

Prediction Method of El Nino Southern Oscillation: ENSO by Means of Wavelet Based Data Compression with Appropriate Support Length of Base Function

Kohei Arai

Graduate School of Science and Engineering
Saga University
Saga City, Japan

Abstract— Method for El Nino/Southern Oscillation: ENSO by means of wavelet based data compression with appropriate support length of base function is proposed. Through the experiments with observed southern oscillation index, the proposed method is validated. Also a method for determination of appropriate support length is proposed and is validated.

Keywords- Prediction; Time series analysis; wavelet; ENSO; support length of mother wavelet function; base function

I. INTRODUCTION

El Nino is worldwide meteorological phenomena [1]-[5]. There are not so small number of papers which deals with prediction and forecasting of the phenomena. Most of these papers proposes methods for prediction and forecasting methods which are mainly focusing on extrapolations with numerical models and system equations expressing the phenomena. One of the paper proposed by Japanese Meteorological Agency describes that more than 10 months of observed data are required for prediction and forecasting of El Nino phenomena.

This paper describes a method for prediction of El Nino phenomena based on data compression utilizing wavelet Multi Resolution Analysis: MRA based feature extraction. Detection of El Nino phenomena can be done with MRA based time series analysis with the Southern Oscillation Index: SOI data and the Sea Surface Temperature: SST data. Also the paper describes a method for determination of support length of the mother wavelet of base function for MRA.

In accordance with Japanese Meteorological Agency: JMA, El Nino is defined as the ocean area of Eastern Pacific of Equatorial region at the latitude ranges from 4 degree North to 4 degree South and at the longitude ranges from 90 West to 150 West at which is called El Nino ocean area. Also El Nino is defined as the phenomena of which the difference between moving averaged Sea Surface Temperature: SST for five months is greater than 0.5 degree Celsius for 6 months compared to the averaged SST for 30 years starting from 1961 to 1990 as a reference. Therefore, JMA needs at least $2 + 6 + 2 = 10$ months for detection of El Nino phenomena.

In order to shorten the time required for detection of El

Nino phenomena, the prediction and forecasting method for detection of El Nino phenomena based on wavelet MRA is proposed. MRA is a kind of filter bank which allows extraction of wavelet frequency components. Therefore, there is a possibility to detect El Nino phenomena by using MRA in an efficient manner.

MRA is based on Discrete Wavelet Transformation: DWT with mother wavelet function. Mother wavelet function can be expressed with order and support length. In particular, support length is important parameter for time series analysis. Therefore, a method which allows determination of appropriate support length is proposed.

Through experiments with SOI and SST data, the proposed prediction and forecasting method for detection of El Nino phenomena as well as determination of appropriate support length of mother wavelet function are validated.

In the following section, the proposed method is described followed by some experimental results. Then conclusion is described together with some discussions.

II. PROPOSED METHOD

A. Discrete Wavelet Transformations

Discrete wavelet transformation for the given time series of scalar variables, η_i are defined in equation (1) with square matrix, C_n (wavelet transformation matrix) which consists of low wavelet frequency component coefficients, p_i and high wavelet frequency component coefficients, q_i .

$$C_n \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix} \quad (1)$$

Then the given time series of scalar variables can be divided into two parts, low frequency and high frequency components. There is a variety of wavelet transformation matrix with the different parameters of the order and the support length of the base function. For instance, the wavelet transformation matrix for the 8th order and two of the support

length is shown in equation (2). In the case of the wavelet transformation matrix for the 8th order with two support length is expressed in equation (3).

In the case of Daubechies base function, the wavelet transformation matrix with two of the support length can be calculated with equation (4) while that with four of the support length can be calculated with equation (5). Meanwhile, the wavelet transformation matrix with the arbitrary support length, (sup) can be calculated with equation (6).

$$C_8^{[2]} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \\ \eta_7 \\ \eta_8 \end{bmatrix} = \begin{bmatrix} p_0 & p_1 & & & & & & \\ q_0 & q_1 & & & & & & \\ & & p_0 & p_1 & & & & \\ & & q_0 & q_1 & & & & \\ & & & & p_0 & p_1 & & \\ & & & & q_0 & q_1 & & \\ & & & & & & p_0 & p_1 \\ & & & & & & q_0 & q_1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \\ \eta_7 \\ \eta_8 \end{bmatrix} = \begin{bmatrix} p_0x_1 + p_1x_2 \\ q_0\eta_1 + q_1\eta_2 \\ p_0\eta_3 + p_1\eta_4 \\ q_0\eta_3 + q_1\eta_4 \\ p_0\eta_5 + p_1\eta_6 \\ q_0\eta_5 + q_1\eta_6 \\ p_0\eta_7 + p_1\eta_8 \\ q_0\eta_7 + q_1\eta_8 \end{bmatrix} \quad (2)$$

$$C_8^{[4]} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \\ \eta_7 \\ \eta_8 \end{bmatrix} = \begin{bmatrix} p_0 & p_1 & p_2 & p_3 & & & & \\ q_0 & q_1 & q_2 & q_3 & & & & \\ & & p_0 & p_1 & p_2 & p_3 & & \\ & & q_0 & q_1 & q_2 & q_3 & & \\ & & & & p_0 & p_1 & p_2 & p_3 \\ & & & & q_0 & q_1 & q_2 & q_3 \\ p_2 & p_3 & & & & & p_0 & p_1 \\ q_2 & q_3 & & & & & q_0 & q_1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \\ \eta_7 \\ \eta_8 \end{bmatrix} = \begin{bmatrix} p_0\eta_1 + p_1\eta_2 + p_2\eta_3 + p_3\eta_4 \\ q_0\eta_1 + q_1\eta_2 + q_2\eta_3 + q_3\eta_4 \\ p_0\eta_3 + p_1\eta_4 + p_2\eta_5 + p_3\eta_6 \\ q_0\eta_3 + q_1\eta_4 + q_2\eta_5 + q_3\eta_6 \\ p_0\eta_5 + p_1\eta_6 + p_2\eta_7 + p_3\eta_8 \\ q_0\eta_5 + q_1\eta_6 + q_2\eta_7 + q_3\eta_8 \\ p_0\eta_7 + p_1\eta_8 + p_2\eta_1 + p_3\eta_2 \\ q_0\eta_7 + q_1\eta_8 + q_2\eta_1 + q_3\eta_2 \end{bmatrix} \quad (3)$$

$$\begin{aligned} (C_n^{[2]})^T C_n^{[2]} &= I_n \\ p_0 + p_1 &= \sqrt{2} \\ q_0 &= p_1 \\ q_1 &= -p_0 \\ 0^0 q_0 + 1^0 q_1 &= 0 \end{aligned} \quad (4)$$

$$\begin{aligned} (C_n^{[4]})^T C_n^{[4]} &= I_n \\ p_0 + p_1 + p_2 + p_3 &= \sqrt{2} \\ q_0 &= p_3 \\ q_1 &= -p_2 \\ q_2 &= p_1 \\ q_3 &= -p_0 \end{aligned} \quad (5)$$

$$\begin{aligned} 0^0 q_0 + 1^0 q_1 + 2^0 q_2 + 3^0 q_3 &= 0 \\ 0^1 q_0 + 1^1 q_1 + 2^1 q_2 + 3^1 q_3 &= 0 \\ (C_n^{[sup]})^T C_n^{[sup]} &= I_n \\ \sum_{j=0}^{sup-1} p_j &= \sqrt{2} \\ q_j &= (-1)^j p_{(sup-1)-j} \quad (j=0,1,2,\dots,(sup-1)) \\ \sum_{j=0}^{sup-1} j^r q_j &= 0 \quad \left(r=0,1,2,\dots,\left(\frac{sup}{2}-1\right) \right) \end{aligned} \quad (6)$$

Where n denotes the number of observed data while sup denotes support length of mother wavelet which is determined by the analysts.

B. Generalized Inversion Matrix

Cn is also described as follows,

$$C_n = \begin{bmatrix} P_n \\ Q_n \end{bmatrix} \quad (7)$$

Where Pn and Qn which are n/2 by n matrices can be expressed from the generalized inverse matrices as follows,

$$P_n^+ = P^T(P P^T)^{-1} \quad (8)$$

$$Q_n^+ = Q^T(Q Q^T)^{-1} \quad (9)$$

If the rank M of Pn and Qn is within a range from 0 to n/2, then both are decomposed with Singular Value Decomposition: SVD as follows,

$$P_n = \sum_{i=1}^M \kappa_i w_i v_i^T \quad (\kappa_1 \geq \kappa_2 \geq \dots \geq \kappa_M > 0) \quad (10)$$

$$Q_n = \sum_{i=1}^M \kappa_i w_i v_i^T \quad (\kappa_1 \geq \kappa_2 \geq \dots \geq \kappa_M > 0) \quad (11)$$

Then

$$P_n^+ = \sum_{i=1}^M \kappa_i^{-1} v_i w_i^T \quad (12)$$

$$Q_n^+ = \sum_{i=1}^M \kappa_i^{-1} v_i w_i^T \quad (13)$$

Where v and w denotes orthogonal matrix while k denotes singular matrix which consists of square root of eigen values.

C. Multi Resolution Analysis: MRA

The input data of scalar variables can be transformed to high (H) and low (L) wavelet frequency components with the wavelet transformation matrix. This Discrete Wavelet Transformation: DWT is called as the first level of DWT. L component can be divided into H and L components. This DWT is called as the second level of DWT. Furthermore, these transformations can be repeatedly applied to the L components again. These DWT is called as decomposition. The level is corresponding to the frequency components. In other words, arbitrary frequency component can be extracted with the different level of wavelet frequency component.

Because of the $C^T C = C^{-1} C$, it is possible to reconstruct original input data of scalar variables with the all levels of H components and the highest level of L component. This process is called with reconstruction, or Inverse Discrete Wavelet Transformation: IDWT.

DWT and IDWT can also be defined to the two dimensional image data as well as three dimensional moving pictures. Furthermore, these can be applied to arbitrary dimensional data f as follows,

$$(fC^T)^T C^T \dots C^T \quad (7)$$

For instance, DWT divides two dimensional input data into LL, HL, LH, and HH. The first and the second characters denote horizontal and vertical directions, respectively.

D. Process Flow of the Proposed Method

Process flow of the proposed method is as follows,

- 1) Determination of base function,
- 2) Determination of support length
- 3) Decompose the given time series of data by using DWT
- 4) Select the most nearest data compared to the most recent data from the low wavelet frequency components
- 5) Reconstruct the data by using the selected low frequency components based on Inverse DWT: IDWT.

III. EXPERIMENTS

A. The Data Used

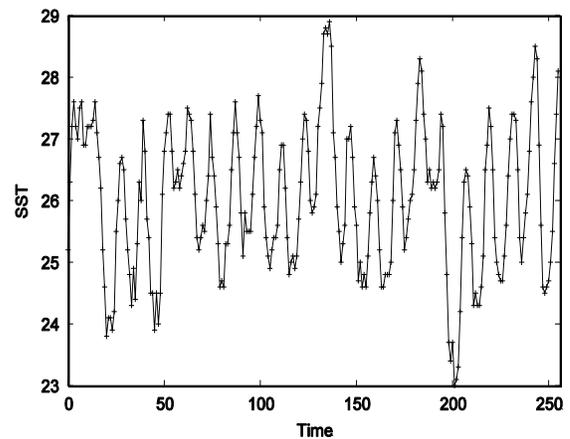
Monthly averages data of SOI and SST data are used for the experiments. 21.3 years starting from January, 1972 of data are used. SOI and SST are usually varied together and can be confirmed the changes of SOI and SST.

Time series of SST data are shown in Figure 1 (a) while those of SOI data are shown in Figure 1 (b), respectively. SST data is for the ocean area of Eastern Pacific of Equatorial region at the latitude ranges from 4 degree North to 4 degree South and at the longitude ranges from 90 West to 150 West at which is called El Nino ocean area. From Figure 1 (a), it is found that SST ranges from 23 to 29 degree Celsius.

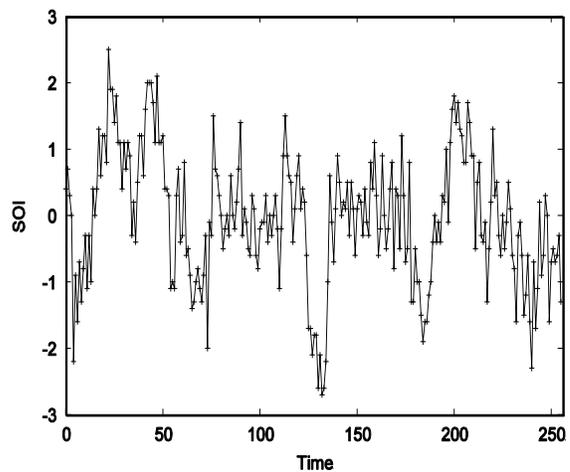
Meanwhile, SOI data is derived from the surface atmospheric pressure measured at the Tahiti and Darwin in Australia. SOI indicate jet stream changes. From Figure 1 (b), it is found that the SOI ranges from -3 to +3.

B. Method for Evaluation of Effect of the Support Length

Effect of the support length of mother wavelet function of DWT and IDWT used for decomposition and reconstruction is confirmed with the data used. Applying DWT with the different support length of mother wavelet, then IDWT with the same support length of mother wavelet with only low frequency components is performed. Then Root Mean Square: RMS error between both the original and the reconstructed time series data is evaluated. Figure 2 (a) and (b) shows the RMS errors for SST and SOI data.



(a) Relatively calm changes of SST



(b) Relatively rapid changes of SOI

Fig.1. Two different types of time series of data

C. Effect of the Support Length

In order to clarify the effect of the support length, relatively calmly changed Southern Oscillation Index: SOI is compared to relatively rapidly changed SOI which are shown in Figure 1 (a) and (b), respectively. DWT is applied at once (first level) to the data with the different support length. After that IDWT is applied to the transformed wavelet frequency

component with L component only. Root Mean Square Error: RMSE between the reconstructed data and the original data is evaluated. The results are shown in Figure 2.

D. Evaluation of Effect of the Support Length

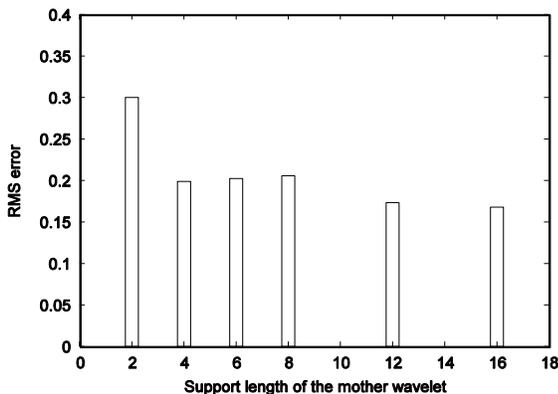
From Figure 2 (a) and (b), it is found that the appropriate support length for SST is 16 while that for SOI is 8.

In order to check the relation between appropriate support length and the changes of the time series of data, the following equation is proposed.

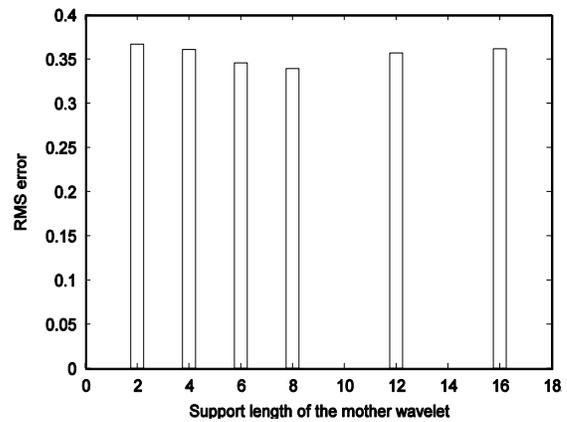
$$\begin{aligned}
 P(i, j) &= \frac{\overline{\Delta\epsilon(i, j)}}{\sqrt{\frac{U^2(i, j)}{N}}} \\
 U^2(i, j) &= \frac{1}{N-1} \sum_{k=1}^N (\alpha^{[k]}(i, j) - \overline{\Delta\epsilon(i, j)})^2 \\
 \alpha^{[k]}(i, j) &= \epsilon^{[k]}(j) - \epsilon^{[k]}(i) \\
 \overline{\Delta\epsilon(i, j)} &= \frac{\epsilon(j) - \epsilon(i)}{j - i}
 \end{aligned}
 \tag{14}$$

Where $\epsilon^{[k]}(sup)$ denotes square error at the support length is “sup” and at the time “k” while $\overline{\epsilon}(sup)$ denotes average of the $\epsilon^{[k]}(sup)$. Namely, if $P(i, j)$ is large, then the effect of support length is also large. In the evaluation N is set at 256.

$P(16,8)$ for SST is 1.774 while $P(8,16)$ for SOI is 1.624, respectively. It is found that $P(i, j)$ for SST is confidential at 95% of confidence level while $P(i, j)$ for SOI is confidential at the confidence level of 90% through Chi square test..If $P(i, j)$ is greater than 1.65, then $P(i, j)$ is 95 % confidential.



(a) Relatively calm changes of SST



(b) Relatively rapid changes of SOI

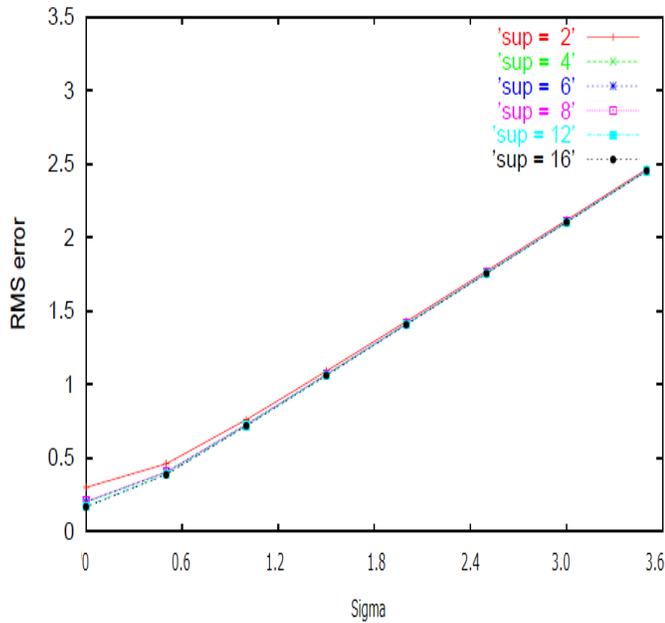
Fig.2. Root Mean Square error between original and reconstructed time series data through DWT with different support length (2-16) of mother wavelet.

E. In Case When the Low Frequency Components of the Past Data Does not Match to That of Present Data

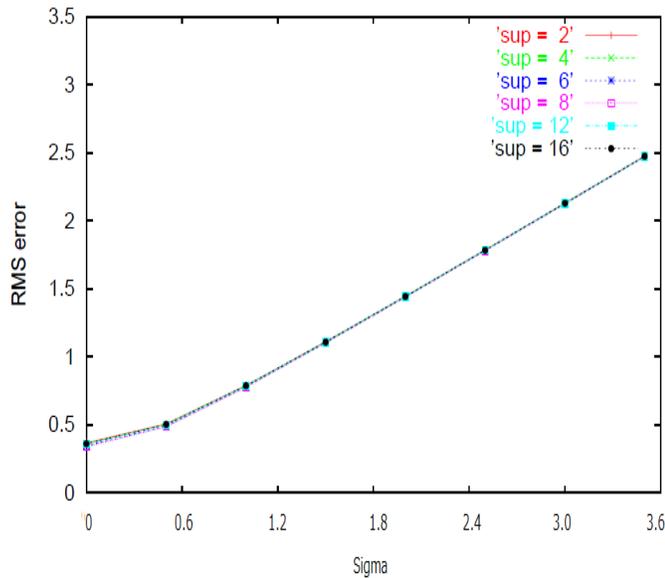
It is not always that the low frequency components of the past data are matched to those of the current data. In such case, normal distributed random number of noise (zero mean and σ of standard deviation) is added to the low frequency component derived from the past time series data through DWT of decomposition. Then the noise added low frequency component is used for reconstruction. It has to be matched to the low frequency component of the current data. After that, the reconstructed data is compared to the original time series of data before the decomposition. RMS difference between both shows a goodness of the restoration.

Figure 3 shows RMS error between the reconstructed data and the original time series of data before the decomposition as a function of sigma of standard deviation of added noise to the low frequency component derived from the past time series of data through DWT, decomposition. Figure 3 (a) shows RMS error for SST while Figure 3 (b) shows that for SOI.

In accordance with increasing of added noise, RMS error is increased for the SST and the SOI cases. It is also found that RMS error is increased in accordance with decreasing of support length in particular for the SST case, relatively calm changes of time series data. As is mentioned already, it is found that the appropriate support length for SST is 16 while that for SOI is 8.



(a)SST



(b)SOI

Fig.3. RMS error between the reconstructed data and the original time series of data before the decomposition as a function of sigma of standard deviation of added noise to the low frequency component derived from the past time series of data through DWT, decomposition.

IV. CONCLUSION

Method for El Nino/Southern Oscillation: ENSO by means of wavelet based data compression with appropriate support length of base function is proposed. Through the experiments with observed southern oscillation index, the proposed method is validated. Also a method for determination of appropriate support length is proposed and is validated.

In accordance with increasing of added noise, RMS error is increased for the SST and the SOI cases. It is also found that RMS error is increased in accordance with decreasing of support length in particular for the SST case, relatively calm changes of time series data. As is mentioned already, it is found that the appropriate support length for SST is 16 (relatively calm changes) while that for SOI is 8 (relatively rapid changes).

ACKNOWLEDGMENT

The authors would like to thank Arai's laboratory members for their valuable discussions and suggestions through this research.

REFERENCES

- [1] T.Yasanari, Global structure of the El Nin/ Southern Oscillation, Part-I, Meteorological Journal, 65, 1, 67-80, 1987.
- [2] T.Yasanari, Global structure of the El Nin/ Southern Oscillation, Part-II, Meteorological Journal, 65, 1,81-102, 1987.
- [3] T.Nita and T.Motoki, Abrupt enhancement of convective activity and low-level westerly first during the onset phase of the 1986-87 El Nino, Meteorological Journal, 65, 3, 497-506, 1987.
- [4] T.Maruyama and Y.Tsuneoka, Anomalous short duration of the easterly wind phase of QBO at 50 hPa in 1987 and its relationship, Meteorological Journal, 66, 4, 629-634, 1988.
- [5] T.Nitta, Development of a twin cyclone and westerly burst during the initial phase of the 1986-87 El Nino, Meteorological Journal, 67, 4, 677-681, 1989.

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission "A" of ICSU/COSPAR since 2008. He wrote 30 books and published 322 journal papers