Monthly values of the standardized precipitation index in the State of São Paulo, Brazil: trends and spectral features under the normality assumption

Gabriel Constantino Blain (*)

Instituto Agronômico (IAC), Centro de Pesquisa e Desenvolvimento de Ecofisiologia e Biofísica, Caixa Postal 28, 13012-970 Campinas (SP), Brazil.
(*) Corresponding author: gabriel@iac.sp.gov.br

Received: April 5, 2011; Accepted: Sept. 5, 2011

Abstract
The aim of this study was to describe monthly series of the Standardized Precipitation Index obtained from four weather stations of the State of São Paulo, Brazil. The analyses were carried out by evaluating the normality assumption of the SPI distributions, the spectral features of these series and, the presence of climatic trends in these datasets. It was observed that the Pearson type III distribution was better than the gamma 2-parameter distribution in providing monthly SPI series closer to the normality assumption inherent to the use of this standardized index. The spectral analyses carried out in the time-frequency domain did not allow us to establish a dominant mode in the analyzed series. In general, the Mann-Kendall and the Pettitt tests indicated the presence of no significant trend in the SPI series. However, both trend tests have indicated that the temporal variability of this index, observed at the months of October over the last 60 years, cannot be seen as the result of a purely random process. This last inference is due to the concentration of decreasing trends, with a common beginning (1983/84) in the four locations of the study.

Key words: Pearson type III distribution, drought, climate trends.

1. INTRODUCTION

Strongly related to a deficit in the rainfall amounts, the drought events can cause severe damages to human society. Under this consideration, the standardized precipitation index (SPI) was developed by McKee et al. (1993) as a statistical procedure (normalized in time-space domain) capable of representing drought and flood events in a similar probabilistic way, even when different rainfall regimes (observed in distinct regions and periods) are being evaluated (Wu et al., 2007). Also according to Wu et al. (2007), being normally distributed is the major reason that leads the SPI to be widely accepted in operational mode as well as in academic studies all over the world.

For instance, after have applied the SPI in Nebraska, USA, Tadesse et al. (2005) indicated that the most occurrences of drought events (1950 to 1999) was preceded, among others oceanic-atmospheric parameters, by positive values of the Southern Oscillation Index (SOI; related to the El Niño/southern oscillation - ENOS). Livida and Assemaopoulos (2007) used this drought...
index to evaluate, on regional basis, the frequency, the duration and the intensity of the drought events detected in Greece from 1951 to 2000. Mishra et al. (2007) employed the SPI to evaluate, among other features, the transition probabilities of drought events in the Kansabati River basin; India. By using the SPI, Zhang et al. (2009) observed changes in the events of drought and wetness in Pearl River Basin; China (1960-2005). Zhai et al. (2010) demonstrated that the SPI can be used to describe possible trends in the dryness and wetness severity in the regions of China, as well as in other countries. Manatsha et al. (2010) used the SPI to address several features of agricultural droughts in Zimbabwe, Africa. Li et al. (2008) observed that the values of the SPI computed over southern Amazon region decreased in the period of 1970 to 1999 by 0.32 per decade. According to these authors this result indicates an increase trend in dryness conditions.

In Brazil, the SPI is largely used on operational mode by governmental agricultural institutions, such as Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA; www.agritempo.gov.br), Instituto Agronômico (IAC; www.infoseca.sp.gov.br) and, Instituto Nacional de Meteorologia (INMET; www.inmet.gov.br), in order to monitoring (possible) drought conditions in different regions of the country. Furthermore, the studies of Fernandes et al. (2010), Macedo et al. (2010) and, Sansigolo (2004), address the use of the SPI in the climate conditions of Brazil. In order to assess the robustness of the SPI (among other drought indexes) in estimating the upland rice yield (1983/1984 - 2005/2006; Goiânia- State of Goiás), Fernandes et al. (2010) have addressed important features (including its limitations) of the use of this standardized index as an agricultural drought indicator. Macedo et al. (2010) calculated the SPI in order to monitoring drought events within three rainfall homogeneous sub-regions of the State of Pará. The main purpose of Sansigolo (2004) was to establish a comparative performance analysis between the Palmer Drought Index (Palmer, 1965) and the SPI in Piracicaba, State of São Paulo.

However, despite all the concerns related to the possible consequences of the global warming over the temporal variability of drought events, such as a possible increasing in its intensity and frequency, the evaluation of the presence of trend components in time series of this drought index seems to be still neglected in the Brazilian literature. Thus, considering that the analysis of long SPI series provides important information to the climate literature, which may be helpful in reducing the human vulnerability to drought events, the aim of the study was to describe the temporal variability of monthly SPI series obtained from four weather stations of the State of São Paulo. This description was carried out by evaluating (i) the normality assumption of the SPI distributions, (ii) the spectral features of these series (including their possible relationship with the ENOS) and, (iii) the presence of climate trends in these datasets.

2. MATERIAL AND METHODS

Monthly rainfall series available from four weather stations of the State of São Paulo, Brazil (Campinas, 1890 to 2010; Jundiaí 1942 to 2010; Pindorama, 1951 to 2010 and; Ribeirão Preto 1937-2010; figure 1) were chosen for this study. After Guttmann (1994) has investigated the effect of the length of the data records on rainfall distributions, he stated that at least 40-60 years are required for parameter estimation stability. Thus, the series were chosen in order to provide 60 years of data records. The selected series did not have missing values and their consistencies were assessed in Blain (2009). Also according to Blain (2009) the main feature of the monthly rainfall series (PRE) of the State of São Paulo is its great variability in time-space domain. For a given location it can be observed strongly skewed rainfall distributions during the winter season; however, during the summer season the shape of these monthly distributions becomes similar to the “bell shape” of the Gaussian distribution.

The gamma 2-parameter distribution (Gam) is frequently used in the State of São Paulo (as well as in others regions of Brazil) in describing monthly rainfall records (and, therefore, in the SPI calculation algorithm). However, according to Guttmann (1999) the Pearson type III (PE3) distribution, also known as gamma 3-parameter, is ‘the best universal model for the SPI calculations’. Furthermore, Wu et al. (2007) stated that the confidence in SPI results, obtained from the Gam distribution (undefined for the 0 value), can be affected since this probability density function has only two free parameters. According to these authors, the use of alternatives distributions, such as the PE3, is worth to be evaluated. In this study, the cumulative probability associated with the Gam model was estimated as described in Blain (2009) and Wu et al. (2007), while the PE3 was estimated as recommended by Guttmann (1994, 1999) and Yue and Hashino (2007). In order to obtain the SPI, the [H(PRE)] values were subject to the standardization procedure based on the following equations proposed by Abramowitz and Stegun, (1965).

\[
SPI = \begin{cases} 
\frac{co+ct+ct^2}{d+dt+d^2t^2} & \text{when } 0 < H(PRE) \leq 0.5 \\
+ \frac{co+ct+ct^2}{d+dt+d^2t^2} & \text{when } 0.5 < H(PRE) < 1 
\end{cases}
\]

(1)

where

\[
t = \left\lfloor \frac{\ln\left( \frac{1}{(H(PRE))^2} \right) }{2} \right\rfloor \]

(2)

when \(0 < H(PRE) \leq 0.5\)
where $0 < H(PRE) < 1$

$$t = \sqrt{\frac{1}{\ln\left(1 - H(PRE)\right)^2}}$$

$$c_0 = 2,515517; \quad c_1 = 0,802853; \quad c_2 = 0,010328;$$

$$d_1 = 1,432788; \quad d_2 = 0,189269; \quad d_3 = 0,001308$$

Although the use of the Gam distribution in characterizing monthly rainfall series of the State of São Paulo has already been extensively evaluated (Blain et al., 2007; Blain et al., 2009; Blain 2009; 2010), the use of PE3 distribution for this purpose seems to be little assessed. Since the SPI is a probabilistic index, assessing the robustness of the PE3 distribution in describing rainfall series may be seen as the first step in the study of the possibility of adopting this distribution in the SPI calculations.

As discussed by Wilks (2006), Steinskog et al. (2007) and Vlcek and Huth (2009), if (and only if) the parameters of the theoretical distribution have not been estimated from the same dataset used to evaluate the fit of the parametric distribution, the original algorithm of the Kolmogorov-Smirnov test is applicable. Thus, since the three parameters of the PE3 were fitted using all available data, the test had to be modified. Hereafter this adapted method will be referred as Kolmogorov-Smirnov/Lilliefors test (KS-L). The statistical simulations required for calculating the KS-L test were based on the procedure called “nonuniform random number generation by inversion”. It were generated $N_s=100000$ synthetic data samples. More information about the KS-L can be found in Wilks (2006).

According to Edwards and McKee (1997) and Wu et al. (2007) despite the presence of equations 1 and 2 within the SPI algorithm, for dry climates or those with a distinct dry season associated with strongly skewed rainfall distributions (in which the zero value is common), the SPI series may not be normally distributed. At this point, it is worth emphasizing that also according to Wu et al. (2007) the SPI is widely accepted throughout the world because, conceptually, ‘it is normalized to a location and is normalized in time’. Thus, a normality test was applied on the SPI values obtained from the Gam and from the PE3 models. The parametric distribution capable of providing the greatest number of normally distributed SPI series will be seen as the best probabilistic model for the calculations of this standardized index. The SPI series will be considered as coming from a standard normal distribution if at least one of the following criteria is met: (a) Shapiro-Wilks statistic ($W$) equal or greater than 0.96 (p-value equal or greater than 0.10) or (b) the absolute value of the median is equal or less than 0.05. For further information about this procedure see Wu et al. (2007).

Following Torrence and Compo (1998) the wavelet analysis was used in order to decomposing the SPI series into time-frequency domain. Consequently, this sort of spectral analysis has allowed us to observe the variance peaks in the frequency domain and also to verify how
those peaks vary in time. Detailed explanation of the continuous wavelet transform (CWT) used in this study can be found in Torrence and Compo (1998), Grinsted et al. (2004), Rebouca et al. (2006), Blain (2009), Pezzi and Kayano (2009) and, Kayano and Sansigolo (2009). The wavelet analysis (including its statistical significance testing) was estimated based on the computational procedure described by Torrence and Compo (1998) and available at http://paos.colorado.edu/research/wavelets (accessed at November 30, 2010).

As pointed out by Grinsted et al. (2004), it is frequently desirable to evaluate the variability of two time series that may be linked to each other by some physical reason. In this view, and also according to Grinsted et al. (2004), it is of great interest to verify whether the CWT of these time series have a consistent phase relationship and common periods of large wavelet power. Thus, in order to evaluate possible influences of the ENOS over the temporal variability of drought and/or floods events observed in the four locations under analysis, it was applied the squared wavelet coherence, as described in Grinsted et al. (2004), between the SPI and the SOI series. The cross-wavelet calculation algorithm can be obtained at http://www.pol.ac.uk/home/research/waveletcoherence (copyright 2002-2004; Aslak Grinsted). The SOI data is available at http://www.bom.gov.au/climate. It is worth emphasizing that values of the SOI above 8 may be seen as representative of La Niña events. Values below -8 may be seen as representative of El Niño events, while values between -8 and 8 usually indicate neutral conditions.

The Mann-Kendall (MK) test (Kendall and Stuart, 1967) is largely used for evaluating the presence of trends in meteorological time series. The null hypothesis (Ho) associated with this test assumes that the dataset under evaluation is free from trends. The absence of significant serial correlation is also assumed. The probability of occurrence of the type I error (p-value) can be estimated from the Gaussian standard distribution. The MK test was applied in each one of the 12 monthly datasets (obtained from each weather station) in order to avoid possible influences of temporal persistence in the MK final values. In this case, the time span between two subsequent values within each monthly dataset is one year (e.g., two consecutive Januaries). Following Kundzewicz and Robson (2004) and Blain (2010), the Pettitt test (Pettitt, 1979) was used to detect the initial year of the climate trends observed in each one of the series. The p-value associated with this test was also estimated. Further information about the MK and the Pettitt test can be found in Blain (2010).

The MK test was initially applied considering all available data in each location. After that, the MK and the Pettitt tests were applied in all locations considering the common period of 1951-2010. Following Blain et al. (2007) and Blain (2010), this last procedure was carried out in order to enable the comparison of the temporal variability of the SPI values observed in each weather station during the last 60 years. Figure 2 depicts the fundamental steps of this section.

3. RESULTS AND DISCUSSION

Before verifying the normality assumption inherent to the SPI, it is worth emphasizing that after have evaluated the SPI calculation algorithm, Wu et al. (2007) stated that ‘in order to have balanced negative and positive SPI values, the factor t’ (equation 2) ‘must be the same under the two situation’: 0 < H(PRE) ≤ 0.5 and 0.5 < H(PRE) < 1. By

---

**Figure 2.** Schematic representation of the steps proposed as a procedure for evaluating the normality assumption of the SPI distributions, the spectral features of these series and, the presence of climate trends in these datasets.
following this last consideration it becomes evident that \( H(PRE) \) values are critical in providing symmetric SPI distributions (Wu et al., 2007). Moreover, the statements of Edwards and Mckee (1997) and Wu et al. (2007) described on the previously section (equations 1 and 2 failing in provide SPI series that can be seen as coming from a standard normal distribution), agree with the results presented in Table 1, since practically all non normal distributions of the SPI are observed during the dry (winter) season of the State of São Paulo (Table 2; months of May to August). The high frequency of zeros resulted in a high value of \( H(PRE=0) \). In such cases, and especially for the SPI obtained from the Gam model, the equations 1 and 2 have produced asymmetric distributions, since it was not possible to achieve a balance between the cases in which \( H(PRE) \leq 0.5 \) and the cases in which \( H(PRE) > 0.5 \).

It also can be easily noted from table 1, that the 3-parameter PE3 model was better than the 2-parameter gamma model in providing SPI monthly distributions that can meet the conceptual assumption that a given value of this probabilistic index would represent the same climatic condition (related to possible rainfall deficits or rainfall excess) for different precipitations regimes. Finally, considering the information of table 1 and the fact that the PE3 model is more flexible than the Gam model in fitting a data sample (Guttman, 1999), it becomes expected that the PE3 distribution will fit the rainfall data substantively well. This last consideration was statistically proven by the KS-L test, since all p-values obtained from this goodness-of-fit procedure were greater than 0.05. Thus, the following analyses presented in this study were carried out by considering only the SPI series obtained from this 3-parameter distribution. The PE3 function was considerate the best model in describing the observed data used in SPI calculations.

Based only on the frequency domain, the spectral analysis of the SPI signals, carried out for the series of Campinas, Pindorama and, Ribeirão Preto, has indicated a concentration of power in the (≈) 4 year band (note that only the global wavelet power of Campinas - Figure 3b - shows a variance peak above the 5% significance level). In addition, it is also worth emphasizing the presence of a significant peak of wavelet energy in the 4-6 year band in the SPI signal of Jundiaí. However, extending this approach to the time-frequency domain, it can be observed (Figure 3a,c,e,g) a decrease in the wavelet energy after the beginning of the 1990’s. It is also worth emphasizing that in the location of Campinas, there is another period in which a low level of wavelet energy can also be verified (1935-1960; Figure 3a). Thus, considering that after the beginning of the 1990’s no variance peak was observed in the 3-6 year band, and also considering that in the longest data record of this study no significant wavelet power (≈ 4 band) was observed during (approximately) 25 years (1935-1960), the establishment of a dominant mode for the SPI temporal variability becomes (at least) unreliable. Finally, it is interesting to realize that although only

Table 2. Evaluating the normality assumption of the Standardized Precipitation Index series (State of São Paulo, Brazil). N is associated with standard normal distribution. NN is associated with nonnormal distribution. The index calculations were based on Pearson type III distribution (PE3) and on gamma 2-parameters distribution (Gam)

<table>
<thead>
<tr>
<th>Month</th>
<th>Campinas</th>
<th>Jundiaí</th>
<th>Pindorama</th>
<th>Ribeirão Preto</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PE3</td>
<td>Gam</td>
<td>PE3</td>
<td>Gam</td>
</tr>
<tr>
<td>January</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>February</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>March</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>April</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>May</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>June</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>July</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>August</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>September</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>October</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>November</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>December</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 1. Evaluating the normality assumption of the Standardized Precipitation Index series (State of São Paulo, Brazil). N is associated with standard normal distribution. NN is associated with nonnormal distribution. The index calculations were based on Pearson type III distribution (PE3) and on gamma 2-parameters distribution (Gam)

G.C. Blain

3-parameter PE3 model was better than the 2-parameter gamma model in providing SPI monthly distributions that can meet the conceptual assumption that a given value of this probabilistic index would represent the same climatic condition (related to possible rainfall deficits or rainfall excess) for different precipitations regimes. Finally, considering the information of table 1 and the fact that the PE3 model is more flexible than the Gam model in fitting a data sample (Guttman, 1999), it becomes expected that the PE3 distribution will fit the rainfall data substantively well. This last consideration was statistically proven by the KS-L test, since all p-values obtained from this goodness-of-fit procedure were greater than 0.05. Thus, the following analyses presented in this study were carried out by considering only the SPI series obtained from this 3-parameter distribution. The PE3 function was considerate the best model in describing the observed data used in SPI calculations.

Based only on the frequency domain, the spectral analysis of the SPI signals, carried out for the series of Campinas, Pindorama and, Ribeirão Preto, has indicated a concentration of power in the (≈) 4 year band (note that only the global wavelet power of Campinas - Figure 3b - shows a variance peak above the 5% significance level). In addition, it is also worth emphasizing the presence of a significant peak of wavelet energy in the 4-6 year band in the SPI signal of Jundiaí. However, extending this approach to the time-frequency domain, it can be observed (Figure 3a,c,e,g) a decrease in the wavelet energy after the beginning of the 1990’s. It is also worth emphasizing that in the location of Campinas, there is another period in which a low level of wavelet energy can also be verified (1935-1960; Figure 3a). Thus, considering that after the beginning of the 1990’s no variance peak was observed in the 3-6 year band, and also considering that in the longest data record of this study no significant wavelet power (≈ 4 band) was observed during (approximately) 25 years (1935-1960), the establishment of a dominant mode for the SPI temporal variability becomes (at least) unreliable. Finally, it is interesting to realize that although only
as previously discussed, according to Grinsted et al. (2004) the squared wavelet coherence of two processes that may be linked to each other by some geophysical reason tends to exhibit a constant phase relationship within several time scales. In this view, the results depicted in figure 4 (Campinas and Ribeirão Preto) indicate no remarkable relationship between the temporal variability of the SPI and the ENSO signal. It can be verified that regions of figure 4 with significant coherence occur during short periods of time. In addition, even within these regions no consistent phase relationship can be observed between these both physical signals. The same features were observed (not illustrated) in the locations of Jundiaí and Pindorama. Thus, at least under a statistical point of view, the use of the ENSO signal in predicting possible drought conditions in these four locations can result in unreliable information. This last statement seems to agree with previous studies that also indicate the lack of a statistically significant ENSO influence over the rainfall variability of the State of São Paulo (Rao and Hada, 1984; Blain et al., 2009; Blain and Kayano, 2011).

Concerning the long term trends, it is worth emphasizing that considering solely the significance level associated with the MK test, the analysis of tables 3 and 4 indicates, in general, the presence of no important change in the SPI variability. The results obtained in most series did not reach the critical significance levels usually adopted in the scientific literature (5 and 10%). All MK values obtained from the SPI series of Campinas (Table 3) are associated with p-values greater than 0.10. Results of this trend test associated with p-values ≤ 0.05 are observed on the months of May and September in Jundiaí and Ribeirão Preto, and on the month of January in Pindorama. All of them indicate a change to wetter conditions. The only negative MK value, associated with
a p-value \( \leq 0.10 \), has occurred on the months of June in Pindorama. Considering the last 60 years (Table 4), only four cases associated with p-values \( \leq 0.10 \) are observed: Jundiaí (January and November), Pindorama (January) and Ribeirão Preto (October).

However, before indicating the absence of any evidence of climate trends in the SPI series, it is important to recall that according to Yue and Hashino (2003), although the critical level usually adopted for rejecting the \( H_0 \) associated with the MK is p-value=0.05, for (really) trend-free samples the final value of this test should be near to zero. The corresponding p-value should be close to 0.5. In addition, Radziejewski and Kundzewicz (2004) states that even when no significant trend has been detected by statistical tests, this cannot ensure the absence of any form of change in the time series. According to these authors, statistical evaluations may fail in detecting weak changes and/or changes which have occurred in the final years of the time series under analysis.

Considering these last two statements, it becomes important to verify that according to the Pettitt test (Table 4) the beginning of a (decreasing) trend in the SPI series of Campinas and Ribeirão Preto (months of October; p-value < 0.10) was detected as being the years of 1983/84. The same feature is observed in the series of Jundiaí and Pindorama (months of October), although with p-values greater than 0.20. The MK final values obtained from these four weather stations for the monthly series of October were always negative. Furthermore, it is worth emphasizing that a climate change that can severely impact the
agricultural activities of the State of São Paulo, has already been described by DuFek and Ambrizzi (2007). Based on the evaluation of indices derived from 59 daily precipitation series, these authors described an increasing trend in the number of consecutive dry days observed in the aforementioned State. Also according to DuFek and Ambrizzi (2007) this climate change has occurred after 1985 (very close to the change point -1983/84- observed in this study).

Thus, it can be pointed out that the variability observed in the SPI series during the transition period between the dry and the rainy seasons of the State of São Paulo cannot be seen as the result of a purely random process. In other words, it is unlikely that this concentration of negative MK values (Tables 2, 3 and 4; months of October), is just a product of chance.

4. CONCLUSION

The Pearson type III distribution was better than the gamma 2-parameter distribution, in providing monthly SPI series closer to the normality assumption inherent to the use of this probabilistic standardized index.

The spectral analyses used in this study did not allow us to establish a dominant mode in the SPI time series. Moreover, no remarkable relationship between the temporal variability of this drought index and the ENSO signal was observed.

In general, the trend tests indicated the presence of no significant trend components in the SPI series. However, the results obtained from these trend tests have also indicated that the temporal variability of the SPI, observed at the months of October over the last 60 years, cannot be seen as the result of a purely random process. This last inference is due to the concentration of decreasing trends, with a common beginning (1983/84), in the four SPI series of the study. Thus, the hypothesis of the presence of climate changes in these series cannot be rejected.

REFERENCES


BLAIN, G.C. Considerações estatísticas relativas à oito séries de precipitação pluvial da Secretaria de Agricultura e Abastecimento.


KAYANO, M.T.; SANSSíGOLo, C. Interannual to decadal variations of precipitation and daily maximum and daily minimum temperatures in southern Brazil. Theoretical and Applied Climatology, v.97, p.81-90, 2009.


