EnvironmentAsia
The international journal by the Thai Society of Higher Education Institutes on Environment

EnvironmentAsia 12(3) (2019) 32-42
DOI 10.14456/ea.2019.43
ISSN 1906-1714; ONLINE ISSN: 2586-8861

Appropriate Transformation Techniques to Determine a Modified Standardized Precipitation Index for the Ping River in Northern Thailand

Tanachot Chaito[1], Manad Khamkong[1, 2*] and Pennapar Murnta[1]

[1]*Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand*
[2]*Environmental Science Research Center, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand*

*Corresponding Author: manad.k@cmu.ac.th
Received: October 2, 2018; Revised: December 11, 2018; Accepted: June 16, 2019

Abstract

The standardized precipitation index (SPI) is used to characterize precipitation when evaluating drought intensity over a range of timescales. From a statistical point of view, an appropriate standardization method depends on the limitations of the input data to make them effective as SPI values. For this reason, a modified SPI (TSPI) based on selecting an appropriate transformation for the distribution of the precipitation data is introduced in the present study. Firstly, a number of appropriate distributions were found to fit the seasonal rainfall data during the period of 1957–2014 for the Ping River in northern Thailand, and secondly, numerical analysis for various situations was carried out to compare the SPI and TSPI for each selected distribution. The results show that the TSPI performed well for all of the situations in the study. Finally, the TSPI was applied to identify rainfall characteristics in the data from three rain gauging stations on the Ping River in northern Thailand. The TSPI is recommended as an appropriate alternative to the SPI for drought analysis when limited to a small sample size such as the precipitation distribution of interest in this study.

Keywords: Non-normal distribution; Normalization; Seasonal rainfall data; Transformation

1. Introduction

A drought is a natural disaster caused by below-average precipitation in a region, resulting in water shortage in an area for an extended time, which can have impact on agriculture, local economy and ecosystem. In Thailand, drought occurs two periods, the first being the winter season to the summer season and second period being the middle of rainy season, which main cause of drought is insufficient rain. A drought in Thailand has a direct impact on agriculture including the soil lacks moisture and dehydrated plants, resulting in low quality agricultural produce and reduced quantities.

The Standardized Precipitation Index (SPI) developed by Mckee *et al.* (1993) is a popular and widely used method to indicate severe drought intensity and to monitor drought levels. Computation of the SPI uses the distribution of precipitation data during the specific time-period of interest and transforms it to a standard normal distribution after first fitting it to a gamma distribution. The transformed precipitation data are used to compute the SPI by the difference in precipitation (P) and the mean divided by the standard deviation (SD) from past records:

$$SPI = \frac{(P\text{-}mean)}{SD} \quad (1)$$

In general, researchers try to find an appropriate distribution to specify the precipitation distribution under each area and time period in the study data of interest. MeKee *et al*., (1993) fitted rainfall data to a gamma distribution. Yue and Hashino (2007) investigated the probability of seasonal precipitation in Japan; spring season, the Pearson type III distribution the best fitted for spring precipitation, Log- Pearson type III the best fitted for summer and winter season and 3-parameter lognormal the best fitted for autumn season. Zhang *et al*. (2009) identified the lognormal distribution as the best fit for the rainfall data from the Pearl River in China. In the state of Sao Paulo, Brazil, Gabriel (2011) fitted the rainfall data as a Pearson Type III distribution, while Khamkong and Bookamana (2011) applied generalized extreme value distributions to the annual monthly maximum rainfall data in upper northern Thailand (the first study on modeling annual maximum daily rainfall for this area). Moreover, Yusof and Hui-Mean (2012) fitted the rainfall data for the state of Johor, Malaysia as a Weibull distribution.

When applying the SPI to evaluate drought intensity, it is a necessary condition to transform a non-normal distribution to a normal distribution. The SPI is transformation data from non-normal distribution to normal distribution when the sample is large. But distribution of the amount of rain in Thailand has right –skewed distribution because there is no rain in some months. Moreover, long-tail or right-skew distribution, if using SPI transforms data to normal distribution, data may not have normal distribution. Therefore, we should find methods to transform the data that is appropriate for the data before calculating the SPI values. There are a variety of ways to achieve this, such as Krishnamoorthy *et al*. (2008)'s suggestion to use either a cube-root or fourth-root for transformation of gamma distribution data to normal data. In addition, Yeo and Johnson (2000) proposed an effective method to transformed right-skewed data to a normal distribution, while Watthanacheewakul (2012) improved the Box–Cox power transformation for transforming skewed data to a normal distribution. Later on, Chaito and Khamkong (2018) proposed a transformation method that improved the Box and Cox power transformation and found that it was effective in transforming Weibull data to a normal distribution. Furthermore, Arkadiusz *et al*. (2014) applied the Box–Cox power transformation to calculate the SPI value in Eastern Kujawy (Central Poland) and found that the Box–Cox power transformation was effective in transformation monthly rainfall data.

In this paper, seasonal rainfall data for the Ping River in northern Thailand during the period from February 1957 to September 2014 from three meteorological stations was used in the analysis. The aims of this study were as follows: 1) to find an appropriate distribution in representing the seasonal rainfall data for the Ping River in northern Thailand, 2) to modify the SPI (TSPI) based on a selected appropriate transformation for each precipitation distribution of the data of interest and 3) to apply the TSPI in order to evaluate the rainfall data characteristics of the Ping River in northern Thailand.

2. Materials and methods

2.1 Study area and data

Monthly precipitation data during the period from February 1957 to September 2014 used in this study were obtained from the Hydrology and Water Management Centre for the Upper Northern Region of Thailand (2015, in mm). The Ping River in northern Thailand is the origin of the major tributaries of the Chao Phraya River in the central region of Thailand (Figure 1a). Data from three rain gauging stations at Doisaket, Maetang and Mueang Chiang Mai (Figure 1b) was selected for the study. The rainfall data were divided into four months using the criterion for Thai seasons determined by the Meteorological Department of Thailand (2015): the summer season (February to May), the rainy season (June to September) and the winter season (October to January of the following year)

2.2 Types of Distribution

The motivation for the current work was to find an appropriate distribution to represent the seasonal rainfall data for the Ping River in northern Thailand, which was found to be right-skewed. Consequently, a selection of right-skewed distributions and their parameter estimations using the maximum-likelihood method are reported in Table 2.

Note that it is difficult to select an appropriate distribution for rainfall data, and many researchers have studied the goodness of fit for selecting an appropriate model. In this paper, the model selection criteria used for selecting an appropriate distribution of rainfall were the Akaike information criterion (AIC: Akaike, 1973) and the Anderson-Darling (AD: Anderson and Darling, 1952) test. AIC is defined as

$$AIC = 2k - 2\ln L \qquad (2)$$

where k is the number of parameters in the model and L is the maximized value of the likelihood function for the model.

The AD test is written as

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln F(x_i) + \ln(1-F(x_{n+1-i}))] \qquad (3)$$

where n is the sample size, $F(.)$ is the expected CDF and x_i are the ordered data.

Moreover, Laio *et al.* (2009) reported that the AD test outperformed AIC when testing three parameters of the distribution whereas AIC was better for two parameters. Consequently, the best model for rainfall data with the smallest AD value and smallest AIC indicates that a distribution is appropriated.

Table 3 and Figure 2 indicate the best fitting distributions for the rainfall data from the Doisaket and Maetang stations: a Weibull distribution for the summer season and a gamma distribution for the rainy and winter season. For the Mueang Chiang Mai station, a lognormal distribution and a gamma distribution were the best fits for the summer and winter season rainfall data, respectively. Furthermore, a normal distribution was appropriate for the rainy season rainfall data from this station. Subsequently, the parameter estimates for each distribution are shown in Table 4.

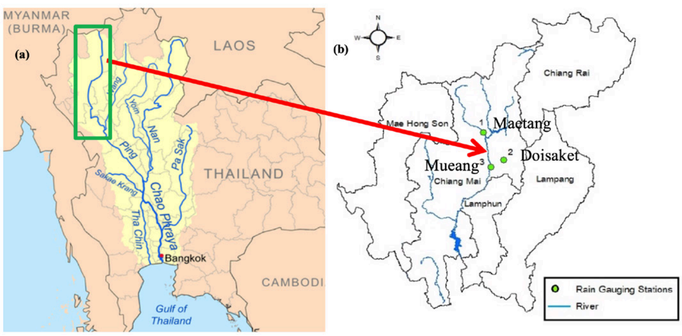

Figure 1. (a) Geographical location of northern Thailand. (b) Position of the three meteorological stations on the Ping River in northern Thailand.

Table 1. Information about rain gauges

Station	Latitude	Longitude	Observation period	Month Precipitation (mm)	
				Mean (mm)	St.dev
Doisaket	18°52'08"N	99°08'22"E	1957 - 2014	92.8	98.41
Maetang	19°07'08"N	98°56'52"E	1957 - 2014	96.5	99.03
Mueang	18°47'21"N	99°01'01"E	1957 - 2014	97.6	99.64

Table2. Summary of selected distributions in this study

Distribution	Abbreviations	Cumulative Distribution Function (cdf)	Range
Normal	norm	$F(x)=\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$; $(\sigma>0\,,\,-\infty<\mu<+\infty)$ $erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt$ is an error function	$-\infty<x<\infty$
Log-normal	lnorm	$F(x)=\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{\ln(x)-\mu}{\sigma\sqrt{2}}\right)\right]$; $(\sigma>0\,,\,-\infty<\mu<+\infty)$ $erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt$ is an error function	$x>0$
Gamma	Gamma	$F(x)=\Gamma_{x/\sigma}(a)/\Gamma(\alpha)$; $(\alpha,\sigma>0)$ $\Gamma(.)$ is the gamma function and $\Gamma_x(.)=\int_0^x t^{\alpha-1}e^{-t}dt$ is the incomplete gamma function	$x>0$
Weibull	Weibull	$F(x)=1-\exp\left(-(x/\sigma)^{\alpha}\right)$; $(\alpha,\sigma>0)$	$-\infty<x<+\infty$
Frechet	Frechet	$F(x)=\exp\left(-(\sigma/x)^{\alpha}\right)$; $(\alpha,\sigma>0)$	$x>0$
Generalized extreme value	GEV	$F(x)=\exp\left(-(1-(\alpha(x-\mu))/\sigma)^{1/\alpha}\right)$ $,\alpha\neq 0$ $F(x)=\exp(-\exp(-((x-\mu)/\sigma)))$ $,\alpha=0$; $(\sigma>0\,,\,-\infty<\mu<+\infty)$	$\mu+\sigma/\alpha\le x<+\infty\,,\alpha<0$ $-\infty\le x<+\infty\,,\alpha=0$ $-\infty<x\le\mu+\sigma/\alpha\,,\alpha>0$
Pearson type III	P3	$a=4/\alpha^2\;,\;c=\mu-2\sigma/\alpha$ $F(x)=\Gamma_{(\ln(x)-c)/\sigma}(a)/\Gamma(a)\quad,\alpha>0$ $F(x)=1-\Gamma_{(\ln(x)-c)/\beta}(a)/\Gamma(a)\;,\alpha<0$; $(\sigma>0\,,\,-\infty<\mu<+\infty)$	$x\ge\mu$
Log-Pearson type III	LP3	$a=4/\alpha^2\;,\;c=\mu-2\sigma/\alpha$ $F(x)=\Gamma(a,(x-c/\sigma))/\Gamma(a)\quad,\alpha>0$ $F(x)=1-\Gamma(a,(x-c/\sigma))/\Gamma(a)\;,\alpha<0$; $(\sigma>0\,,\,-\infty<\mu<+\infty)$	$x\ge e^{\mu}$

Note: α is the shape parameter, σ is the scale parameter and μ is the location parameter.

2.3 Transformation Methods

The results in Section 3.1 indicate that appropriate distributions for the seasonal rainfall data for the Ping River in northern Thailand are gamma, Weibull and lognormal distribution, depending on the rain gauging station and season. In order to compare the performance of the transformation of each distribution to a normal distribution based on the results of Chaito *et al*. (2016), the possibilities are as follows.

- Box and Cox transformation:

$$Y=\begin{cases}\dfrac{x^{\lambda}-1}{\lambda}, & \lambda\neq 0\\ \ln(x), & \lambda=0\end{cases}\qquad(4)$$

where λ is a transformation parameter

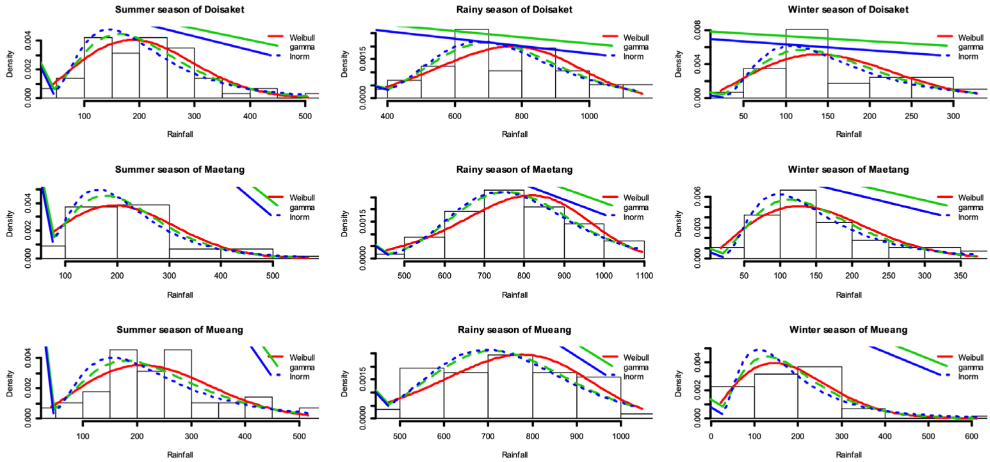

Figure 2. Comparison of histogram and theoretical densities for the tested distributions for the three rain gauging stations on the Ping River.

Table 3. Summary of selected distribution using the AIC and AD statistics for three stations on the Ping

Station	Distribution	Summer		Rainy		Winter	
		AIC	AD	AIC	AD	AIC	AD
Doisaket	norm	685.151	0.171	763.664	0.077	658.877	0.994
	lnorm	685.380	0.420	764.253	0.114	657.827	0.312
	Gamma	680.850	0.092	**762.980**	**0.045**	**653.315**	**0.196**
	Weibull	**680.781**	**0.091**	764.431	0.065	653.377	0.348
	Frechet	706.510	2.544	777.095	1.204	681.431	2.673
	GEV	682.794	0.110	764.286	0.062	655.351	0.367
	P3	682.475	0.108	764.863	0.064	655.031	0.304
	LP3	682.255	0.105	764.305	0.064	654.144	0.444
Maetang	norm	691.499	0.996	**734.883**	**0.041**	663.116	1.050
	lnorm	**677.496**	**0.115**	736.762	0.102	658.163	0.182
	Gamma	678.519	0.158	735.468	0.043	**653.720**	0.113
	Weibull	684.580	0.194	736.853	0.056	655.119	0.169
	Frechet	686.657	1.035	752.513	1.531	684.282	3.136
	GEV	679.783	0.123	736.291	0.055	655.273	**0.092**
	P3	678.828	0.161	736.809	0.046	655.465	0.130
	LP3	679.422	0.142	736.585	0.053	655.737	0.175
Mueang	norm	700.534	0.545	741.143	0.433	693.137	0.286
	lnorm	706.449	1.181	740.161	0.328	688.827	0.556
	Gamma	698.920	0.309	**739.939**	**0.320**	**683.442**	**0.154**
	Weibull	**696.999**	**0.260**	742.312	0.522	683.911	0.198
	Frechet	732.241	5.238	746.511	0.595	711.139	2.736
	GEV	698.894	0.288	741.090	0.506	686.076	0.228
	P3	699.032	0.310	741.508	0.412	685.286	0.172
	LP3	699.219	0.507	741.633	0.515	685.693	0.216

- The Cube-root transformation for the gamma distribution (Krishnamoorthy *et al.*, 2008):

$$Y = x^{1/3} \tag{5}$$

- The lognormal distribution using log transformation:

$$Y = \ln(x) \tag{6}$$

Table 4. Maximum-likelihood estimates of the selected distribution's parameters and descriptive statistics for the three stations on the Ping River

Station	Season	Best-fitting Distribution	Parameter Estimates				Precipitation (mm)						
			Shape	St. error	Scale	St. error	Min.	Max.	Q_1	Q_2	Q_3	Mean	St. dev
Doisaket	Summer	Weibull	2.35	0.237	237.50	14.150	40.9	504.8	129.7	208.0	270.0	210.2	96.06
	Rainy	Gamma	15.00	2.653	49.77	0.004	394.9	1,158.0	629.6	734.5	884.1	746.5	191.26
	Winter	Gamma	4.12	0.736	38.69	0.005	22.2	328.0	114.5	136.3	205.9	159.2	76.28
Maetang	Summer	Lognormal	5.29[1]	0.059	0.44	0.042	73.4	567.2	158.1	202.6	264.4	219.1	101.55
	Rainy	Normal	783.10[1]	19.510	147.30	13.790	453.4	1,096.0	694.6	781.6	879.4	783.1	148.59
	Winter	Gamma	3.71	0.661	41.30	0.004	16.9	372.6	92.1	133.5	193.0	153.2	79.17
Mueang	Summer	Weibull	2.30	0.234	267.90	16.240	42.3	515.3	171.7	232.6	271.1	237.4	109.93
	Rainy	Gamma	22.63	4.078	32.85	0.006	468.5	1,048.0	613.0	738.7	877.1	743.2	156.97
	Winter	Gamma	3.19	0.562	58.79	0.003	21.7	611.4	112.3	175.2	251.0	187.5	103.03

Note: [1] Location parameter, Q_i = the i^{th} quartile of the data

Table 5. SPI categories based on SPI values

SPI	Category
2.00 and above	Extremely wet
1.50 to 1.99	Severely wet
1.00 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.00 to -1.49	Moderately dry
-1.50 to -1.99	Severely dry
−2.00 and less	Extremely dry

- The Weibull distribution transformation using modified Box and Cox transformation technique (Chaito and Khamkong, 2018):

$$Y = \begin{cases} \dfrac{(x+c)^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \ln(x+c), & \lambda = 0 \end{cases} \qquad (7)$$

where λ is a transformation parameter, $c = \dfrac{(\bar{x}-Q_2)}{2n(\tilde{s})}$ with $\bar{x}$ as the mean, $\tilde{s}$ is $\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$, Q_2 is the 2nd quartile and n is the sample size. In this paper, $\tilde{s}$ is expressed as $\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-Q_2)^2}{n-1}}$.

Parameter estimation of Box and Cox transformation (4) and modified Box and Cox transformation (7) are carried out using the maximum likelihood estimate of the transformation parameter (λ) as

$$\frac{\partial \ln L(\lambda|x_i)}{\partial \lambda} = \frac{-n\left(\sum_{i=1}^{n} x_i^{2\lambda}\ln x_i - \left(\frac{1}{n}\right)\left(\sum_{i=1}^{n} x_i^{\lambda}\right)\left(\sum_{i=1}^{n} x_i^{\lambda}\ln x_i\right)\right)}{\left[\sum_{i=1}^{n} x_i^{2\lambda} - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^{\lambda}\right)^2\right]} + \frac{n}{\lambda} + \sum_{i=1}^{n}\ln(x_i) = 0 \qquad (8)$$

$$\frac{\partial \ln L(\lambda|x_i)}{\partial \lambda} = \frac{-n\left(\sum_{i=1}^{n}(x_i+c)^{2\lambda}\ln(x_i+c) - \left(\frac{1}{n}\right)\left(\sum_{i=1}^{n}(x_i+c)^{\lambda}\right)\left(\sum_{i=1}^{n}(x_i+c)^{\lambda}\ln(x_i+c)\right)\right)}{\left[\sum_{i=1}^{n}(x_i+c)^{2\lambda} - \frac{1}{n}\left(\sum_{i=1}^{n}(x_i+c)^{\lambda}\right)^2\right]} + \frac{n}{\lambda} + \sum_{i=1}^{n}\ln(x_i+c) = 0 \cdot \qquad (9)$$

2.4 Standardized Precipitation Index (SPI) and Trend Analysis

Classifying dry and wet events during the seasonal rainfall data for the Ping River in northern Thailand was carried out by applying the SPI, the criteria for which are shown in Table 5.

Dry and wet periods were then detected using the Mann–Kendall trend test (Mann, 1945; Kendall, 1975) given by

$$K = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_{ij}, \quad a_{ij} = sign = (x_j - x_i)\begin{cases} 1 & ; x_i < x_j \\ 0 & ; x_i = x_j \\ -1 & ; x_i > x_j \end{cases} \qquad (10)$$

For which the average and variance of the K test are

$$E(K) = 0 \text{ and } Var(K) = \frac{n(n-1)(2n+5) - \sum_{j=1}^{m} t_j(t_j-1)(2t_j+5)}{18}, \text{ repectively.} \qquad (11)$$

where m indicates the number of groups of tied ranks, each with tied observations.

The standardized statistics (Z) for the trend analysis is defined as

$$Z = \begin{cases} \dfrac{K-1}{\sqrt{V_0(K)}} & ;K>0 \\ 0 & ;K=0 \\ \dfrac{K+1}{\sqrt{V_0(K)}} & ;K<0 \end{cases} \qquad (12)$$

A positive Z value indicates a drought trend as opposed to a wet trend. In this research, statistical significance at the 95% confidence level ($p < 0.05$) was used in the trend analysis.

2.5 Numerical Studies Methodology

In order to compare the performance of for each distribution based on the discussion in section 3.2, simulation studies were conducted using the R statistical program (R Core Team, 2013) to generate random samples from gamma distribution, lognormal distribution

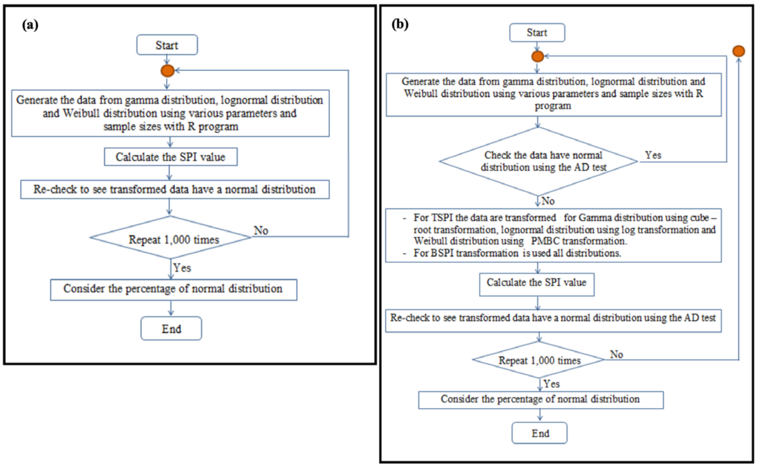

Figure 3. (a) simulation study of SPI method and (b) simulation study of TSPI and BSPI method

Table 6. Comparisons percentage normal distribution of SPI, BSPI and TSPI method from various distributions

n	Gamma					Lognormal					Weibull				
	α	σ	SPI	BSPI	TSPI	μ	σ	SPI	BSPI	TSPI	α	σ	SPI	BSPI	TSPI
15	1.0	1.0	42.1	97.9	**100.0**	1.0	0.3	88.1	98.8	**100.0**	1.0	1.0	40.3	**96.6**	95.5
	2.0	1.0	72.2	99.2	**100.0**	2.0	0.3	88.1	98.8	**100.0**	1.5	1.0	79.1	90.3	**95.4**
	4.0	1.0	87.0	99.5	**100.0**	1.0	0.5	68.5	99.2	**100.0**	2.1	237.5	95.0	60.3	**95.5**
	4.1	39.7	87.2	99.5	**100.0**	2.0	0.5	68.5	99.2	**100.0**	2.3	237.5	**96.4**	41.0	95.5
	15.0	49.7	97.8	99.0	**100.0**	5.3	2.1	0.4	98.2	**100.0**	2.0	267.9	93.9	69.1	**95.5**
	22.6	32.9	98.5	98.9	**100.0**	5.3	2.3	0.2	98.2	**100.0**	2.5	267.9	97.5	**98.5**	95.5
30	1.0	1.0	6.0	93.5	**99.6**	1.0	0.3	73.6	98.2	**100.0**	1.0	1.0	6.7	**99.4**	97.3
	2.0	1.0	35.3	**100.0**	**100.0**	2.0	0.3	73.6	98.2	**100.0**	1.5	1.0	50.0	**99.1**	97.9
	4.0	1.0	65.7	99.6	**100.0**	1.0	0.5	35.8	98.2	**100.0**	2.1	237.5	87.0	97.0	**97.5**
	4.1	39.7	66.7	99.6	**100.0**	2.0	0.5	35.8	98.2	**100.0**	2.3	237.5	91.9	**95.2**	94.1
	15.0	49.7	91.8	99.3	**100.0**	5.3	2.1	0.0	94.0	**100.0**	2.0	267.9	83.6	**97.7**	97.5
	22.6	32.9	95.4	99.2	**100.0**	5.3	2.3	0.0	94.0	**100.0**	2.5	267.9	94.4	91.2	**97.5**
50	1.0	1.0	0.2	96.5	**99.2**	1.0	0.3	52.6	97.1	**100.0**	1.0	1.0	0.20	**98.9**	95.3
	2.0	1.0	8.7	99.0	**100.0**	2.0	0.3	52.6	97.1	**100.0**	1.5	1.0	23.6	**98.7**	95.6
	4.0	1.0	43.7	99.4	**100.0**	1.0	0.5	12.5	93.1	**100.0**	2.1	237.5	77.6	95.0	**95.7**
	4.1	39.7	44.6	99.5	**100.0**	2.0	0.5	12.5	93.1	**100.0**	2.3	237.5	87.0	91.4	**95.7**
	15.0	49.7	85.2	99.5	**100.0**	5.3	2.1	0.0	87.4	**100.0**	2.0	267.9	70.2	**96.1**	95.7
	22.6	32.9	90.9	99.3	**100.0**	5.3	2.3	0.0	87.4	**100.0**	2.5	267.9	91.9	85.8	**95.7**
100	1.0	1.0	0.0	**99.0**	97.5	1.0	0.3	20.2	98.7	**100.0**	1.0	1.0	0.00	**93.7**	91.9
	2.0	1.0	0.0	99.6	**99.9**	2.0	0.3	20.2	98.7	**100.0**	1.5	1.0	1.50	**93.7**	92.0
	4.0	1.0	9.6	99.7	**100.0**	1.0	0.5	0.1	96.7	**100.0**	2.1	237.5	51.9	89.2	**92.0**
	4.1	39.7	11.0	99.6	**100.0**	2.0	0.5	0.1	96.7	**100.0**	2.3	237.5	68.5	82.2	**92.2**
	15.0	49.7	68.6	99.6	**100.0**	5.3	2.1	0.0	86.5	**100.0**	2.0	267.9	81.8	90.7	**92.2**
	22.6	32.9	78.6	99.6	**100.0**	5.3	2.3	0.0	96.5	**100.0**	2.5	267.9	40.6	69.0	**92.2**

Note: The values in bold indicate the best transformation performance.

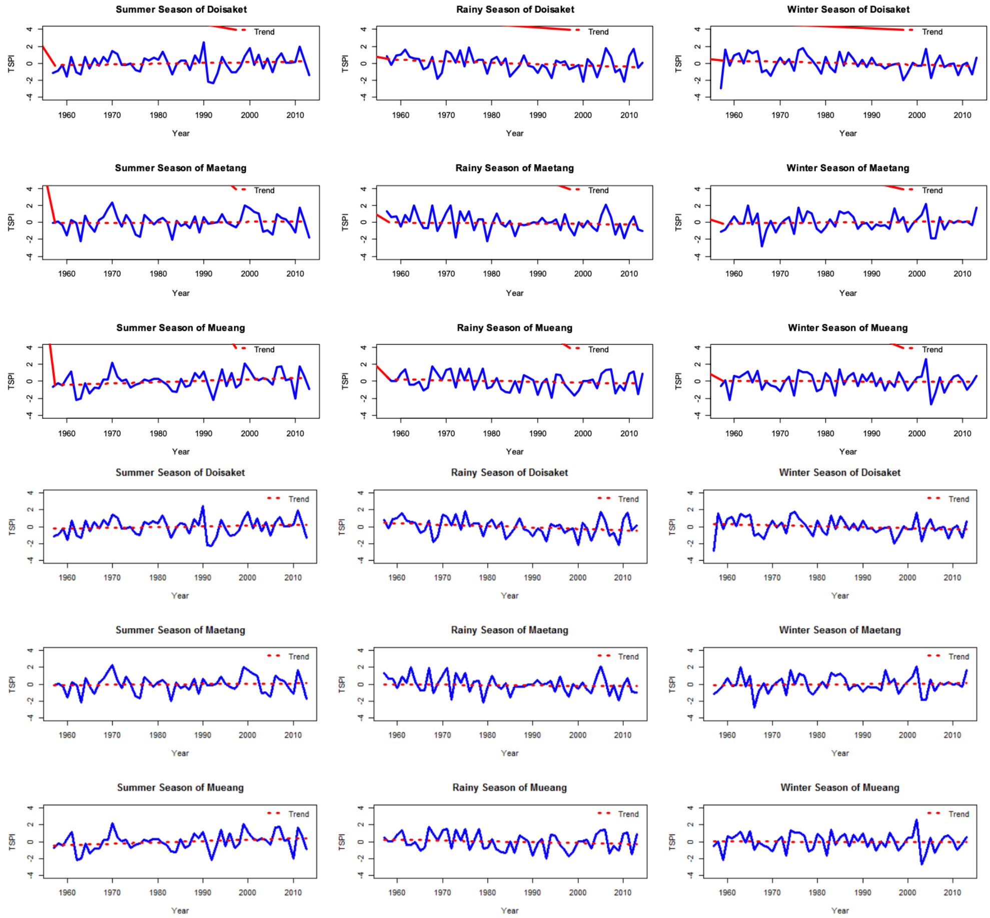

Figure 4. TSPI and linear trend for the three rain gauging stations on the Ping River

Table 7. TSPI percentages for the three rain gauging stations on the Ping River

TSPI	Doisaket			Maetang			Mueang		
	Summer	Rainy	Winter	Summer	Rainy	Winter	Summer	Rainy	Winter
Extremely wet	1.75	0.00	0.00	3.51	0.00	3.51	3.51	0.00	1.75
Severely wet	3.51	8.77	8.77	5.26	7.02	5.26	5.26	5.26	0.00
Moderately wet	7.02	5.26	8.77	3.51	7.02	10.53	8.77	12.28	14.04
Near normal	66.67	68.42	64.91	68.42	70.18	66.67	70.18	63.16	64.91
Moderately dry	15.79	7.02	12.28	10.53	7.02	8.77	5.26	14.04	12.28
Severely dry	1.75	7.02	3.51	5.26	3.51	3.51	3.51	5.26	3.51
Extremely dry	3.51	3.51	1.76	3.51	5.25	1.75	3.51	0.00	3.51

Table 8. Mann-Kendall trend testing for the three rain gauging stations on the Ping

Station	Season	Best Distribution	Mann-Kendall Trend Test	
			tau	p-value
	Summer	Weibull	0.1000	0.2737
Doisaket	Rainy	Gamma	-0.2040	0.0253
	Winter	Gamma	-0.1340	0.1445
	Summer	Lognormal	0.0326	0.7255
Maetang	Rainy	Normal	-0.1990	0.0291
	Winter	Gamma	0.0508	0.5818
	Summer	Weibull	0.1900	0.0376
Mueang	Rainy	Gamma	-0.0990	0.2798
	Winter	Gamma	-0.0301	0.7463

and Weibull distributions with samples and parameters using the rainfall data from the three rain gauging stations (Table 4) based on 1,000 replications. All of the experiments were performed for different sample sizes (n = 15, 30, 50, and 100). The criterion for comparing the performance of SPI, BSPI and TSPI was the transformation percentage to a normal distribution. The simulation process shows in Figure 3.

3. Results and Discussion

3.1 Simulation Study

The results in Table 6 indicate that overall, the performance of the TSPI was better than the SPI. For gamma data, the SPI method had a decreased percentage of data transformation to a normal distribution when the sample size increased whereas the opposite was true when the scale parameter increased. For the TSPI, using the fourth root transformed the data to a normal distribution by more than 97%, thus when sample size increased, the TSPI transformed the data to a normal distribution more so than the SPI.

For lognormal data, the SPI showed a decrease in percentage transformation to a normal distribution when the sample size, location parameter and scale parameter increased. The TSPI logarithmically transformed the data, after which they fitted perfectly to a normal distribution under all experimental conditions.

For Weibull data, the SPI showed an increase in percentage transformation to a normal distribution when the sample size decreased and scale parameter increased. The TSPI using modified Box and Cox transformed the data to a normal distribution by more than 90%. When the sample size increased, the data transformation decreased.

Furthermore, the results of simulation between BSPI and TSPI showed that TSPI had more effective in transformation to normal distribution than BSPI for gamma data and lognormal data. Weibull data, the percentage transformation to a normal distribution of BSPI and TSPI were not difference in some cases and TSPI had percentage normal distribution more than 90%.

3.2 Rainfall Data Study

The evaluation of drought by applying the TSPI to each distribution is exhibited in Figure 4. The Doisaket station had extremely dry conditions during the summer season in 1991 and 1992, the rainy season in 2000 and 2009 and the winter season in 1957. The Maetang station had extremely dry conditions during the summer season in 1963 and 1983 and the rainy and winter seasons in 1979 and 1966, respectively. The Mueang Chiang Mai station had extremely dry conditions during the summer season in 1962 and 1992 and during the winter season in 2003.

Table 7 shows season classification and drought analysis using the TSPI. The results indicate that all of the rain gauging stations had the highest TSPI value at near normal for all three seasons (63.16 -70.18%). The rainy season data from Maetang station and the summer season data from Mueang Chiang Mai station had the highest TSPI at near normal. For the Doisaket station, the summer and rainy season data had the lowest TSPI (3.51%), indicating extremely dry conditions, as was also the case for the winter season (1.76%). For the Maetang station, the lowest TSPI values classified the summer, rainy and winter seasons as extremely dry (3.51%), severely dry (3.51%) and extremely dry (1.75%), respectively. Lastly, the lowest results from the Mueang Chiang Mai station characterized conditions as extremely dry (3.51%) in the summer and winter seasons.

A trend analysis of the seasonal rainfall data from the three rain gauging stations was carried out using the Mann–Kendall trend test at the 95% confidence level, the results which are shown in Table 8 and correspond with linear trend in Figure 4. These show that the Doisaket and the Maetang stations had a drought trend in the rainy season and the Mueang Chiang Mai station had a rain trend in the summer season.

4. Conclusions

The simulation results showed that right-skewed distributions could be transformed to a normal distribution by various data transformations. Furthermore, SPI could transform data to normal distribution less than BSPI and TSPI, Which was a way to find transformation methods to fit data, every various parameters and sample sizes. Thus, we should find methods to transform before calculating the SPI values.

Rainfall data from three rain gauging stations on the Ping River at Doisaket, Maetang and Mueang Chiang Mai, Thailand, classified by season had mostly right-skewed distributions. The evaluation of drought conditions of each of the TSPI-transformed distributions found that the summer season had extremely dry conditions in 1962, 1963, 1983 and 1992; the rainy season had extremely dry conditions in 1979, 2000 and 2003; and in 1957 and 1966, there were extremely dry conditions in the winter season. These findings are consistent with Agro-Meteorological Academic Group Meteorological Development (2011), Thailand who reported that in 1967 to 1993, there were drought conditions in northern and central Thailand because of insufficient rainfall. Moreover, the effects of drought also cause damage to Thailand in many sectors, such as water shortage for consumption, agriculture and husbandry. In 1979, damage from drought affected the agriculture and industries such as power generation and affecting the people in the country because of lack of water for consumption. In 1990–1993, drought affected the agriculture because water in various dams and reservoirs are below levels, causing the agricultural sector could not be able to use water for farming and husbandry. For drought conditions to be noted in Thailand, two periods of insufficient rainfall are required, the first being the winter season to the summer season (after October to May of next year) and second period being the middle of rainy season (June to July). EL Nina and La Nina of precipitation of Thailand; which EL Nina had annual precipitation lower standard precipitation conditions in 1991-1992, 1997-1998, 1987 and 1982-1983, which La Nina had annual precipitation over standard precipitation conditions in 1999 – 2000, 1988, 1974 – 1975 and 1954 – 1956. Therefore, as demonstrated in this study, an appropriate transformation of rainfall data for analysis by the standardized precipitation index is recommended.

Acknowledgements

This research work was partially supported by Chiang Mai University.

References

Agro-Meteorological Academic Group Meteorological Development. Study on Drought Index in Thailand. Available from: http://www.tmd.go.th/info/info.php?FileID=71. 2017. [Accessed 13 February 2017].

Akaike H. Information theory and an extension of the maximum likelihood principle. Proceedings of 2nd International Symposium on Information Theory. 1973; 267-281.

Anderson TW, Darling DA. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Annals of Mathematical Statistics. 1952; 23(2): 193– 212.

Box GEP, Cox DR. An analysis of transformations. Journal of the Royal Statistical Society: Series B. 1964; 26(2): 211-252.

Chaito T, Khamkong M. A modified Box and Cox power transformation to determine the standardized precipitation index. Songklanakarin Journal of Science and Technology. 2018; 40(4): 867-877.

Chaito T, Khamkong M, Bookkamana P. Application of transformation techniques to evaluate drought. Burapha Science Journal. 2016; 21(2): 86-98.

Gabriel CB. Standardized precipitation index based on Pearson type III distribution. Revista Brasileira de Meteorologia. 2011; 26(2): 167-180.

Hydrology and Water Management Center

for Upper Northern Region Chiang Mai Thailand. Rainfall data, Available from: http://hydro-1.net. 2017. [Accessed 13 February 2017]

Kendall MG. Rank Correlation Methods. Griffin: Brooks/Cole Publishing; 1975.

Khamkong M, Bookamana P. Development of statistical models for maximum daily rainfall in upper northern region of Thailand. Chiang Mai Journal of Science. 2011; 42(4): 1044-1053.

Krishnamoorthy K, Mathew T, Mukherjee S. Normal-based methods for a gamma distribution: Prediction and tolerance intervals and stress-strength reliability. Technometrics. 2008; 50(1): 69-78.

Laio F, Baldassarre GD, Montanari A. Model selection techniques for the frequency analysis of hydrological extremes. Water Resource Research. 2009; 45: 29 –40.

Mann HB. Nonparametric tests against trend. Econometrica. 1945; 13(3): 245–259.

Mckee TB, Doesken NJ, Kleist J. The relationship of drought frequency and duration to time scale. In: Eighth Conference on Applied Climatology, 17-22 January 1993, Anaheim, California: American Meteorological Society; 1993. p. 179-184.

Meteorological Department of Thailand. The Climate of Thailand. Available from: http://www.tmd.go.th/en /archive/thailand_climate.pdf. 2015. [Accessed 13 February 2017].

R Core Team. R: Language and Environment for Statistical Computing. Available from: https://www.gbif.org/tool /81287/r-a-language-and-environment-for-statistical-computing. 2016. [Accessed 19 January 2016].

Watthanacheewakul L. Transformations with right skewed data. Proceedings of the World Congress on Engineering. 2012; 308-311.

Yue S, Hashino M. Probability distribution of annual, seasonal and monthly precipitation in Japan. Hydrological Sciences. 2007; 52(5): 863 – 877.

Yeo I, Johnson NR. 2000. A new family of power transformations to improve normality or symmetry. Biometrika. 2000; 87(4): 954-959.

Yusof F, Hui-Mean F. Use of statistical distribution for drought analysis. Applied Mathematical Sciences. 2012; 6(21): 1031–1051.

Zhang Q, Xu CY, Zhang Z. Observed changes of drought wetness episodes in the Pearl River Basin, China using the standardized precipitation index and aridity index. Theoretical and Applied Climatology. 2009; 98(1): 89–99.