STOCHASTIC PROCESS IN THE TIME SERIES MODEL OF PACIFIC DECADAL OSCILLATION (PDO)

*),**) Sandro Wellyanto Lubis
*) Climate Modeling Division of Center for Climate Risk and Opportunity Management in Southeast Asia and the Pacific (CCROM-SEAP) Bogor 16143, Indonesia
**) Department of Geophysics and Meteorology, Bogor 16680, Indonesia
e-mail: lubis.sandro@yahoo.de
Diterima 18 Mei 2011; Disetujui 23 Agustus 2011

ABSTRACT

Pacific Decadal Oscillation (PDO) is long-lived El Niño-like pattern of Pacific climate variability generated by coupled ocean-atmosphere interaction in the Northern Pacific Ocean. The best way to acquire a signal of PDO evidence is by determining the index of PDO. In this study, the PDO indexes are accurately modeled with time series methods through exponential smoothing analysis (Single and Holts Double exponential smoothing model) and Box-Jenkins analysis (ARIMA {1,1,1}, {2,1,1}. {3,1,1} and {4,1,1}). Nicholas's PDO Model (ARMA 9, 7) is also considered as comparative model in order to obtain the level of the reliability models that have been produced. The best selected prediction model that close to the real PDO index is ARIMA (2,1,1) Z_t = 1.574* Z_{t-1} - 0.427* Z_{t-2} -0.147* Z_{t-3} -0.976* a_{t-1} which means the forecast of PDO in the future depending on three months earlier data and a month earlier error of PDO index. Mean absolute error (MAE) of this model is 0.5283 and with root mean square error (RMSE) 0.6661. The predicted and observed PDO indexes are significantly correlated with r =0.76.

Keywords: *PDO, Box-Jenkins analysis, Exponential smoothing analysis*

ABSTRAK

Decadal Pacific Oscillation (PDO) adalah variabilitas iklim Pasifik yang menyerupai pola hidup El-Nino jangka panjang yang dibangkitkan oleh interaksi laut-atmosfer di bagian utara Samudra Pasifik. Cara terbaik untuk mendeteksi PDO adalah dengan cara menentukan indeks PDO. Pada kajian ini, indeks PDO dimodelkan secara akurat melalui penerapan metode runtun waktu dalam analisis pemulusan eksponensial (*Single* dan *Holts Double Exponential Smoothing Model*) dan analisis Box-Jenkins (ARIMA {1,1,1}, {2,1,1} {3., 1,1} dan {4,1,1}). Model PDO Nicholas (ARMA 9, 7) juga digunakan sebagai pembanding untuk melihat tingkat keandalan model yang telah dibuat. Hasil model prediksi terbaik yang mendekati nilai aktual indeks PDO adalah ARIMA

(2,1,1) = Zt = 1.574* Z_{t-1} -0.427* Z_{t-2} -0.147* Z_{t-3} -0.976* a_{t-1} bahwa untuk memprediksi nilai PDO di masa yang akan datang tergantung pada data indeks PDO tiga bulan sebelumnya dan error satu bulan sebelumnya. *Mean absolut error* (MAE) dari model ini adalah 0,5283 dan dengan *root mean square error* (RMSE) 0,6661. Observasi dan model PDO memiliki korelasi yang signifikan pada r = 0,76.

Kata kunci: *PDO, Analisis Box-Jenkins, Analisis pemulusan eksponensial*

1 INTRODUCTION

Pacific Decadal Oscillation (PDO) is one of the climate variability generated by coupled interaction between ocean and atmosphere, primarily occurs in the Northern hemisphere of Pacific Ocean. This phenomenon was discovered by fisheries scientist Steven Hare in the mid-1990's, based on observations of Pacific fisheries cycles [Mantua N.J 1999]. He named it when he studied about salmon production pattern in the Northern Pacific Ocean [Mantua *et al* 1997]. PDO cycle is characterized by the presence of warm and cool surface waters in the Pacific Ocean and the regimes shift from warm (positive phase) to cool (negative phase) in decadal time scale [Mantua and Hare 2002]. The cool period, for instance, is actually associated with extremely high sea surface temperatures in the Northern Pacific and the warm period is reversed (Figure 1-1).

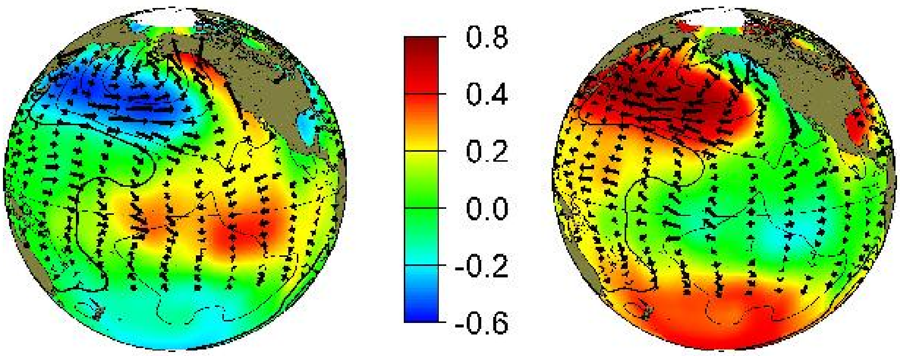

Figure 1-1:Sea Level Pressure (contours) and surface windstress (arrows) anomaly patterns during warm and cool phases of PDO [http://jisao.washington.edu/pdo]

PDO is almost similar to the El-Nino Southern Oscillation (ENSO) and some call as long-lived El-Nino-like pattern of pacific climate variability because both of these 'event's occur in the Pacific Ocean and highly detected in the variability of Pacific Ocean SST, but some scientists argue that based on collective body of research, there are three main characteristics distinguishing PDO from ENSO [Mantua *et al* 1997]: first, Period of oscillation, PDO has shifts phase on at least in inter-decadal time scale, persisted about 20 to 30 years. While ENSO is commonly known as inter-annual climate variability, persisted for 6 to 18 months in this region; second, visibility of climatic fingerprints of the PDO; and the third the mechanisms causing PDO variability were not known well, while causes for ENSO variability were relatively well-understood [Zhang *et al.* 1997, Mantua *et al.* 1997, NRC 1998].

The best way to acquire a signal of PDO event is by determining the index of PDO which is defined as the leading principal component of North Pacific monthly sea surface temperature variability from an un-rotated empirical orthogonal function analysis [http://jisao.washington.edu/pdo/]. The index is measured poleward of 20 degrees north latitude and very useful to describe the climate variation attributed to the Pacific Decadal Oscillation [Mantua *et al* 1997, Zhang *et al* 1997]. The most commonly PDO index is developed by Mantua [ftp://ftp.atmos.washington.edu/mantua/], but there are other PDO indexes developed from outside North America such as PDO index introduced by Evans *et al* [2000] and Linsley *et al* [2000] which are interesting because they substantiated a robust PDO to tropical and southern hemisphere climate [Evan *et al* 2000].

The research that was conducted by Mantua and Hare 2002, assertively explained some impacts emerging by strengthening and weakening of PDO event. The existence of PDO event affects the surface climate anomaly in some regions, for example the warm phase of PDO coincide with anomalously dry period in eastern Australia, Korea, Japan, Interior Alaska and in a zonally elongated belt from Pacific Northwest to northern South America; warm PDO phases also tend to coincide with anomalously with wet periods in southwest US, Mexico, southeast Brazil, south central South America and western Australia; the PDO ‘event’ is also broadly affected temperature from northwestern North America to Northwestern Australia [Mantua *et al* 2002, Willmott and Robeson 1995]. Mantua *et al* [2002] also noted that PDO ‘event’ has widespread impacts on natural system, including water resources in the Americas and many marine fisheries in North Pacific. In other words, the decadal variation of these shift phases of PDO ‘event’ has influenced the large-scale spatiotemporal patterns of large-scale wildfires occurrence in Northern American and also in the other areas geographically affected by this phenomenon [Nairn-Birch 2008]. A study of PDO which was) researched by Faqih *et al* [2008] showed that PDO and IPO are linked to low-frequency of rainfall variability in the Indonesian-Australia region through regulating regional SST patterns (Figure 1-2). He indicated that there is a significant relationship between interdecadal SST anomaly and rainfall variability.

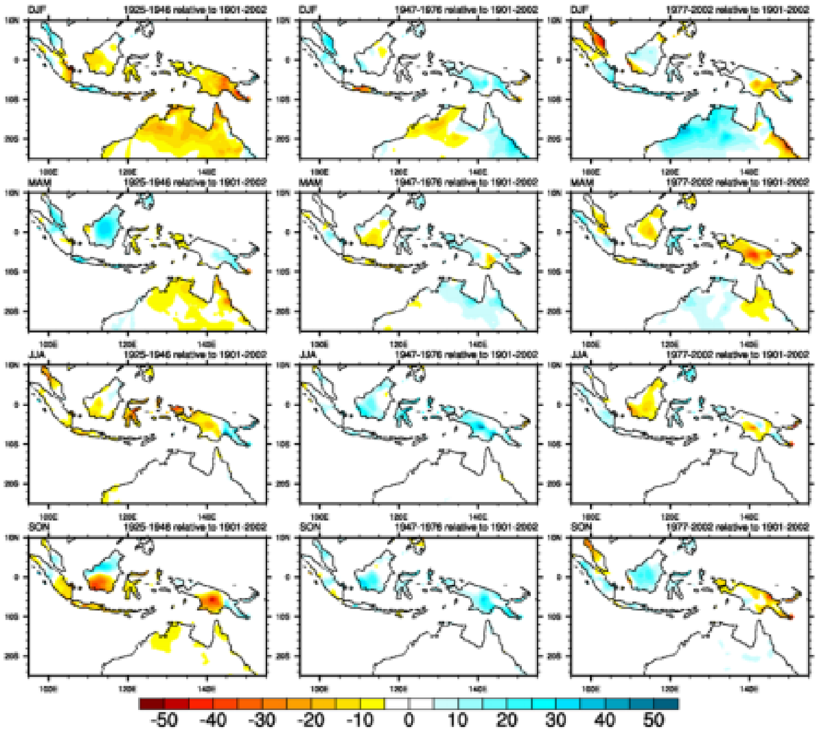

Figure 1-2: The anomalies of seasonal rainfall climatology during different phases of interdecadal variability. All relative to 1901-2002 climatology and rainfall climatology differences [Faqih *et al.,* 2008]

According to above information, this needs to be considered to climatologists either to develop a prediction model of the PDO index or to know the probability of occurrence of a particular pattern of PDO variation. In this paper, one of methods to accomplish that needs is to apply the procedure consists of time series analysis and exponential smoothing analysis to existing monthly PDO time series data from 1900 to 2010. This study describes various probability models for time series of PDO such as ARIMA, ARMA and exponential smoothing analysis, which are collectively called stochastic process. This stochastic process can be described as 'a statistical phenomenon that evolves in time according to probabilistic law'. The goal of this research is to develop a prediction model of PDO that can be broadly used to analyze the effect of PDO event to the rainfall variability in Indonesia and other purposes regarding to the climate variability assessments.

2 DATA AND METHODS

2.1 Data

In this study, we only use PDO index data produced through the first principal component from an un-rotated empirical orthogonal function analysis. This data was acquired from the website of the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) at the University of Washington (http://jisao.washington.edu/pdo/); the time series data begin in January of 1900 and end in December of 2010, totaling 1332 monthly observations. The data were divided into two groups: raw data from 1900-2007 which was used to build a prediction model of PDO and data from 2008-2010 for model verification.

2.2 Methods

The methods applied in this study are univariate analysis grouped into two main procedures namely the *Exponential Smoothing Procedure* and *the Box-Jenkins Procedure*. Box Jenkins model is developed from integrated of autoregressive model (AR) and moving average model (MA) or integrated of ARMA model for non-seasonal data of PDO index. Due to non-seasonal character of the data, first-order differencing is usually sufficient to analyze PDO model. The main stages in the setting up a Box-Jenkins forecasting are generally followed by four steps: first, model identification, to examine data to see which member of the class of ARIMA process appears to be most appropriate; second, estimation, to estimate the parameters of the chosen model; third, diagnostic checking, to examine the residual from the fitted model; and the last is to consider alternative model if necessary [Chatfield, 1989].

3 MODEL

Box-Jenkins model (ARIMA) is derived from general autoregressive integrated moving average process. In the time series analysis, we must understand how to decide the appropriate model for a time series data. The chosen model is generally determined by examining the APC and PACF curve or interpreting the correlogram (Table 3-1).

Table 3-1: IDENTIFICATION TIME SERIES MODEL OF AR (p), MA (q), AND ARMA (p, q) [Chatfield 1988, Evana, 2009]

	AR (p) or ARIMA (1,0,0)	**MA (q) or ARIMA (0,0,1)**	**ARMA (p,q) or ARIMA (1,0,1)**
ACF	Dies Down (Decreasing exponentially or sinusoidally)	Cuts-off after lag-q	Decreasing exponentially after lag - p
PACF	Cuts-off after lag - p	Dies Down (Decreasing exponentially or sinusoidally)	Decreasing exponentially after lag - q

Suppose that $\{Z_t\}$ is a purely random process with mean zero and variance σ^2_z. Then $\{X_t\}$ is a moving average process of order q (MA (q) process):

$$X_t = \beta_o Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q} \quad (3\text{-}1)$$

$$B^j X_t = X_{t-j}, \; X_t = \theta(B) Z_t \quad (3\text{-}2)$$

Where $\{\beta_i\}$ are constants, B is backward shift operator and θ (B) is a polynomial of order q in B. Now, suppose that $\{Z_t\}$ is a purely random process with mean zero and variance σ^2_z , then $\{X_t\}$ is an autoregressive process of order p (AR (q) process):

$$X_t = \alpha_o Z_t + \alpha_1 Z_{t-1} + \ldots + \alpha_p Z_{t-p} + Z_t \quad (3\text{-}3)$$

$$X_t = Z_t (1 - \alpha B)^{-1} = f(B) Z_t \quad (3\text{-}4)$$

Where $f(B) = (1-\alpha_1 B- \ldots - \alpha_p B^p)^{-1}$, if we mix equation (3-2) and (3-4), then we will find a new equation of mixed autoregressive/moving average process containing p AR terms and q MA terms or its commonly called as ARMA process of order (p,q):

$$X_t = \alpha_1 Z_{t-1} + \ldots + \alpha_p Z_{t-p} + Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q} \quad (3\text{-}5)$$

By using shift operator B, equation (3-5) can be written as:

$$\phi(B) X_t = \theta(B) Z_t \quad (3\text{-}6)$$

Where,

$$\phi(B) = 1 - \alpha_1 B - \ldots - \alpha_p B^p \quad (3\text{-}7)$$

$$\theta(B) = 1 + \beta_1 B - \ldots - \beta_q B^q \quad (3\text{-}8)$$

If the ARMA is non-stationary, the solution to produce a stationary model is by summing or integrating the data to provide a model for non-stationary data:

$$W_t = \nabla^d X_t = (1-B)^d X_t \quad (3\text{-}9)$$

$$X_t = \alpha_1 W_{t-1} + ... + \alpha_p W_{t-p} + Z_t + ... + \beta_q Z_{t-q} \quad (3\text{-}10)$$

$$\phi(B) W_t = \theta(B) Z_t \quad (3\text{-}11)$$

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t \quad (3\text{-}12)$$

Equations (3-9)-(3-12) are describing the *d*th differences of Xt or called as ARIMA process of order (p, d, and q). Differencing process introduced by Box and Jenkins (1970) is one of the methods for removing a trend in a given time series data until it becomes stationary.

$$Y_t = X_{t+1} - X_t = \nabla X_{t+1} \quad (3\text{-}13)$$

Or for second order differencing:

$$\nabla^2 X_{t+2} = \nabla X_{t+2} - \nabla X_{t+1} \quad (3\text{-}14)$$

To compare the above models, we use exponential analysis models which are Single and Holts Double exponential smoothing analysis, these methods can measure and remove a nonsymmetric trend or in other words these methods can be used if the data is not significantly influenced by seasonal factor.

Data smoothing with single exponential smoothing requires a parameter called the smoothing constant (α). Each data point is given a certain weighting, α for the newest data, (1-α) for older data and etc. The value of α must be between 0 and 1. The following is the equation of smoothed value:

$$X_t = \alpha\left[Y_t + (1-\alpha)Y_{t-1} + (1-\alpha)^2 Y_{t-2} + ...\right] \quad (3\text{-}15)$$

$$X_t = \alpha Y_{t-1} + (1-\alpha) X_{t-1} \quad (3\text{-}16)$$

Where, α is the smoothing factor, and $0 < \alpha < 1$, X_t is output of the exponential smoothing algorithm and Y is raw data. Holts Double exponential smoothing method uses different parameters than the one used in original series to smooth the trend value.

$$X_t = \alpha Y_t + (1-\alpha)\left(X_{t-1} + T_{t-1}\right) \quad (3\text{-}17)$$

$$T_t = \gamma\left(X_t - X_{t-1}\right)(1-\gamma)\ T_{t-1} \quad (3\text{-}18)$$

$$\hat{Y}_{t+m} = X_t + T_t m \quad (3\text{-}19)$$

Equation (3-17) calculates smoothing value X_t from the trend of the previous period T_{t-1} added by the last smoothing value X_{t-1}. Equation (3-18) calculates trend value T_t from S_t, S_{t-1}, and T_{t-1} and equation (3-19) (forward prediction) is obtained from trend, T_t, multiplied with the amount of next period forecasted, m, and added to basic value S_t.

4 RESULT AND DISCUSSION

4.1 Test for Stationary Time Series

Stationary test needs to be done as a pre-processing process before building a model because the time series forecasting requires a condition that the data must be stationary. A set of data is called stationary if both of mean value and variance are constant in time. In other words, a stationary series contains no trend (systematic change in mean) and systematic change in variance, or strictly periodic variations. Non-stationary data must be modified to be stationary using a type of linear filter called differencing.

Figure 4-1 shows a time plot of the monthly PDO index values from January 1900 through December 2007. An initial examination roughly shows any systematic increase or decrease in variance with time; the mean appears to remain constant over time. It indicates that the data is not stationary in variance, and transformations of the data are necessary. The proof of non-stationary data can be also examined through analysis of correlogram.

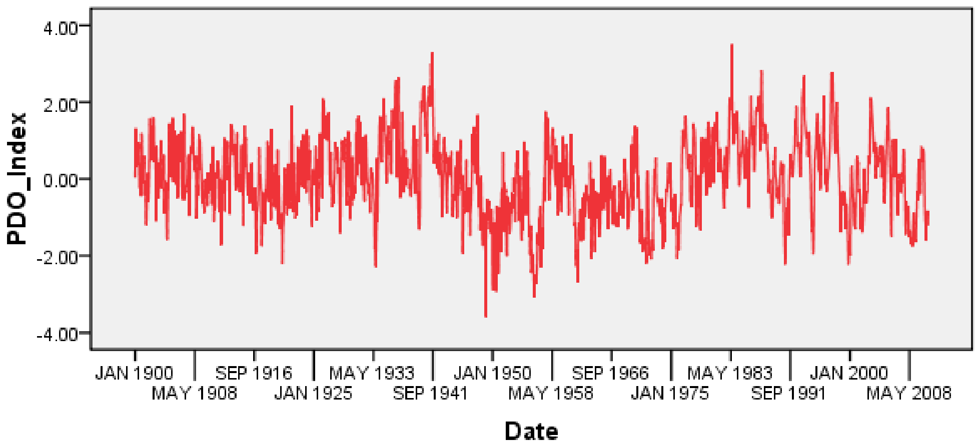

Figure 4-1: Time series plot of monthly PDO Index values from January 1900 to December 2007

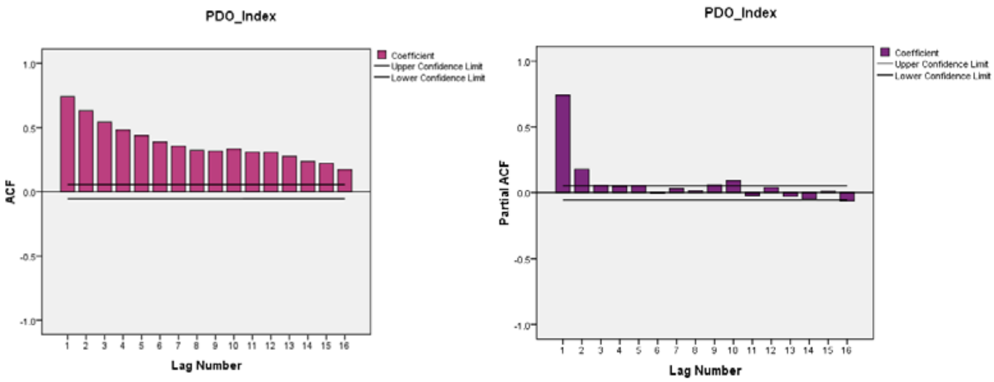

Figure 4-2: Autocorrelation (ACF) and partial autocorrelation (PACF) of the raw PDO data with upper/lower confidence limits

ACF curve decreases at the first lag and meanwhile the PACF curve is significantly cut-off at the first lag. It indicates that the data is not stationary in variance. To coerce the data to be stationary, the first order of differencing technique should be applied. After applying the first order difference, the variance value decreased from 1.044 to 0.537 and the standard deviation from 1.0218 to 0.7327, which mean(s) that the data is ready to produce reasonable forecast. The following table is a statistical description of raw data before differencing and after first order differencing applied.

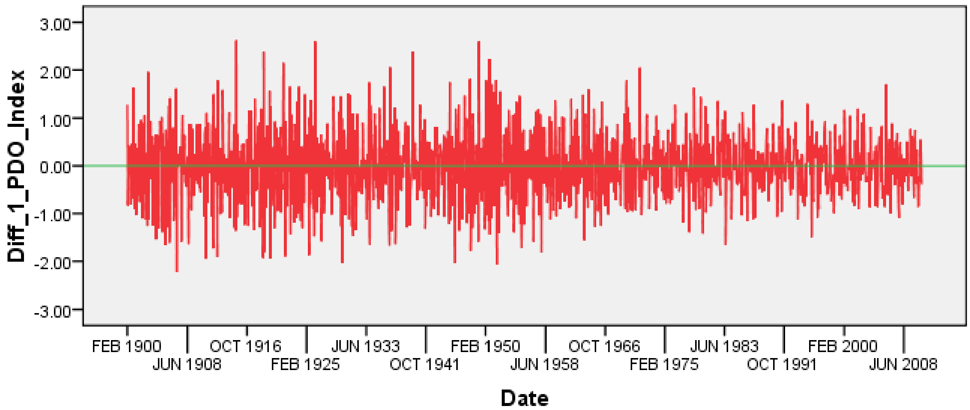

Figure 4-3: Time series plot of monthly PDO Index values from January 1900 to December 2007 from first differencing technique

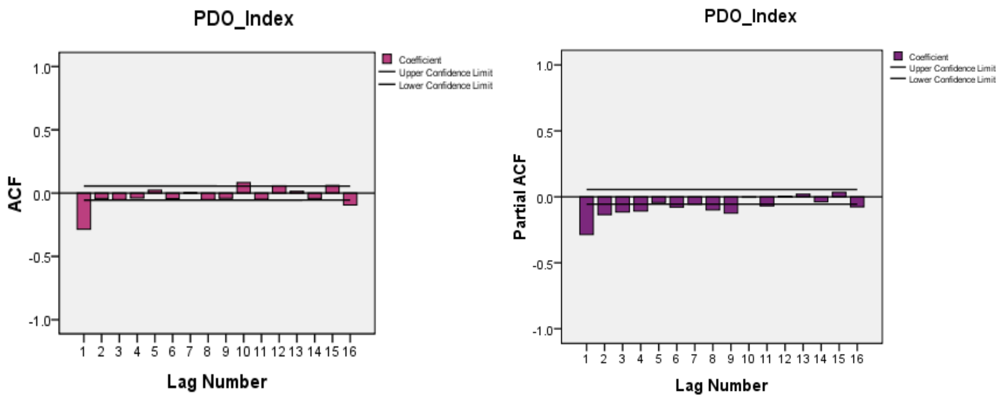

Figure 4-4: Autocorrelation (ACF) and partial autocorrelation (PACF) from first differencing technique

After differencing process, we found that the time series data has been stationary because there is no sign that the ACF curve is dumping out and the PACF is cut-off at the first lag. So, next we try to identify the best model by combining the p, d, and q order in ARIMA models.

Table 4-1: DESCRIPTIVE STATISTICS OF PDO INDEX BEFORE AND AFTER FIRST DIFFERENCE

Variable	PDO Index	First Order Difference of PDO Index
N	1296	1295
Mean	0.0511	-0.0005
Std. Error of	0.02838	0.02036
Std. Deviation	1.02175	0.73272
Variance	1.044	0.537
Minimum	-3.60	-2.23
Maximum	3.61	2.62

4.2 Model Identification

The chosen significant coefficients decide whether the autocorrelation coefficients is significantly different from zero or its modulus exceeds from $2/\sqrt{N}$, where N is n-observations. According to ACF curve, the critical value is 0.06 and the significant coefficient is only at lag 1 (Figure 4-2). Moreover, the significant coefficient in the PACF curve whose moduli exceed 0.06 is at lag 1, 2, 3, and 4. (Figure 4-4). Thus, to identify an appropriate ARIMA model to forecast of PDO index, we should asses the combination of non-seasonal values of p, q and d in the model. The values of p and q are assessed by looking at the first few values of the significant coefficients of ACF and PACF. In this case, we have p=1, 2, 3, and 4, d=1 (∇X_t) and q=1. To avoid the subjectivity of the model, we prefer to combine the model index to be some ARIMA model equations (Table 4-2) and then in the diagnostic checking step, we will decide which one of the models that performance the lowest of average error.

Table 4-2: EQUATIONS OF SELECTED ARIMA MODELS d=1 (∇X_t) WITH SOME COMPARATIVES MODELS

ARIMA (p,d,q) Model	Model Equations
ARIMA (1,1,1)	X_t=1.607*Z $_{t-1}$-0.607*Z $_{t-2}$-0.935*a_{t-1}
ARIMA (2,1,1)	X_t= 1.574* Z $_{t-1}$-0.427* Z $_{t-2}$-0.147* Z $_{t-3}$-0.976* a_{t-1}
ARIMA (3,1,1)	X_t =1.572* Z $_{t-1}$-0.408* Z $_{t-2}$-0.13* Z $_{t-3}$-0.034* Z $_{t-4}$-0.979* a_{t-1}
ARIMA (4,1,1)	X_t =1.572* Z $_{t-1}$-0.397* Z $_{t-2}$-0.127* Z $_{t-3}$-0.018* Z $_{t-4}$-0.03* Z $_{t-5}$-0.981* a_{t-1}
Nicholas ARMA (9.7)	X_t =0.794* Z $_{t-1}$-0.366* Z $_{t-2}$+0.44* Z $_{t-3}$-0.346* Z $_{t-4}$+0.361* Z $_{t-5}$-0.371* Z $_{t-6}$+0.99* Z $_{t-7}$-0.516* Z $_{t-8}$-0.054* Z $_{t-9}$-0.202* a_{t-1}+0.371* a_{t-2}-0.207* a_{t3}+0.271* a_{t-4}-0.163* a_{t-5}+0.25* a_{t-6}-0.808* a_{t-7}
Single Exp. Smoothing Model	X_t=α*Y $_{t-1}$+(1-α)*X $_{t-1}$ with α =0.602
Holts Double Exp Smoothing Model	X_t=0.551* Y $_{t-1}$+0.449*(X_t +T_{t-1}), and T_t=1.68.10^{-5} *(X_t- X_{t-1})*0.99 T_{t-1} Ft= X_{t-1+} $T_{t-1,}$ with α=0.551 and γ=1.68.10^{-5}

4.3 Diagnostic Checking

In this stage, we should check the diagnostic statistics to see which of ARIMA (p, d, and q) model is adequate. Other comparative models are included in this diagnostic checking such as high-order mixed ARMA (9, 7) model and either single or double exponential smoothing model. The checking of the model will be examined through assessment of Mean Absolute Error (MAE), Correlation (r) and Root Mean Square Error (RMSE) to each of models.

The MAE measures the average magnitude of the errors in a set of PDO forecasts, without considering their direction. It measures accuracy for continuous PDO variables, or simply, it is the average over the verification sample of the absolute PDO values of the differences between forecast and the corresponding observation. While, RMSE is a quadratic scoring rule which measures the average magnitude of the PDO forecast error. This value is very useful when large errors are particularly undesirable. The MAE and the RMSE can be used together to diagnose the variation in the errors in a set of forecasts. Both of these values describe the average of model-performance error.

Table 4-3: MEAN ABSOLUTE ERROR (MAE), ROOT MEAN SQUARE ERROR (RMSE) AND CORRELATION (r) FOR SELECTED ARIMA MODELS AND COMPARATIVE MODELS

MODEL	**MAE**	**RMSE**	**r**
ARIMA (1,1,1)	0.533038682	0.672549282	0.75
ARIMA (2,1,1)	0.528363866	0.666179088	0.76
ARIMA (3,1,1)	0.528522385	0.666638291	0.75
ARIMA (4,1,1)	0.528536327	0.665655729	0.76
ARMA (9.7)*	0.528536327	0.665655729	0.76
Single Exponential Smoothing*	0.546579559	0.686730484	0.76
Holts Double Exponential Smoothing*	0.548719451	0.687716433	0.75

*Comparative models

According to Table 4-3, all selected models have nearly the same predictive skill. Therefore, the determination of the best selected model is not only from the statistical performance but also based on the composition model which has the simplest form. The best selected model which has the smallest MAE and RMSE values is ARIMA (2,1,1). This model is only consisting of few parameters (more efficient) than the ARMA (9,7) model. This model is constructed by three months earlier data and the previous error of PDO index, or in mathematical expressions can be written as $Zt = 1.574^* Z_{t-1} -0.427^* Z_{t-2} -0.147^* Z_{t-3} - 0.976^* {}_{at-1}$. In this case, Zt is forward prediction of PDO's index. MAE of this model is 0.5283 with RMSE 0.6661. Insignificantly difference between RMSE and MAE indicates that all the errors are in the same magnitude. Detail composition of ARIMA (2, 1, 1) model can be seen in the Table 4-4.

Table 4-4:PARAMETER OF ARIMA (2, 1, 1) MODEL

			Estimate	SE	t	Sig.
ARIMA (2,1,1)	Constant		.000	.002	-.110	.912
	AR	Lag 1	.574	.029	19.948	.000
		Lag 2	.147	.029	5.158	.000
	Difference		1			
	MA	Lag 1	.976	.008	129.613	.000

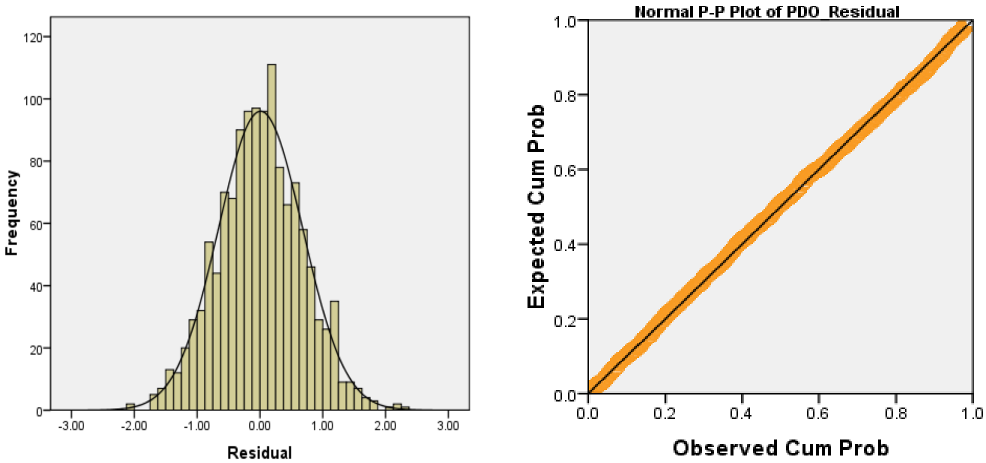

Figure 4-5: Tests for residual normality

Further analysis used to examine the skill of the model is by applying the residual test. Figure 4-5 shows that distribution of ARIMA (2,1,1) residual model is classical bell-shaped, symmetric histogram with most of the frequency counts bunched in the middle and with the counts dying off out in the tails. The mean of residual is nearly zero 0.01. Then the recommended next step is to do a normal probability plot to confirm approximate normality as we did before. The normal probability plot produces an approximately straight line which means that the distribution of residual of the model meets normality assumption. From this diagnostic checking, we can conclude that the ARIMA (2, 1, 1) PDO Model is a good model for data forecasting.

4.4 Forecasting of PDO Model

The model diagnostics indicate that the ARIMA (2, 1, 1) model successfully captures the variation of the stationary series. Thus, forecasting the future values of an observed time series is only conducted by using the best selected PDO model. Calibration of the model was done by using PDO index started from 1900 to 2007 and the

PDO forecasting (verification of the model) was started from January 2008 to December 2010 (Figure 4-6). Figure 4-6 roughly shows that both of best selected ARIMA models and comparative models simultaneously fit the observed PDO index data. But as we have done in diagnostic checking step, the best forecasting model is just selected by examining the lowest of MAE, RMSE, and the highest correlation of the good fitness of the model. The highest correlation of the model fitting is 76% produced by ARIMA (2, 1, 1), ARIMA (4, 1, 1) and single exponential model.

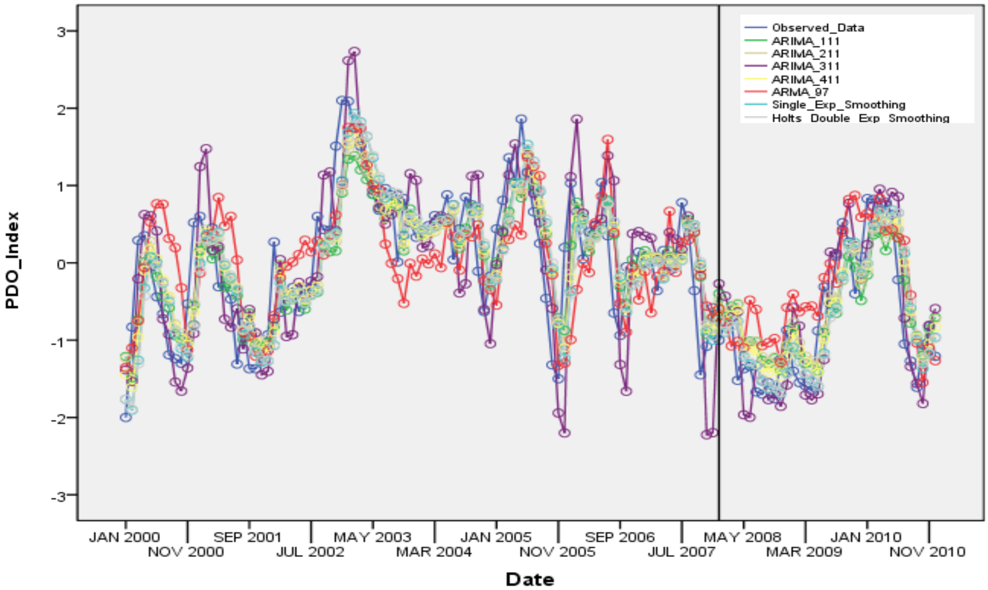

Figure 4-6: Calibration and Forecasting Models Based on PDO Index from an Un-rotated Empirical Orthogonal Function (EOF) Analysis of JISAO

The decision of final model is based on capability of the model to minimize the error and has the simplest form. On this PDO research, it's possible to decide that our reliable model to predict of PDO index is ARIMA (2, 1, 1). Performance of the model to determine forward PDO index (ability to resemble of PDO variation) is by using three months earlier data and a month earlier error of PDO index (Figure 4-7) with 76% of Pearson's correlation. Because forecasts are conditional case about future based on specific assumption, so this model is not the end of PDO prediction model, the analysis should always be prepared and developed to modify them as necessary in the light of any external or internal information is available.

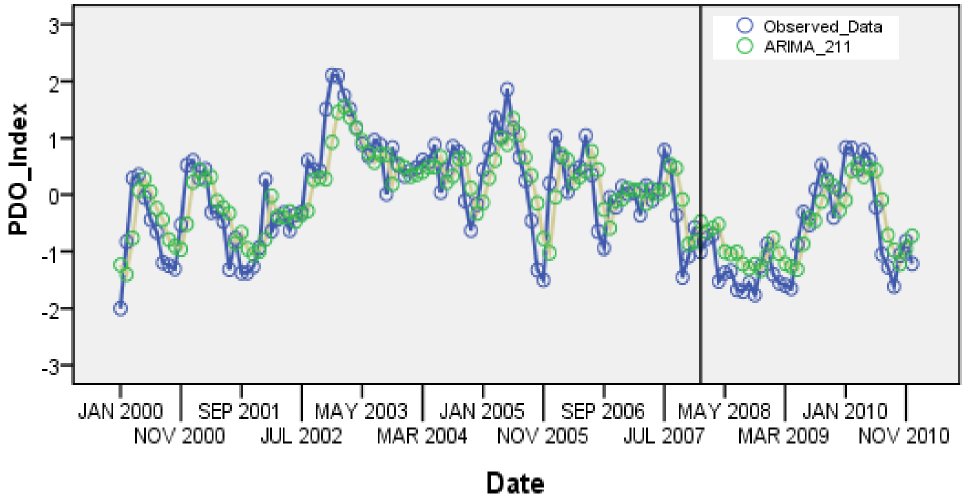

Figure 4-7: Time series plot of validation and verification of PDO model (ARIMA (2, 1, 1)) and raw data

Table 4-5: PREDICTED PDO INDEX FROM JULY 2009 TO JULY 2011

Period	Observed PDO	Predicted PDO	Error
Jul-09	-0.53	-0.41385	-0.11615
Aug-09	0.09	-0.45913	0.549129
Sep-09	0.52	-0.12241	0.64241
Oct-09	0.27	0.230968	0.039032
Nov-09	-0.4	0.151615	-0.55161
Dec-09	0.08	-0.28295	0.362954
Jan-10	0.83	-0.09721	0.927213
Feb-10	0.82	0.4261	0.3939
Mar-10	0.44	0.540064	-0.10006
Apr-10	0.78	0.318072	0.461928
May-10	0.62	0.468458	0.151542
Jun-10	-0.22	0.430235	-0.65024
Jul-10	-1.05	-0.09105	-0.95895
Aug-10	-1.27	-0.71397	-0.55603
Sep-10	-1.61	-0.9756	-0.6344
Oct-10	-1.06	-1.21833	0.158326
Nov-10	-0.82	-0.94881	0.128806
Dec-10	-1.21	-0.7271	-0.4829
Jan-11	-0.92	-0.92727	0.007274
Feb-11	-0.83	-0.81797	-0.01203

Mar-11	-0.69	-0.72397	0.033968
Apr-11	-0.42	-0.62956	0.209563
May-11	-0.37	-0.44897	0.078973
Jun-11	-0.69	-0.37869	-0.31131
Jul-11	-1.86	-0.56249	-1.29751

5 CONCLUSION

The PDO models are accurately developed in the time domain analysis. The projection methods of time domain analysis help to interpret the variation and behavior of PDO series as being dependent on time, where the value of a current observation is determined by a regression on past values. Time domain analysis used in this research is defining into exponential smoothing analysis (Single and Holts Double exponential smoothing model) and Box-Jenkins analysis. The best selected model developed in this research is ARIMA (2, 1, 1) or X_t= 1.574* Z_{t-1}-0.427* Z_{t-2}-0.147* Z_{t-3}-0.976* a_{t-1} which means that the prediction of the PDO index variation in the future is depending on three months earlier data and a month earlier error of PDO index. The reasons of this determination are based on statistical performance showed by the model through the best assessment of Mean Absolute Error (MAE), Correlation (r) and Root Mean Square Error (RMSE) to each of models. This model has capability to produce the reliable prediction with minimum MAE (0.5293) and RMSE (0.6663). The accuracy of the fitting model is 76%. In addition, the ARIMA (2, 1, 1) model only needs few parameters to predict the PDO index in the future than the ARIMA (9, 7) model. However, this model still has a weakness where the resulting error is still fluctuating and the accuracy of the predicted results is very dependent on the quality of the previous data (historical data). This model also has only a good skill to predict the data for the next few months and the quality is decreased when used to predict the PDO index for a long term. The authors highly recommend to use the others statistical methods (e.g. adding seasonal effect in the ARIMA model or multivariate methods) and dynamical methods as the comparative models in order to develop the PDO model with higher accuracy value. The authors expect that this paper can especially explain the implementation of Box Jenkins methods in the development of PDO model.

6 ACKNOWLEDGMENT

Several of the ideas presented in this paper are extensions of concepts previously considered by Nairn-Birch S. N. 2008. Sandro also acknowledge his gratefulness to Mr. Gonzalo Ovalles, Minitab Pty Technical Support Specialist Australia (Minitab Pty Ltd) for his discussion and valuable suggestion regarding the estimation the back forecast and residuals of the back forecast on determination of PDO index. The anonymous reviewers of this paper provided valuable interpretations, and we truly appreciate them.

REFERENCES

Chatfield C.,1984. *The Analysis of Time Series*: *An Introduction.* London, Chapman and Hall.

Evana L., 2009. *Pengembangan Model MJO Berbasis Data Multivariate RMM1 dan RMM2*. Skripsi. GFM IPB.

Evans, M. N., Kaplan, A., Cane, M. A. and Villalba, R., 2000. *Globality and Optimality in Climate Field Reconstructions from Proxy Data*. p. 53–72. In *Present and Past Interhemispheric Climate Linkages in the Americas and Their Societal Effects*, ed. by V. Markgraf, Cambridge University Press, Cambridge, U.K.

Faqih A, Ribbe, J. and Meinke H., 2008. *Interdecadal Variability of Regional Sea Surface Temperature and Rainfall in the Austral-Indonesian Region.*

Linsley, B. K., Wellington, G. M. and Schrag, D. P., 2000. *Decadal Sea Surface Temperature Variability in the Subtropical South Pacific from 1726 to 1997 A.D.* Science, Vol. 290, 1145–1148.

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. and Francis, R. C., 1997. *A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production*, Bull. Amer. Meteor. Soc., Vol. 78, 1069–1079.

Mantua, N.J., 1999. *The Pacific Decadal Oscillation,* A brief overview for non-specialists, to appear in the Encyclopedia of Environmental Change.

Mantua, N.J. and Hare, S.R., 2002. *The Pacific Decadal Oscillation*, Journal of Oceanography, Vol. 58 (No. 1), pp. 35-44.

National Research Council, 1998. *Decade-to-Century Scale Climate Variability and Change: A Science Strategy*. National Academy Press, Washington D.C., 141pp. (http://www.nap.edu).

Nairn-Birch S. N., 2008. *A Time Series Analysis of the Pacific Decadal Oscillation*, Thesis, University of California: Los Angles.

Willmott, C. J. and Robeson, S.M., 1995. *Climatologicallyaided Interpolation (CAI) of Terrestrial Air Temperature*, Int. J. Climatol., Vol. 15, 221–229.

Zhang, Y., Wallace, J.M. and Battisti, D.S. 1997. *ENSO-like Interdecadal Variability*. J. Climate, Vol. 10, 1004–1020.