WALAILAK JOURNAL
http://wjst.wu.ac.th Applied Mathematics

The Southern Oscillation Index as a Random Walk

Mayuening ESO[1,2], Metta KUNING[1,2,*], Hilary GREEN[3], Attachai UERANANTASUN[1] and Somporn CHUAI-AREE[1,2]

[1]*Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Mueang, Pattani 94000, Thailand*
[2]*Centre of Excellence in Mathematics, CHE, Si Ayutthaya Rd., Bangkok 10400, Thailand*
[3]*Department of Statistics, Faculty of Science and Engineering, Macquarie University, New South Wales 2109, Australia*

([*]Corresponding author's e-mail: metta.kuning@gmail.com)

Received: 24 June 2015, Revised: 12 October 2015, Accepted: 16 November 2015

Abstract

The Southern Oscillation Index (SOI) has been used as a predictor of variables associated with climatic data, such as rainfall and temperature, and is related to the El Nino and La Nina phenomena, also called the El Nino Southern Oscillation (ENSO). The present study aims to describe the characteristics of the SOI between 1876 and 2014 using statistical methods. The graph of the cumulative monthly SOI in the period 1876 - 2014 shows that the data can be divided into 4 periods. The first period, from 1876 to 1919, shows no trend. An increasing trend is apparent in the second period from 1920 until 1975, while a decreasing trend is apparent in the third period, 1976 to 1995. In the last period, between 1996 and 2014, the SOI appears fairly stable. In order to investigate those trends, the linear regression and autoregressive (AR) model have been fitted. For the linear regression model, the outcome, SOI, is regressed against boxcar function, where the functions model the trends of the SOI. An autoregressive process is used to account for serial correlation in the residuals. The conclusion is that the SOI is quite similar to a random noise process.

Keywords: Autoregressive model, boxcar function, serial correlation, Southern Oscillation Index, white noise

Introduction

The southern oscillation (SO) is normally identified by the southern oscillation index (SOI) [1], that is, the large-scale fluctuations of the Mean Sea Level Pressure (MSLP) in the tropical Pacific between Tahiti and Darwin, Australia. The SOI is expressed as a number which ranges from about −30 to +30. Prolonged periods of negative SOI values are called *El Niño* events, where air pressure is below-normal at Tahiti and above-normal at Darwin. The opposite condition occurs when there are prolonged periods with strong positive SOI values, called *La Niña* events. Those events are also called El Nino Southern Oscillation (ENSO) events [2-4].

The SOI time series is useful for research into climatic data. There are several studies on the relationship between SOI and rainfall (e.g. Hadiani *et al*. [5]; Tigona and de Freitas [6]), as well as SOI and temperature (e.g. Jones and Trewin [7]; Halpert and Ropelewski [8]). Some studies have indicated that positive values of SOI correlate with above average rainfall and negative values of SOI correlate with below average rainfall, for example, in Australia (Stone and Auliciems [9]) and Ghana (Adiku and Stone [10]), whereas the correlation between SOI and rainfall was shown to be weaker in Western Australia (Carberry *et al*. [11]). In addition, Carbone *et al*. [2] reported that the total ozone in southern Brazil correlated with SOI, such that a reduction in total ozone occurred during *El Niño* episodes (negative SOI) and an increase in ozone occurred during *La Niña* episodes (positive SOI) from 1997 to 2003. On the other hand, Al-Zuhairi *et al*. [12] studied the correlation between temperature and SOI over Iraq, where there was no relationship detected in the period from 1900 to 2008.

That "the weather is not static", Carberry *et al*. [11], implies that the SOI is also continuously fluctuating. The SOI index is calculated using a moving 30-day average in order to try to eliminate some *Noise*, or small, random changes. A section of these data is illustrated in **Figure 1**, which shows the monthly average SOI from 1876 to 2014.

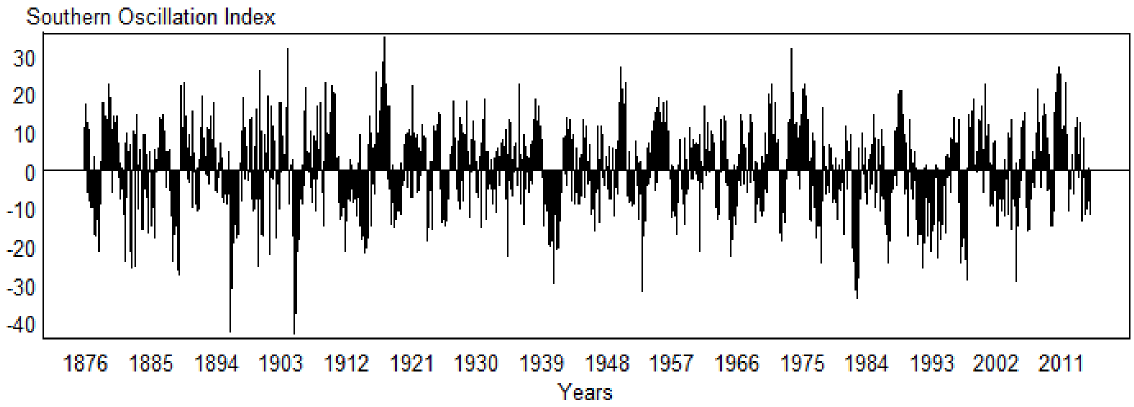

Figure 1 The variation of SOI from 1876 to 2014.

The graph does not show the trend of SOI because the data fluctuate. However, when we look at the graph of the cumulative SOI values in **Figure 2**, we can see an apparent trend from which Watts [13] deduced that "...from 1920 went into a long La Nina-dominate trend that ended with the Great Pacific Climate Shift of 1976" and "The subsequent El Nino-dominated trend from 1976 to 1995 was almost 3 times as fast as the rise". Thus, the focus of this paper is to evaluate the evidence for these claims, using appropriate scientific methods.

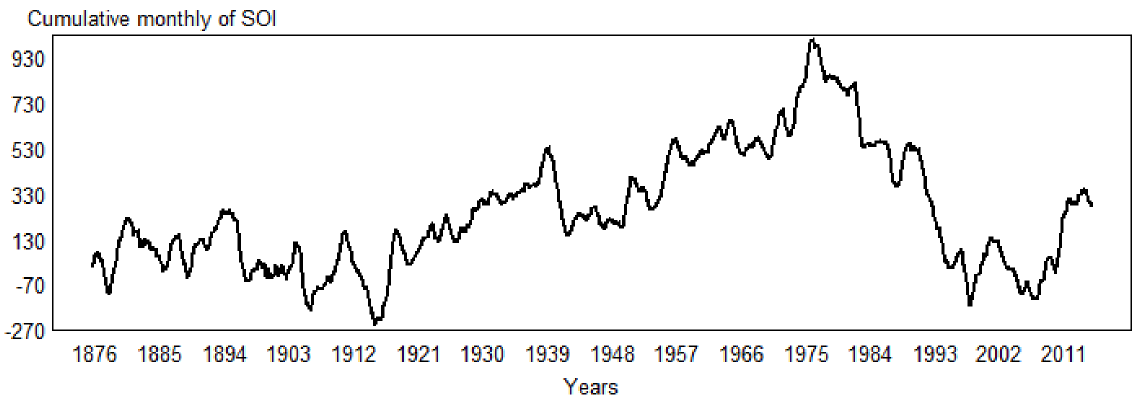

Figure 2 Cumulative monthly of SOI.

The data and methodology used in this study are described in the following section. Results and discussion are then presented, and summarized in section 4.

Materials and methods

Data

The data of the present study were downloaded from the Australian Government Bureau of Meteorology website, http://www.bom.gov.au/climate/current/soihtm1.shtml. The data are the average monthly SOI calculated over 139 years, so that there are 1,664 observations from the year 1876 to 2014. There are several methods used to calculate the SOI; the method used by the Australian Bureau of Meteorology is the Troup SOI [14] (see Australian Government Bureau of Meteorology [15]).

Methodology

The null hypothesis of this analysis is that the data are just noise, and in turn, the fluctuations in the SOI are random. In order to test this hypothesis, firstly, the scientific approach involves fitting an appropriate model to the data, and then applying an appropriate statistical test of the null hyphothesis, resulting in a p-value. The p-value is the probability that a data confuguration is at least as unusual as that observed could have arisen purely by chance, that is, assuming that the null hypothesis is true. By convention, p-values smaller than 0.05 provide sufficient evidence to reject the null hypothesis.

The first model fitted to the data is a linear regression model (see, Gill [16]; Venables and Ripley [17]); taking the formulation below, Eq. (1);

$$y_t = b_0 + b_i x_{it} + z_t \,. \tag{1}$$

In this model y_t represent the SOI for month t where t equals 1 for January 1876, while x_{it} represent boxcar function (see, Weisstein [18]) taking the values of 1 on intervals starting with a changing point and 0 elsewhere. The z_t constitute successive values which may be a series of auto-correlated normally-distributed errors (noise).

For a time series [19], the residuals z_t are often called the noise, and the model which does not include the noise is called the signal. If the residuals arise from uncorrelated errors, the noise is called white noise; otherwise, it is coloured. A commonly implemented model for coloured noise follows an autoregressive (AR) process. The AR model takes the form;

$$z_t = \sum\nolimits_i a_i z_{t-s_i} + w_t \tag{2}$$

where w_t is white noise. This means that each value of z_t is expressed as a linear combination of previous values at specified lags (s_1, s_2, s_3, etc.) plus an independent white noise series.

The linear regression model, which assumes independent errors using the function *lm*() in R [20], is used to determine the number of parameters in the AR model. The validity of a model can be checked by using the *acf*() function in R to plot the auto-correlation function (ACF) of the z_t Eq. (1). The *pacf*() function, the partial auto-correlation function (PACF) of the residuals, indicates which lagged terms need to be included in the AR model Eq. (2) to ensure that the w_t component represents white noise. The function *arima*() is used to analyse an AR model with the same predictor as in the linear regression model Eq. (1). From the AR model, the p-values of each lag are considered; these should be significant (< 0.05), for all lagged terms. Finally, the Analysis of Variance (ANOVA) procedure is used to test the predictor from the final model to confirm that at least one of the parameters from Eq. (1) is not equal to zero.

Results and discussion

The cumulative data from 1876 to 2014 are shown in **Figure 3**, and the patterns suggest 4 periods. The first period from 1876 to 1919 seems to show no trend. An increasing trend is apparent between the years 1920 and 1975, followed by a decreasing trend between the years 1976 and 1995. Again, it appears that there is no trend in the period from 1996 to 2014.

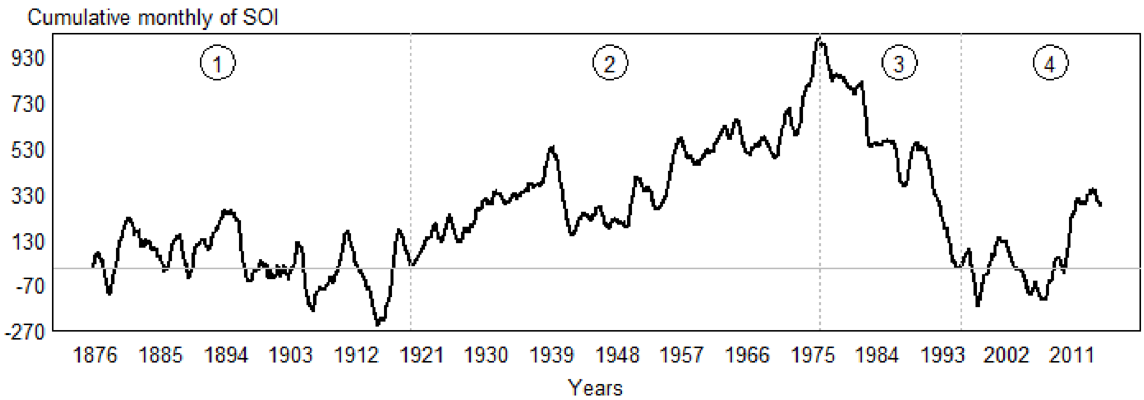

Figure 3 The 4 periods of the data.

The same statistical methods are applied to 3 subsets of the raw data. The first series includes all the data, from 1876 to 2014. The second series includes the data from 1920 to 2014, while the third series includes the data from 1920 to 1995. The results from the analyses of the data from the 3 different periods are presented in the next section.

Analysis of 4 periods of data

Firstly the 4 periods are analysed using the linear model on 3 boxcar functions. The first boxcar function takes a value of 1 starting at the changing point during the year 1920 to 1975 and 0 elsewhere (x_1), the second boxcar function takes a value of 1 during the year 1976 to 1995 and 0 elsewhere (x_2) and the last boxcar function takes a value of 1 during the year 1996 to 2014 and 0 elsewhere (x_3).

Table 1 shows the coefficient of parameters b_0, b_1, b_2, and b_3 from the linear model when b_0, b_1, b_2, and b_3 refer to the slopes of the first, second, third, and last period, respectively. The model indicates that the estimates of the coefficients of b_0 and b_3 are not statistically significant, but those of for b_1 and b_2 are significant, with p-values less than 0.05. In other words, there is no significant trend in either the first or fourth period, but there is a significant increasing trend in the second period, (b_1 = 1.373, p-value = 0.023) and a decreasing trend in the third period, (b_2 = –4.045, p-value < 0.001).

Table 1 The coefficient of parameters from linear model for 4 periods.

Parameters	Coefficients	SE	t-value	P-value
b_0	0.037	0.450	0.081	0.935
b_1	1.373	0.601	2.284	0.023
b_2	-4.045	0.804	-5.028	< 0.001
b_3	1.130	0.814	1.389	0.165

The p-value of ANOVA test is < 0.001

The ACF graph of the residuals from the linear regression model above shows that the data are highly serially correlated, and the PACF graph suggests the correlation of terms at lags 1, 2 and 3 months (**Figure 4**).

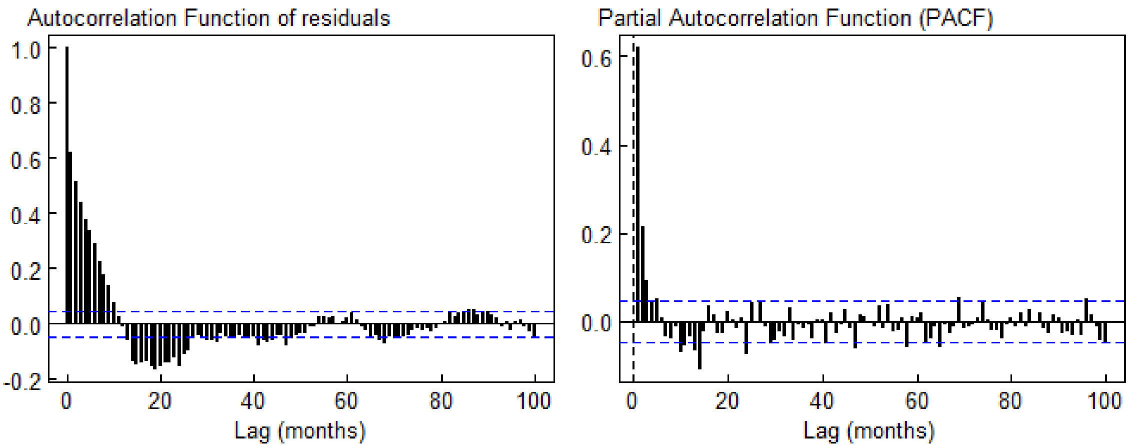

Figure 4 ACF graph (left panel) shows serial correlation and PACF graph (right panel) shows lagged terms of serial correlation from the linear model for 4 periods.

Then, these lagged terms are included in the function *arima*(), giving the results as shown in **Table 2**. As all AR terms at lags 1, 2, and 3 are statistically significant, we conclude we have derived an appropriate model.

Table 2 The result from AR(3) for 4 periods.

Parameters	Coefficients	SE	P-value
ar1	0.467	0.024	< 0.001
ar2	0.168	0.027	< 0.001
ar3	0.095	0.024	< 0.001
b_0	-0.083	1.249	0.947
b_1	1.645	1.652	0.319
b_2	-4.142	2.185	0.058
b_3	1.367	2.234	0.541
			The p-value of ANOVA test is 0.078

The ACF graph from this model in the right panel of **Figure 5** indicates that successive values w_t are not correlated, and the graph of the PACF in the right panel of **Figure 5** shows that the model fits the data quite well. However, the p-value from the ANOVA test for parameters b_0, b_1, b_2, and b_3 is 0.078, indicating that there is insufficient evidence against the null hypothesis, and that the coefficients of each parameter may indeed equal zero. It can be interpreted that there is no trend for all periods, which is in accordance with the p-value of more than 0.05 for those 4 parameters, as shown in **Table 2**.

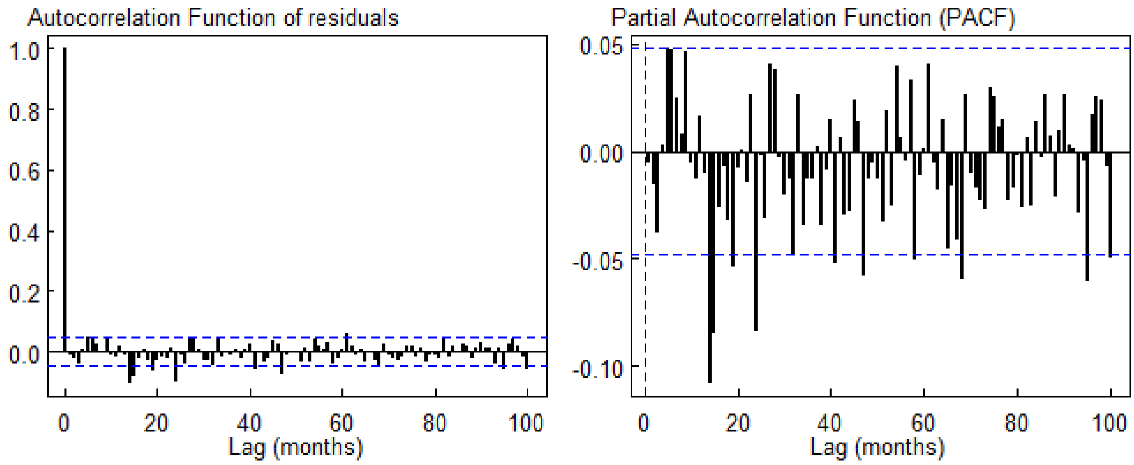

Figure 5 ACF graph in the left panel and PACF graph in the right panel from AR(3) for 4 periods.

Analysis of 3 periods of data

Next, we apply the same method as above, but only include the data from 3 periods, from 1920 to 2014, which includes the data from the second, third, and fourth periods, as shown in **Figure 3**. The first period is presumed to have no trend. In this analysis, 2 boxcar functions are considered, that is, during the year 1976 to 1995 (x_1) and 1996 to 2015 (x_2). **Table 3** shows the estimated coefficients of the paramters b_0, b_1, and b_2 from the linear model. These results are consistent with those we found from the previous analysis, and shows significant increasing and decreasing trends in the second and third periods respectively, and no trend in the fourth period. After the ACF and PACF graphs presented in **Figure 6** are examined, it is again apparent that 3 lagged terms should be in the AR model.

Table 3 The coefficient of parameters from the linear model for 3 periods.

Parameters	Coefficients	SE	t-value	P-value
b_0	1.409	0.373	3.779	< 0.001
b_1	-5.417	0.727	-7.452	< 0.001
b_2	-0.242	0.736	-0.329	0.742

The p-value of ANOVA test is < 0.001

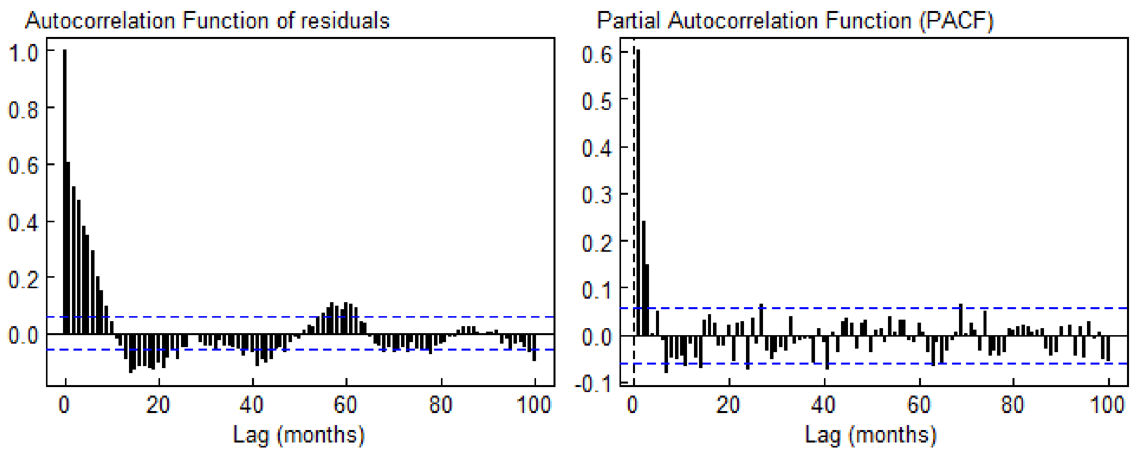

Figure 6 ACF graph (left panel) shows serial correlation and PACF graph (right panel) shows lagged terms of serial correlation from the linear model for 3 periods.

The p-value from the ANOVA test of parameters b_1, b_2, and b_3 from the subsequent AR(3) model is 0.037, indicating that there is at least one parameter not equal to zero. The results of this model are shown in **Table 4**, and we see there is insufficient evidence of a trend in the period 1920 to 1975, but there has been a significant decreasing trend between 1976 and 1995, and there is no trend in the period 1996 to 2014.

Table 4 The coefficient of parameters from AR(3) for 3 periods.

Parameters	Coefficients	SE	P-value
ar1	0.420	0.029	< 0.001
ar2	0.173	0.031	< 0.001
ar3	0.151	0.029	< 0.001
b_0	1.442	1.096	0.188
b_1	-5.675	2.039	0.005
b_2	-0.159	2.122	0.940
			The p-value of ANOVA test is 0.037

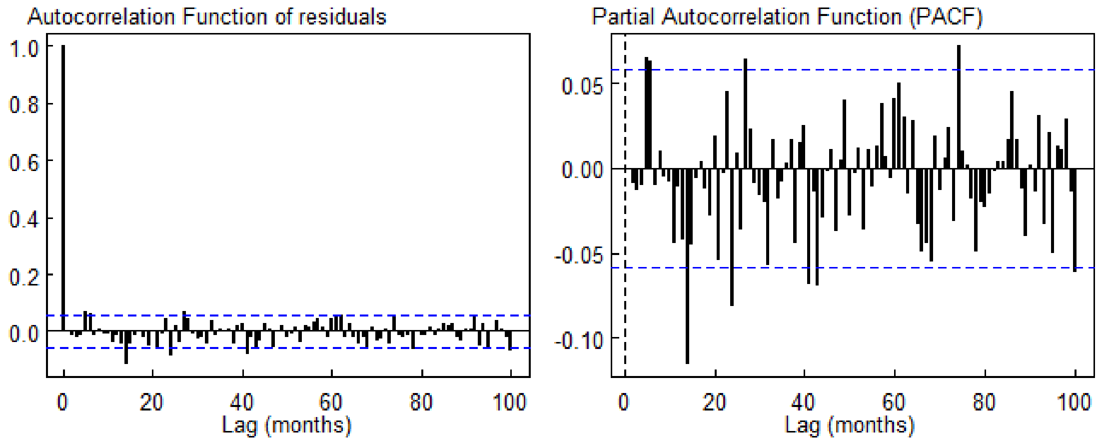

Figure 7 ACF graph in the left panel and PACF graph in the right panel from AR(3) for 3 periods.

Analysis of 2 periods of data

In this final analysis, only the data from the second and third periods, between 1920 and 1995, are considered, while the first and last periods are assumed to be constant. One boxcar function starting at the change point, that takes the value of 1 in the period 1976 to 1995 and 0 elsewhere (x_1), is the determinant in the linear model. **Table 5** shows that there are significant upward trends and downward trends in the second and third periods, respectively. The result from PACF graph presented in **Figure 8** also suggests that 3 lagged terms should be the order in the AR model.

Table 5 The coefficient of parameters from the linear model for 2 periods.

Parameters	Coefficients	SE	t-value	P-value
b_0	1.409	0.363	3.888	< 0.001
b_1	-5.417	0.707	-7.667	< 0.001
				The p-value of ANOVA test is < 0.001

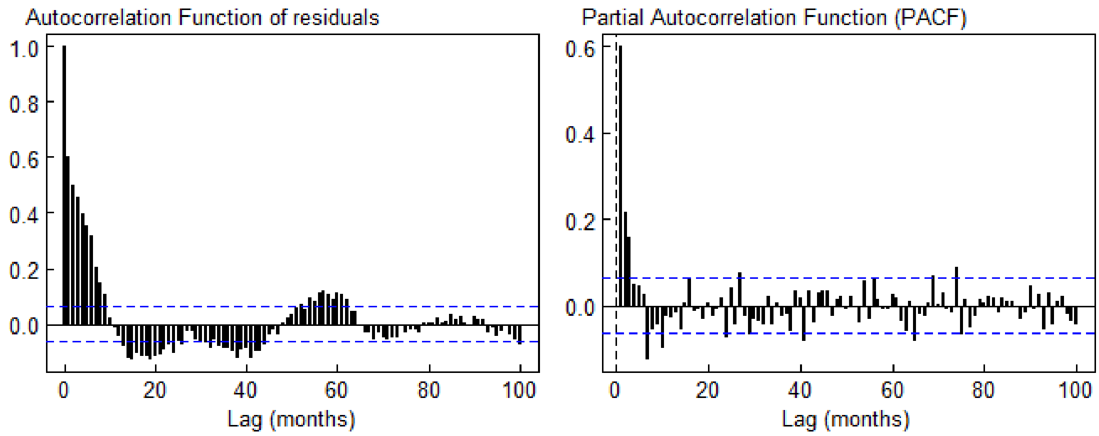

Figure 8 ACF graph (left panel) shows serial correlation and PACF graph (right panel) shows lagged terms of serial correlation from the linear model for 2 periods.

The ANOVA test of the parameters b_1 and b_2 from the AR(3) model is 0.021. Thus, we can conclude there is no significant trend in the second period, and a significant decreasing trend in the third period (**Table 6**).

Table 6 The coefficient of parameters from AR(3) for 2 periods.

Parameters	Coefficients	SE	P-value
ar1	0.437	0.033	< 0.001
ar2	0.139	0.035	< 0.001
ar3	0.158	0.033	< 0.001
b_0	1.433	1.038	0.168
b_1	-5.437	1.961	0.006
			The p-value of ANOVA test is 0.021

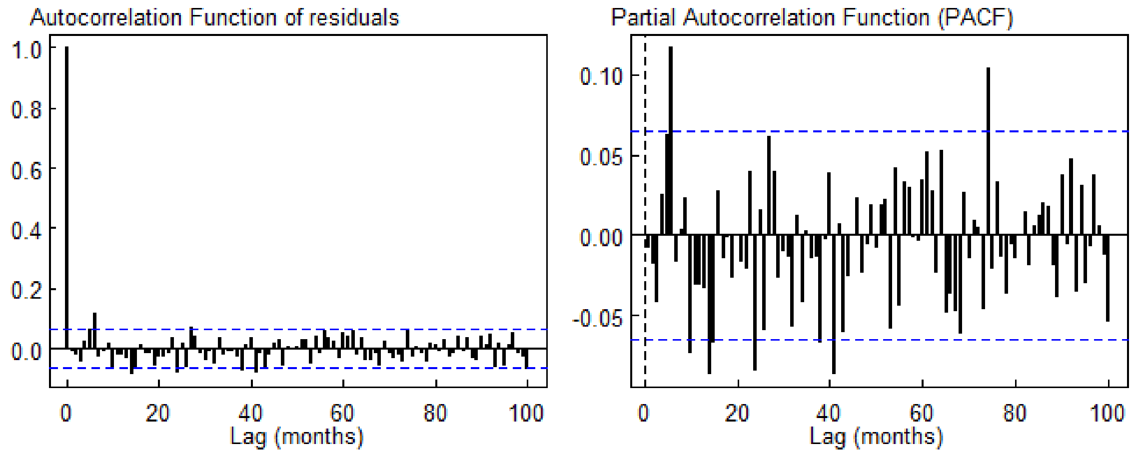

Figure 9 ACF graph in the left panel and PACF graph in the right panel from AR(3) for 2 periods.

Accumulating time series data over time can give a misleading impression, suggesting that a purely random process appears to have an informative structure. As the results have shown, there is insufficient evidence to conclude that the fluctuations in the SOI are due to anything other than random noise. Likewise, the study of Wu and Huang [21] used the SOI as noisy data, to illustrate their developed method. However, some researches have suggested that the SOI might be used to predict rainfall in the Fiji Islands. (Dennett and *et al*. [22]), southern Ghana (Adiku and Stone [10]), and some parts of Australia (Chiew and *et al*. [23]), and also predict itself (Chu and Katz [24]), indicating that some other correlation structure may be involved with other variables.

Given that the selected periods of possible changes are based on the data, the p-value is not strictly correct, since the null hypothesis is postulated from the same data used to test it. Even so,though we could test the null hypotheses from different periods with this study to define the boxcar function on the data by other criteria, such as using the period of every 10, 20 or 30 years, etc., to determine whether or not the variability in SOI is just a random process. An appropriate method needs to be based on sound statistical methodology; it is almost impossible to duplicate this method with the same data, because the SOI signal is unique. We could use other methods, including bootstrapping, to test the null hypotheses. However, in this study, we emphasize that the cumulative data has appeared to show some trends, but that we do not have sound statistical evidence to verify this.

Conclusions

It seems reasonable that the cumulative monthly SOI data can be divided into 4 periods, comprising the years 1876 - 1919, 1920 - 1975, 1976 - 1995, and 1996 - 2014. We have investigated the apparent increasing and decreasing trends seen in the second and third periods, respectively. To examine these trends, the 3 data sets have been analysed using the linear regression and autoregressive model on the boxcar function. The results from this study show that the appropriate autoregressive model for each of the 3 data sets is AR(3). For the model which included only the data from periods 2 and 3, we found no significant trend in the period 1920 to 1975, but a significant decreasing trend in the period 1976 to 1995. However, the result from the analysis of the data including all 4 periods shows no evidence of trend for any periods. Consequently, the evidence suggests that the fluctuations in the SOI seem to be similar with a random noise process.

Acknowledgements

This research is supported by the Centre of Excellence in Mathematics, Commission on Higher Education, Thailand. We gratefully acknowledge Emeritus Professor Don McNeil and Assistant Professor Phattrawan Tongkumchum for their helpful suggestions.

References

[1] SG Philander. *Encyclopedia of Global Warming and Climate Change*. 2nd ed. Sage Publication, Thousand Oaks, 2012, p. 1453.
[2] S Carbone, RL Manfro, AMP Schuch, MB da Rosa, DK Pinheiro, MT Kayano, and NJ Schuch. Correlation between southern oscillation index and total ozone column using Brewer and Toms at southern Brazil (29.4°S, 53.8°W) from 1997 to 2003, EGU General Assembly 2005, Available at: http://mtc-m18.sid.inpe.br/col/sid.inpe.br/ePrint@1905/2005/09.26.17.44/doc/correlation%20 between.pdf, accessed April 2015.
[3] YH Jin, A Kawamura, K Jinno and Y Iseri. On the long-term variability of Southern Oscillation Index. *In*: Proceedings of the 2003 Korea Water Resources Association. 2003, p. 151-8.
[4] JI Salisbury and M Wimbush. Using modern time series analysis techniques to predict ENSO events from the SOI time series. *Nonlinear Proc. Geoph*. 2002; **9**, 341-5.
[5] M Hadiani, SJ Asl, MR Banafsheh, Y Dinpajouh and E Yasari. Investigation the Southern Oscillation index effect on dry/wet periods in north of Iran. *Int. J. Agric. Crop Sci*. 2012; **4**, 1291-9.
[6] RS Tigona and CR de Freitas. Relationship between the Southern Oscillation and rainfall in Vanuatu. *Weather Clim*. 2012; **32**, 43-50.
[7] DA Jones and BC Trewin. On the relationships between the El nino-southern oscillation and Australian land surface temperature. *Int. J. Climatol*. 2000; **20**, 697-719.
[8] MS Halpert and CF Ropelewski. Surface temperature patterns associated with the southern oscillation. *J. Clim*. 1992; **5**, 577-93.
[9] R Stone and A Auliciems. SOI phase relationships with rainfall in eastern Australia. *Int. J. Climatol*. 1992; **12**, 625-36.
[10] SGK Adiku and RC Stone. Using the Southern Oscillation Index for improving rainfall prediction and agricultural water management in Ghana. *Agr. Water Manag*. 1995; **29**, 85-100.
[11] P Carberry, D George and D Buckley. The southern oscillation index and southern Australia, NSW department of primary industries, Available at: http://www.dpi.nsw.gov.au/__data/assets/pdf_file/0007/138139/the-southern-oscillation-index-and-southern-australia.pdf, accessed April 2015.
[12] MF Al-Zuhairi, JH Kadhum and AJ Farhan. Time series analysis of Southern Oscillation index and temperature over Iraq. *Diyala J. Pure Sci*. 2013; **9**, 57-68.

[13] A Watts. The story told by the Southern Oscillation Index, Available at: http://wattsupwiththat.com/2010/12/09/the-story-told-by-the-southern-oscillation-index/, accessed April 2015.
[14] AJ Troup. The Southern Oscillation. *Q. J. R. Meteorol. Soc*. 1965; **91**, 490-506.
[15] Australian Government Bureau of Meteorology. Sothern Oscillation Index, Available at: http://www.bom.gov.au/climate/glossary/soi.shtml, accessed May 2015.
[16] J Gill. *Generalized Linear Models*: *A Unified Approach*. Sage Publications, Thousand Oaks, 2001, p. 6-7.
[17] WN Venables and BD Ripley. *Modern Applied Statistics with S*. 4th ed. Springer, New York, 2002, p. 144-5.
[18] EW Weisstein. Boxcar function, Available at: http://mathworld.wolfram.com/BoxcarFunction.html, accessed April 2015.
[19] RH Shumway and DS Stoffer. *Time Series Analysis and Its Applications with R Examples*. Springer, New York, 2011, p. 11-37.
[20] R Development Core Team. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna. Austria: R Foundation for Statistical Computing, Available at: http://www.R-project.org, accessed March 2015.
[21] Z Wu and NE Huang. A study of the characteristics of white noise using the empirical mode decomposition method. *In*: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, London, 2004, p. 1597-611.
[22] MD Dennett, J Elston and PC Prasad. Seasonal rainfall forecasting in Fiji and the Southern Oscillation. *Agr. Meteorol*. 1978; **19**, 11-22.
[23] FHS Chiew, TC Piechota, JA Dracup and TA McMahon. E1 Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *J. Hydrol*. 1998; **204**, 138-49.
[24] P Chu and RW Katz. Modeling and forecasting the Southern Oscillation: A time-domain approach. *Mon. Weather Rev*. 1985; **113**, 1876-88.