STANDARDIZED PRECIPITATION INDEX BASED ON PEARSON TYPE III DISTRIBUTION

GABRIEL CONSTANTINO BLAIN

Instituto Agronômico de Campinas (IAC), Campinas, SP, Brasil

gabriel@iac.sp.gov.br

ABSTRACT

The initial step in calculating the Standardized Precipitation Index (SPI) is to determine a probability density function (pdf) that describes the precipitation series under analysis. Once this pdf is determined, the cumulative probability of an observed precipitation amount is computed. The inverse normal function is then applied to the cumulative probability. The result is the SPI. This article assessed the changes in SPI final values, when computed based on Gamma 2-parameters (Gam) and Pearson Type III (PE3) distributions (SPIGam and SPIPE3, respectively). Monthly rainfall series, available from five weather stations of the State of São Paulo, were chosen for this study. Considering quantitative and qualitative assessments of goodness-of-fit (evaluated at 1-, 3-, and 6-months precipitation totals), the PE3 distribution seems to be a better choice than the Gam distribution, in describing the long-term rainfall series of the State of São Paulo. In addition, it was observed that the number of SPI time series that could be seen as normally distributed was higher when this drought index was computed from the PE3 distribution. Thus, the use of the Pearson type III distribution within the calculation algorithm of the SPI is recommended in the State of São Paulo.
Keywords: drought, Gamma distribution, SPI

RESUMO: ÍNDICE PADRONIZADO DE PRECIPITAÇÃO BASEADO NA DISTRIBUIÇÃO PEARSON TIPO III
O cálculo do Índice Padronizado de Precipitação (IPP) inicia-se com a adoção de uma distribuição paramétrica (dp) utilizada para a estimativa das probabilidades de ocorrência associadas a uma série de precipitação pluvial. Após essa escolha, a probabilidade acumulada de ocorrência de um determinado valor de precipitação é calculada. O IPP é obtido após a aplicação da função normal inversa a essa probabilidade acumulada. O artigo avaliou as alterações nos valores finais do IPP, quando calculado com base nas distribuições Gama com dois parâmetros (Gam) e Pearson tipo III (PE3; IPPGam e IPPPE3, respectivamente). Utilizaram-se dados de precipitação pluvial de cinco localidades do Estado de São Paulo, nas escalas de 1, 3 e 6 meses. Aplicando testes de aderência, tanto qualitativos, quanto quantitativos, verificou-se que as probabilidades associadas às séries utilizadas foram melhores estimadas por meio da PE3. Foi também observado, que o número de séries temporais do IPP que podem ser consideradas normalmente distribuídas, foi maior quando este índice de seca foi calculado com base na PE3. Dessa forma, recomenda-se a adoção da PE3 no cálculo do IPP no Estado de São Paulo.
Palavras-Chave: seca, distribuição gama, IPP

1. INTRODUCTION

The Standardized Precipitation Index (SPI) was developed by McKee et al. (1993, 1995) as a drought indicator which standards the rainfall deficits/excess on temporal and regional basis. As described by Keyantash and Dracup (2002) the SPI represents observed rainfall as a standardized departure with respect to a rainfall probability distribution function.

Since McKee et al. (1993), the SPI model has been used by several authors on climate variability evaluations. Hayes et al. (1999) applied the SPI algorithm in describing drought conditions in the State of Texas, USA. Zhai et al. (2009) calculated this

drought index for 483 meteorological stations, covering the whole territory of China. Blain (2005) recommended the SPI model as a drought indicator of the State of São Paulo, Brazil. According to Wu et al. (2007) the SPI is widely accepted and used throughout the world in both research and operational modes because it is normalized, or standardized, to location and to time.

As pointed out by Guttman (1998, 1999), the first step in the calculation of the SPI is to determine a probability density function (pdf) that describes the long-term series of observations. Once this pdf is determined, the cumulative probability (cdf) of an observed precipitation amount is estimated. The standard inverse normal function, which has zero mean and unitary variance, is then applied to the cumulative probability. The result is the SPI. Theoretically, the SPI is unbounded. Practically, however, the number of precipitation data, which is generally less than 100 for given month, suggests bounds of -3.09 ≤ SPI ≤ 3.09. It is also worth mentioning that although the SPI can be seen as a meteorological drought index – which uses only precipitation data within its calculation algorithm – this index is used, on operational mode, by government agricultural institutes, such as *Instituto Agronômico (IAC), Empresa Brasileira de Pesquisa Agropecuária* (EMBRAPA), and *Instituto Nacional de Meteorologia* (INMET). In this last view, and considering agricultural purposes, the SPI calculated at the monthly time scale is always the most assessed.

Fowling Guttman (1999), a standard method must be used for computing this drought index. If different pdf are used to describe an observed series of precipitation data, then different SPI values may be obtained. Mckee et al. (1993), Hayes et al. (1999), Sansigolo (2004) and Blain (2005) used the 2-parameters gamma distribution (Gam) in SPI calculation. Guttmann (1999), analyzing the fit of several distributions to monthly precipitation series of the United States, recommends the Pearson type III distribution (PE3), as a universal model in SPI calculation. Kumar et al. (2009) applying the SPI for a low rainfall and a high rainfall districts of Andra Pradesch State, India, suggested that the PE3 distribution may be a better choice than the Gam distribution, on SPI calculations. Soto et al. (2005) using monthly precipitation data of Mexico, concluded that the differences between SPI final values, based on both PE3 and Gam distributions, were non significant and averaged zero. According to Wu et al. (2007) the Gam distribution can affect the confidence in the SPI results, because the Gam pdf has (only) two free parameters, which would not give the best goodness-of-fit for a given (rainfall) series. Also according to Wu et al. (2007), other alternative distributions, e.g. PE3, are worth to be studied. Furthermore, adopting a standard probability model may be seen as an important step in developing a standard software package for which the computational complexities, within the SPI calculation algorithm, are transparent to the user (Guttman, 1999).

Thus the aim of this study was to evaluate the changes in SPI final values, when computed based on Gamma 2-parameters and Pearson Type III distributions in the State of São Paulo, Brazil.

2. MATERIAL AND METHODS

Monthly rainfall series, available from five weather stations (Campinas, 1890 to 2008; Pindorama, 1951 to 2008; Presidente Prudente, 1960 to 2008; Ribeirão Preto, 1937 to 2008 and Ubatuba 1935 to 2008) were chosen for this study. These five stations represents climatically dissimilar areas of the State of São Paulo, ranging from the coast of the State (Ubatuba), where there is the absence of a dry season, to the western region of the State (Presidente Prudente), where there is a remarkable dry season during the winter months (Figure 1).

The pdf of the Gam distribution is defined by:

$$gam(X) = \frac{X^{\alpha-1} * e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \qquad x, \alpha, \beta > 0 \qquad (1)$$

Since 1 is undefined at x=0 (no rainfall event), the final cumulative probability becomes:

$$Hg(x) = qg+(1-qg)Gam(x) \qquad (2)$$

qg is the empirical probability of x=0, and
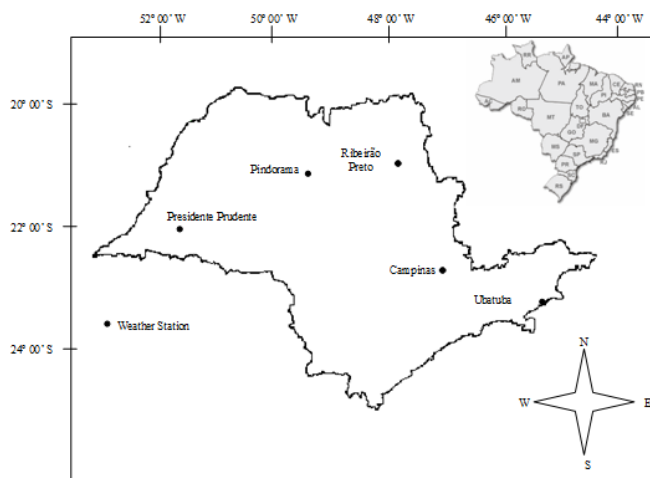
$$Gam(x) = \int_0^x gam(x)d(x) \qquad (3)$$

Figure 1 - Selected weather stations across the State of São Paulo, Brazil

The qg empirical factor can be estimated by the ratio between the number of data within the sample (N) and the number of zeros (n) within this sample (qg=N/n). The two parameters of the Gam distribution are α, the shape parameter; and β, the scale parameter. The quantity $\Gamma(\alpha)$ is the gamma function (Wilks, 2006). Following Thom (1966), Blain (2005) and Wilks (2006), α and β estimation was based on the maximum likelihood method.

The PE3 distribution has a three parameters pdf described by:

$$pe(x) = \frac{(x-\xi)^{\alpha-1}\exp[(-x+\xi)/\beta]}{\beta^{\alpha}\Gamma(\alpha)}$$

for $\gamma > 0$; and $\xi=\mu-2\sigma/\gamma$ (4)

where μ, σ, γ are, respectively, the location (mean value of the series), the scale, and the shape parameters.

According to Hosking and Wallis (1997), the quantile function pe(x) has no explicit analytical form. Since 4 is undefined at $x<\mu-2\sigma/\gamma$, the cumulative probability becomes:

$$Hp(x) = qp+(1-qp)PE3(x)$$

(5)

qp is the empirical probability of $x<\mu-2\sigma/\gamma$ values, and

$$PE(x) = \frac{G\left(\alpha, \frac{x-\xi}{\beta}\right)}{\Gamma(\alpha)}$$

(6)

G(.) is the incomplete gamma function

As pointed out by Hosking and Wallis (1997) if γ equals zero, then the distribution is Normal and the range of x becomes $-\infty<x<\infty$. Following Hosking and Wallis (1997) and Guttmann (1999) the parameters estimation was based on the methods of L-moments. According to Hosking (1990) L-moments fitting tend to be preferred for small data samples. For moderate and large samples sizes the results of the two parameters estimation methods (L-moments and maximum likelihood) are usually similar. On this sense, Guttmann (1999) recommends at least 50 years of rainfall data for SPI calculations.

Although several author, such as Thom (1966), Blain (2005), Blain et al. (2007), Blain et al. (2009) and Blain (2009), have already assessed the robustness of the Gam distribution in describing monthly precipitation series of the State of São Paulo; the use of PE3 distribution in describing these series seems to be neglected. Since the SPI is a probability index, assessing the robustness of the PE3 distribution in describing

precipitation series may be seen as the first (and perhaps the most important) step in evaluating the possibility of adopting this three parameters distribution within SPI calculation algorithm. Two frequently used tests of the goodness-of-fit are the Kolmolgorov-Smirnov test (KS) and the chi-squared test (χ^2). According to Wilks (2006) the χ^2 operates more naturally for discrete random variables since to implement it the range of the data must be divided into discrete classes. For continuous distribution, the KS test, which compares the empirical and theoretical cdfs, will be more powerful than the χ^2. As pointed out by Panofsky and Brier (1968) if the parametric/theoretical distribution is far different from the empirical distribution, the KS and/or χ^2 values will be large. The null hypothesis (H_0) associated with these both tests, affirms that the empirical distribution is statistically similar to the parametric distribution under analysis. Because the samples being tested in this work are the same samples being used in deriving the distributions parameters, the KS test is also known as the lilliefors test. The KS/lilliefors value (here after referred just as KS) for rejection/acceptance of H_0 will depend on the sample size being used to test the parametric distribution, on the shape of the parametric distribution and, of course, on the significance level chosen by the user (Crutcher, 1975). The confidence in accepting or rejecting the theoretical distributions was measured by the p-value (p). The KS and χ^2 tests were applied in order to evaluate the goodness-of-fit of Gam and PE3 distributions to the observed rainfall series. A p-value smaller than 0,05 was taken as an evidence that the H_0 cannot be accepted. It is also worth mentioning that the χ^2 limiting value depends, among others, on the number of categories in which each monthly series was divided.

Graphical comparisons of the rainfall data and the fitted PE3 distributions were also carried out by using the percentil-percentil (pp) plots. The pp plots compare the empirical cumulative probability as a function of the fitted parametric cdf (Wilks, 2006). The KS, χ^2, and pp plots calculation algorithms can be found on Wilks (2006).

The SPI values were estimated based on PE3 and Gam distributions for three time scales (1, 3, and 6 months). It is worth mentioning that the time scale associated with an SPI value is described by the month at the end of the analysis period. For instance, the 3-months SPI value observed at the end of the month of May is related to the precipitation total observed during the months of March, April and May. From both probability distributions (PE3 and Gam) the cumulative probability of an observed precipitation amount was computed. The inverse normal (Gaussian) function, with mean zero and variance one, was then applied to the cumulative probability [Hp(x) and/or Hg(x), respectively]. The result is the SPI.

$$SPI = -\left(t - \frac{co + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \text{for } 0 < \text{Hg(x) or Hp(x)} \le 0.5$$

$$(7)$$

$$SPI = +\left(t - \frac{co + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \text{for } 0.5 < \text{Hg(x) or Hp(x)} < 1$$

$$t = \sqrt{\left(\ln\left(\frac{1}{(H(x)^2)}\right)\right)} \quad \text{for } 0 < \text{Hg(x) or Hp(x)} \le 0.5$$

$$(8)$$

$$t = \sqrt{\left(\ln\left(\frac{1}{(1 - H(x)^2)}\right)\right)} \text{ for } 0.5 < \text{Hg(x) or Hp(x)} < 1$$

c_o = 2.515517; c_1 = 0.802853; c_2 = 0.010328; d_1= 1.432788; d_2 = 0.189269; d_3 = 0.001308

The SPI drought categories, defined by McKee et al. (1993), are shown on Table 1.

As pointed out by Hayes et al. (1999) because SPI values fit a normal distribution, one can expect these values to be within one standard deviation 68% of the time, within two

standard deviations 95% of the time, and within three standard deviations 99% of the time. A related interpretation would be that an SPI value of less than 1.0 occurs 16 times in 100 year, an SPI of less than 2.0 occurs two to three times in 100 year, and an SPI of less than 3.0 occurs once in approximately 200 year. Thus, as also pointed out by Hayes et al. (1999), because of this characteristics associated with the standard normal distribution (Equations 7 and 8), the SPI by itself cannot identify regions (or seasons) that may be more "drought prone" than others.

The regression analyses were used to compare SPIPE3 and SPIGam final values. Since the SPI are (at least approximately) normally distributed, the regression coefficients were obtained by using linear least square fitting. The significance of the coefficient of determination (R^2), was assessed by using the F test.

According to Wu et al. (2007), a SPI distribution can be considerate non-normal when its variables, related to the SPI time series, meet three criteria simultaneously: i) Shapiro-Wilk statistic (W) less than 0.960; ii) p-values less than 0.100; and iii) the absolute value of the median greater than 0.05. Otherwise, the SPI distribution is normal. The W statistic is the ratio of the best estimator of the variance, based on the square of a linear combination of the order statistic, to the usual corrected sum of squares estimator of the variance. The p-value is the probability that is associated with W (Wu et al., 2007). Since being normalized, or standardized on regional and temporal basis, is an important feature of the SPI algorithm, this normality test was applied on both SPIPE3 and SPIGam series.

It is worth mentioning that the presence of non-random components within a time series (such as trends and/or serial correlation) may affect the stability of the underlying statistic within the data sample. In this view, the Mann-kendall (MK)

Table 1 - Standardized Precipitation Index (SPI) values and the associated drought categories (MCKEE et al., 1993)

SPI values	Drought Category
0 to -0.99	mild drought
-1.00 to -1.49	moderate drought
-1.50 to -1.99	severe drought
≤-2.00	extreme drought

Table 2 - Runs test (Z) and Mann-Kendall test (MK) applied to five monthly precipitation series of the State of São Paulo, Brazil

Month	Campinas		Pindorama		Pres.Prudente		Rib.Preto		Ubatuba	
	Z	MK	Z	MK	Z	MK	Z	MK	Z	MK
January	0.85	0.51	-0.07	0.03	0.95	0.36	1.34	0.91	-0.29	-1.00
February	0.59	-0.87	1.59	0.58	-0.14	-0.03	1.37	0.05	0.20	-1.29
March	-1.79	-0.16	-0.53	0.90	0.35	0.78	0.36	-0.92	0.35	-0.96
April	0.85	0.59	-0.10	0.21	-0.62	1.21	1.02	-0.13	-1.42	0.77
May	-0.95	1.28	1.04	0.15	0.05	1.37	-1.27	2.05*	0.35	0.79
June	0.84	-0.80	1.55	0.09	-0.14	-0.82	-0.04	-0.81	0.05	0.06
July	-0.32	0.74	-0.37	0.09	0.30	0.96	-0.30	0.99	0.20	-0.04
August	1.76	-1.67	0.34	0.32	0.24	-0.28	0.55	-0.60	-0.29	-2.32*
September	-0.83	-0.82	-0.17	0.36	1.39	1.58	-0.04	1.82	0.36	1.45
October	-0.01	0.06	0.67	0.49	0.20	-0.56	1.14	-1.45	0.36	0.46
November	0.93	-0.93	-1.34	0.42	1.91	1.19	-1.05	1.39	1.12	0.10
December	-0.52	-0.25	-0.36	0.99	1.93	-1.03	0.86	-0.34	0.20	-1.31

* Significant at 5% level

Table 3 - Median, Standard deviation (Sd) and number of cases (months) in which the pearson type III distribution (#PE3) and the gamma distribution (#Gam) could not be defined at five weather station of the State of São Paulo, Brazil. The median and the Sd are given in millimeters

Monthy	Campinas (1890/2008)				Pindorama(1951/2008)			
	Median	Sd	#PE3	#Gam	Median	Sd	#PE3	#Gam
January	229.5	100.97	0	0	242.5	103.03	0	0
February	194.4	88.23	0	0	192.8	108.18	0	0
March	148.6	70.34	0	0	147.2	75.96	0	0
April	49.8	44.37	0	0	54.3	53.22	0	2
May	49.7	48.07	0	3	52.3	52.39	0	3
June	35.4	44.00	0	8	19.5	35.98	0	10
July	16.4	37.68	0	14	9.7	30.61	0	13
August	25.8	33.52	0	16	9.5	30.72	0	11
September	61.5	52.51	0	1	46.3	47.79	0	3
October	114.4	62.89	0	0	112.3	57.11	0	0
November	142.2	67.57	0	0	114.4	67.19	0	0
December	206.4	96.33	0	0	211.4	90.38	0	0
	Presidente Prudente(1960/2008)				Ribeirão Preto(1937/2008)			
January	198.20	85.78	0	0	276.4	107.54	0	0
February	160.90	125.64	1	0	208.7	102.51	0	0
March	113.10	85.85	1	0	161.0	77.22	0	0
April	57.10	88.80	0	1	60.0	57.33	0	1
May	71.60	75.91	0	2	48.9	47.47	0	3
June	46.30	76.68	0	2	15.9	32.17	0	14
July	24.10	77.45	0	5	10.0	28.54	0	15
August	24.20	88.98	0	10	5.4	29.11	0	22
September	56.70	86.36	1	0	42.7	49.63	0	3
October	121.40	100.76	0	0	117.8	60.58	0	0
November	131.00	126.05	0	0	169.2	72.88	0	0
December	162.20	109.03	0	0	248.0	113.64	0	0

trend test (Kendall and Stuart, 1967, and Sneyers, 1975) is frequently used for detections of trends in time series data (Hamed and Hao, 1998). A positive MK value is observed under the presence of increasing trends. A negative MK value is observed under the presence of decreasing trends. If the MK value falls within [-1.96:1.96] the trend is taken as non-significant (The Ho is accepted at 5% level of significance). The Runs test (Z) for randomness was also calculated for the precipitation monthly series. If $-1.96 < Z < 1.96$, the hypothesis that the sequence is free of serial correlation (Ho) cannot be rejected at the 5% significance level.

3. RESULTS AND DISCUSSION

3.1 The monthly time scale

  As indicated by the Runs test, all monthly precipitation series can be seen as free of significant serial correlations. The Ho associated with the MK test could not be accepted only

Table 4 - Median, Standard deviation (Sd) and number of cases (months) in which the pearson type III distribution (#PE3) and the gamma distribution (#Gam) could not be defined at the weather station of Ubatuba, SP, Brazil (1935/2008). The median and the Sd are given in millimeters

Monthy	Median	Sd	#PE3	#Gam
January	328.7	176.13	0	0
February	297.8	206.14	0	0
March	280.4	180.04	0	0
April	180.9	130.60	4	0
May	115.8	70.80	0	0
June	70.0	70.69	0	0
July	75.5	82.22	0	0
August	80.6	71.94	4	0
September	169.8	78.53	0	0
October	220.8	98.63	0	0
November	260.1	124.54	0	0
December	306.8	134.88	0	0

for the monthly series of May (Ribeirão Preto; MK=2.05) and August (Ubatuba; MK=-2.32). All others 108 monthly series can be seen as free of trends (at the 5% significance level).

The inspection of Equations 7 and 8 indicates that if Hg differs significantly from Hp, then the SPIGam values may also differ significantly from the SPIPE3 values. Thus, as pointed out by Guttman (1999), standardization of this drought index calculation algorithm is necessary in order to provide for all users, a common basis for both spatial and temporal comparison of the SPI values.

According to Equations 1 and/or 4, while the Gam distribution has a fixed (x = 0) bound, the PE3 distribution bound depends on the values of its own parameters (x<μ-2σ/γ). Tables 3 and 4 show for each monthly series the median, the Standard deviations, and the number of cases in which the theoretical/parametric PE3 and/or Gam distributions could not be defined [x=0, gam(x); or x<μ-2σ/γ, pe3(x)].

Table 3 indicates that the Gam distribution [$gam(x)$] could not be computed for a higher number of (no) rainfall events than the PE3 distribution [pe (x)] could be. On this view,

Table 5 - Lilliefors/Kolmolgorov-Smirnov test applied to fitted gamma (KSGam) and Pearson type III (KSPE) distributions to rainfall series. The p-value (p) is also illustrated. μ, σ, γ, α, β and are the parameters of the distributions.

Monthy	KSPE (p)	KSGam (p)	γ	σ	μ	α	β
			Campinas				
January	0.04 (0.23)	0.04 (0.24)	0.8	100.8	249.0	6.0	41.4
February	0.06 (0.17)	0.09 (0.02)	0.6	88.0	200.5	4.6	43.2
March	0.04 (0.51)	0.06 (0.15)	0.5	71.6	153.4	3.8	40.1
April	0.05 (0.25)	0.05 (0.15)	1.4	45.2	63.4	1.9	33.0
May	0.05 (0.37)	0.07 (0.26)	1.3	47.9	58.1	1.3	47.0
June	0.08 (0.13)	0.07 (0.09)	1.8	44.5	47.1	1.4	36.6
July	0.06 (0.33)	0.13 (0.03)	1.8	36.4	30.9	0.9	37.0
August	0.08 (0.06)	0.15 (0.02)	1.8	34.2	33.1	1.2	30.8
September	0.05 (0.10)	0.08 (0.03)	1.3	52.6	66.8	1.4	49.6
October	0.03 (0.35)	0.06 (0.16)	0.8	63.6	120.0	3.0	39.5
November	0.04 (0.51)	0.05 (0.25)	0.7	67.8	148.4	4.5	33.3
December	0.04 (0.35)	0.03 (0.48)	0.8	98.0	220.0	5.0	44.1
			Pindorama				
January	0.08 (0.16)	0.08 (0.14)	0.7	103.7	257.3	6.0	42.8
February	0.08 (0.28)	0.08 (0.27)	1.0	108.9	213.0	3.9	54.3
March	0.05 (0.46)	0.05 (0.44)	0.9	78.0	155.9	4.1	37.9
April	0.07 (0.20)	0.08 (0.09)	1.5	55.0	70.2	2.0	36.8
May	0.05 (0.32)	0.05 (0.66)	1.5	53.8	61.7	1.5	43.2
June	0.11 (0.13)	0.17 (0.08)	1.5	34.9	31.6	1.1	36.5
July	0.14 (0.11)	0.08 (0.16)	1.5	29.5	24.4	0.9	34.5
August	0.14 (0.07)	0.07 (0.17)	1.5	29.8	24.7	0.8	36.7
September	0.05(0.31)	0.06 (0.38)	1.3	49.6	57.4	1.3	47.9
October	0.08 (0.20)	0.10 (0.07)	0.5	58.7	113.6	3.4	33.1
November	0.07 (0.27)	0.07 (0.15)	1.2	66.6	134.1	4.4	30.7
December	0.05 (0.50)	0.05 (0.46)	0.6	92.6	216.8	5.2	41.6
			Presidente Prudente				
January	0.05 (0.43)	0.04 (0.39)	1.3	125.4	221.0	3.4	64.8
February	0.10 (0.28)	0.09 (0.10)	1.3	103.6	181.5	3.2	56.2
March	0.06 (0.42)	0.06 (0.21)	1.3	74.4	128.7	3.1	41.4
April	0.08 (0.12)	0.13 (0.01)	0.8	43.8	67.9	1.9	36.3
May	0.08 (0.14)	0.09 (0.03)	1.3	55.4	80.2	2.5	32.8
June	0.06 (0.50)	0.10 (0.06)	1.6	51.6	53.9	1.0	55.6
July	0.10 (0.66)	0.05 (0.28)	1.6	45.2	39.8	0.9	48.6
August	0.09 (0.14)	0.08 (0.11)	1.6	45.5	39.1	1.0	49.5
September	0.07 (0.68)	0.05 (0.20)	1.7	55.0	69.4	1.8	39.6
October	0.05 (0.41)	0.06 (0.39)	0.7	70.7	131.1	2.8	46.9
November	0.04 (0.65)	0.05 (0.40)	0.7	68.0	136.3	4.0	34.3
December	0.09 (0.11)	0.10 (0.11)	1.1	108.9	185.1	3.1	59.0

the empirical factor (qg and/or qp) has a smaller contribution (weight) to SPIPE3 values than it has to SPIGam values. Table 4 indicates the opposite feature.

According to KS test (Tables 5 and 6) the monthly rainfall series are, in general, consistent with the proposition of their having been drawn from a population with PE3 and/or Gam distributions. The same feature was also pointed out by the χ^2 (not illustrated). However, for the monthly series of Ribeirão Preto, the PE3 seems to be a better choice than the Gam distribution, since the Ho, associated with this last parametric function, could not be accepted for June.

Figures 2 to 4 show the pp plots for the weather station of Campinas, Presidente Prudente and Ubatuba at the months of January, April, July, and October. These four months represents, respectively, the summer, the autumn, the winter, and the spring seasons.

Figures 2 to 4 indicate that the fitted PE3 distribution corresponds well to the data through most of its range, since the parametric cumulative probability, evaluated at the corresponding data value, is quite close to the empirical cumulative probability. The pp points are very close to the line 1:1. The same feature (not illustrated) is also observed at the others monthly series and for the others weather stations. Considering the quantitative (KS and χ^2 tests) and the qualitative (pp plots) assessments of the goodness-of-fit, the PE3 distribution is adequate for SPI calculations in the State of São Paulo, Brazil.

Figures 5 and 6 show the regression analysis between SPIPE3 and SPIGam time series. The regression analysis was estimated for every data at each weather station.

The linear regression models (coefficient of determination, R^2, slope and intercept) (Figures 5 and 6) show no significant

Table 6 - Lilliefors/Kolmolgorov-Smirnov test applied to fitted gamma (KSGam) and Pearson type III (KSPE) distributions to monthly rainfall series. μ, σ, γ, α, β and are the parameters of the distributions. The p-value is shown in brackets.

|	KS	KS	γ	σ	μ	α	β
Monthy		Ribeirão Preto					
January	0.05 (0.35)	0.05 (0.51)	0.7	109.0	278.9	6.6	42.4
February	0.08 (0.26)	0.09 (0.11)	0.5	105.7	220.9	3.9	56.5
March	0.05 (0.33)	0.07 (0.20)	0.7	77.6	166.7	4.0	41.4
April	0.08 (0.17)	0.08 (0.03)	1.6	56.8	72.6	1.6	45.2
May	0.10 (0.06)	0.12 (0.01)	1.1	48.6	55.7	1.0	58.1
June	0.16 (0.05)	0.23 (0.01)	1.1	30.1	27.1	0.9	36.8
July	0.16 (0.05)	0.06 (0.23)	1.1	26.1	21.5	0.8	35.4
August	0.15 (0.05)	0.14 (0.06)	1.1	24.8	19.2	0.8	32.9
September	0.07 (0.18)	0.07 (0.16)	1.8	50.4	51.2	1.1	49.6
October	0.06 (0.20)	0.07 (0.10)	0.7	62.5	125.2	4.0	31.6
November	0.03 (0.35)	0.05 (0.33)	0.6	73.6	173.0	5.1	33.9
December	0.06 (0.21)	0.05 (0.20)	0.8	112.3	270.4	5.4	49.7
		Ubatuba					
January	0.07 (0.22)	0.07 (0.20)	0.9	179.4	362.0	3.9	92.9
February	0.05 (0.36)	0.07 (0.33)	1.0	211.8	326.2	2.2	151.6
March	0.06 (0.21)	0.07 (0.12)	1.3	178.8	320.7	3.5	92.3
April	0.10 (0.16)	0.08 (0.06)	1.7	131.8	222.3	3.4	65.1
May	0.03 (0.54)	0.05 (0.53)	1.0	71.3	125.7	2.9	43.6
June	0.05 (0.53)	0.06 (0.20)	1.7	72.5	89.9	1.7	52.4
July	0.07 (0.62)	0.06 (0.21)	1.7	78.3	101.2	2.0	50.9
August	0.10 (0.09)	0.07 (0.11)	1.9	71.8	99.7	2.3	42.6
September	0.07 (0.12)	0.09 (0.12)	0.6	81.1	173.6	3.8	45.2
October	0.07 (0.21)	0.08 (0.11)	0.6	101.0	234.1	5.3	44.0
November	0.06 (0.12)	0.06 (0.11)	1.1	123.4	269.3	5.1	53.3
December	0.07 (0.12)	0.09 (0.03)	1.1	135.2	323.5	6.2	52.2
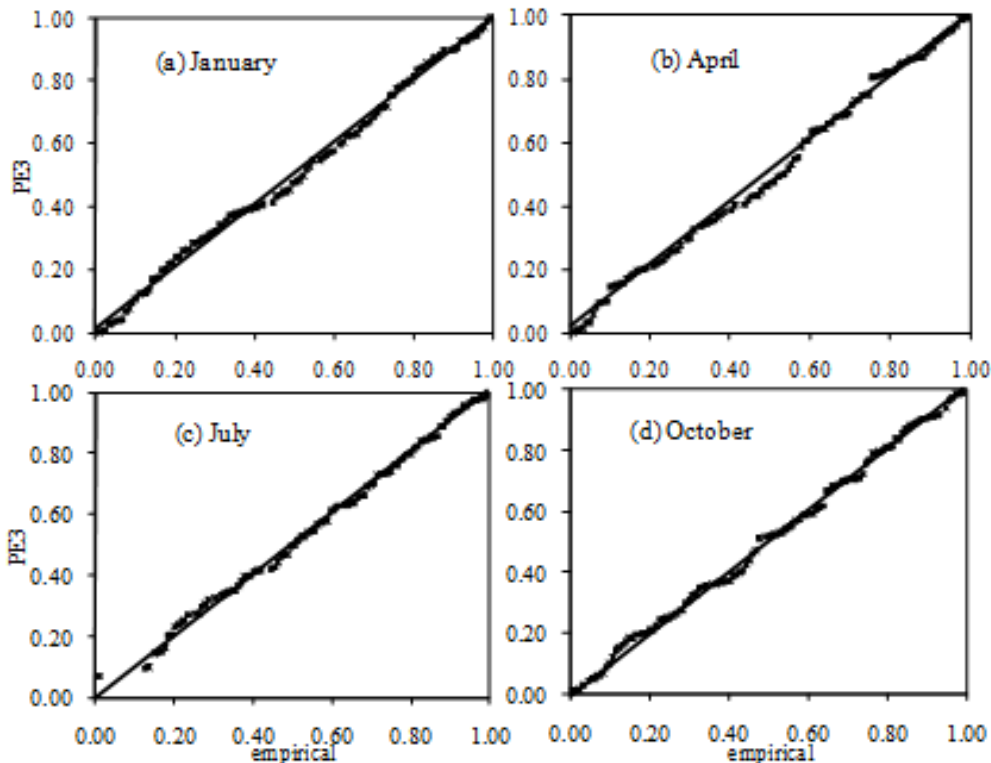
Figure 2 - Percentil-percentil plots between fitted pearson type III (PE3) cumulative probability and empirical rainfall cumulative probability. Campinas, SP-Brazil (1890 to 2008)
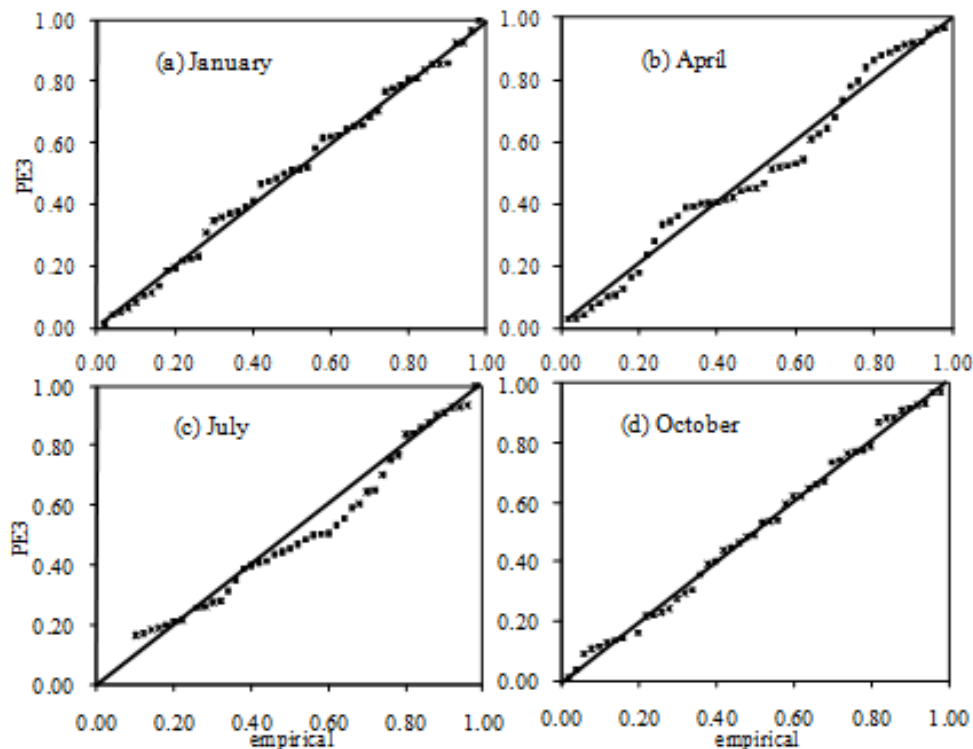

Figure 3 - Percentil-percentil plots between fitted pearson type III (PE3) cumulative probability and empirical rainfall cumulative probability. Presidente Prudente, SP-Brazil (1960 to 2008)
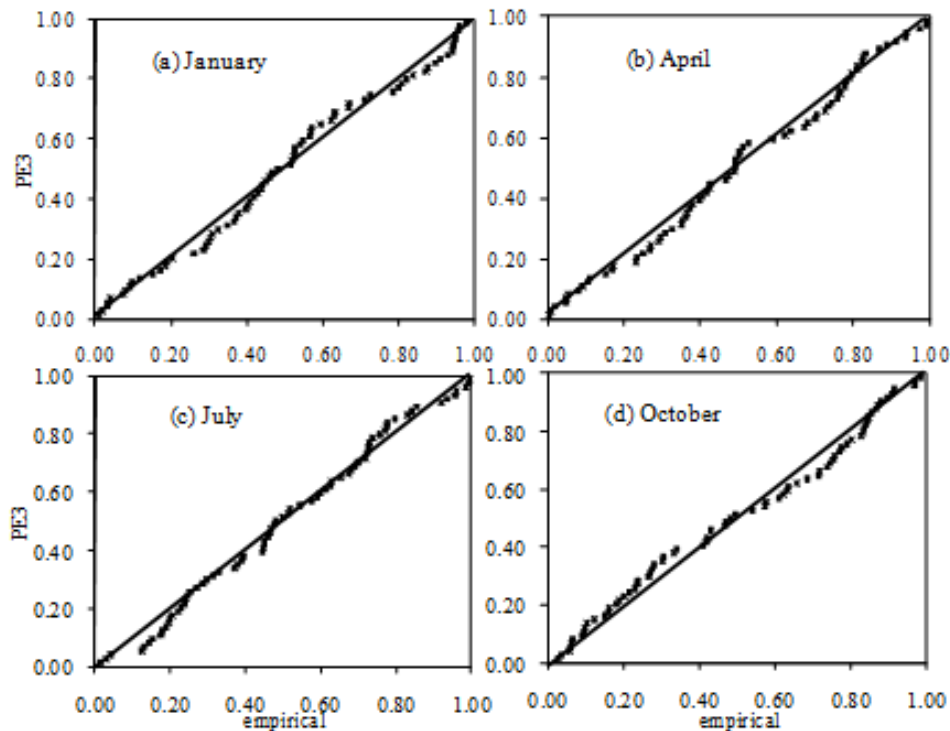
Figure 4 - Percentil-percentil plots between fitted pearson type III (PE3) cumulative probability and empirical rainfall cumulative probability. Ubatuba, SP-Brazil (1935 to 2008)

difference (p<0.05 for all series) between SPIPE3 and SPIGam temporal variability. All R^2 and slope values are close to unit. All intercept values are also close to zero. However, for few extreme SPI negative values, there are remarkable differences between SPIPE3 and SPIGam values.

For the weather station of Campinas, the major difference between SPIPE3 and SPIGam occurred during March, 1908 for a SPIPE3= -2.5 and a SPIGam= -3.6. However, these both SPI values indicate the same drought categories (Table 1). Considering drought categories, the major difference between SPIPE3 and SPIGam occurred when there was no rainfall amount during the months of June (1929, 1936, 1963, 1966, 1979, 1984, 1986 and 2002) and during the months of September (1986 and 2002). In these both cases (months), while the SPIPE3 was indicating severe drought conditions, the SPIGam was indicating extreme drought conditions.

For the weather station of Pindorama, while the SPIPE3 was indicating, during the months of October (1984 - 24mm; and 1985 – 18mm), severe drought conditions, the SPIGam was indicating extreme drought conditions. The opposite feature occurred during the months of April (1952 and 1953) when no precipitation was observed. For these last events, while the SPIPE3 was indicating extreme drought conditions, the SPIGam was indicating severe drought conditions.

For the weather station of Presidente Prudente, the SPIPE3 indicated during the months of May, when no precipitation was observed (1967 and 1981), extreme drought condition. For these same (no) precipitation events, the SPIGam indicated severe drought conditions. The observed precipitation data of November 1970 was 32mm. For this rainfall amount, the SPIPE3 and SPIGam values were, respectively, -1,85 (severe drought) and -2,16 (extreme drought).

For the weather station of Ubatuba (Figure 6), while the SPIPE3, during the months of April, 1963 (27mm) and August, 1988 (8mm), was indicating severe drought conditions, the SPIGam was indicating extreme drought conditions. For the weather station of Ribeirão Preto, during the months of October, 1944 (26,4mm) and February, 2005 (50,2mm), while the SPIPE3 was indicating severe drought conditions, the SPIGam was indicating extreme drought conditions.

Figures 7 and 8 show the frequency distribution of dry SPI categories (Table 1) for the five weather stations. It is worth mentioning that the limitation of the SPI in identifying regions that may be more drought prone than others becomes evident since the frequencies of occurrence of the dry events were approximately the same among the five weather stations. Also according to Figures 7 and 8, the frequency of the SPIPE3 extreme drought events tends to be lower than the frequency
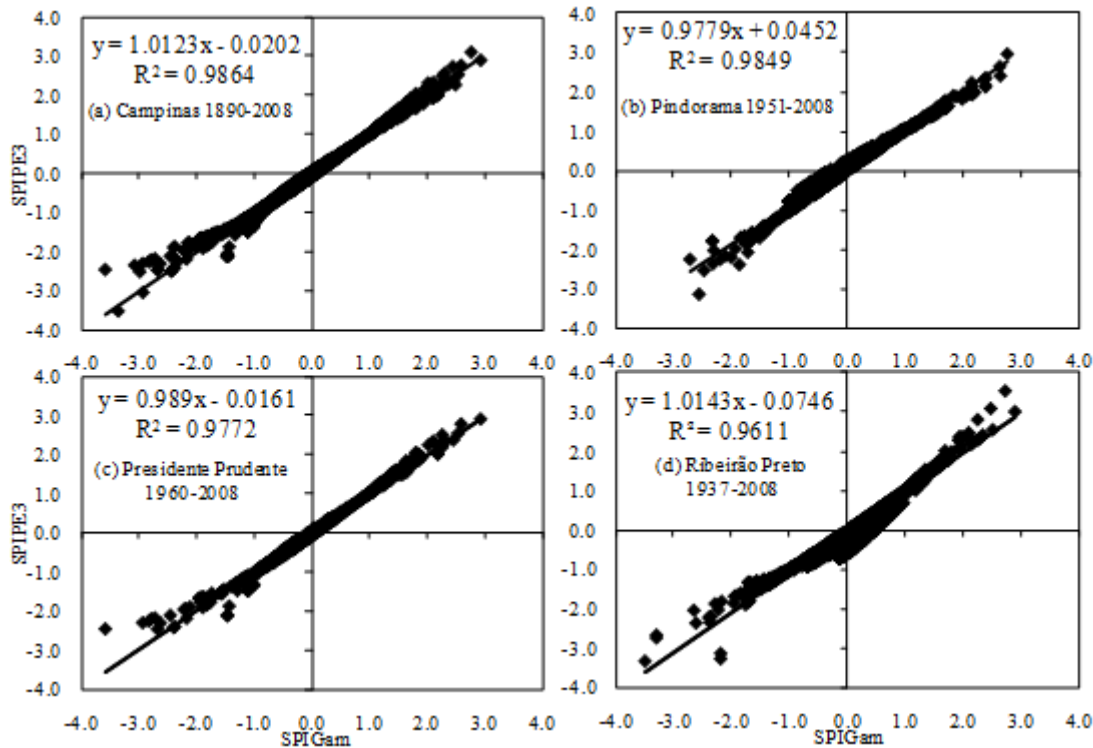
Figure 5 - Linear regression analysis between Standardized Precipitation Index values, when computed based on pearson type III (SPIPE3) and Gam (SPIGam) distributions

of the same SPIGam event. For the station of Ribeirão Preto, the difference between the frequencies of SPIPE3 and SPIGam mild drought event is greater than 5%.

As indicated by Tables 7 and 8, almost all SPI non-normal distributions occur during the winter season. Further studies should better investigate the reasons for this non-normality, which seems to be related with the high probability of zero rainfall values during the dry season of the State of São Paulo. In this view, following Wu et al. (2007), it is worth mentioning that expression 7 computes both negative and positive SPI values. Consequently, in order to have balanced negative and positive value within a SPI time series, t (Equation 8) must be the same under the situations (i) $0 < Hg(x)$ or $Hp(x) \leq 0.5$ and (ii) $0.5 < Hg(x)$ or $Hp(x) \leq 1$.

Comparing the values within Tables 3 and 4 (number of no precipitation events), with the values within Tables 7 and 8, it becomes evident that in the locations (and months) where there is a high probability of zero rainfall values (Pindorama, Presidente Prudente and Ribeirão Preto), the lowest possible value of Hg(x) and/or Hp(x) is far from zero. Thus the high number of zeros, observed during the winter season led to an unbalanced Equation 8. As a result, the SPI time series were non-normally distributed (Table 7). On the contrary, for the weather station of Ubatuba, because all the monthly precipitation

totals are greater than zero, the departures of Equation 8 were symmetric distributed. Thus the SPI was normally distributed during all months (Table 8).

Despite the reasons, the number of non-normal SPIPE3 distribution (3) was lower than the number of non-normal SPIGam distribution (9). In this view, once again, the PE3
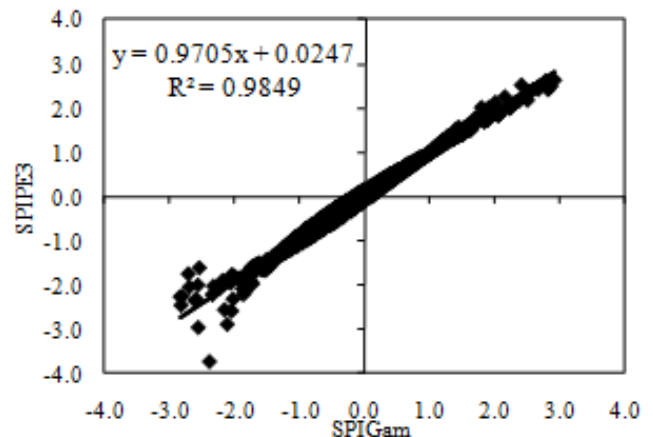

Figure 6 - Linear regression analysis between Standardized Precipitation Index values, when computed based on pearson type III (SPIPE3) and Gam (SPIGam) distributions. Ubatuba, State of São Paulo, Brazil.

distribution seems to be a better choice than the Gam distribution, in developing a drought indicator which standards the rainfall deficits/excess on temporal and regional basis.

3.2 Others time scales (3 and 6 months)

The SPI was designed as a drought index that recognizes the importance of times scales in the analysis of water availability and water use. Thus, it becomes necessary to assess

the robustness of Gam and PE3 distributions in describing precipitation series for larger time scales (larger than the monthly time scale).

For 3-months time scale, (Table 9) the Ho, associated with the 2-parameters gamma distribution, could not be accepted for the period ending on the months of July and August (Campinas), September (Ribeirão Preto), January, and December (Ubatuba). For 6-months time scale, the Ho, associated with this former distribution (Table 8) could not be accepted for the periods

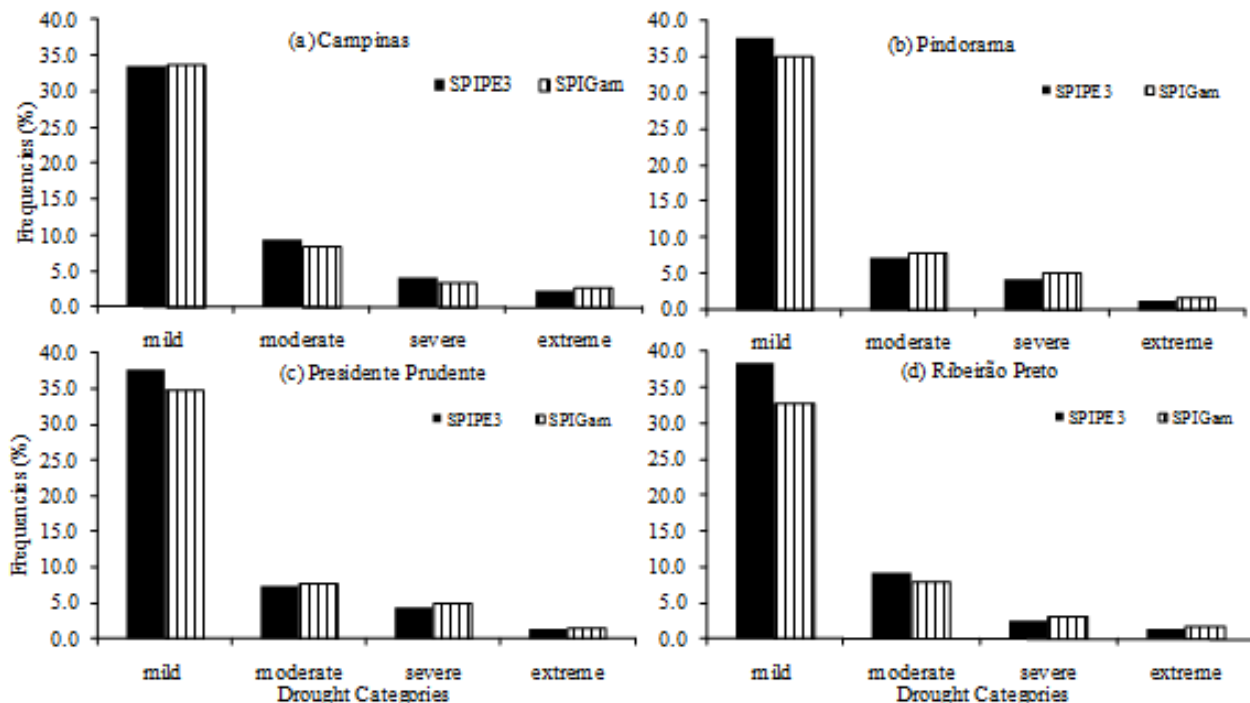**Figure 7 -** Frequency distributions of monthly dry SPI categories for four weather stations of the State of São Paulo, Brazil. Campinas (1890 to 2008), Pindorama (1951 to 2008), Presidente Prudente (1960 to 2008), Ribeirão Preto (1937 to 2008).
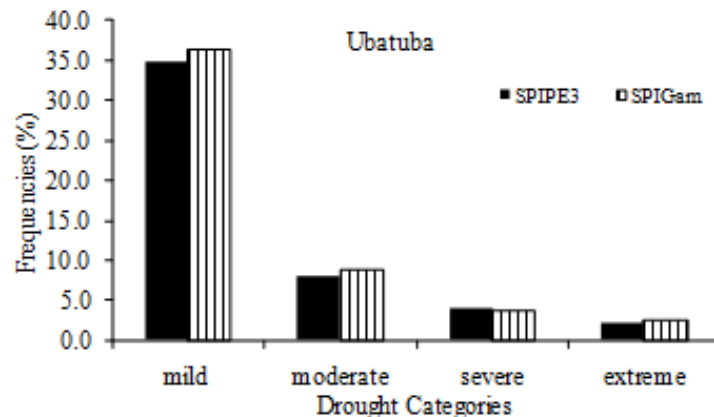

Figure 8 - Frequency distributions of monthly dry SPI categories for the weather station of Ubatuba (1935 to 2008), State of São Paulo, Brazil.

Table 7 - Lilliefors/Kolmolgorov-Smirnov test applied to fitted gamma (KSGam) and Pearson type III (KSPE) distributions to monthly rainfall series. μ, σ, γ, α, β and are the parameters of the distributions. The p-value is shown in brackets.

Month	SPIPE3											
	Campinas			Pindorama			Pres. Prudente			Rib. Preto		
	W	p	Med	W	p	Med	W	p	Med	W	p	Med
January	0.99	0.20	-0.10	0.99	0.68	0.04	0.99	0.30	0.03	0.99	0.95	0.07
February	0.99	0.30	0.01	0.98	0.62	0.03	0.96	0.09	0.00	0.98	0.18	0.12
March	0.99	>0.70	0.01	0.99	0.81	0.01	0.99	0.40	0.01	0.99	>0.70	0.06
April	0.97	0.03	0.09	0.99	>0.70	0.08	0.97	0.51	0.10	0.97	0.11	0.03
May	0.98	0.15	0.06	0.98	0.55	0.07	0.96	0.20	0.06	0.95	0.01	0.05
June	0.98	0.04	0.01	0.94	<0.01	0.03	0.97	0.49	0.11	0.92	<0.01	0.04
July	0.97	<0.01	0.02	0.90	<0.01	0.16	0.97	0.70	0.05	0.90	<0.01	0.00
August	0.96	<0.01	0.05	0.91	<0.01	0.08	0.91	<0.01	0.03	0.89	<0.01	0.06
September	0.98	0.18	0.12	0.97	0.16	0.02	0.98	0.66	0.03	0.98	0.21	0.11
October	0.99	>0.70	0.04	0.97	0.24	0.05	0.99	0.18	0.00	0.97	0.12	0.01
November	0.99	>0.70	0.03	0.98	0.38	0.13	0.99	0.36	0.02	0.99	>0.70	0.04
December	0.99	0.66	0.01	0.99	>0.70	0.03	0.96	0.27	0.02	0.97	0.07	0.02
Month	SPIPGam											
	W	p	Med	W	p	Med	W	p	Med	W	p	Med
January	0.99	>0.70	0.06	0.98	>0.70	0.03	0.98	>0.70	0.03	0.99	0.10	0.08
February	0.98	0.13	0.09	0.98	>0.70	0.03	0.98	>0.70	0.04	0.96	0.09	0.11
March	0.97	0.02	0.10	0.99	0.19	0.05	0.98	>0.70	0.02	0.97	0.13	0.12
April	0.99	0.39	0.06	0.98	>0.70	0.09	0.96	0.14	0.02	0.98	0.65	0.04
May	0.98	0.14	0.14	0.98	0.67	0.09	0.97	0.65	0.03	0.95	0.01	0.23
June	0.97	0.03	0.01	0.93	<0.01	0.04	0.98	0.70	0.20	0.92	<0.01	0.04
July	0.94	<0.01	0.10	0.90	<0.01	0.09	0.96	0.14	0.01	0.91	<0.01	0.08
August	0.95	<0.01	0.10	0.91	<0.01	0.11	0.89	<0.01	0.01	0.85	<0.01	0.12
September	0.98	0.09	0.22	0.97	0.37	0.12	0.98	0.61	0.01	0.98	0.74	0.17
October	0.97	0.04	0.11	0.97	0.47	0.16	0.94	0.04	0.06	0.98	0.48	0.04
November	0.99	>0.70	0.08	0.98	>0.70	0.14	0.99	0.25	0.03	0.98	0.58	0.08
December	0.99	0.69	0.01	0.98	>0.70	0.10	0.97	0.39	0.01	0.96	0.07	0.11

ending on the month of May (Campinas), May and October (Pindorama), April and November (Ribeirão Preto), May and August (Ubatuba). Also according to Table 8, the Ho, associated with the Pearson Type III distribution, could be accepted for all periods at both time scales. Furthermore, in the most of the cases, the p-value associated with the PE3 distributions was numerically greater than the p-value associated with the Gam distribution. Thus, it becomes evident from the p-values in Table 9 that the Gam distribution fits the precipitation data less well than the PE3 does. In this last view, as it was the case for the monthly time scale, the use of the PE3 distribution is once again recommend**.**

For both 3- and 6-moths times scales all SPIPE3 series can be seen as normally distributed. In the location of Ribeirão Preto, the SPIGam with time scales 3- and 6-months calculated for the period ending on the month of June are distributed non-normally.

Table 8 - Normality test of Standardized Precipitation Index values (1-month time scale), based on Pearson type III distribution (SPIPE3) and on gamma 2-parameters distribution (SPIGam). Ubatuba, State of São Paulo, Brazil.

Month	W	p	Med	W	p	Med
January	0.98	0.27	0.01	0.98	0.44	0.07
February	0.99	>0.70	0.05	0.99	>0.70	0.11
March	0.99	0.61	0.02	0.99	>0.70	0.06
April	0.99	0.58	0.00	0.97	0.24	0.19
May	0.99	>0.70	0.00	0.99	>0.70	0.03
June	0.99	>0.70	0.02	0.99	0.43	0.06
July	0.98	0.33	0.01	0.98	0.44	0.16
August	0.99	0.66	0.05	0.98	0.44	0.11
September	0.98	0.44	0.02	0.95	0.03	0.12
October	0.98	0.17	0.03	0.98	0.49	0.00
November	0.98	0.38	0.11	0.98	0.69	0.02
December	0.98	0.43	0.06	0.98	>0.70	0.02

Table 9 - The p-value associated with the Lilliefors/Kolmolgorov-Smirnov test applied to fitted gamma and Pearson type III distributions to 3- and 6-months rainfall series.

Time scale	Campinas		Pindorama		Pres.Prudente		Rib.Preto		Ubatuba	
	3-	6-	3-	6-	3-	6-	3-	6-	3-	6-
	Pearson Type III distribution									
January	0.07	0.19	0.05	0.45	0.06	0.42	0.06	0.35	0.04	0.60
February	0.05	0.26	0.05	0.44	0.06	0.42	0.06	0.36	0.05	0.34
March	0.06	0.15	0.07	0.33	0.08	0.23	0.07	0.12	0.05	0.34
April	0.08	0.05	0.09	0.15	0.10	0.10	0.09	0.23	0.08	0.13
May	0.09	0.08	0.07	0.46	0.08	>0.70	0.08	0.25	0.07	0.21
June	0.05	0.23	0.05	0.45	0.06	0.40	0.07	0.20	0.05	0.35
July	0.05	0.67	0.09	0.16	0.08	0.22	0.07	0.19	0.08	0.12
August	0.05	0.23	0.07	0.29	0.08	0.21	0.08	0.18	0.07	0.3
September	0.08	0.09	0.05	0.49	0.13	0.06	0.04	0.63	0.05	0.36
October	0.05	0.51	0.07	0.27	0.08	0.39	0.04	0.54	0.05	0.44
November	0.06	0.23	0.07	0.3	0.08	0.33	0.10	0.27	0.08	0.14
December	0.05	0.34	0.09	0.15	0.08	0.26	0.06	0.47	0.05	0.45
January	0.07	0.1	0.07	0.26	0.06	0.39	0.06	0.33	0.04	0.53
February	0.05	0.23	0.07	0.28	0.06	0.38	0.07	0.20	0.07	0.2
March	0.08	0.05	0.07	0.28	0.08	0.22	0.10	0.06	0.07	0.2
April	0.08	0.05	0.10	0.07	0.13	0.05	0.11	0.02	0.07	0.21
May	0.09	0.02	0.09	0.14	0.17	0.08	0.12	0.11	0.03	
June	0.07	0.09	0.05	0.45	0.06	0.39	0.08	0.12	0.08	0.12
July	0.09	0.35	0.09	0.15	0.08	0.22	0.06	0.32	0.08	0.09
August	0.05	0.35	0.09	0.14	0.06	0.4	0.07	0.09		
September	0.08	0.05	0.09	0.15	0.10	0.11			0.05	0.35
October	0.05	0.35	0.07	0.26	0.13	0.04	0.06	0.33	0.04	0.34
November	0.05	0.22	0.07	0.27	0.06	0.39	0.11	0.03	0.08	0.12
December	0.05	0.23	0.10	0.07	0.10	0.10	0.07	0.2	0.04	0.53

4. CONCLUSIONS

Since the SPI was developed as a standardized (on regional and temporal basis) drought index, the detected differences between SPIPE3 and SPIGam final values indicate the need of adopting uniform distribution model in calculation the SPI in the State of São Paulo, Brazil. Considering the applied goodness-of-fit tests the Pearson type III distribution seems to be a better choice than the Gamma distribution, in describing the long-term rainfall series of the State of São Paulo. In addition, it was observed that the number of SPI time series that could be seen as normally distributed was higher when this drought index was computed from the Pearson Type III distribution. Thus, the use of the Pearson type III distribution within the calculation algorithm of the SPI is recommended in the State of São Paulo, Brazil.

5. REFERENCES

BLAIN, G. C. **Avaliação e adaptação do Índice de Severidade de Seca de Palmer (PDSI) e do Índice Padronizado de Precipitação (SPI) às condições climáticas do Estado de São Paulo**, 2005, 120p, Dissertação (Mestrado em Agricultura Tropical e Sub-tropical) – Instituto Agronômico, Campinas – SP.

BLAIN, G. C.; PIEDADE, S. M. S.; CAMARGO, M. B. P.; GIAROLLA, A. Distribuição temporal da precipitação pluvial mensal observada no Posto Meteorológico do Instituto Agronômico, em Campinas, SP. **Bragantia**, v. 66, n.2, p. 347-355, 2007.

BLAIN, G. C. Considerações estatísticas relativas à oito séries de precipitação pluvial da Secretaria de Agricultura e Abastecimento do Estado de São Paulo, **Revista Brasileira de Meteorologia**, v.24, n. 1 p.12-23, 2009.

BLAIN, G. C.; KAYANO, M. T.; CAMARGO, M. B. P.; LULU, J. Variabilidade amostral das séries mensais de precipitação pluvial em duas regiões do Brasil: Pelotas-RS e Campinas-SP. Revista Brasileira de Meteorologia, v.24, n.11 p.1-04, 2009.

CRUTCHER, H.L., A note on the possible misuse of the Kolmogorov-Smirnov test. **J. Appl. Meteor., v.**14, p. 1600-1603, 1975.

GUTTMAN, G.B. Comparing the Palmer Drought Index and the "Standardized Precipitation Index". **Journal of the American Water Resources**, v.16, n.2, p. 113-121, 1998.

GUTTMAN, G.B. Accepting the "Standardized Precipitation Index": A calculation algorithm Index. **Journal of the American Water Resources**, v.35, n.2, p. 311-322, 1999.

HAMED, K.H.; RAO, A.R. A modified Mann-Kendall trend test for auto-correlated data. **Journal of Hydrologic**, 204, p. 182-196. 1998.

HAYES, M. J.; SVOBODA, M.D.; WILHITE, D.A.; VANYARKHO, O.V. Monitoring the 1996 drought using the Standardized Precipitation Index. **Bulletin of the American Meteorological Society**. v. 80, n.3, p.429-438, 1999.

HOSKING, J.R.M.; L-moments: analysis and estimation of distribution using linear combination of order statistics, **Journal of the Royal Statistical Society**, v. 58, n.1, p. 105-124, 1990.

HOSKING, J.R.M.; WALLIS, J.R. **Regional Frequency Analysis**, Cambridge: Cambridge University Press, 1997. 224p.

KEYANTASH, J.; DRACUP, J.A. The quantification of drought: An evaluation of drought indices. **Bulletin of the American Meteorological Society,** V. 83, n.8, p.1167-1180, 2002.

KUMAR, M.N.; MUTHLY, C.S.; SAI, M.V.R.S.; ROY, P.S. On the use of Standardized Precipitation Index (SPI) for drought intensity assessment, **Meteorological applications**, v.16, p.381-389, 2009.

KENDALL, M. A.; STUART, A. **The advanced theory of statistics**, Londres: Charles Griffin & Company Limited, 1967 v2 ed 2, 690p.

MCKEE, T.B.; DOESKEN, N.J.; KLEIST, J. The relationship of drought frequency and duration to times scale. In: CONFERENCE ON APPLIED CLIMATOLOGY, 8.,1993, Boston. **Anais…** American Meteorological Society, Boston: PREPRINTS, p.179 – 184, 1993.

MCKEE, T.B.; DOESKEN, N.J. e KLEIST,J. Drought monitoring with multiple times scales. In: CONFERENCE ON APPlIED CLIMATOLOGY, 9.,1995, Boston. **Anais…** American Meteorological Society, Boston: PREPRINTS, p.233 – 236, 1995.

PANOFSKY, H.A.; BRIERG, W. **Some applications of statistics to meteorology,** Pennsylvania: Pennsylvania State University, 1968. 224p.

SANSIGOLO, C.A. Análise comparativa do desempenho dos Índices de Seca de Palmer (PDSI) e da Precipitação Normalizada (SPI) em Piracicaba, SP (1917 – 2001). **Revista Brasileira de Meteorologia**, v.19, n.1, p. 51-56, 2004.

SNEYERS, R. Sur l´analyse statistique des séries d´observations. **OMM Note Technique, n. 143**, 192p. 1975.

SOTO, L.G.M.; RUTHERFORD, B.M.; MAAROUF, A. Standardized Precipitation Index Zones for México**, Atmosfera,** v.18, n.1, p.33-56, 2005

THOM, H. C. S. Same methods of climatological analyse*s*. **World Meteorological Organization**, Geneva, 53p, 1966. (WMO, 199; TP, 103; Technical note. ,81.)

WILKS, D.S. **Statistical methods in the atmospheric sciences,** San Diego: Academic Press, 2nd ed, 2006, p, 102-104.

WU, H.; SVOBODA, M.D.; HAYES, M.J.; WILHITE, D.A.; WEN, F. Appropriate application of the Standardized Precipitation Index in arid locations and dry seasons, **International Journal of Climatology**, V.27, P.65–79, 2007

ZHAI, J.; SU, B.; KRYSANOVA, V.; TOBIAS, VETTER, T.; GAO, C.; JIANG, T. Spatial variation and trends in PDSI and SPI indices and their relation to streamflow in ten large regions of China, **Journal of Climate**, v.23, n.3, p. 649-663, 2010