NONPARAMETRIC APPROACH FOR DETECTION OF CORRELATION BETWEEN SOI AND PRECIPITATION

Young-Hoon JIN[1], Akira KAWAMURA[2] and Kenji JINNO[3]

[1]Student Member of JSCE, Doctoral student, Institute of Environmental Systems, Kyushu University
[2]Member of JSCE, Dr. of Eng., Associate Professor, Institute of Environmental Systems, Kyushu University
[3]Member of JSCE, Dr. of Eng., Professor, Institute of Environmental Systems, Kyushu University
(6-10-1, Hakozaki, Higashi-ku, Fukuoka City, 812-8581, Japan)

During the latest several decades, there has been considerable interest in revealing the relationship between El Niño/Southern Oscillation (ENSO) and hydro-meteorological variables. The oscillation is characterized by a simple index, the Southern Oscillation Index (SOI). Various approaches have been proposed for detection of correlation between SOI and hydro-meteorological variables. However, some of the approaches present a limited applicability for the data, which cannot be normalized by usual method such as power transformation. When the relationship between pairs is strongly nonlinear, usual Pearson's correlation coefficient also results in the spurious evidence for the relationship.

In the present study, we propose two nonparametric approaches with the overall purpose to obtain meaningful results from the relationship between SOI and precipitation. A data transformation is proposed for the monthly precipitation data in Fukuoka, Japan. The method transforms the precipitation data into nonexceedance probability time series. Also, Kendall's correlation coefficient, which is a rank-based procedure and can represent a nonlinear relationship between pairs, is used and verified through the comparison with the results by usual methods at the same station.

Key Words: *El Niño/Southern Oscillation (ENSO), Southern Oscillation Index (SOI), nonparametric approach, nonexceedance probability, Kendall's correlation coefficient*

1. INTRODUCTION

A large scale weakening of the trade winds and warming of sea surface temperature in the eastern and central equatorial Pacific Ocean defines El Niño, which typically lasts 12-18 months and occurs irregularly at 2-7 year intervals. The opposite situation, La Niña, refers to the condition when sea surface temperature is lower than normal. The two situations define an inter-annual seesaw phenomenon called the Southern Oscillation (SO), in tropical sea level pressure between eastern and western hemispheres. This oscillation is characterized by a simple index, the Southern Oscillation Index (SOI) which is used by NOAA (National Oceanic and Atmospheric Administration) to judge whether the El Niño or the La Niña event is occurring 1). The features are collectively known as the El Niño/Southern Oscillation (ENSO) phenomenon. This phenomenon is, consequently, a result from the interaction between large-scale oceanographic and atmospheric circulation processes in the equatorial Pacific Ocean.

During the latest several decades there has been considerable interest in the influence of ENSO on regional or local meteorological/hydrological variables, such as temperature, precipitation, and streamflow, etc. 2), 3), 4). These studies showed that the influence of ENSO on hydro-meteorological variables in the lower to mid-latitudes appears evident. For middle to high latitudes the impact of ENSO on hydrological variables is not clear. However, some studies have also shown effects of La Niña and SO on hydrometeorology for the region 5), 6).

For south-east Asia, several studies have been made. 7) studied relationships between SOI and precipitation in the Philippines/Malaysia and Japan, respectively. Although no quantitative relationship between SOI and precipitation was calculated, a clear pattern similarity between SOI and precipitation in the Philippines and Malaysia could be found. However, no clear relationship between SOI and precipitation in Japan was found. Kawamura et al. 8), 9), on the other hand, detected quantitative and statistically significant correlation
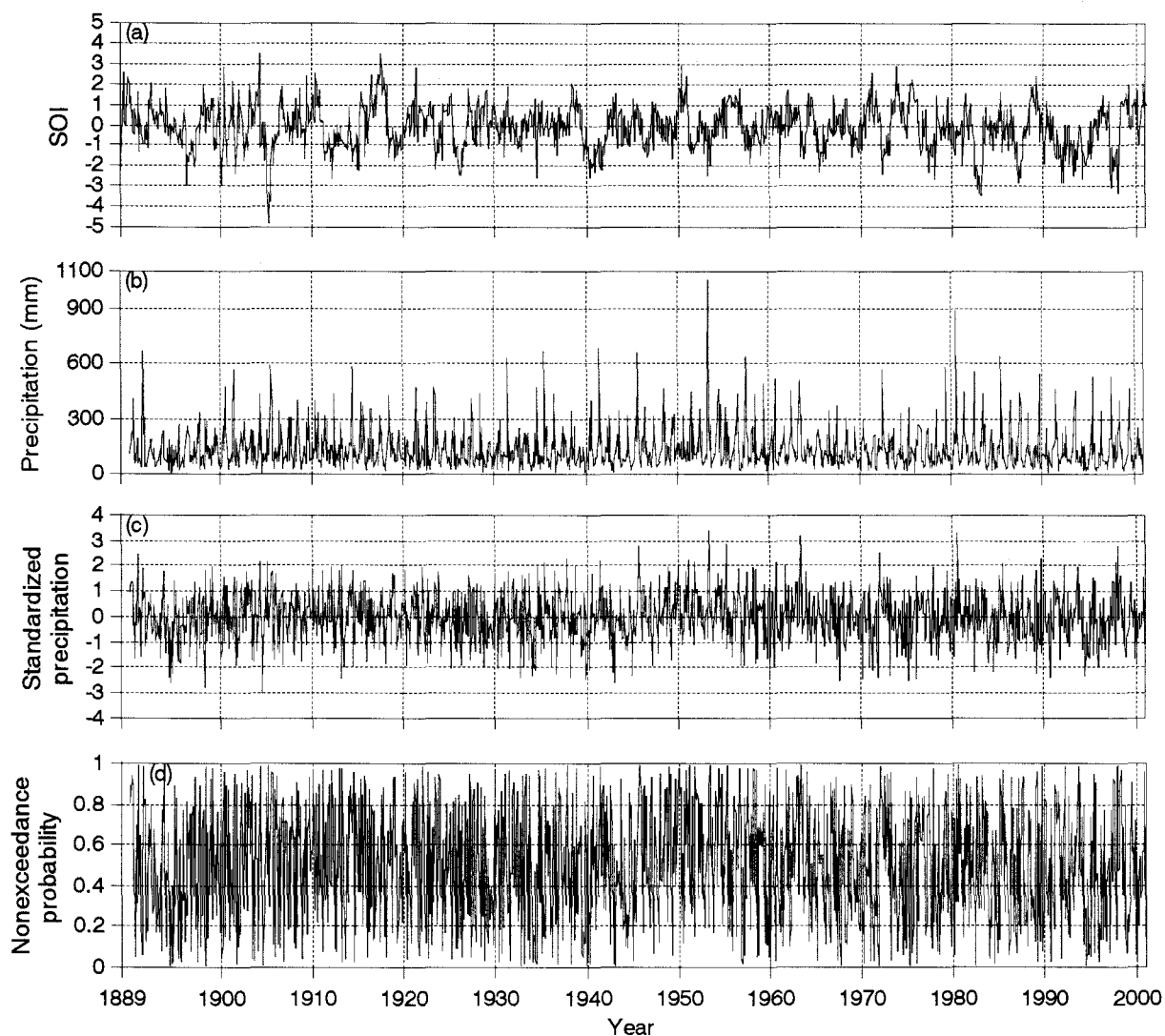
Fig. 1 Time series plots for (a) SOI values, (b) raw precipitation data, (c) normally standardized precipitation after cubic root transformation, and (d) nonexceedance probability time series. Tick marks on the time axis refer to January.

between SOI and precipitation/ temperature in Japan, using a simple method in which SOI data were categorized into five groups according to their magnitude.

As seen from the above studies, various approaches have been applied to reveal relationships between SOI and hydro-meteorological variables. Evaluating various methods to reveal influence of ENSO in middle to high latitudes is important because there is little evidence of El Niño/La Niña phenomena for these regions. Alternative techniques should, therefore, be evaluated in order to reveal hidden patterns and covariance in data. From this viewpoint, an appropriate pre-processing of raw data is essential. Normalization of raw data is usually performed because of skewness of data. However, when normalization is not possible with the usual transformations such as power transformation, an alternative one should be

performed instead and verified.

In the present paper, nonparametric approaches to reveal nonlinear relationships between SOI and precipitation are used. The approaches are used in estimating the correlation coefficient between the variables as well as data pre-processing.

An approximately 100-year monthly precipitation time series observed in Fukuoka, Japan is used to reveal the correlation with SOI. After providing general properties of data used, two different transformation techniques, which are cubic root transformation as a usual method and nonexceedance probability time series as a nonparametric approach, are applied to the monthly precipitation data. The cubic root transformation for the monthly precipitation at Fukuoka has been considered as a proper method. Therefore, the results from the method can be a basis to compare with the estimation of correlation by a newly

proposed method in the present study.

Besides the commonly used Pearson's r (Pearson product moment correlation coefficient), also the Kendall's τ is used. This method is robust with regard to effects of extreme values and deviations from a linear variation 10).

The results are analyzed in comparison according to the applied methodology. We close with conclusions of practical aspects of results obtained.

2. DATA USED AND TRANSFORMATION

The SOI data were calculated using the monthly mean sea level pressure (MSLP) at Papeete, Tahiti in central Pacific Ocean and Darwin in northern Australia. There are two commonly used methods to compute the SOI from the MSLP data at both stations: Troup's method and the Climate Prediction Centre's method. In the present study, we used Troup's method 11). The $SOI(y,m)$ in year y, month m (m = January to December) is calculated by the following equation;

$$SOI(y,m) = \frac{\{P_T(y,m) - P_D(y,m)\} - M_{30}(m)}{S_{30}(m)} \quad (1)$$

Here, $P_T(y,m)$ and $P_D(y,m)$ are MSLP (hPa) data at Tahiti and Darwin, respectively; $M_{30}(m)$ and $S_{30}(m)$ are the mean value (hPa) and its standard deviation (hPa) of MSLP differences between Tahiti and Darwin for the base period of 30 years (usually 1951-1980). **Fig. 1 (a)** shows the time series of the monthly SOI values calculated by the above Eq. (1) for the period from January 1889 to December 2000.

For the precipitation data, we selected Fukuoka station. The station is located in the northern part of Kyushu Island at 130.4°E and 33.6°N, southern Japan. It has measured the precipitation since 1890. The precipitation data from 1890 to 2000 was used for the present study. The time series is shown in **Fig. 1 (b)** and the basic statistics of the data were depicted by box-whisker plot, which indicates mean, standard deviations, maximum and minimum values on a monthly basis **(Fig. 2 (a))**. Apparent for the precipitation on a monthly basis is the skew data for all months.

As pointed out by Kawamura et al. 8), 9) and Jin et al. 12), no strong correlation between SOI and monthly precipitation at Fukuoka was revealed for all lag times up to 24-month when using the SOI data directly without any manipulation such as categorization. Therefore, for obtaining of the better understanding of the relationship between SOI and precipitation, we classified the SOI values into five groups according to the magnitudes, such as "Strong
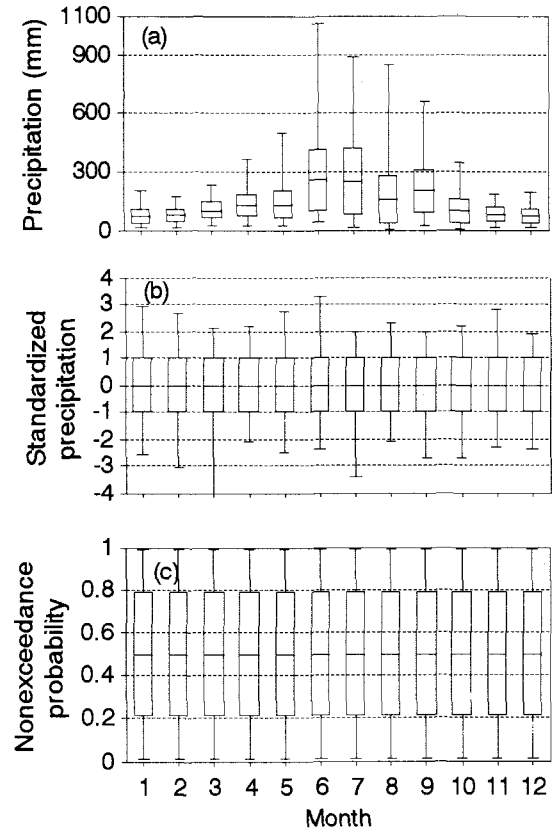

Fig. 2 Box-whisker plots for (a) raw precipitation data, (b) normally standardized precipitation after cubic root transformation, and (c) nonexceedance probability on a monthly basis

El Niño (SOI<-2)", "Weak El Niño (-2 ≤ SOI<-1)", "Normal Condition (-1 ≤ SOI ≤ 1)", and "Weak La Niña (1<SOI ≤ 2)", "Strong La Niña (2<SOI)". This naming for each categories of SOI is for easy association with the El Niño and La Niña phenomena. The statistical characteristics of SOI (e.g., the categorical occurrence and long-term fluctuation) can refer to the study by Kawamura et al. 13), 14) and Jin et al., 15).

The monthly precipitation data were, on the other hand, transformed by two calculations: cubic root transformation and nonexceedance probability time series. The two methods were applied on a monthly basis to remove deterministic components such as seasonality and an annual periodicity. **Fig. 1 (c)** shows the normally standardized precipitation with the cubic root transformation. To survey the distribution of the transformed data, the box-whisker plot is shown in **Fig. 2 (b)**. As is seen in the figure, the normally standardized data with cubic root transformation show mean of zero and standard deviation of one, on a monthly basis. Even though the transformation is appropriate for the precipitation data at Fukuoka, it is not applicable for the precipitation data, which are including several zero values.

Therefore, the nonexceedance probability time series is proposed to overcome the limitation of usual transformation. The nonexceedance probability of the ith-smallest precipitation can be obtained using $\alpha=0$ from the general formula proposed by Cunnane 16);

$$q_i = \frac{i - \alpha}{n + 1 - 2\alpha} \qquad (2)$$

where, q_i is the nonexceedance probability of the ith-smallest precipitation, n is the number of data on a monthly basis, and α is a plotting position parameter.

Fig. 1 (d) represents the nonexceedance probability time series of monthly precipitation and the distribution of data is depicted in **Fig. 2 (c)**. In addition, the transformation into nonexceedance probability time series can standardize the raw data so as to remove the deterministic components, as is seen **Fig. 2 (c)**.

The two transformed datasets are used for the cross-correlation analysis in the following section and the results are compared with each other.

3. CROSS-CORRELATION ANALYSIS

For the cross-correlation analysis, we calculated two types of correlation coefficient: Pearson and Kendall's correlation coefficients. The former is usually applied to compute correlation between pairs, while the latter is a rank-based procedure. We used the Kendall's correlation coefficient (Kendall's τ) because it has several advantages as one of nonparametric techniques comparing with the usual Pearson's r. In particular, nonparametric procedures require few assumptions about the underlying populations from which data are obtained and are relatively insensitive to outlying observations. Kendall's τ is, therefore, resistant to the effect of extreme values and to deviations from a linear relationship. Thus, it is well suited to use with dependent variables for which the variation around the general relationship exhibits a high degree of skewness or kurtosis 10).

Respective results from the application of two data transformations and two correlation coefficients are described as followings in the order of the type of correlation coefficients. As mentioned earlier, when using the SOI values without categorization and transformed precipitation, no significant correlation was found. Therefore, in the following description, we are focusing on the results from the categorization of SOI values.

Fig. 3 shows the correlation coefficients by Pearson's r with lag time up to 12-month between categorized SOI and cubic root transformed precipitation at Fukuoka. As are seen in the figure,
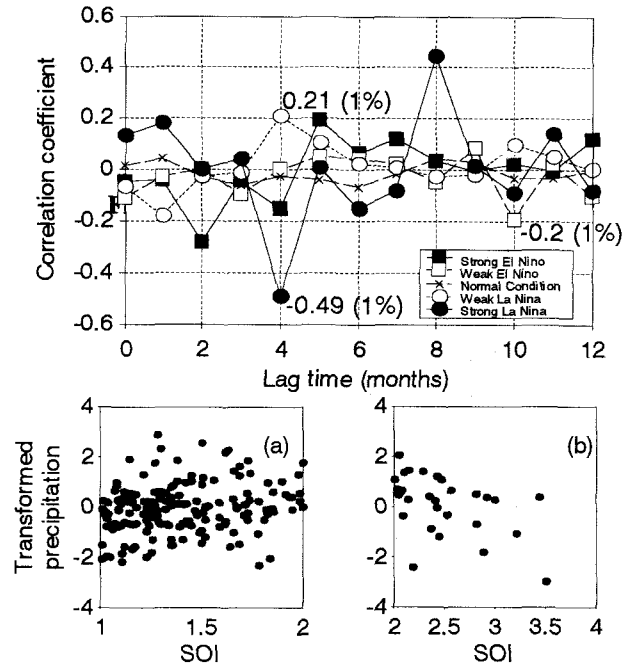

Fig. 4 Scatter plots under (a) "Weak La Niña" condition and (b) "Strong La Niña" condition

there are several significant correlations at 1 % level. However, under the weak condition as well as normal condition, considerable tendency was not revealed, even though the significance level is very high. Here, significance level shows that the correlation has the percentage of not being true 10). As an example, the scatter plot under the "Weak La Niña" category with lag time 4-month is shown in **Fig. 4 (a)**. On the other hand, the tendency under the "Strong La Niña" category with the lag time 4-month reveals very clear tendency (**Fig. 4 (b)**). The magnitude of the correlation is −0.49 with 1 % significance level and the tendency shows the stronger the La Niña event, the less precipitation is expected four months later. This results are the basis for the verification of applicability of nonexceedance probability time series and the Kendall's τ as nonparametric techniques to reveal the correlation between SOI and precipitation.

Correlation coefficients by Pearson's r between the categorized SOI and the nonexceedance probability revealed similar strengths and the same lag time with those obtained above (**Fig. 5**). As is seen in the scatter plot, which is corresponding to the lag time 4-month (**Fig. 6**), the same tendency with the above one was found. However, the significance levels became lower than those above because the magnitudes of correlations were slightly smaller with the same number of data.

From comparison of the results from the two data transformation, no big difference was found. Therefore, the nonexceedance probability time
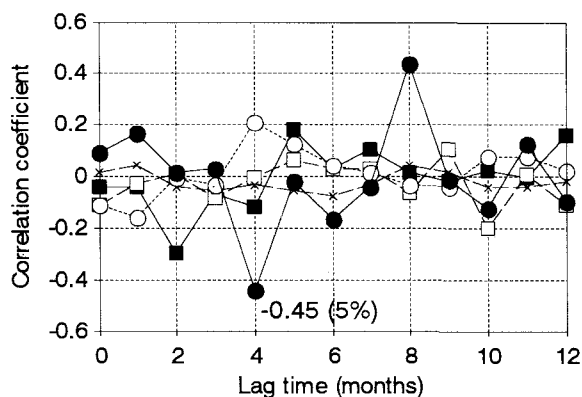
Fig. 5 Correlation coefficients by Pearson's *r* between categorized SOI and nonexceedance probability time series
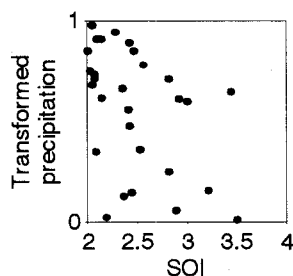

Fig. 6 Scatter plot for the lag time 4-month under the "Strong La Niña" condition

Fig. 7 Correlation coefficients by Kendall's τ between categorized SOI and cubic root transformed precipitation

Fig. 8 Correlation coefficients by Kendall's τ between categorized SOI and nonexceedance probability time series

series was proved as an alternative of usual transformation.

For the next step, we applied Kendall's τ to reveal the correlation between categorized SOI and two transformed precipitation. **Fig. 7** depicts the correlation coefficient by Kendall's τ when using the cubic root transformed precipitation with the categorized SOI. The very similar pattern with the above results by Pearson's *r* is shown in the figure. However, comparing with the results by Pearson's *r* when using the cubic transformed data, the strengths of significant correlation became weaker. Also, the even weaker correlation coefficients are shown in **Fig. 8**, which represents the correlations between categorized SOI and nonexceedance probability time series of precipitation. Generally, the strengths from Kendall's correlation coefficients are weaker than those from Pearson's correlation coefficient. Also, the magnitudes of the correlation when using nonexceedance probability time series are smaller than those from cubic root transformation in the same type of correlation coefficient. However, their significance levels are consistent through all procedures with higher than 5 % level with lag time four-month under the "Strong La Niña" category.
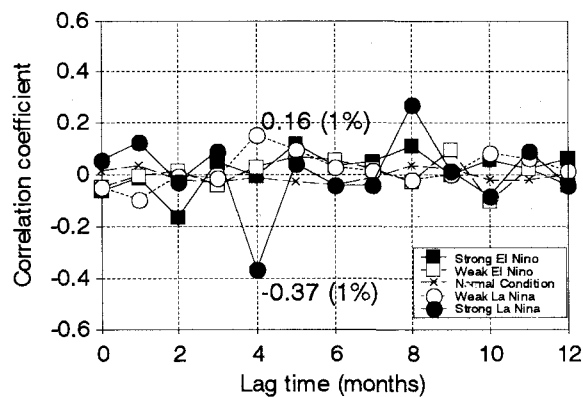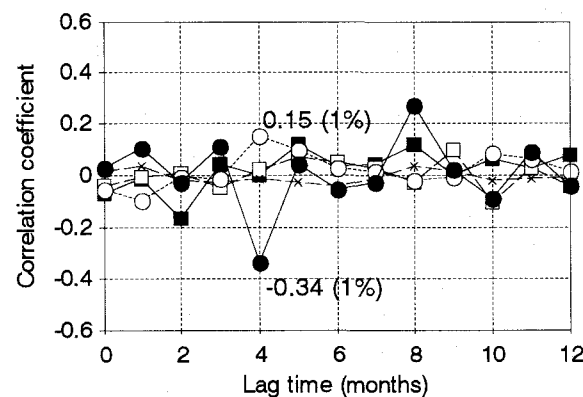
Even though the correlations under the "Strong

La Niña" condition showed relatively high coefficients at lag time 8-month in **Fig. 7** and **Fig. 8**, their significance levels were low, therefore, they have no significant meaning.

4. CONCLUSIONS

The cross-correlation analysis was carried out for the detection of relationship between SOI and precipitation at Fukuoka. For the analysis, we used several approaches, e.g., categorization of SOI, data transformation by two methods, and two types of correlation coefficients.

The categorization method for SOI has shown a considerable applicability from a series of researches and it was also used for the present study. On the other hand, the monthly precipitation data were changed by two transformations such as the cubic root transformation as a usual one and nonexceedance probability time series as an alternative nonparametric approach. The cubic root transformed data were standardized into mean of one and standard deviation of one on a monthly basis. Nonexceedance probability time series was

also calculated on a monthly basis. Therefore, deterministic components were removed from the data. For the calculation of correlation coefficient, we applied Kendall's τ as well as Pearson's r. Kendall's τ was used for the nonparametric approach, which has several advantages against Pearson's r as mentioned earlier.

The results from cross-correlation analysis with the two data transformations and the two correlation coefficients were compared with each other to verify the applicability of newly proposed data transformation and a rank-based correlation coefficient. For the verification, the results from the correlations between categorized SOI and cubic root transformed data by Pearson's r were referred as a standard basis. The comparison between the correlations with different data transformations by the same type of correlation coefficient showed a slight difference in magnitude of correlation. The strengths between categorized SOI and cubic root transformed data were slightly higher than those from nonexceedance probability time series of precipitation.

The significance level from Pearson's correlation coefficient was decreasing from the results of cubic root transformation to those of nonexceedance probability time series. However, the significance in Kendall's τ have the same significance levels with the same lag time four-month.

Consequently, applicability of nonexceedance probability time series and Kendall's τ as nonparametric techniques was successfully verified through the comparison of the respective results from the cross-correlation analyses. The newly proposed approaches can be applied to the various data type of hydro-meteorological variables and practically used for detection of correlation between SOI and precipitation in other study area, for the further study.

REFERENCES

1) Japanese Study Group for Climate Impact & Application.: *El Niño & Global Environment*, Seizando, Japan, 1999.

2) Poveda, G., Jaramillo, M. M. Gil, Quiceno, N.: Seasonality in ENSO-related precipitation, river discharge, soil moisture, and vegetation index in Colombia. *Water Resour. Res.*, Vol. 37, No. 8, pp. 2169-2178, 2001.

3) Gutiérrez F, Dracup J. A.: An analysis of the feasibility of long-range streamflow forecasting for Colombia using El Niño-Southern Oscillation indicators. *J. Hydrol.*, Vol. 246, pp. 181-196, 2001.

4) Chiew F. H. S., Piechota T. C., Dracup J. A., and McMahon T. A.: El Niño/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *J. Hydrol.*, Vol. 204, pp. 138-149, 1998.

5) Dracup J. A., and Kahya E.: The relationships between U. S. streamflow and La Niña events. *Water Resour. Res.*, Vol. 30, No. 7, pp. 2133-2144, 1994.

6) Rodo X., Baert E., and Comin F. A.: Variations in seasonal rainfall in Southern Europe during the present century: relationships with the North Atlantic Oscillation and the El Niño-Southern Oscillation. *Clim. Dynamics*, Vol. 13: pp. 275-284, 1997.

7) Yoshino, F.: On the relation between the monthly rainfall in Shikoku island and the El-Niño event, *Proc. of Fifth Symposium of Shikoku Branch of JSCE*, pp. 90-91, 1999.

8) Kawamura, A., Eguchi, S. and Jinno, K.: Cross-correlation between Southern Oscillation Index and precipitation/temperature in Fukuoka, Japan, *Proc. of Fresh Perspectives on Hydrology and Water Resources in Southeast Asia and the Pacific*, Christchurch, New Zealand, pp, 32-39, 2000.

9) Kawamura, A., Eguchi, S. and Jinno, K.: Correlation between Southern Oscillation and monthly precipitation in Fukuoka, *Journal of Hydraulic, Coastal and Environmental Engineering, JSCE*, Vol, 691(II-57), pp. 153-158, 2001a.

10) Hirsh R. M., Helsel D. R., Cohn T. A., and Gilroy E. J.: Statistical analysis of hydrological data. In: *Handbook of Hydrology*, Maidment DR (eds)., McGraw-Hill; New York: 28-30, Chapter 17, 1993.

11) Troup, A.J.: The "southern oscillation", *Quarterly Journal of the Royal Meteorological Society*, Vol. 91(390), pp. 490-506, 1965.

12) Jin, Y.-H., Kawamura, A. and Jinno, K.: Comparison of correlation between categorized SOI and monthly precipitation at Pusan in Korea and at Fukuoka in Japan, *Proc. of Korea Water Resources Association*, pp. 1251-1256, 2002.

13) Kawamura, A., Eguchi, S. and Jinno, K.: Statistical characteristics of Southern Oscillation Index and its barometric pressure data, *Annual Journal of Hydraulic Engineering, JSCE*, Vol. 45, pp. 169-174, 2001b.

14) Kawamura, A., Eguchi, S. and Jinno, K.: Long-term fluctuation characteristics of Southern Oscillation, *Annual Journal of Hydraulic Engineering, JSCE*, Vol. 46, pp. 103-108, 2002.

15) Jin, Y.-H., Kawamura, A., Jinno, K., and Iseri, Y.: On the long-term variability of Southern Oscillation Index, *Proc. of Korea Water Resources Association*, pp. 151-158, 2003.

16) Cunnane C.: Unbiased plotting positions – A review. *J. Hydrol.*, Vol. 37, pp. 205-222, 1978.