 7th, and 16th percentiles. The range of values between the 16th and 84th percentiles are considered near-normal. These thresholds are shown with the ADI EDF for the CV in Figure 8.

11. Conclusion

[48] A multivariate, aggregate drought index (ADI) has been developed to describe drought based on local water deficiencies within the hydrologic cycle. Surface water storage, which is important in California to mitigate the effects of drought, is also considered as an anthropogenic variable by the ADI. In this way, the ADI uniquely describes drought beyond the traditional meteorological, hydrological, and agricultural subcategories. The broad drought perspective of the ADI is considered a strength of the approach; an aggregate description of drought intensity is an alluring complement to the regular distillation of drought into its meteorological, hydrological, and/or agricultural aspects.

[49] The ADI is constructed separately for each month, based on five to six hydrologic variables: precipitation P, evaporation E, streamflow Q, surface reservoir storage

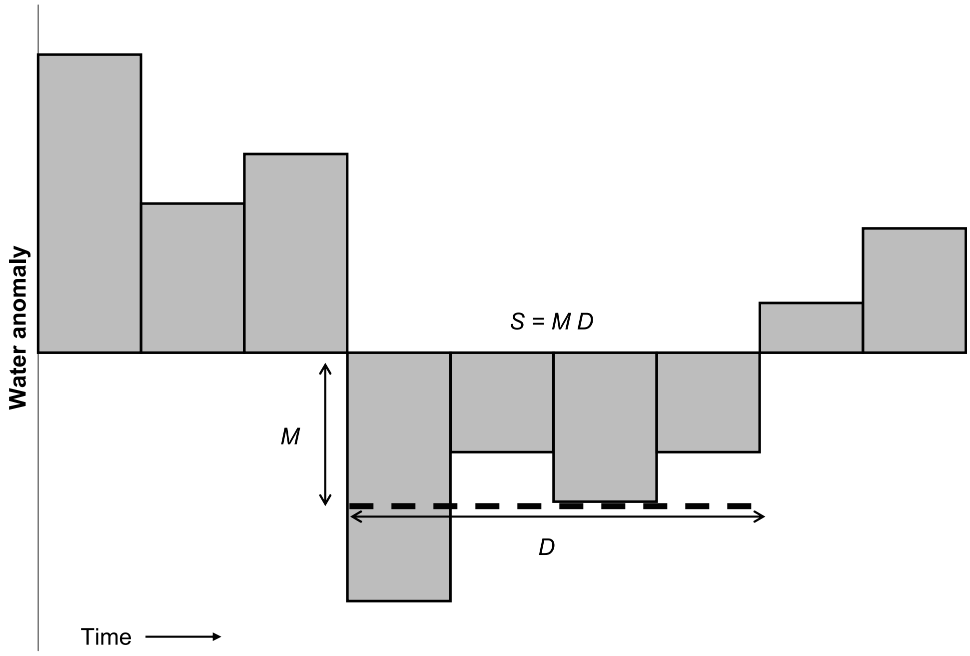

Figure 6. Drought severity *S* is the product of the average magnitude *M* and the drought duration *D*. Reprinted by permission from *Keyantash and Dracup* [2002].

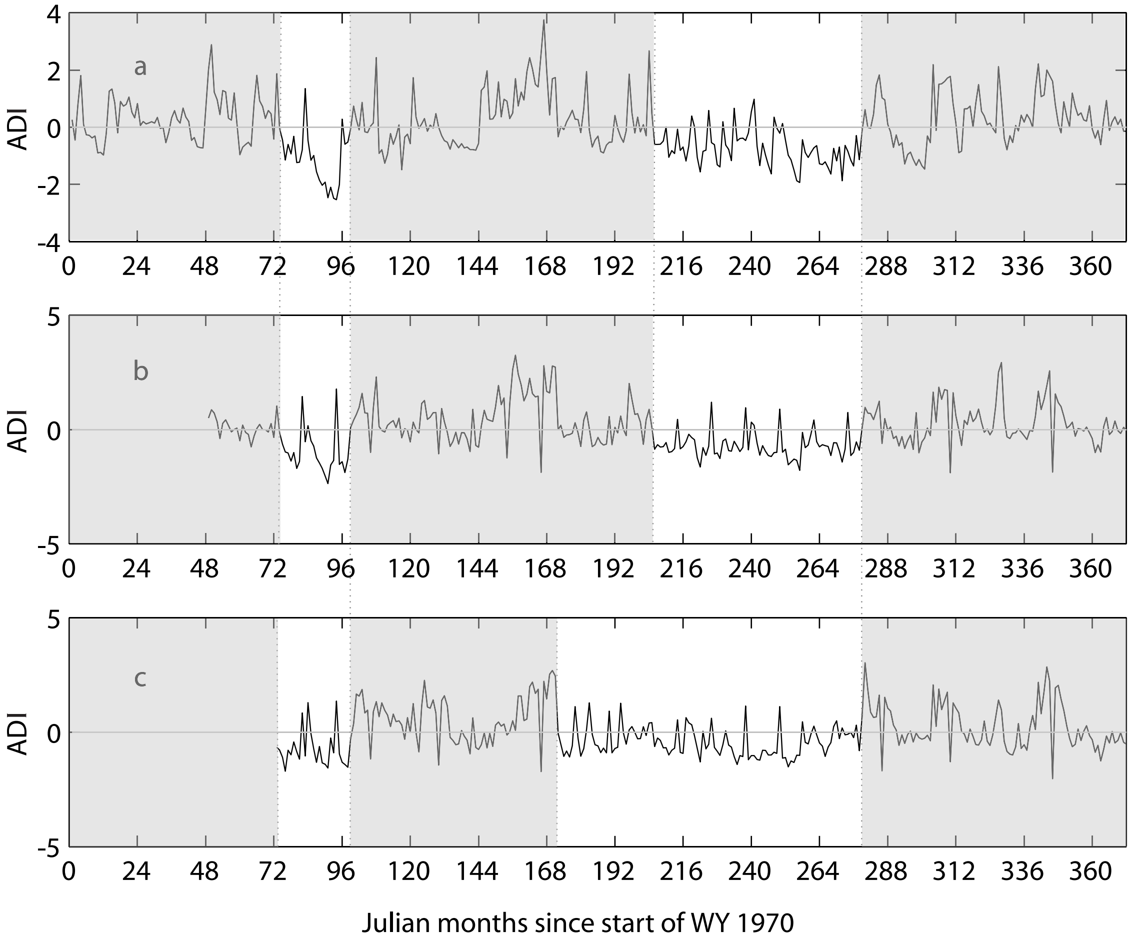

Figure 7. ADI time series for the (a) NC, (b) CV, and (c) SC climate divisions, beginning with water year 1970. Months within the 1976–1977 and 1987–1992 droughts have white backgrounds bordered by vertical dotted lines; the remaining portions of the time series are shaded gray. Where dotted lines cross multiple plots, it indicates that the divisional start/stop points were synchronous.

Table 3. ADI Properties of the 1976–1977 and 1987–1992 Droughts

	NC	CV	SC
	1976–1977 Drought		
Start month	Nov. 1975	Nov. 1975	Oct. 1975
End month	Dec. 1977	Dec. 1977	Dec. 1977
Maximum month	July 1977	April 1977	Jan. 1976
D, months	26	26	27
S	−27.9	−24.2	−19.7
M	−1.1	−0.9	−0.7
$M_{\max}$	−2.5	−2.4	−1.7
	1987–1992 Drought		
Start month	Nov. 1986	Nov. 1986	Jan. 1984
End month	Dec. 1992	Dec. 1992	Dec. 1992
Maximum month	Feb. 1991	Feb. 1991	Oct. 1990
D, months	74	74	108
S	−53.5	−50.9	−44.3
M	−0.7	−0.7	−0.4
$M_{\max}$	−1.9	−1.8	−1.5

V, soil moisture W, and snow water content s (if applicable). Principal component analysis is used to extract dominant hydrologic signals from correlations among the observational data. The ADI is the first principal component, normalized by the standard deviation of the monthly series. Twelve ADI series for three diverse California climate divisions were chronologically ordered and compared against state hydrologic records. The ADI successfully delineated two critical droughts in California, the 1976–1977 and 1987–1992 droughts, with severities and magnitudes qualitatively appropriate to the recorded severities of the events. The ADI was also compared against the PDSI and produced rank correlations ranging between 0.78 and 0.65. The distinct advantages of the ADI include its assessment of drought from the aggregate perspective of meteorological, hydrological, and agricultural water shortages, and its direct mathematical formulation, which can be rapidly applied to new observational data in a straightforward manner.

Appendix A: Mean and Standard Deviation of ADI

[50] The ADI is computed from standardized anomalies of hydrologic data with means of zero and unit variances. The ADI similarly possesses a mean of zero but generally has a nonunit standard deviation. These properties can be explained as follows.

A1. Mean of Zero

[51] Each column j of $\mathbf{X}$ is a column vector $\mathbf{x}_j$ of standardized anomalies for the jth hydrologic

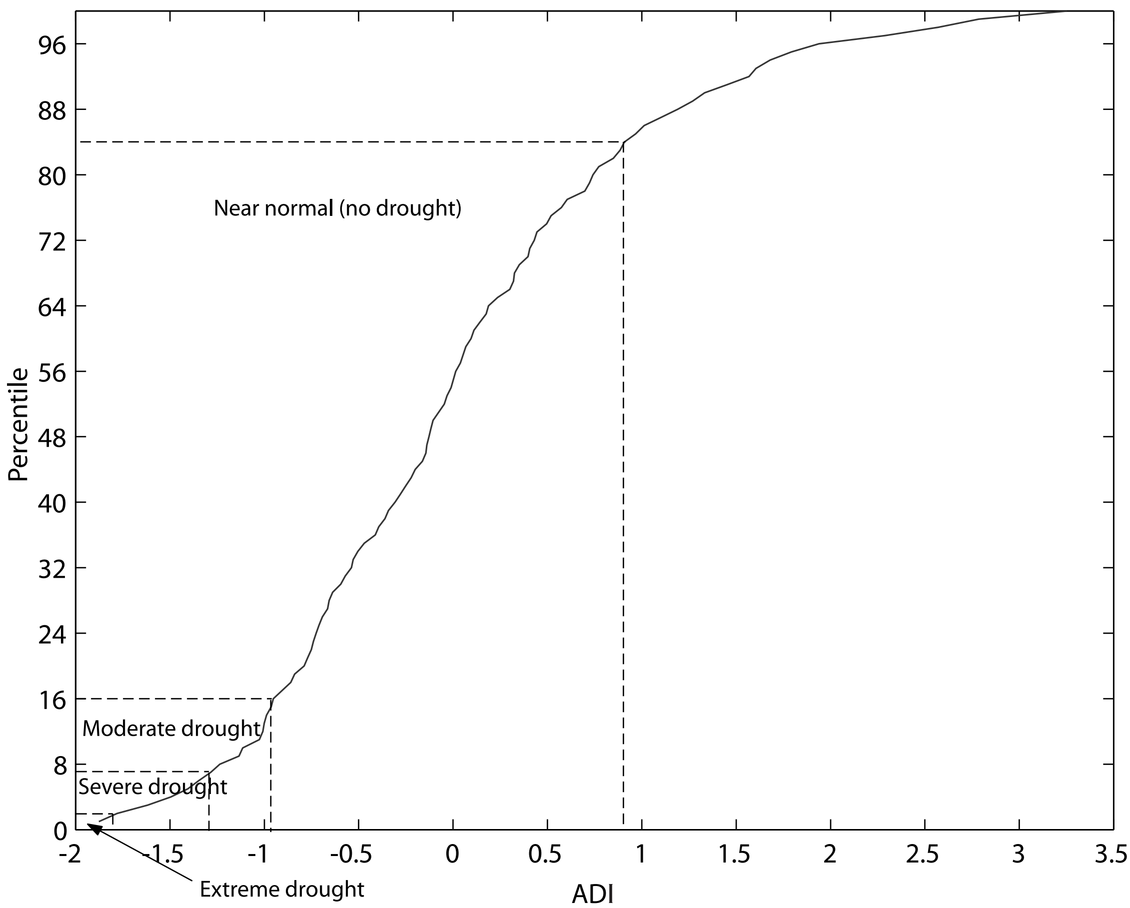

Figure 8. ADI percentiles for the CV. The horizontal dotted lines intersecting at the 2nd, 7th, 16th, and 84th percentiles represent the thresholds for extreme, severe, and moderate drought, and near-normal conditions, respectively. Vertical dotted lines show the corresponding ADI values in the CV.

variable. The sum of the anomalies is zero over all n observations:

$$\sum_{i=1}^{n} x_{ij} = 0. \tag{A1}$$

[52] The computation of a PC for time i is the dot product between a row of **X** and a retained eigenvector. Consider the calculation of the first two elements for PC1 (indicated as $\mathbf{z}_1$):

$$\begin{bmatrix} z_{11} \\ z_{21} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & x_{15} & x_{16} \\ x_{21} & x_{22} & x_{23} & x_{24} & x_{25} & x_{26} \end{bmatrix} \cdot \begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{41} \\ e_{51} \\ e_{61} \end{bmatrix}$$

$$z_{11} = x_{11}e_{11} + x_{12}e_{21} + x_{13}e_{31} + x_{14}e_{41} + x_{15}e_{51} + x_{16}e_{61}$$

$$z_{21} = x_{21}e_{11} + x_{22}e_{21} + x_{23}e_{31} + x_{24}e_{41} + x_{25}e_{51} + x_{26}e_{61}$$

Sum the terms and factor out the eigenvector elements:

$$z_{11} + z_{21} = e_{11}(x_{11} + x_{21}) + e_{21}(x_{12} + x_{22}) + e_{31}(x_{13} + x_{23}) + e_{41}(x_{14} + x_{24})e_{51}(x_{15} + x_{25}) + e_{61}(x_{16} + x_{26})$$

By induction, extend the pattern to include all n elements of $\mathbf{z}_1$:

$$\sum_{i=1}^{n} z_{i1} = \sum_{j=1}^{6} e_{j1} \sum_{i=1}^{n} x_{ij}$$

The $\sum_{i=1}^{n} x_{ij}$ term is given in equation (A1) as zero. Therefore

$$\sum_{i=1}^{n} z_{i1} = 0 \tag{A2}$$

The mean of the ADI is

$$\overline{a} = \frac{\sum_{i=1}^{n} a_i}{n} = \frac{\sum_{i=1}^{n} z_{i1}}{\sigma n}.$$

The numerator was given by equation (A2) as zero. Therefore the mean of the ADI vanishes:

$$\overline{a} = 0 \tag{A3}$$

A2. Nonunit Standard Deviation

[53] Consider the ADI sample standard deviation σ_a:

$$\sigma_a = \sqrt{\frac{\sum_{i=1}^{n} (a_i - \overline{a})^2}{n-1}}$$

Equation (A3) states $\overline{a} = 0$. For σ_a to be unit,

$$\sum_{i=1}^{n} a_i^2 = n - 1$$

This is possible, but unlikely, so σ_a is generally not 1.

Appendix B: Unit Vector Expression of the ADI

[54] The ADI was given by equation (3) as

$$\mathbf{a} = \frac{\mathbf{z}_1}{\sigma}. \tag{B1}$$

The sample standard deviation σ equals

$$\sigma = \sqrt{\frac{\sum_{i=i}^{n} (z_{i1} - \bar{z}_1)^2}{n-1}}$$

Equation (A2) demonstrated that the sum of all z_{i1} terms is zero. Since this sum forms the numerator in the computation of a mean for z_1, it is necessary that $\bar{z}_1 = 0$. Also note that

$$\sqrt{\sum_{i=1}^{n} z_{i1}^2} = \| \mathbf{z} \|$$

Therefore σ reduces to

$$\sigma = \frac{\| \mathbf{z}_1 \|}{\sqrt{n-1}}.$$

and the ADI is

$$\mathbf{a} = \frac{\sqrt{n-1}\mathbf{z}_1}{\| \mathbf{z}_1 \|}.$$

Any vector divided by its magnitude defines a unit vector in the direction of the original vector. Calling this unit vector $\hat{\mathbf{z}}_1$,

$$\mathbf{a} = \sqrt{n-1}\hat{\mathbf{z}}_1 \tag{B2}$$

To usefully compute the ADI through the alternate approach of equation (B2), $\hat{\mathbf{z}}_1$ should be calculated independent of $\mathbf{z}_1$:

$$\hat{\mathbf{z}}_1 = \frac{\mathbf{z}_1}{\| \mathbf{z}_1 \|} = \frac{\mathbf{X}\mathbf{e}_1}{\| \mathbf{X}\mathbf{e}_1 \|} = \frac{(\mathbf{O} - \mathbf{M})\mathbf{D}\mathbf{e}_1}{\| (\mathbf{O} - \mathbf{M})\mathbf{D}\mathbf{e}_1 \|}, \tag{B3}$$

where

- **X** $= n \times p$ matrix of standardized observational data;
- **O** $= n \times p$ matrix of original observational data (may be nonstandardized, and in multiple measurement units);
- **M** $= n \times p$ matrix of observational data means (every row is identical);
- **D** $= p \times p$ diagonal matrix with the inverse of the sample standard deviation for each hydrologic variable x_i (i.e., $\sigma_{x_i}^{-1}$, $1 \le i \le p$) residing upon each diagonal element $d_{i,i}$;
- $\mathbf{e}_1$ $= p \times 1$ first eigenvector of the observational data anomalies.

Table C1. Vector Formulae for the ADI

			ADI Formula Incorporating Vector (Equation Numbers Given in Parentheses)	
Vector	Dimension	Magnitude	Baseline $\mathbf{a}$ =	Update $\mathbf{a_{new}}$ =
$\mathbf{e_1}$	p	1	$\frac{\mathbf{Xe_1}}{\sigma}$ (4)	$\frac{\mathbf{X_{new}e_1}}{\sigma}$ (5)
$\hat{\mathbf{z}}_1$	n	1	$\sqrt{n-1}\hat{\mathbf{z}}_1$ (B2)	$\mathbf{X_{new}}\ (\mathbf{X^TX})^{-1}\mathbf{X^T}\sqrt{n-1}\hat{\mathbf{z}}_1$ (C2)
$\mathbf{z}_1$	n	$\sigma\sqrt{n-1}$	$\frac{\mathbf{z}_1}{\sigma}$ (3)	none
$\mathbf{a}$	n	$\sqrt{n-1}$	$\mathbf{a}$	$\mathbf{X_{new}Ca}$ (C3)

The need to compute $\mathbf{M}$ directly can be obviated by recognizing

$$\mathbf{M} = \frac{1}{n}\mathbf{1O}, \tag{B4}$$

where
$\mathbf{1}$ = $n \times n$ matrix of ones.

[55] Inserting equation (B4) into (B3) and factoring out $\mathbf{O}$,

$$\hat{\mathbf{z}}_1 = \frac{\left(\mathbf{I}-\frac{1}{n}\mathbf{1}\right)\mathbf{ODe_1}}{\left\|\left(\mathbf{I}-\frac{1}{n}\mathbf{1}\right)\mathbf{ODe_1}\right\|} = \frac{\left(\mathbf{I}-\frac{1}{n}1\right)\mathbf{ODe_1}}{\sqrt{\left[\left(\mathbf{I}-\frac{1}{n}\mathbf{1}\right)\mathbf{ODe_1}\right]^{\mathbf{T}}\left(\mathbf{I}-\frac{1}{n}\mathbf{1}\right)\mathbf{ODe_1}}}, \tag{B5}$$

where
$\mathbf{I}$ = $n \times n$ identity matrix;
$\mathbf{1}$ = $n \times n$ matrix of ones;
$\mathbf{O}$ = $n \times p$ matrix of original observational data (may be nonstandardized, and in multiple measurement units);
$\mathbf{D}$ = $p \times p$ diagonal matrix with the inverse of the sample standard deviation for each hydrologic variable x_i (i.e., $\sigma_{x_i}^{-1}$, $1 \le i \le p$) residing upon each diagonal element $d_{i,i}$;
$\mathbf{e}_1$ = $p \times 1$ first eigenvector of the observational data anomalies.

Appendix C: Operational ADI in Terms of $\hat{\mathbf{z}}_1$

[56] The reference and operational versions of the ADI may be computed using either of two unit vectors: $\mathbf{e}_1$ (p terms) or $\hat{\mathbf{z}}_1$ (n terms). The $\mathbf{e}_1$ and $\hat{\mathbf{z}}_1$ unit vector expressions for the reference ADI are given by equations (4) and (B2), respectively. The operational formula for the ADI given in equation (5) invokes the first eigenvector, but $\mathbf{a}_{\text{new}}$ does not have an immediate $\hat{\mathbf{z}}_1$ analogy. However, such a relationship can be derived. Begin by setting the two unit vector expressions for $\mathbf{a}$ (equations (4) and (B2)) equal:

$$\frac{\mathbf{Xe_1}}{\sigma} = \sqrt{n-1}\hat{\mathbf{z}}_1$$

Premultiply by $\mathbf{X^T}$ to create a square matrix, and then premultiply by $(\mathbf{X^TX})^{-1}$ to produce an identity matrix $\mathbf{I}$ on the left-hand side:

$$\mathbf{I}\frac{\mathbf{e_1}}{\boldsymbol{\sigma}} = \left(\mathbf{X^TX}\right)^{-1}\mathbf{X^T}\sqrt{n-1}\hat{\mathbf{z}}_1 \tag{C1}$$

$\mathbf{I}$ may be omitted without altering the expression. Use equation (C1) to substitute for $\mathbf{e_1}/\sigma$ in equation (5):

$$\mathbf{a_{new}} = \mathbf{X_{new}}\left(\mathbf{X^TX}\right)^{-1}\mathbf{X^T}\sqrt{n-1}\hat{\mathbf{z}}_1 \tag{C2}$$

The $(\mathbf{X^TX})^{-1}\mathbf{X^T}$ term can be considered a $p \times n$ matrix of constants $\mathbf{C}$. Using equation (B2),

$$\mathbf{a_{new}} = \mathbf{X_{new}Ca} \tag{C3}$$

Table C1 summarizes the vector expressions for the ADI.

References

Alley, W. M. (1984), The Palmer Drought Severity Index: Limitations and assumptions, *J. Clim. Appl. Meteorol.*, *23*, 1100–1109.

Barnston, A. G., and R. E. Livezey (1987), Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, *Mon. Weather Rev.*, *115*, 1083–1126.

Cattell, R. B. (1952), *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist*, 462 pp., Greenport Press, Westport, Conn.

Department of Water Resources (DWR) (1993), California's 1987–92 drought: A summary of six years of drought, report, 38 pp., Calif. Dep. of Water Resour., Sacramento.

Department of Water Resources (DWR) (2003), Water conditions in California, *Bull. 120*(1), 16 pp., Calif. Dep. of Water Resour., Sacramento.

Dracup, J. A., K. S. Lee, and E. G. Paulson Jr. (1980), On the definition of droughts, *Water Resour. Res.*, *16*, 297–302.

Draper, A. J., M. W. Jenkins, K. W. Kirby, J. R. Lund, and R. E. Howitt (2003), Economic-engineering optimization for California water management, *J. Water Resour. Plann. Manage.*, *129*(3), 155–164.

Garen, D. C. (1992), Revised surface-water supply index for the western United States, *J. Water Resour. Plann. Manage.*, *119*, 437–454.

Gibbs, W. J., and J. V. Maher (1967), Rainfall deciles as drought indicators, *Bull. 48*, 37 pp., Aust. Bur. of Meteorol., Melbourne, Aust.

Guttman, N. B., J. R. Wallis, and J. R. M. Hosking (1992), Spatial comparability of the Palmer Drought Severity Index, *Water Resour. Bull.*, *28*, 1111–1119.

Haan, C. T. (1977), *Statistical Methods in Hydrology*, 378 pp., Iowa State Univ. Press, Ames.

Heim, R. R., Jr. (2000), Drought indices: A review, in *Drought: A Global Assessment*, *Hazards Disasters Ser.*, vol. I, edited by D. A. Wilhite, pp. 159–167, Routledge, New York.

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

Keyantash, J. A. (2001), Formulation of a multivariate drought index, with analysis for hydroclimatic regions of California, Ph.D. dissertation, 244 pp., Univ. of California, Los Angeles.

Keyantash, J., and J. A. Dracup (2002), The quantification of drought: An analysis of drought indices, *Bull. Am. Meteorol. Soc.*, *83*(8), 1167–1180.

Kinninmonth, W. R., M. E. Voice, G. S. Beard, G. C. de Hoelt, and C. E. Mullen (2000), Australian climate services for drought management, in *Drought: A Global Assessment*, vol. I, edited by D. A. Wilhite, pp. 210–222, Routledge, New York.

Lins, H. F. (1997), Regional streamflow regimes and hydroclimatology of the United States, *Water Resour. Res.*, *33*(7), 1655–1667.

McKee, T. B., N. J. Doesken, and J. Kleist (1993), Drought monitoring with multiple timescales, in *Proceedings of the Eighth Conference on Applied Climatology*, pp. 179–184, Am. Meteorol. Soc., Boston, Mass.

National Climatic Data Center (NCDC) (2000a), Time bias corrected division temperature-precipitation-drought index data, *TD-9640*, Asheville, N. C., Oct.

National Climatic Data Center (NCDC) (2000b), Climate division: Temperature-precipitation-drought data, Asheville, N. C., Oct. (Available at http://www.ncdc.noaa.gov/onlineprod/drought/images/ca.gif.)

National Climatic Data Center (NCDC) (2001a), Monthly surface data (primarily U.S.), *TD-3220*, Asheville, N. C., June.

National Climatic Data Center (NCDC) (2001b), Daily surface data (primarily U.S.), *TD-3200*, Asheville, N. C., July.

Palmer, W. C. (1965), Meteorological drought, *Res. Pap. 45*, 58 pp., Weather Bur., U.S. Dep. of Commer., Washington, D. C.

Preisendorfer, R. W. (1988), *Principal Component Analysis in Meteorology and Oceanography*, 425 pp., Elsevier Sci., New York.

Redmond, K. T. (2000), Integrated climate monitoring for drought detection, in *Drought: A Global Assessment*, *Hazards Disasters Ser.*, vol. I, edited by D. A. Wilhite, pp. 145–158, Routledge, New York.

Ross, T., and N. Lott (2000). A climatology of recent extreme weather and climate events, *Tech. Rep. 2000–02*, Natl. Clim. Data Cent., Asheville, N. C.

Shafer, B. A., and L. E. Dezman (1982), Development of a Surface Water Supply Index (SWSI) to assess the severity of drought conditions in snowpack runoff areas, in *Proceedings of the Western Snow Conference*, pp. 164–175, Colo. State Univ., Fort Collins.

Wilhite, D. A. (2000), Drought as a natural hazard: Concepts and definitions, in *Drought: A Global Assessment*, *Hazards Disasters Ser.*, vol. I, edited by D. A. Wilhite, pp. 3–18, Routledge, New York.

Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences: An Introduction*, *Int. Geophys. Ser.*, vol. 59, 467 pp., Academic, San Diego, Calif.

World Meteorological Organization (WMO) (1975), Drought and agriculture: Report of the CAgM working group on the assessment of drought, 127 pp., *Tech. Note 138*, Geneva.

J. A. Dracup, Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA 94720, USA.

J. A. Keyantash, Department of Earth Sciences, California State University, Dominguez Hills, 1000 E. Victoria Street, Carson, CA 90747, USA. (jkeyantash@csudh.edu)

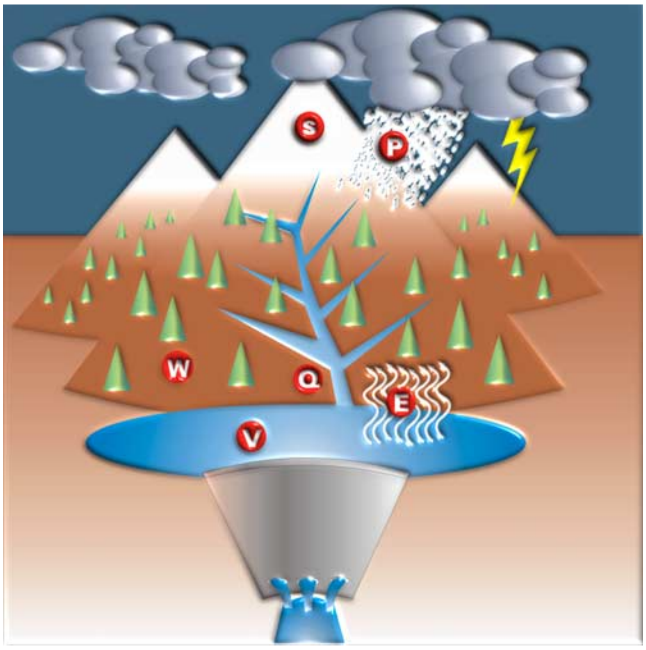

Figure 1. Surface water balance as described by the aggregate drought index (ADI). Precipitation P and snow water content s contribute to streamflow Q, some of which is captured as reservoir storage V. Evapotranspiration E stems from surficial evaporation and vegetative transpiration, both of which diminish the surface soil water content W.