**MEMORANDUM**

11 January 2006

**Submitted To:** Michelle Light, MP-700

**By:** Levi Brekke, D-8520

**Subject:** **Drought Reoccurrence Analysis for the Stanislaus River**

---

## 1.0 INTRODUCTION

Reclamation's Stanislaus Policy Group has expressed an interest in understanding the reoccurrence intervals of critical droughts in the Stanislaus Basin. Such information is expected to support a decision on whether a New Melones Revised Plan of Operation should be based on the critical drought observed during 1928-1934 or 1987-1992.

Framing questions:

1. What are the reoccurrence intervals of 6-year droughts of varying severity in the Stanislaus Basin?
2. How do these reoccurrence intervals change if the analysis is based on an earlier period of record (e.g., pre-operations hydrologic record)?
3. How do these reoccurrence intervals change if the analysis is based on a precipitation- rather than a runoff-defined drought?

For question 1, the focus is on 6-year drought reoccurrence rather than 7-year drought reoccurrence (i.e. similar in duration to the 1928-1934 "drought"). The 6-year duration was selected mainly because it coincides with the 1987-1992 duration. Also, review of 1928-1934 annual Stanislaus runoff sequence (**Figure 1**) shows that the 1928-1934 drought might actually be described as back-to-back droughts (i.e. a 4-year drought and a 2-year interrupted by a wetter year in 1932), rather than as a 7-year drought.

For question 2, it reasoned that when New Melones construction was being completed, an original plan of operations was still being developed. Planning at that time was informed by hydrologic observations that did not include the severe 1987-1992 drought. It is of interest to understand the severity of the 1987-1992 relative to this pre-operations understanding of hydrology.

For question 3, it is reasoned that as multi-year droughts persist, the runoff response to precipitation may decay as basin infiltration potential increases. Thus, a normal precipitation year might produce less runoff if the given year follows multiple dry years rather than a wet year.

The remainder of this memorandum is organized as follows:

- Section 2.0 – methodology for assessing drought reoccurrence.

- Section 3.0 – description of cases that were studied, varying by basis period and whether the drought is defined relative to precipitation or runoff observations.
- Section 4.0 – results and discussion on the framing questions.
- Section 5.0 – analysis limitations.
- Section 6.0 – summary.

Supplemental details on the methodology are provided in Appendices A and B. Supplemental results graphics for each of the case studies are provided in Appendix C.

## 2.0 METHODOLOGY

The methodology is designed to reveal observed and theoretical probabilities of drought reoccurrence, given either a runoff- or precipitation-basis for defining drought. The methodology includes three primary steps:

- Define drought.
- Analyze drought reoccurrence based on observed flow data.
- Analyze drought reoccurrence based on synthetic flow data modeled during a longer period, where the flow model is designed to produce a synthetic flow time series exhibiting statistical consistency with the observed flow time series.

### 2.1 Drought Definition

Droughts can be described as meteorological or hydrological phenomena. They are often measured relative to median conditions, and can be expressed in terms of spell (i.e. duration of below-median conditions), severity (i.e. cumulative deficit during spell), or intensity (i.e. severity divided by spell) (**Frick et al. 1990**). Both meteorological and hydrologic droughts are considered in this analysis, with each consistently defined as severity during several predetermined durations: 2-, 3-, 4-, 5- and 6-year spells. Grant it, only the results from studying 6-year droughts are used to address the framing questions. However, additional context for the 6-year drought reoccurrence is provided by also studying 2- to 5-year droughts.

To illustrated subsequent methodology steps, consider *hydrologic* drought defined as *severity* relative observed annual "full-natural" flow in the Stanislaus River at Goodwin (**Figure 1**). These data are reported by the California Data Exchange Center (http://cdec.water.ca.gov/, station I.D: SNS). Flow data from water years 1901-2004 are shown (N = 104).

### 2.2 Drought Analysis based on Observed Data

The methodology to compute reoccurrence distributions for the various assumed drought spells is as follows:

- from the observed time series with N years of data, compute running n-year sums of observed annual flow, where n = 2, 3 … 6 years. The reporting year for the n-year sum is the end-year of the sum.
- compute running n-year deficits relative to respective median n-year sums.
- compute rank-based return-period (RP) plotting positions, where:

$$\textbf{(1)}\quad \begin{aligned} rank &= 1 \ldots (N-n+1) \\ RP &= (N-n+1)/rank \end{aligned}$$

- Sort running n-year deficits from largest surplus (i.e. most negative "deficit") to largest deficit (i.e. most positive deficit) and plot versus return-period (**Figures 2-6**). The plots show drought reoccurrence distributions illustrating the "observed" return periods of the assumed n-year drought severity.
- On each n-year curve, plot several notable observed droughts to provide a sense for how historical droughts rate within the observed distribution. In the case of this example, notable observed droughts include those of n-year duration ending on 1931, 1934, 1977, 1990, 1991, 1992, and 1994.

Since the analysis is based on N=104 years of observations, the worst drought of record for each n-year duration would have an observed return-period of (104-n+1) years. For example, **Figure 6** might be used to support the statements:

- The 6-year drought ending on 1992 has a severity of 3971 TAF and might be expected to occur once in every 99 years.
- The 6-year drought ending on 1934 has a severity of 3016 TAF and might be expected to occur once in every 50 years.

These statements are only true if we can assume that the observed flows from 1901-2004 represent the *true* distribution of Stanislaus annual flow. This assumption is challenged in the next analysis step.

## 2.3 Drought Analysis based on Synthetic Data

In the example of Section 2.2, it is difficult to assess droughts having return periods of 50 to 100 years because such episodes would only have infrequent or singular occurrence in the observed record (i.e. N=104 years). Given this, it is assumed that the reoccurrence distributions of **Figures 2-6** and observed return periods of the notable droughts may be inaccurate. To test this assumption, the preceding methodology was re-applied using a synthetic time series of Stanislaus annual flow, modeled to be statistically consistent with the observed data.

To do this analysis, a synthetic flow model must be developed. The initial model concept was as follows:

**(2)** Modeled flow = Explanatory Term(s) + Error, where "Explanatory Term(s)" accounts for non-random flow variations and "Error" accounts for the random component of the flow.

<u>2.3.1 Synthetic Model Development – Explanatory Term:</u>

It is common in hydrologic time series data to find observations that are correlated with their own values from previous time periods (Haan 1977). Such a phenomenon is referred to as auto-correlation, and varies with the "lag" between time periods. Given significant auto-correlation, one might say that "persistence" exists in the hydrologic time series, or that the time series has "memory".

Using the example of Section 2.2, auto-correlation was analyzed in the Stanislaus annual flow time series at p-year lags, where p = 1, 2 … 6 **(Figure 7).** The correlations were tested for statistical significance. The hypothesis for these tests is that the true p-year lag correlation is zero despite the computed correlation. This hypothesis can be rejected at a specified level of confidence. A 95% level of confidence was used in this analysis, and in this example leads to rejection of the hypothesis for only the 6-year lag condition. Thus, only a lag-6 year auto-regressive variable was retained for further consideration in flow model development.

The next step was to graphically and statistically analyze the lag-6 relationship (**Figure 8**). Statistically, the annual flow from 6-years ago explains very little of the current year's flow variability (r-square = 0.04). Based on this result, it appears that the basis for including a lag-6 auto-regressive Explanatory Term in the model is weak. Consequently, model development proceeds in this example with omission of the Explanatory Term.

<u>2.3.2 Synthetic Model Development – Error Term:</u>

Given results from the preceding discussion, the synthetic flow model can be simplified:

**(3)** Modeled flow = Error,

In this case, there is no need to isolate the random component of the observed time series; the entire time series is treated as the random component. The model is then constructed and applied as follows, understanding the flow time series to be a random variable:

(a) treat the flow time series as a "data pool" and fit a probability density function (PDF) to the data. Convert the PDF into a cumulative distribution function (CDF).

(b) construct an M-period time series of randomly sampled values from a uniform distribution between 0 and 1. Treat these values as sampling probabilities.

(c) construct an M-period time-series of synthetic flow values by sampling from the cumulative distribution function from (a) at the sampling probabilities from (b).

The synthetic flow period (M) should be far greater than the observed flow period (N). Specifying the distribution fit in (a) and applying it in (c) requires some judgment (see Appendix A for details).

Continuing with the example from Section 2.3.1, the approach for Error modeling was implemented with M=100,000 years and with a nonparametric approach to assuming the

PDF (see Appendix A). The observed data's distribution (i.e. a histogram), the fitted PDF, and the PDF converted into a CDF are shown, respectively, on **Figures 9-11**.

The reasonability of the synthetic flow time series was then judged by plotting century time-slices from the synthetic series with an overlay of the observed series. Doing this with our example shows that the synthetic data spread and variability is comparable to observed (**Figure 12**). Also, a re-generation of the PDF and CDF based on the synthetic rather than observed data suggests that the sampling procedure produces a synthetic flow distribution that is comparable to observed (**Figures 13 and 14**).

<u>2.3.3 Drought Analysis on the Modeled Synthetic Flow</u>:

The drought-analysis methodology of Section 2.2 was applied to the synthetic flow data to reveal synthetic reoccurrence distributions for n-year droughts (**Figures 15-19**). On each n-year curve, the notable *observed* n-year droughts from **Figures 2-6** are also shown as an overlay. These results support the following types of statements:

- The 6-year drought ending on 1992 has a severity of 3971 TAF and might be expected to occur once in approximately 450 years.
- The 6-year drought ending on 1934 has a severity of 3016 TAF and might be expected to occur once in approximately 50 years.

In general, the most extreme, observed, 5- and 6-year droughts appear to have larger return periods according to the synthetic reoccurrence distributions (**Figures 18-19**) compared to return periods according to the observed reoccurrence distributions (**Figures 5-6**).

## 3.0 APPLICATION

The methodology was applied for several cases varying by drought definition and period of observed record:

- Case A - Flow1: Based on annual full-natural flow in the Stanislaus River during WY1901-2004 (**Figure 1**) (data i.d. SNS on the California Data Exchange Center (CDEC) at http://cdec.water.ca.gov).
- Case B - Flow2: Based on annual full-natural flow in the Stanislaus River during WY1901-1980.
- Case C - Flow3: Based on annual full-natural flow in the Stanislaus River during WY1906-2003.
- Case D - PrecipSOR: Based on annual precipitation amount at station "Sonora RS" (CDEC i.d. SOR) during WY1906-2003.
- Case E - PrecipYSV: Based on annual precipitation amount at station "Yosemite Headquarters" (CDEC i.d. YSV) during WY1906-2003.
- Case F - PrecipNFR: Based on annual precipitation amount at station "North Fork R.S." (Upper San Joaquin Basin, CDEC i.d. NFR) during WY1906-2003.

- Case G – PrecipIndex1: Based on annual precipitation index (Appendix B) representing stations spanning the American to Upper San Joaquin basins during WY1906-2003 (CDEC i.d. CLF, AUB, PCV, FLD, SOR, NFR, YSV).
- Case H – PrecipIndex2: Based on annual precipitation index (Appendix B) representing stations spanning the Stanislaus to Upper San Joaquin basins during WY1906-2003 (CDEC i.d. SOR, NFR, YSV).

The framing questions are addressed by considering results from the following cases: question 1 – collectively consider results from Cases A and C; question 2 – compare results between Cases A and B; question 3 – compare results between Case C and the collective of Cases D-H. For each case, a standard set graphics was produced (i.e. **Figures 1-19**). Case A results are depicted by **Figures 1-19**. Graphics for Cases B-H are provided in Appendix C

## 4.0 RESULTS

Tables 1 and 2 summarize case-specific results on observed and synthetic reoccurrence of the historical 6-year droughts ending on 1934 and 1992.

**Table 1**: Observed Reoccurrence Interval of 6-Year Droughts (Years)

| Case | Description | 1929-1934 Drought | 1987-1992 Drought |
|---|---|---|---|
| A | Flow1 | 50 | 99 |
| B | Flow2 | 75 | n/a (1) |
| C | Flow3 | 50 | 93 |
| D | PrecipSOR | 31 | 93 |
| E | PrecipNFR | 47 | 93 |
| F | PrecipYSV | 31 | 47 |
| G | PrecipIndex1 | 47 | 93 |
| H | PrecipIndex2 | 47 | 93 |

Notes:
(1) Period of observed record was WY1901-1980 and did not include this drought.

**Table 2:** Synthetic Reoccurrence Intervals of 6-Year Droughts (Years)

| Case | Description | 1929-1934 Drought | 1987-1992 Drought |
|---|---|---|---|
| A | Flow1 | 50 | 433 |
| B | Flow2 | 67 (1) | 719 (1) |
| C | Flow3 | 36 | 258 |
| D | PrecipSOR | 25 | 199 |
| E | PrecipNFR | 53 | 68 |
| F | PrecipYSV | 20 | 23 |
| G | PrecipIndex1 | 49 | 56 |
| H | PrecipIndex2 | 46 | 108 |

Notes:
(1) The 1929-1934 and 1987-1992 droughts defined in Case A were overlaid on the synthetic reoccurrence distributions of Case B to arrive at these synthetic reoccurrence intervals.

Discussion of results in relation to the framing questions:

- Question 1: based on a hydrologic drought definition, the apparent reoccurrence of the 1987-1992 drought appears to be approximately once in every 250 to 450 years. In contrast, the apparent reoccurrence of the 1929-1934 drought appears to be once in every 30 to 50 years.
- Question 2: staying with the hydrologic drought definition, truncation of the period of observed record from WY1901-2004 to WY1901-1980 leads to a greater apparent reoccurrence interval for both the 1929-1934 and 1987-1992 droughts. The apparent reoccurrence of the 1987-1992 drought increases to as much as once in approximately 700 years.
- Question 3: switching to the precipitation drought definition, the apparent reoccurrence of the 1987-1992 drought is less than the reoccurrence based on runoff-drought (Case C). For example, the station-based precipitation definitions led to reoccurrence estimates of once in every 199 years at Sonora, once in every 68 years at North Fork San Joaquin, and once in every 23 years at Yosemite Valley. The region-based precipitation definitions led to reoccurrence estimates of once in every 56 to 108 years, with the reoccurrence appearing greater as the index reflected relatively more influence from the Sonora station.

For precipitation-based drought, it is interesting to note how the results depended on station locations (i.e. Cases D-F in Table 2). A rank-percentile analysis of annual precipitation amounts from these stations during 1906-2003 (shown on Figures D1, E1 and F1 in Appendix C) reveals that the dryness relative to station-specific variability was more persistent at the Sonora gage than at the other two (**Table 4**).

**Table 4**: Rank-Percentile of Annual Precipitation Amounts relative to 1906-2003 Record

| Water Year | Station Name (I.D.) [1] | | |
|---|---|---|---|
| | Sonora R S (SOR) | North Fork R S (NFR) | Yosemite Headquarters (YSV) |
| 1987 | 3 | 4 | 4 |
| 1988 | 6 | 18 | 16 |
| 1989 | 27 | 28 | 33 |
| 1990 | 20 | 15 | 29 |
| 1991 | 8 | 27 | 43 |
| 1992 | 25 | 38 | 35 |

Notes:
(1) California Data Exchange Center (http://cdec.water.ca.gov)

## 5.0 LIMITATIONS

The synthetic return periods computed in this analysis are sensitive to a number of factors, including:

- choice of drought definition,
- procedure of drought measurement,
- assumptions in synthetic flow and precipitation modeling,
- decision in synthetic flow and precipitation modeling to constrain sampling so that the fitted distribution was not sampled at probabilities less than 0.01
- decision in synthetic flow and precipitation modeling to generate synthetic record of 100,000 years rather than a longer period of some other duration.
- quality of underlying flow and precipitation data

The first two limitations can be regarded as caveats for this analysis. The analysis could be repeated with a different drought definition and method of measurement.

The third limitation could be explored by a more exhaustive survey of potential synthetic flow models. It is possible that a superior synthetic flow model could be identified. However, it is not expected that another synthetic flow model would affect the conclusion from this analysis that there is significant reoccurrence difference between the 1929-1934 and 1987-1992 droughts.

The fourth limitation leads to a synthetic flow time series that when subjected to nonparametric density fitting, has anomalous probability assignment at the flow associated with the 0.01 cumulative probability (**Figure 13**). However**,** this effect on the synthetic PDF does not seem to create a synthetic CDF that differs significantly CDF fit to observed data (**Figure 14**). Thus, the results seem benign to this limitation.

To explore sensitivity to the fifth limitation, the random sequence of sampling probabilities used to generate Case C and D was permuted 7 times. The resultant range of synthetic reoccurrence of the 1987-1992 precipitation drought (Case D) was 186 to 212 years with a median of 198 years. For runoff drought (Case C), the range was 232 to 284 with a median of 252 years. Thus, the sampling uncertainty interval is approximately +/- 15 years for the precipitation droughts and +/- 25 years for the runoff droughts.

Finally, it is necessary to acknowledge the sixth limitation that this analysis assumes accurate annual precipitation measurements at the surveyed CDEC stations, and accurate estimates of annual full natural flow at Goodwin (also as reported by CDEC). Quality review of these data was not scoped in this analysis.

## 6.0 SUMMARY

A drought reoccurrence analysis for the Stanislaus River was conducted. Drought was defined relative to Stanislaus flow and regional precipitation observations. The analysis was repeated for numerous cases: cases A-C using annual flow observations but from different periods of record (i.e. 1901-2004 (information to support modern-day planning), 1901-1980 (information that would have supported New Melones pre-operations planning), and 1906-2003 (the period coinciding with available precipitation data)), and cases D-H using annual precipitation station and index observations.

Drought was measured by severity (i.e. cumulative deficit measured relative to median condition during the period of record). Drought severity was assessed for 2-year to 6-year spells, with framing questions based on results from the 6-year drought analyses. Reoccurrence intervals for droughts were first evaluated against observed data and then again against synthetic data in an effort to analyze a longer period of record The synthetic data were modeled to be statistically consistent with the observed data.

Results from Cases A-H were used to address the three questions from the introduction. If drought is defined by Goodwin flow during 1906-2003 (Case C), then the 1987-1992 drought could be expected to reoccur once in every 233 to 283 years based on results from Section 4 and the estimated sampling uncertainty of Section 5. However, this range of reoccurrence is sensitive to the period of observed record (Cases A, B, and C). For example, ignoring post-1980 observations suggests a reoccurrence interval that is significantly greater.

If drought is defined by Sonora precipitation rather than Goodwin flow, then the 1987-1992 drought could be expected to have a reduced reoccurrence interval (i.e. once in every 184-214 years, based on Case D results from Section 4 and the estimated sampling uncertainty of Section 5. The fact that the precipitation-based reoccurrence estimate is less than the runoff-based estimate supports the reasoning behind framing question 3 (i.e. that runoff response to precipitation might decay as multi-year droughts persist).

Limitations on this analysis include assumptions related to drought definition, drought measurement, synthetic flow modeling, model application, and data.

## 6.0 REFERENCES

Frick, D.M , Bode, D. and Salas, J.D., 1990. "Effect of Drought on Urban Water Supplies. I: Drought Analysis," *Journal of Hydraulic Engineering*, 116(6), 733-753.

Haan, C.T., 1977. Statistical Methods in Hydrology, *The Iowa State University Press*, Ames, IA, 378pp.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences, *Academic Press*, San Diego, CA, 467pp.

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lutkepohl, and T.-C. Lee. 1988. Introduction to the Theory and Practice of Econometrics, *Wiley*, New York, NY, 1064pp.

Silverman, B.W., 1986. Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability, *Chapman and Hall*, New York, NY, 175pp.

Scott, D.W., 1992. Multivariate Density Estimation: theory, practice, and visualization. Probability and Mathematical Statistics, *Wiley*, New York, NY, 317 pp.

Lall, U., B.R. Rajagopalan, D.G. Tarboton, 1996. "A nonparametric wet/dry spell model for resampling daily precipitation," *Water Resources Research* 32(9), 2803-2823.

## Appendix A – On Parametric and Nonparametric Distribution Fitting

In this analysis (Memorandum Section 2.3.2), the random variations in synthetic flow are modeled by randomly sampling from the observed flow, assuming that it is a random variable. To accomplish this, a distribution must be identified to describe the range and variation of flow as a random variable. Parametric or nonparametric rules for distribution fitting may be applied. Ultimately, a nonparametric approach was adopted. However several parametric distributions were initially considered, as explained below.

A.1 Parametric Distributions:

In a "parametric" specification, the overall distribution function is assumed to display an expected form (e.g., Normal distribution appearing as a "bell-curve", or Gamma distribution appearing as a skewed "bell-curve"). Parametric distributions can be fully described based on statistics derived from their underlying observations.

The feasibility of fitting the "random" observed flow to a Normal distribution was assessed using (a) a Normal probability plot (**Haan, 1977**)**,** and (b) quantitative tests of Normality (i.e. Komolgorov-Smirnov, Lilliefors **(Wilks, 1995)**, and Jarque-Bera **(Judge et al. 1988)**). On (a), the data on the Normal probability plot should closely plot along the line of Normality.

For our example (results not shown), it was found that the observed annual flow data do not adequately fit to a Normal distribution Normality probability plot, particularly at the extreme flows cases. Also the test of Normality was not successful using the Lilliefors and Jarque-Bera tests. The assessment was repeated using *transformed* observations of annual flow: square-root and natural log. Results from Normal Probability plotting were more encouraging and the three quantitative tests of Normality did not rule out Normality. However, both transformations produced distributions that overestimated annual flow in the low-flow range relative to expected values at low probabilities. This was significant because it would lead to synthetic reoccurrence intervals erring high on return periods associated with key observed droughts. For this reason, the transformed Normal distribution approach to describing the Error term was rejected.

Other parametric distributions were also considered (e.g., Gamma and Log-Normal Error). Similar problems were experienced with distribution fitting at low-flow ranges, and with coincidental distribution fitting at low- and high-flow ranges. For this reason, parametric distribution assumptions for describing the observations' "randomness" were disregarded.

A.2 Nonparametric Distributions:

Nonparametric distributions are not required to exhibit an assumed overall shape or form (e.g., like a Normal distribution depicting a "bell-curve"). Fitting a nonparametric distribution often leads to a more complicated probability density function on appearance. However, once the distribution is fit, the act of sampling values from the distribution at specified probabilities can be completed just as easily as if the distribution had been fit parametrically. Moreover, nonparametric distribution fitting often fares better than parametric distribution fitting when trying to coincidentally fit the distribution to observed high- and low-flow cases.

Fitting a nonparametric distribution requires adopting a kernel function that relates a single case to the overall probability distribution. Mathematically, the following steps occur:

- Begin with the given pool of data to which the distribution will be fit (e.g., the Stanislaus annual flow time series from 1901-2004, illustrated in Figure 1 of the Memorandum).
- Choose an estimation range within which the nonparametric probability distribution will be estimated. The estimation range should at least bracket the given pool of data and ideally include a buffer beyond the data extremes. The buffer is subjective and does not affect distribution fitting within the estimation range.
- Choose a kernel function that defines a single data case's influence on the overall distribution estimate. The kernel has two attributes: (a) *shape* and (b) *bandwidth*. Generally speaking, the kernel function peaks at the value where the estimation range value coincides with the case value, and decays when the estimation range value becomes different from the case value.
- Position N kernel functions within the estimation range. Center each function over a single data case. In our example, there are 104 data cases.
- Compute the overall distribution estimate as the superposition of the N positioned kernel functions.

On the choice of kernel functions, there are several types that may be used (e.g. Gaussian (or Normal), Epanechnikov, Triangular (**Silverman 1986; Scott 1992**)). It has been shown that different kernel choices can be made equivalent by rescaling according to appropriate bandwidths (**Lall et al. 1996**). It has also been suggested that bandwidth selection is the most important consideration when applying kernel density estimation methods (**Lall et al. 1996**). Given these considerations, the following kernel assumptions were made:

- Gaussian kernel function shape
- Optimal Gaussian kernel function bandwith (**Silverman, 1986**).

Equations describing the resultant probability density and "building-block" kernel functions are as follows:

$$p(\hat{x}) = \frac{1}{hN} \sum_i K(a)$$

**(4)** $$a = \left( \frac{\hat{x} - x_i}{h} \right)$$

$$K(a) = (2\pi)^{-0.5} \exp\left(-0.5a^2\right)$$

where $x_i$ is an annual flow case in the data sample, $\hat{x}$ is a discrete flow coordinate in the flow-range at which density is being estimated, $N$ is the number of sample observations (i.e. 104), and $h$ is the optimal Gaussian kernel-function bandwidth (**Silverman 1986**) computed as follows:

$$\textbf{(5)}\quad \begin{aligned} h &= 0.9AN^{-0.2} \\ A &= \min\left\{\sigma, \frac{(x_{i,75\%} - x_{i,25\%})}{1.34}\right\} \end{aligned}$$

where $\sigma$ is the sample standard deviation, and $x_{i,75\%}$ and $x_{i,25\%}$ are the 75th and 25th percentile flow values from the data sample.

<u>A.3 Applying the Nonparametric Approach for our Example:</u>

Revisiting the example from A.1, a nonparametric probability distribution function (PDF) was fit to the flow observations (**Figure 10** of the Memorandum). The PDF was then converted into a cumulative distribution function. This function was evaluated relative to the observed, or empirical, frequency distribution (**Figure 11** of the Memorandum). The shapes of the empirical and nonparametric distributions are similar at the extremes, as desired.

One problem with our example application is that the fitted PDF assigns probability to negative flows. Such negative annual flows might imply net annual depletion in the basin measured at Goodwin, which seems unrealistic (but not impossible). In general, it is expected that application of this methodology could produce a distribution function that "tails" at extreme values, assigning small amounts of probability to unrealistic conditions. To avoid the possibility of sampling unrealistically low-value conditions and impairing our ability to model drought reoccurrence, a probability sampling constraint was imposed in the methodology (Memorandum Section 2.3.2) such that the randomly generated sampling probabilities were confined to be within an arbitrary range (i.e. 0.01 to 0.99) even though they're initially sampled from a uniform distribution between 0 and 1. In the example of **Figure 14**, such a constraint on sampling probability is designed to limit the sampled synthetic flow range, but not so much that the sampled range is less than the observed range.

# Appendix B – Development of Regional Precipitation Indices

Several factors were considered when selecting stations to describe precipitation in the Stanislaus Basin:

- Location
- Elevation
- Scale of variability (e.g. mean and range of historical data)
- Period of record

On location and elevation, the index would ideally represent station observations that are representative of precipitation in our locale of interest (i.e. in the Stanislaus Basin above Goodwin). On period of record, the index would ideally be based on as many years of observations as possible, and certainly include the observed droughts of 1929-1934 and 1987-1992. On scale of variability, it is recognized that higher elevation stations and more northward stations in the Sierra Nevada (from the American to the Upper San Joaquin Basins) tend to experience greater precipitation amounts than lower elevation and more southward stations, respectively.

Station data available from the California Data Exchange Center were surveyed, revealing a number of stations having a common "maximum" period of record (i.e. WY1906-2003) and being geographically proximate to the Stanislaus Basin. These stations are listed in Table B-1.

**Table B-1**: Precipitation Stations near the Stanislaus Basin having 1906-2003 data

| Station I.D. (1) | Station Name (1) | Elevation (2) | Basin Location |
|---|---|---|---|
| CLF | Colfax | 2400 | American |
| AUB | Auburn | 1292 | American |
| PCV | Placerville | 1850 | American |
| FLD | Folsom Dam | 350 | American |
| SOR | Sonora R S | 1749 | Stanislaus |
| NFR | North Fork R S | 2630 | Upper San Joaquin |
| YSV | Yosemite Headquarters | 3966 | Merced |

Notes:
(1) Station I.D. at the California Data Exchange Center (http://cdec.water.ca.gov)
(2) Units in feet above mean sea level

Table B-1 also indicates which stations are included in the two regional indices mentioned in Memorandum Section 3.0. The thought behind developing two regional indices is that the mix of station selection might affect the resultant index.

For each station, a time series of annual precipitation amounts was computed. Regional index construction then proceeded with the philosophy that the index should reflect common "phase of variability" found among the contributing stations, while paying little regard to the stations' "central tendency" and "range of variability". This ensures that the index is not dominated by stations that experience the most precipitation or the greatest range of precipitation.  Instead, then index reveals common relative levels of annual wetness among the stations.

The mechanics of index construction given this philosophy are as follows:

- The annual station time series were converted into standardized station time series, by removing the period mean (based on WY1906-2003) and then dividing by the period standard deviation.
- A principal component analysis (**Haan 1977**) was performed on the collection of standardized time series (first for the collection contributing to PrecipIndex1 and then to the collection contributing to PrecipIndex2).  The principal component analysis serves to transform the "dispersion matrix" of station time series (i.e. time periods as rows, stations as columns) into a matrix principal component (PC) "scores" time series.  The PC scores exhibit two useful characteristics:  (1) each PC scores time series is uncorrelated with the other PC scores time series, and (2) they are hierarchically arranged, with the first PC scores times series (i.e. PC1) explaining the most amount of original data variance in the dispersion matrix, PC2 explaining the next most amount, and so forth.  In this application, characteristic (2) is of interest to us, as PC1 is defined as the regional index.

# Appendix C – Graphical Results

Memorandum Section 3.0 describes figures generated for each of the analysis cases:

- Case A:  Figures A1-A19 (reprint of Figures 1-19 from the memorandum)
- Case B:  Figures B1-B19
- Case C:  Figures C1-C19
- Case D:  Figures D1-D19
- Case E:  Figures E1-E19
- Case F:  Figures F1-F19
- Case G:  Figures G1-G19
- Case H:  Figures H1-H19