SOFTWARE DESIGN TO RISK ANALYSIS OF PADDY PRODUCTION WITH ENSO INDICATORS
(CASE-STUDY: EAST JAVA)

Nisa Miftachurohmah[1], Imam Mukhlash[2], Sutikno[3]

[1,2]*Department of Mathematics, Sepuluh Nopember Institute of Technology Surabaya*
[3]*Department of Statistics, Sepuluh Nopember Institute of Technology Surabaya*

Abstract

Indonesia known as agrarian country, it caused by most its people are farmers. The level of paddy production in East Java was highest compared to other provinces. However, its level of production has decreased in recent years. This was caused by the lack of knowledge about risk management in paddy production. One of knowledge should be known is the change of seasons. In fact, the change of seasons was difficult to predicted due to global climate disruption. In previous study, ENSO was used to climate forecasts for decision-making in agriculture. A risk assessment framework could be used in policy decision-making processes. In this paper, we discuss about architecture design for decision support system (DSS) to risk of paddy production in East Java with ENSO indicators. ENSO indicators are SST (Sea Surface Temperature) anomalies, as predictor variables. Harvested area in East Java as respons variable. Predictor and respons variables is identified their correlation using Copula correlation. Estimation of harvested area is constructed using Robust regression model. The prediction level of risk of paddy production is determined using confidence interval (CI). The last, we construct DSS components, database systems, and user interface design.

Key words: Decision Support Systems (DSS), agriculture, ENSO, copula correlation, robust regression

INTRODUCTION

Tropical country like Indonesia has two seasons, the dry season and the rainy season. The ground state in Indonesia is very suitable for agriculture. The most people of Indonesia are farmers, so that Indonesia is known as an agrarian country. The level of paddy production in East Java was highest compared to other provinces. On the other hand, its level of production was unstable in 2000 to 2013, even it has decreased in recent years (BPS, 2013). This was caused by the lack of knowledge about risk management in paddy production. One of knowledge should be known is the change of seasons. The El-Nino Southern Oscillation (ENSO) phenomenon results from the two-way interaction between ocen and atmosphere in the tropical Pasific Ocean (Podesta, 2002). Podesta also has explain that the phenomenon involves two extreme phases: warm events known as El-Nino, and cold events known as La Nina.

A DSS can be used to help policy makers to decisions making their policy in agriculture. So that, a software design to DSS is very important before implement it. In previous study, ENSO was used to climate forecasts for decision-making in agriculture (Podesta, 2002). A risk assessment framework could be used in policy decision-making processes (Qian, *et.al*, 2009). Maulidiyah (2014) has studied about DSS to risk of paddy production in East Java using rainfall data and GCM data. In her study, Maulidiyah has predicted risk of paddy production for long-term. However, in fact, the change of season is difficult to be predicted due to global climate

disruption. The climate disruption is related with rainfall impact on agricultural activity, especially paddy production, began from paddy planting until harvest time. The indicators that are sensitive to extreme circumstances which impact on the weather and climate ENSO indicators. The indicators are SST (Sea Surface Temperature) anomalies.

The overall goal of this paper are how algorithm to modelling data set for risk analysis of paddy production; and the architecture design for DSS to risk analysis of paddy production in East Java using ENSO indicators.

RESEARCH METHODS

In this study, we use data of harvested area from Department of Agriculture East Java (*Dinas Pertanian Propinsi Jawa Timur*). We use five areas in East Java as case study area, those are Banyuwangi, Jember, Ngawi, Lamongan and Bojonegoro. We also use SST anomalies data set; SST anomaly area 1+2, 3, 4 and 3.4. Those data set obtained from National Oceanic and Atmospheric Administration (NOAA).

Stages for the architectural design of decision support systems include: identification correlation between harvested area and SST anomalies using Copula correlation, constructing estimation of harvested area of paddy using robust regression model, constructing components of DSS.

a.   Copula

Copula Gaussian, known as Copula Normal, obtained by transformation from random variables to multivariate standard normal distribution.

$$C_X(u_1, u_2, \dots, u_m, w) = F_{\mathcal{N}(0,\Sigma)}\left(F_{\mathcal{N}(0,1)}^{-1}(u_1), F_{\mathcal{N}(0,1)}^{-1}(u_2), \dots, F_{\mathcal{N}(0,1)}^{-1}(u_m), F_{\mathcal{N}(0,1)}^{-1}(w)\right) \quad (1)$$

then density of Copula Gaussian as follows:

$$c_{(X_1,X_2,\dots,X_m,Y)}(u_1, \dots, u_m, w)$$
$$= \frac{\partial}{\partial u_1} \cdot \frac{\partial}{\partial u_2} \cdots \frac{\partial}{\partial u_m} \cdot \frac{\partial}{\partial w} \cdot C_X(u_1, u_2, \dots, u_m, w)$$
$$= \frac{f_{\mathcal{N}(0,\Sigma)}(F_{\mathcal{N}(0,1)}^{-1}(u_1), F_{\mathcal{N}(0,1)}^{-1}(u_2), \dots, F_{\mathcal{N}(0,1)}^{-1}(u_m), F_{\mathcal{N}(0,1)}^{-1}(w))}{f_{\mathcal{N}(0,1)}(F_{\mathcal{N}(0,1)}^{-1}(w)) \prod_{j=1}^{m}(f_{\mathcal{N}(0,1)}(F_{\mathcal{N}(0,1)}^{-1}(u_j)))} \quad (2)$$

Copula Archimedean has distributions with tail dependencies shows the conditional probabilities in the extreme regions. If $\phi: [0,1] \to [0,\infty)$ is continues, decreases, and a convex function, then $\phi(0) = \infty$ and $\phi(1) = 0$. $\phi^{-1}: [0,\infty) \to [0,1]$ is inverse of ϕ. Copula Archimedean function $C: [0,1]^m \to [0,1]$ can be written in equation (Nelsen, 1998 dalam 'Anisa', 2015):

$$C(u_1, u_2, \dots, u_m) = \phi^{-1}\{\phi(u_1) + \phi(u_2) + \dots + \phi(u_m)\} \quad (3)$$

with $\phi: [0,1] \to [0,\infty)$ is a perfect monotone, as follows:

$$(-1)\frac{\partial^k}{\partial u^k}\phi^{-1}(u) \geq 0, \qquad k \in \mathrm{N} \quad (4)$$

Generator of Copula C is ϕ and its parameter is θ (Kpanzou, 2007). Several types of Copula Archimedean families namely: Copula Clayton, Gumbel, and Frank along with their generator are presented in Table 1.

Table 1. Family of Copula Archimedean

Copula	Generator $\phi(u)$	Copula Bivariat $C(u_1, u_2)$
Clayton	$\dfrac{u^{-\theta} - 1}{\theta}, \theta \in (0,\infty)$	$\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-\frac{1}{\theta}}$
Gumbel	$(-\log(u))^{\theta}, \theta \in [1,\infty)$	$\exp\left\{-\left[(-\log(u_1))^{\theta} + (-\log(u_2))^{\theta}\right]^{\frac{1}{\theta}}\right\}$
Frank	$\log\left(\dfrac{e^{\theta u} - 1}{e^{\theta} - 1}\right), \theta \in R \setminus \{0\}$	$\dfrac{1}{\theta}\log\left(1 + \dfrac{(e^{\theta u_1} - 1)(e^{\theta u_1} - 1)}{(e^{\theta} - 1)}\right)$

(*Source*: Kpanzou, 2007)

b. Robust Regression

Regression analysis is a model that describes the relationship between the predictor variable (X) and the response variable (Y), such as the following equation:

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i \tag{5}$$

with β is regression parameter and ε is residual with asumption $\varepsilon \sim IIDN\,(0,\sigma^2)$.

The estimation of equation (5) can be expressed in the equation:

$$\widehat{Y_i} = b_0 + \sum_{j=1}^{k} b_j X_{ij} \tag{6}$$

with $\hat{Y}$ is the estimation of response variable, b is the estimation of regression parameter. For equation (5) and (6), $i = 1,2,\dots,n$ with n is number of observation; $j = 1,2,\dots,k$ with k is number of predictors variables.

M-estimation introduced by Huber (1973) and is known as a simple approach to robust regression. The basic idea of the M-estimation is to minimize the objective function in equation (7).

$$\sum_{i=1}^{n} \rho(e_i^*) = \sum_{i=1}^{n} \rho(e_i/\hat{\sigma}) = \sum_{i=1}^{n} \rho((y_i - x_i b)/\hat{\sigma}) \tag{7}$$

The value of $\hat{\sigma}$ (robust estimation scale) obtained by iteration.

$$\hat{\sigma}^{(j)} = \frac{med_{i=1}^{n}\left|y_i - x_i b^{(l-1)}\right|}{\beta_0} \tag{8}$$

with $j = 1,2,\dots$ is the iteration, Φ^{-1} is the inverse of the standard normal cumulative function, $\rho(e_i^*)$ is a symmetric function of residual or functions that contribute to each of the residuals in the objective function. The influence function $\psi = \rho'$ to minimize equation (7) is calculated as the value of the first partial derivatives to 0. It can be shown equation (9).

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - x_i b}{\hat{\sigma}}\right) x_i = 0 \tag{9}$$

Define the weight function $w_i = \psi(e_i^*)/e_i^*$, then the equation (9) becomes equation (10).

$$\sum_{i=1}^{n} w_i\left(\frac{y_i - x_i b}{\hat{\sigma}}\right) x_i = 0 \tag{10}$$

Equation (10) is denoted to a matrix, known as *weighted least squares* (WLS) and presented to equation (11).

$$X^T W X b = X^T W y \tag{11}$$

with X is a matrix for predictor variables, y is a matrix for response variables, b is a matrix for regression parameters, and W is a matrix for weights.

WLS can be used to calculate M-estimation, then the estimation of parameter equation becomes:

$$b = (X^T W X)^{-1} X^T W y \tag{12}$$

Table 2. The Objective Function, The Influence Function,
The Weight Function in M-Estimation Tukey Bisquare

Tukey Bisquare	
The Objective Function	$\rho_B(e^*) = \begin{cases} \frac{k^2}{6}\left[1 - \left\langle 1 - \left(\frac{e_i^*}{r}\right)^2\right\rangle^3\right] & untuk \\ r^2/6 \quad untuk \quad \left|e_i^*\right| > \end{cases}$

The Influence Function	$\psi_B(e^*) = \begin{cases} e_i^* \left\langle 1 - \left(\frac{e_i^*}{r}\right)^2 \right\rangle^2 & untuk \quad \mid \\ 0 & untuk \quad \left\lvert e_i^* \right\rvert > r \end{cases}$
The Weight Function	$w_B(e^*) = \begin{cases} \left\langle 1 - \left(\frac{e_i^*}{r}\right)^2 \right\rangle^2 & untuk \quad \left\lvert e_i^* \right\rvert \\ 0 & untuk \quad \left\lvert e_i^* \right\rvert > r \end{cases}$

(*Source* : Fox, 2002)

c.   *Confidence Intervals*

Confidence intervals (CI) is the prediction intervals accompanied by the level of confidence. How to calculate the confidence interval with 90% opportunities confidence shown in equation (13).

$$CI = \overline{X} \pm 2s \qquad (13)$$

where $\overline{X}$ is mean of population and s is the standard deviation that presented to equation (14).

$$s = \frac{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}}{n-1} \qquad (14)$$

where $i = 1, 2, \dots n$ and is number of observations; X_i is i-th population.

d.   *Components of DSS*

A DSS application can be composed of the subsystems (shown in Figure 1), namely: Data-management subsystem, it includes a database contains relevant data for the situation and is managed by softwaer called the database management system (DBMS); Model management subsystem, this is a software that includes financial, statistical, management science, or quantitative models that provide the system's analytical capabilities and appropriate software management; User interface subsystem, the user communicates with and commands the DSS through this subsystem. The Web browser provides a familiar, consistent graphical user interface structure for most DSS; and Knowledge-based management subsystem, this subsystem can support any of the other subsystems or act as an independent component. It provides intelligence to augment the decision-maker's own (Turban, 2007).
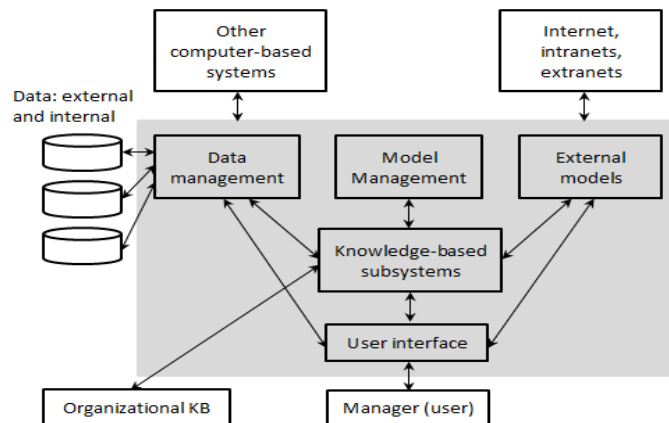
Figure 1. A schematic view of DSS (Turban, 2007)

RESULT AND DISCUSSION

The data set is going to be estimated using regression model should be have predictor data and response data. In this study, we use data set of SST anomalies as predictors and data set of the data harvested area as a response. Predictors and responses must have good correlation, as an initial condition prior to estimation. If the correlation value is not qualified, it is not feasible
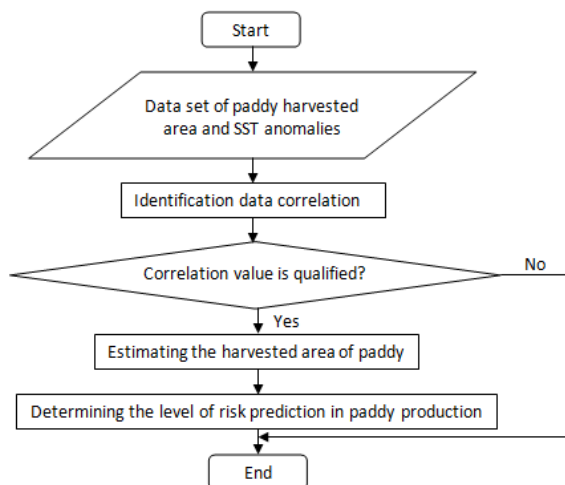
to be estimate.

Figure 2. Flowchart of identification the correlation and estimation

a.   *Identification of the Relationship Between Harvested Area of Paddy and ENSO indicators*

      The data set of harvested area of paddy and SST anomalies tested their normality using Komolgorov-Smirnov dan Anderson Starling. If the results of the normality of one or both methods are declared normal, then we calculate the correlation using Copula Gaussian followed by Copula Archimedean. However, if the results of the normality of the two methods is not normal, we calculate the correlation only using Archimedean Copula. The parameter of estimation in Copula Gaussian using Pearson approach, in Copula Archimedean using Tau-Kendall approach. Maximum Likelihood Estimation (MLE) is used to fitting and choosing copula with the best correlation results.

      In general, the stages that must be passed in the identification of the relationship between the area harvested and ENSO indicators can be seen in Figure 3.
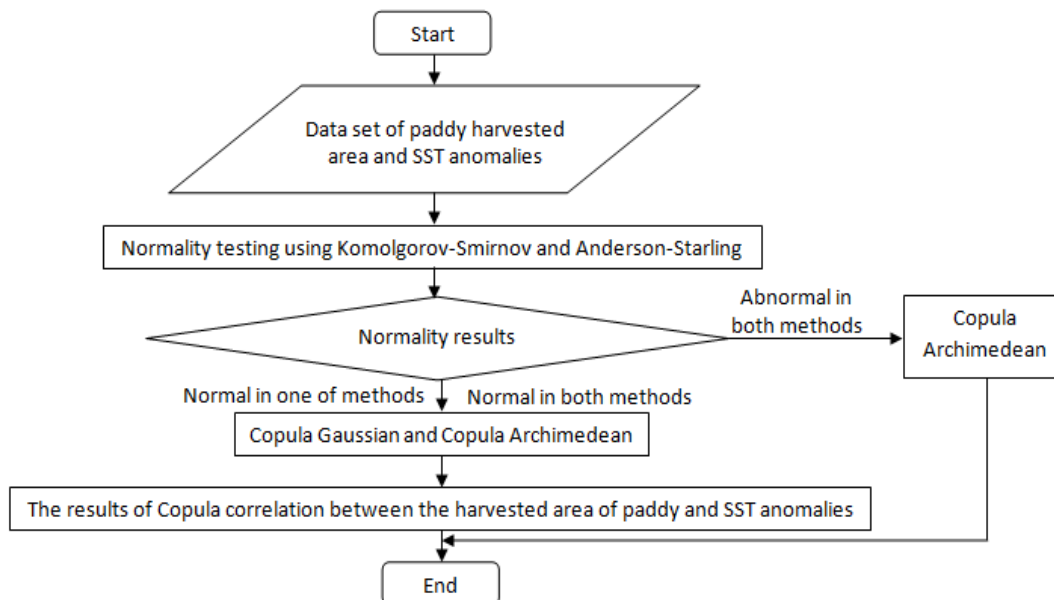

Figure 3. Flowchart of identification the correlation using Copula

b.   Estimation Modeling of the Harvested Area of Paddy Using Robust Regression
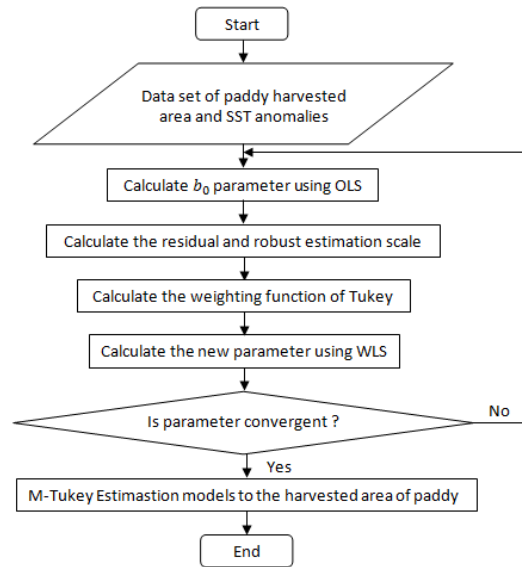

Figure 4. Flowchart of robust regression using M-estimation algorithm

In this study, we use M-estimation methods to estimation modelling in regression robust models. Based on regression equation generally, it always exist the parameter b_0 that calculated using OLS. After b_0 known, then we calculate the residual and robust estimation scale. The next step, we calculate the weight function of Tukey, and then we calculate the new parameter of estimation using WLS. The new parameter of estimation calculated iteratively into the value of parameter of estimation is convergent.

c.   Determining the Level of Risk Prediction in Paddy Production

Paddy production calculated for each period, i.e the first period (January-April), the second period (May-August) and the third period (September to December). In this study, risk level of paddy production also determined for each period (short-term). Flowchart of detemining the level of risk shown in Figure 5, where $\ell = 1,2,3$ are the period of harvested area of paddy.
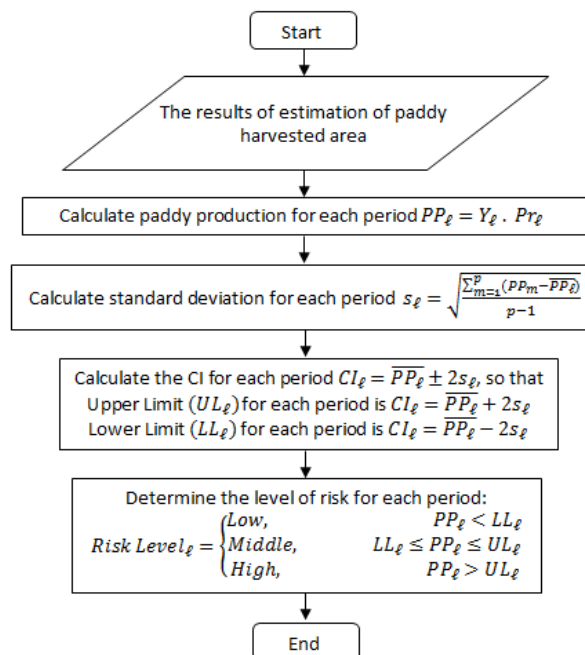

Figure 5. Flowchart of detemining the level of risk

d.   The model of DSS Components

Turban (2007) have explained that there are four fundamental components in DSS, namely: data-management subsystem, model management subsystem, user interface subsystem and knowledge-based management subsystem. Therefore, we follow the architecture, so that we purpose the architecture design for DSS to risk analysis of paddy production shown in Figure 6.
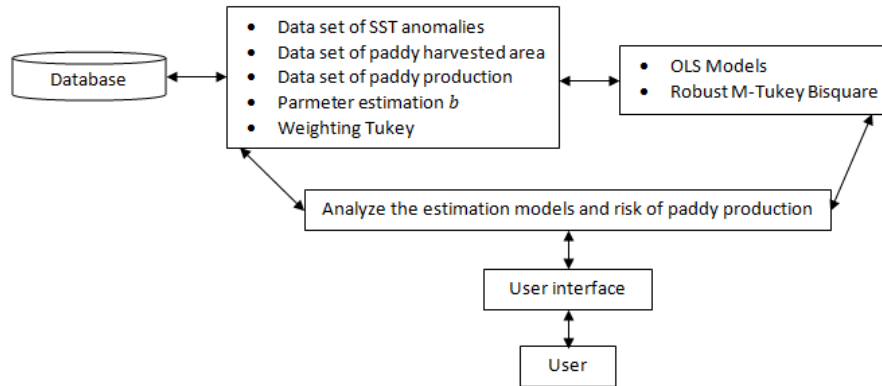

Figure 6. DSS models to risk of paddy production with ENSO indicators

All of data set related with DSS saved in database. Data management subsystem includes: input data set composed of data set of paddy harvested area, SST anomalies and paddy production; estimation and weight data obtained by estimation process using robust regression. Model management subsystem includes: initial estimation model using OLS, then continued using the estimated M-Tukey Bisquare. Knowledge-based subsystem includes: we analyze and determine the level of risk estimates paddy production using CI. The last, user interface subsystem includes: the results of the data information shall be transmitted to end user, and we will represent the result by thematic map.

CONCLUSION AND SUGGESTION

In this study, we have discussed the algoritm to identification the relatioship between harvested area and SST anomalies data; and estimation of harvested area. We also have discussed the architectural design of DSS to risk analysis of paddy production in East Java. We disucussed components of DSS includes: data management subsystem manage input data set composed of data set of paddy harvested area; SST anomalies 1+2, 3, 4 and 3.4; and paddy production, estimation and weight. In model management subsystem use OLS models and M-Tukey Bisquare. Analyzing and determining the level of risk estimates paddy production using CI in knowledge-based subsystem. The results are three levels of risk i.e: low, medium and high.The last, user interface subsystem transmit the data information to end user by graphical user interface, and represent the result by thematic map. In the next study, we continue to discuss about implementation of the software design to risk analysis of paddy production in East Java with ENSO indicators.

REFERENCES

Badan Pusat Statistik (BPS). (2013). *Data Tanaman Pangan – Luas Panen, Produksi dan Produktivitas Padi Provinsi Jawa Timur*, website: http://www.bps.go.id/tnmn_pgn.php.

Fox, John. (2002). Robust Regression, Appendix to An R and S-PLUS Companion to Applied Regression,   accessed   on   Februari   24th,   2015,   *http://*cran.r-project.org/doc/...Companion/appendix-robust-regression.pdf.

Kpanzou, T. A. (2007). Copulas in Statistics, *African Institute for Mathematical Science* (AIMS).

Maulidiyah, Ratna. (2014). Web-based Spatial Decision Support Systems to Predict Risk Level

of Rice Production in East Java, *Surabaya: Department of Mathematics, Sepuluh Nopember Institute of Technology*.

National Oceanic and Atmosphere Administration (NOAA). 2015. http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices, acceessed on Desember 15th, 2014.

Paramitha O., Pratnya. (2012*). A Copula Approach to Construct Vulnerability Rice Puso Maps in East Java with El-Nino Southern Oscillation (ENSO) Indicator, *Surabaya: Department of Statistics, Sepuluh Nopember Institute of Technology.*

Podesta, Guillermo., et. al. (2002). Use of ENSO-related climate information in Agricultural Decision Making in Argentina: A Piot Experience. *Agricultural Systems, Vol. 74, Page. 371-392*.

Qian, *et.al*. (2009). Interactions among ENSO, the Monsoon, and Diurnal Cycle in Rainfall Variability over Java, Indonesia, *American Meteorological Society, Vol. 67, Page. 3509-3524*.

Roberts, *et. al*. (2008*). El-Nino Southern Oscillation Impacts on Rice Production in Luzon, Philipines, *Journal of Applied Meteorology and Climatology, Vol. 48, Page. 1718-1724.*

Turban, Efraim, dan Aronson, Jay E. (2007). Decision Support Systems and Intelligent Systems, *New Jersey: Prentice-Hall International, Inc*.

Xiangzheng, *et. al*. (2010), Impact of El Nino-Southern Oscillation Events on China's Rice Production, *Journal of Geographical Sciences, Vol. 20, Page. 3-16*.

Zubair, Lareef. (2002). El Nino-Southern Oscillation Influences on Rice Production in Sri Lanka, *Intenational Journal of Climatology, Vol. 22, Page. 249-266*.