3.0 METHODOLOGY

3.1 SPI Defined

McKee *et al.* (1993) developed the Standardized Precipitation Index (SPI) for the purpose of defining and monitoring drought. Among others, the Colorado Climate Center, the Western Regional Climate Center, and the National Drought Mitigation Center use the SPI to monitor current states of drought in the United States. The nature of the SPI allows an analyst to determine the rarity of a drought or an anomalously wet event at a particular time scale for any location in the world that has a precipitation record.

Thom (1966) found the gamma distribution to fit climatological precipitation time series well. The gamma distribution is defined by its frequency or probability density function:

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \qquad \text{for } x>0 \tag{3.1}$$

where:

$$\alpha > 0 \qquad \alpha \text{ is a shape parameter} \tag{3.2}$$

$$\beta > 0 \qquad \beta \text{ is a scale parameter} \tag{3.3}$$

$$x > 0 \qquad x \text{ is the precipitation amount} \tag{3.4}$$

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy \qquad \Gamma(\alpha) \text{ is the gamma function} \tag{3.5}$$

For example, figure 3.1 shows the gamma distribution with parameters α=2 and β=1. This distribution is skewed to the right with a lower bound of zero much like a precipitation frequency distribution.

Computation of the SPI involves fitting a gamma probability density function to a given frequency distribution of precipitation totals for a station. The alpha and beta parameters of the gamma probability density function are estimated for each station, for each time scale of interest (3 months, 12 months, 48 months, etc.), and for each month of the year. From Thom (1966), the maximum likelihood solutions are used to optimally estimate α and β:

$$\hat{\alpha} = \frac{1}{4A}\left(1 + \sqrt{1 + \frac{4A}{3}}\right) \tag{3.6}$$

$$\hat{\beta} = \frac{\bar{x}}{\hat{\alpha}} \tag{3.7}$$

where:

$$A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n} \tag{3.8}$$

n = number of precipitation observations (3.9)

The resulting parameters are then used to find the cumulative probability of an observed precipitation event for the given month and time scale for the station in question. The cumulative probability is given by:

$$G(x) = \int_0^x g(x)dx = \frac{1}{\hat{\beta}^{\hat{\alpha}}\Gamma(\hat{\alpha})}\int_0^x x^{\hat{\alpha}-1}e^{-x/\hat{\beta}}dx \tag{3.10}$$

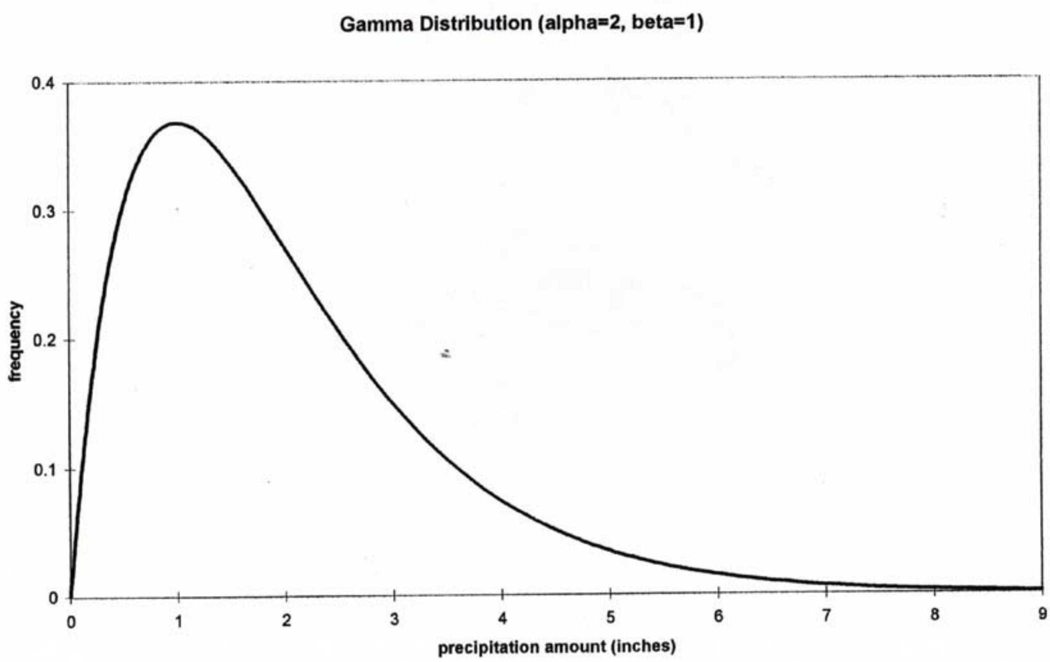

Fig. 3.1 Gamma frequency distribution with parameters alpha=2 and beta=1.

Letting $t = x / \hat{\beta}$, this equation becomes the incomplete gamma function:

$$G(x) = \frac{1}{\Gamma(\hat{\alpha})} \int_0^x t^{\hat{\alpha}-1} e^{-t} dt \qquad (3.11)$$

Since the gamma function is undefined for x=0 and a precipitation distribution may contain zeros, the cumulative probability becomes:

$$H(x) = q + (1-q)G(x) \qquad (3.12)$$

where q is the probability of a zero. If m is the number of zeros in a precipitation time series, Thom (1966) states that q can be estimated by m/n. Thom (1966) uses tables of the incomplete gamma function to determine the cumulative probability G(x). McKee *et al.* (1993) use an analytic method along with suggested software code from Press *et al.* (1988) to determine the cumulative probability.

The cumulative probability, $H(x)$, is then transformed to the standard normal random variable Z with mean zero and variance of one, which is the value of the SPI. This is an equiprobability transformation which Panofsky and Brier (1958) state has the essential feature of transforming a variate from one distribution (*ie.* gamma) to a variate with a distribution of prescribed form (*ie.* standard normal) such that the probability of being less than a given value of the variate shall be the same as the probability of being less than the corresponding value of the transformed variate. This method is illustrated in figure 3.2. In this figure, a 3 month precipitation amount (January through March) is converted to a SPI value with mean of zero and variance of one. The left side of figure 3.2 contains a broken line with horizontal hash marks that designate actual values of 3 month precipitation amounts (x-axis) for Fort Collins, Colorado for the months of January

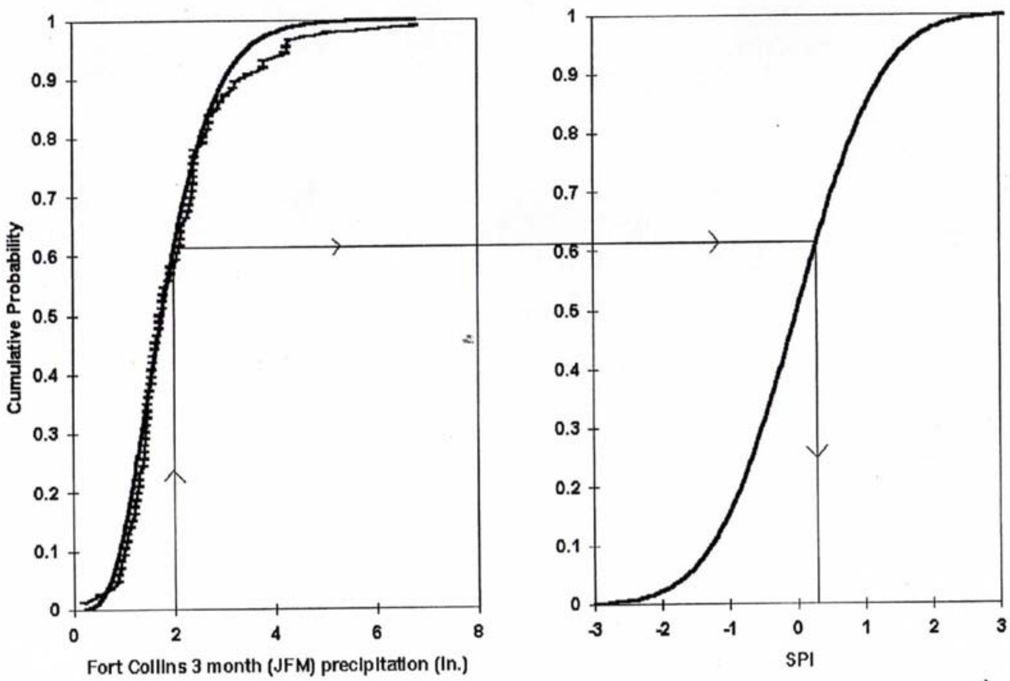

Fig. 3.2 Example of equiprobability transformation from fitted gamma distribution to the standard normal distribution.

through March for the period 1911 through 1995. The broken line also denotes the empirical cumulative probability distribution (y-axis) for the period of record. The empirical cumulative probabilities were found optimally as suggested by Panofsky and Brier (1958) where the precipitation data is sorted in increasing order of magnitude so that the kth value is k-1 values from the lowest and where n is the sample size:

$$\text{empirical cumulative probability} = \frac{k}{n+1} \tag{3.13}$$

The smooth curve on the left hand side of figure 3.2 denotes the cumulative probability distribution of the fitted gamma distribution to the precipitation data. The smooth curve on the right hand side of figure 3.2 denotes the cumulative probability distribution of the standard normal random variable Z using the same cumulative probability scale of the empirical distribution and the fitted gamma distribution on the left hand side of the figure. The standard normal variable Z (or the SPI value) is denoted on the x-axis on the right hand side of the figure. Hence, this figure can be used to transform a given 3 month (January through March) precipitation observation from Fort Collins, Colorado to a SPI value. For example, to find the SPI value for a 2 inch precipitation observation, simply go vertically upwards from the 2 inch mark on the x-axis on the left hand side of figure 3.2 until the fitted gamma cumulative probability distribution curve is intersected. Then go horizontally (maintaining equal cumulative probability) to the right until the curve of the standard normal cumulative probability distribution is intersected. Then proceed vertically downward to the x-axis on the right hand side of figure 3.2 in order to determine the SPI value. In this case, the SPI value is approximately +0.3.

Since it would be cumbersome to produce these types of figures for all stations at all time scales and for each month of the year, the Z or SPI value is more easily obtained computationally using an approximation provided by Abramowitz and Stegun (1965) that converts cumulative probability to the standard normal random variable Z:

$$Z = SPI = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \qquad \text{for } 0 < H(x) \leq 0.5 \qquad (3.14)$$

$$Z = SPI = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \qquad \text{for } 0.5 < H(x) < 1.0 \qquad (3.15)$$

where:

$$t = \sqrt{\ln\left(\frac{1}{(H(x))^2}\right)} \qquad \text{for } 0 < H(x) \leq 0.5 \qquad (3.16)$$

$$t = \sqrt{\ln\left(\frac{1}{(1.0 - H(x))^2}\right)} \qquad \text{for } 0.5 < H(x) < 1.0 \qquad (3.17)$$

$$\begin{aligned} c_0 &= 2.515517 \\ c_1 &= 0.802853 \\ c_2 &= 0.010328 \\ d_1 &= 1.432788 \\ d_2 &= 0.189269 \\ d_3 &= 0.001308 \end{aligned} \qquad (3.18)$$

Conceptually, the SPI represents a z-score, or the number of standard deviations above or below that an event is from the mean. However, this is not exactly true for short time scales since the original precipitation distribution is skewed. Nevertheless, figure 3.3 shows that during the base period for which the gamma parameters are estimated, the SPI will have a standard normal distribution with an expected value of zero and a variance of one. Katz and Glantz (1986) state that requiring an index to have a fixed expected value and variance is desirable in order to make comparisons of index values among different stations and regions meaningful.

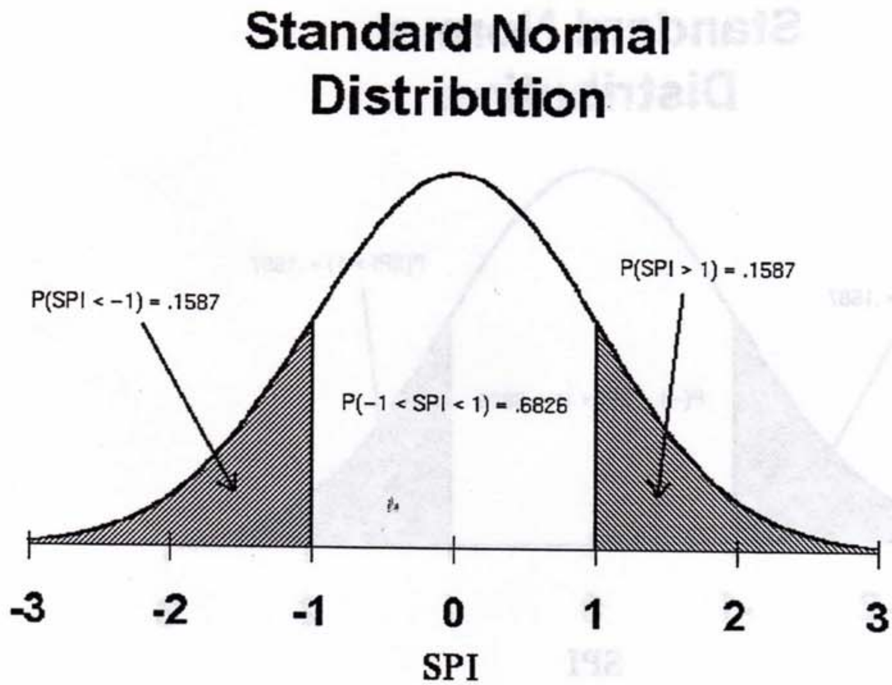

Fig. 3.3 Standard normal distribution with the SPI having a mean of zero and a variance of one.

Tannehill (1947) states that rainfall in the worst drought ever experienced in Ohio would be abundant rainfall in Utah. Akinremi *et al.* (1996) state that the spatial and temporal dimensions of drought create problems in generating a drought index because not only must an anomaly be normalized with respect to location, but the anomaly must also be normalized in time if it is to produce a meaningful estimate of drought. The SPI accomplishes both. The SPI is normalized to a station location because it accounts for the frequency distribution of precipitation as well as the accompanying variation at the station. Additionally, the SPI is normalized in time because it can be computed at any number of time scales, depending upon the impacts of interest to the analyst. Additionally, no matter the location or time scale, the SPI represents a cumulative probability in relation to the base period for which the gamma parameters were estimated. Table 3.1 is a table of SPI and its corresponding cumulative probability.

Table 3.1: SPI and Corresponding Cumulative Probability in Relation to the Base Period

SPI	Cumulative Probability
-3.0	0.0014
-2.5	0.0062
-2.0	0.0228
-1.5	0.0668
-1.0	0.1587
-0.5	0.3085
0.0	0.5000
+0.5	0.6915
+1.0	0.8413
+1.5	0.9332
+2.0	0.9772
+2.5	0.9938
+3.0	0.9986

An analyst with a time series of monthly precipitation data for a location can calculate the SPI for any month in the record for the previous *i* months where i=1, 2, 3, ..., 12, ..., 24, ..., 48, ... depending upon the time scale of interest. Hence, the SPI can be computed for an observation of a 3 month total of precipitation as well as a 48 month total of precipitation. For this study, a 3 month SPI is used for a short-term or seasonal drought index, a 12 month SPI is used for an intermediate-term drought index, and a 48 month SPI is used for a long-term drought index. Therefore, the SPI for a month/year in the period of record is dependent upon the time scale. For example, the 3 month SPI calculated for January, 1943 would have utilized the precipitation total of November, 1942 through January, 1943 in order to calculate the index. Likewise, the 12 month SPI for January, 1943 would have utilized the precipitation total for February, 1942 through January, 1943 while the 48 month SPI would have utilized the precipitation total for February, 1939 through January, 1943.

Figure 3.4 is a graph of the SPI calculated for McPherson, Kansas for the period 1911 through 1995. Three time scales are shown: 3 months, 12 months, and 48 months. As stated by McKee *et al.* (1993) as well as being evident in the figure: the frequency, duration, and intensity of drought at any particular point during the historical record is dependent upon time scale. The long-term drought index (48 month SPI) shows that McPherson was impacted by the long-term droughts of the 1930s and 1950s. Further inspection at the short-term (3 month SPI) shows that the 1930s for McPherson was a series of several short-term droughts with some intermediate normal periods. Skaggs (1975) described this as "waves" of drought. For the 1950s, even though the long-term drought was shorter in duration, the short-term droughts were more consecutive and resulted in a more intense long-term drought (with the 48 month SPI going below a -3.0). Overall, the late 1940s and early 1950s was a long-term wet period. But looking at the short-term drought index, it is evident that some short-term droughts did occur during this period. For example, the summer drought of 1947 is similar in magnitude to droughts that occurred during the 1930s and 1950s. In fact, United States Department of Agriculture (1951) records show that the corn yield per harvested acre was at a 7 year low in Kansas following this dry summer. However, this drought was preceded and followed by anomalously wet conditions and therefore this drought does not show up at the longer

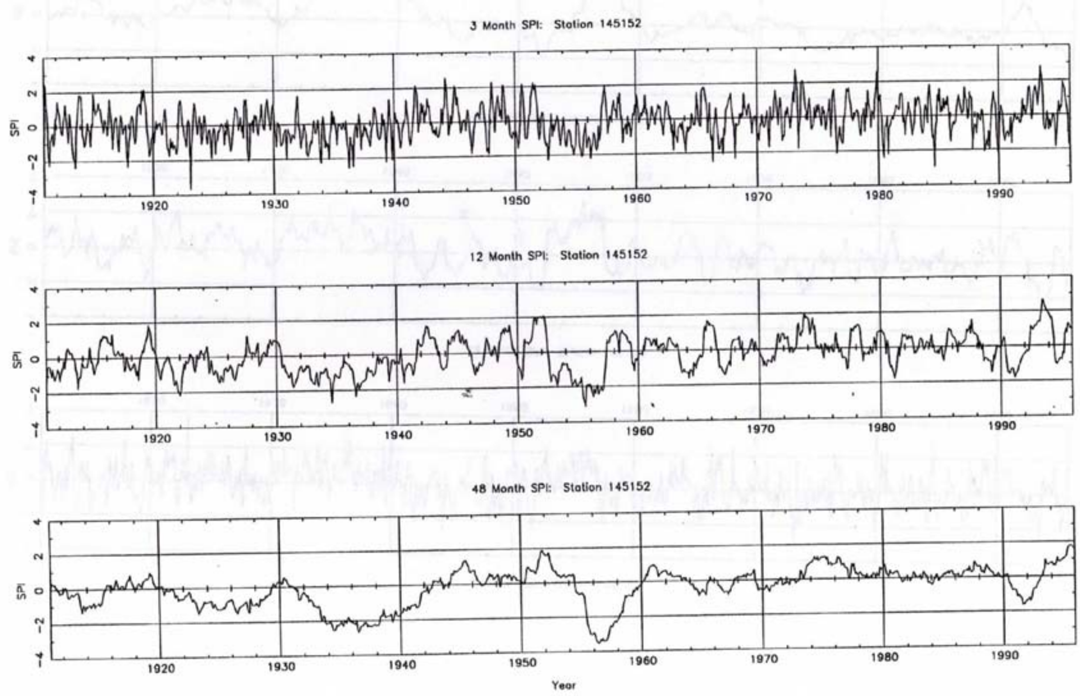

Figure 3.4 Time series of 3 month, 12 month, and 48 month SPI for McPherson, Kansas for the period Jan 1911 through Dec 1995.

time scales. This is a similar situation to the 1980 summer drought in the southern United States. Karl and Quayle (1981) state that the ample rains during the spring of 1980 prevented the 1980 summer drought in the southern United States from being far worse (hence, a short-term drought that didn't translate into a long-term drought). They state that the difference between the summer drought of 1980 and the summer droughts of the 1930s and 1950s was that the summer droughts of the 1930s and 1950s occurred when a very high moisture demand had already developed (in other words, long-term drought).

3.2 Climatological Base Period (1941-1980)

For this analysis, a base period of 1941 to 1980 was utilized to estimate the gamma parameters that are used to calculate the cumulative probabilities of precipitation events. One reason for doing this is that missing data is minimal in the USHCN for this period (figure 2.2). Also, most stations in the USHCN experienced at least one long-term drought and one long-term wet period during this timeframe. Since the cumulative probability is converted to the standard normal random variable Z, the SPI will have a standard normal distribution during the base period. Figure 3.3 shows that about 16% of the time the SPI will be -1.0 or below indicating drought conditions. Similarly, about 16% of the time the SPI will be +1.0 or above indicating anomalously wet conditions. About 68% of the time, the SPI will be between -1.0 and +1.0 indicating normal conditions.