Parameter: Journal of Statistics
Volume 1 Nomor 1 Tahun 2021

CROSS-CORRELATION ANALYSIS BETWEEN SEA SURFACE TEMPERATURE ANOMALIES AND SEVERAL CLIMATE ELEMENTS IN THE INDIAN OCEAN

Fanny Oktaviani[1], Miftahuddin[2*], Ichsan Setiawan[3]
[1,2]Department of Statistics, Faculty of Mathematics and Sciences, Syiah Kuala University
[3]Department of Marine Sciences, Faculty of Marine and Fisheries, Syiah Kuala University

***e-mail:** *miftah@unsyiah.ac.id*

ABSTRACT

Climate change can create a considerable impact in Indonesia. Aceh province is a province located on the island of Sumatra and it are located around in the Indian Ocean. Aceh Province has a considerable impact of climate change caused by the Sea Surface Temperature Anomalies (SSTA). The SSTA in the Indian Ocean is a parameter that can affect climatic conditions in Indonesia. The SSTA changes can cause an extreme climate change on earth. There are several climate elements affected by SSTA including air temperature, rainfall, wind speed, solar radiation, and relative humidity. One of the methods used to look at SSTA's relationship with some climate elements is the Cross-Correlation method. The climate data used in this study was a daily time series data. The purpose of this study is to find out SSTA's relationship with some climate elements. The results showed that using the Pearson correlation, the highest relationship was SSTA and the air temperature was 0.45. Meanwhile, the lowest relationship was SSTA and the rainfall was -0.05. Similarly, the Cross-Correlation method where the highest relationship was SSTA and the air temperature was 0.469, and the lowest close relationship was SSTA and the rainfall was -0.075.

Keywords: *SSTA, cross-correlation, climate elements*

INTRODUCTION

Climate change is a natural phenomenon that can give a considerable impact on the life of the earth. Future climate change impacts can occur in form of floods, tropical storms, prolonged forest fires, or droughts in some regions. The ocean plays an important role in climate change and can affect the movement and circulation of the atmosphere in a certain region. This is because 71% of the Earth's surface is the ocean, and 97% of the water on the earth is seawater (Reid et al., 2009).

Geographical location greatly affects climatic conditions in Indonesia, such as Sea Surface Temperature Anomalies (SSTA)'s conditions in the Indian Ocean. SSTA is a phenomenon of rising or decreased average Sea Surface Temperature (SST). In this case, Aceh province is a province located on the island of Sumatra and it are located around in the Indian Ocean. Thus, Aceh Province has a considerable impact of climate change caused by SSTA in the Indian Ocean.

Changes in SSTA will cause changes in other climatic elements such as air temperature, precipitation, wind speed, solar radiation, and humidity (Miftahuddin, et.al, 2014). Therefore, analysis is needed to determine the relationship between SSTA and climate elements. The method used in this study to look at SSTA's relationship with some climate elements was the Cross-Correlation method. This method is perfect in viewing relationships between two variables in the form of time-series data.

MATERIALS AND METHODS

1. Sea Surface Temperature Anomalies (SSTA)

SSTA is a phenomenon of rising or de decreased sea surface temperature (SST) conditions in a region. SSTA is an important climate indicator in a region because the information contained in the SST data can indicate and suspect the climate on earth (Reynolds et al., 2001).

The status of the earth's atmosphere and surface greatly affects the climatic conditions in Indonesia, such as SSTA conditions in the Indian Ocean. SSTA in the Indian Ocean region is often referred to as the Indian Ocean Dipole (IOD) or Dipole Mode Index (DMI) (Pramudia et al., 2015). The IOD phenomenon occurs due to the very strong interaction between the ocean and the atmosphere in the equatorial region of the Indian Ocean. IOD affects the climate in Indonesia and other countries around the Indian Ocean basin (Saji et al., 1999). As a result of the IOD phenomenon, there are at least some potential disasters that threaten Aceh Province as Indonesia's westernmost province, including landslide, flood on the east and west coast, flash flood, strong wind, lightning, and high wave.

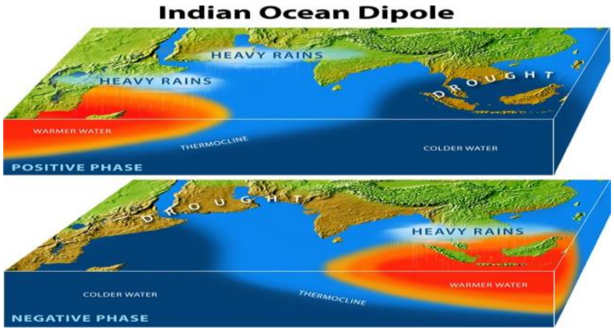

Figure 1. Phenomenon Indian Ocean Dipole (IOD)

IOD is divided into two types, positive IOD and negative IOD. Positive IOD occurred when the west coast of Sumatra experienced an increase in SSTA, while the eastern part of the African Continent experienced a decline in SSTA. Thus, there was a shift in airflow from the western part to the Eastern part of the African Continent that led to the formation of conclusive clouds in the African region and resulted in above-normal rainfall. Meanwhile, in the western region of Sumatra, there is a drought due to a decrease in rainfall intensity.

Similarly, negative IOD occurred when the eastern part of the African Continent experienced an increase in SSTA while in the western part of Sumatra experienced a decrease in SSTA. The transfer of airflow from the eastern part of Africa to the western part of Sumatra led to cloud formation in the western part of Sumatra. This event led to the high rainfall intensity in the western part of Sumatra and drought in the eastern part of the African Continent (Hermawan and Komalaningsih, 2008). SSTA is calculated based on the result of a reduction between the actual data of the SST and the average value of the SST of the place in question. The SSTA equation is as follows:

$$SSTA = SST_a - SST_m \quad (1)$$

SSTA : SST Anomaly value in the relevant place (ºC)
SST_a : actual SST value (ºC)
SST_m : average SST value (ºC)

2. Correlation Analysis

The nature of dependency between one variable and another is already an issue that must be examined. Correlation analysis is one of the tests in statistics to find out the relationship and measure the strong relationship between two variables (Nugroho et al., 2008). The index used to measure the closeness of relationships between two variables is called the correlation coefficient.

In parametric statistics, the correlation coefficient between two variables is obtained by using Pearson correlations that are notified with r. where the lowest scale of observation on Pearson correlation is interval and ratio. If the observation scale is ordinal, nonparametric correlation tests can be conducted with correlation coefficients in the form of Rank Spearman, Kendall, Liliefors, Chi-Square, Phi-Coefficient, Somers, Goodman-Kruskal, and Wilson. And if observations are used in the form of time series data, it can use cross-correlation methods. Pearson's correlation equation can be written as follows:

$$r_{xy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{(n\sum x_i^2 - (\sum x_i)^2) - (n(\sum y_i^2 - \sum y_i)^2)^{1/2}} \quad (2)$$

r_{xy} : correlation coefficient between x and y
x_i : x to i value
y_i : y to i value
n : the number of samples

The strong relationship between response variables and predictors can be measured by the correlation coefficients. The range of correlation coefficient values is -1 to 1 (-1 ≤ r ≤1). If the coefficient is worth -1 or 1, it can be recognized that the relationship between variables is perfect, where events in response variables can be explained by predictor variables without errors. The smaller the correlation coefficient value, the greater the error. Table 1 shows the number of relationships between variables based on criteria (Guilford, 1956):

Table 1. Interpretation of Correlation Coefficient between Variables

No	Correlation Coefficient Value	Description
1	0 < r ≤ 0,2	Relationship are very small or no correlation
2	$0{,}2 < r \leq 0{,}4$	Small relationship/ lack of closeness
3	$0{,}4 < r \leq 0{,}7$	Moderate relationship/ moderate level of closeness
4	$0{,}7 < r \leq 0{,}9$	Close relationship
5	$0{,}9 < r \leq 1$	Very close relationship

3. Cross-Correlation

Cross-correlation has been developed in two different areas, namely signal processing and statistics. In the field of signal processing, the cross-correlation function is used to change one or more signals so that they can be noticeable differences and can produce plots that are useful in identifying hidden signals in data more easily.

The Cross-correlation function is statistically to analyze the relationship between two variables in a system, provided that both variables must be time-series data and assumed to be stationary to mean and variance (Shumway and Stoffer, 2011). Time series data collection by implementing cross-correlation methods must be discrete data and stochastic data. Where discrete data must have the same time interval and stochastic data has different time intervals according to time function (Derrick and Thomas, 2004). The way cross-correlation works are to determine the best time lag at the time of the highest correlation. The optimal time shift is necessary to increase the correlation between the two variables (Hashmi et al., 2009) The greater the correlation the more similar the data patterns of the two variables. Where the time shift is divided into three i.e. positive time shifts which means time will shift to the left (reverse), zero time shift which means no shift, and negative time shift which means time will shift to the right (forward). The basic cross-correlation equations are as follows:

$$r(L) = \frac{\sum_{k=0}^{N-L-1}(x_{(k+L)} - mx)(y_k - my)}{\sqrt{[\sum_{k=0}^{N-1}(x_k - mx)^2][\sum_{k=0}^{N-1}(y_k - my]}} \quad L \geq 0 \quad (3)$$

$$r(L) = \frac{\sum_{k=0}^{N-|L|-1}(x_k - mx)(y_{(k+|L|)} - my)}{\sqrt{[\sum_{k=0}^{N-1}(x_k - mx)^2][\sum_{k=0}^{N-1}(y_k - my]}} \quad L < 0 \quad (4)$$

where:
mx : average rows of data on variable x
my : average rows of data on variable y
r : cross-correlation value
d : the time of the highest correlation between the two variables

4. Procedure

The data used in this study is secondary data. Average daily data with 2948 observations consisting of six variables with a data period from July 5, 2010, to July 31, 2018. The data was obtained from the National Oceanic and Atmospheric Administration (NOAA). The variables used in the study are as shown in Table 2 below.

Table 2. Research Variable

No	Variable	Remarks	Unit
1	Y	Sea Surface Temperature Anomalies	ºC
2	X_1	Temperatures	ºC
3	X_2	Rainfall	mm
4	X_3	Wind Speed	m/s
5	X_4	ShortWave Solar Radiation	w/m^2
6	X_5	Relative Humidity	%

The steps conducted in this study are as follows:
1) Create a statistical summary
2) Analyze the relationship of two variables using Pearson correlation.
3) Create an Autocorrelation Function (ACF) plot
4) Analyze the relationship of two variables using cross correlation

RESULTS AND DISCUSSION

1. Statistical summary

A statistical summary is useful to provide information in the form of data characteristics such as minimum value, maximum value, mean, median, quartile 1, and quartile 3. Here's a summary table of statistics on the variables used in this study.

Table 3. Summary Dataset SST

Summary	SSTA (˚C)	Temperatures (˚C)	Rainfall (mm)	Wind Speed (m/s)	ShortWave Solar Radiation (w/m^2)	Relative Humidity (%)
Min	-1.281	26.01	0.00	0.30	14.73	63.10
Q1	-0.241	28.25	0.00	4.70	193.60	80.20
Median	-0.111	28.42	0.00	5.50	225.57	81.00
Mean	0.000	28.39	3.24	5.49	220.06	80.79
Q3	0.129	28.56	1.44	6.4	259.90	81.80
Maks	2.519	30.61	226.32	11.00	328.72	93.20
Range	3.8	4.6	226.32	10.7	313.99	30.10

Table 3 shows that the lowest SSTA value was -1,281°C and the highest SSTA value was 2,519°C. The lowest SSTA value occurred on December 6, 2010, while the highest SSTA value occurred on April 27, 2016. The lowest air temperature in the Period 2010 to 2018 was 26.01°C, and the highest air temperature reached 30.61°C. The lowest rainfall was 0 mm and the highest rainfall reached 226.32 mm. The lowest wind speed was 0.30 m/s and the highest wind speed reached 11.00 m/s. The lowest solar radiation was 14.73 w/m2 and the highest solar radiation reached 328.72 w/m2. As well as the lowest relative humidity was 63.10% and the highest relative humidity reached 93.20%. Scatterplot Matrix provides information about the form of distribution, relationship patterns, and correlations between two variables, both between response variables and predictor variables and between predictor variables. The following is a scatterplot matrix image that aims to look at histograms, correlations, and regression lines from research data.

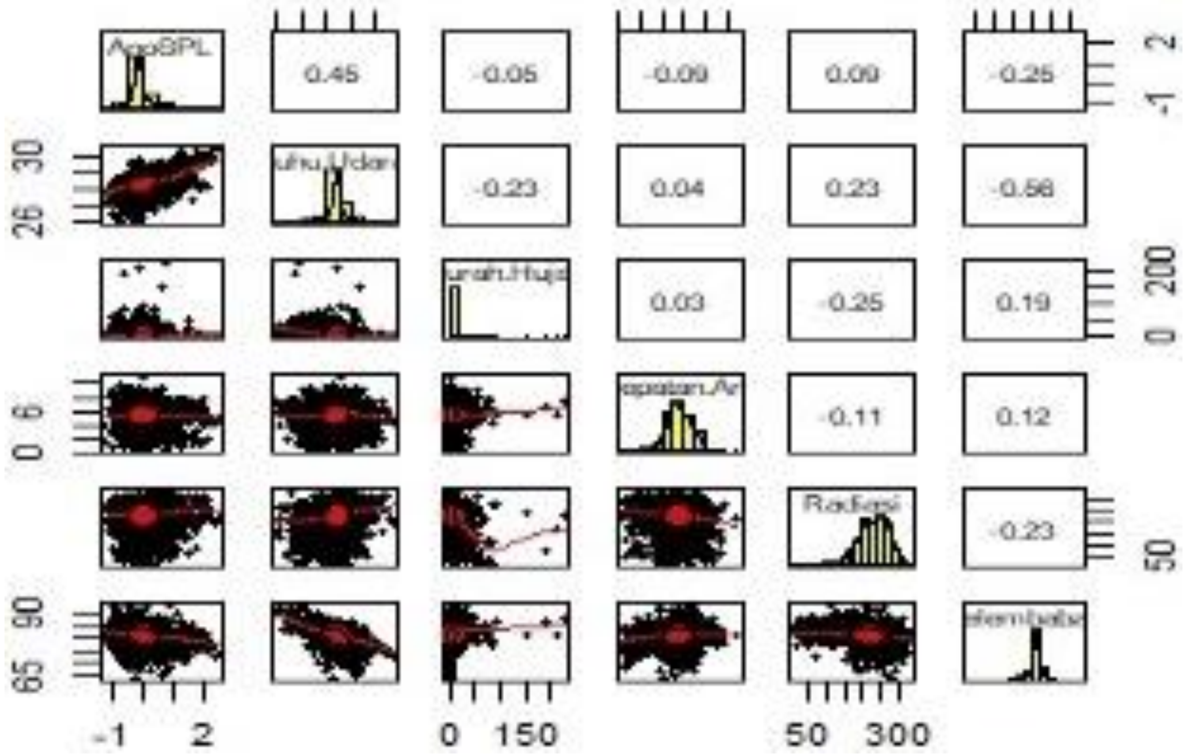

Figure 2. Scatterplot Matrix

Figure 2 is a research matrix scatterplot that can provide information about relationship patterns, distribution forms, and correlations between two variables. Based on the image shows that there are two research variables that have a symmetrical histogram shape, while the other four are not symmetrical. Variables that have symmetrical histogram shapes are X_1 (air temperature) and X_3 (solar radiation) while variables with unsymmetrical histogram shapes are Y (SSTA), X_2 (precipitation), X_4 (solar radiation), and X_5 (relative humidity).

The correlation in the scatterplot matrix above is the Pearson correlation which is to know the pattern of relationship between the response variable (Y) with all predictor variables (X) and also between predictor variables. Based on the correlation value in the matrix scatterplot, it can be seen that overall variables X_1 (air temperature) and X_6 (humidity) have the highest closeness relationship compared to other variables of -0.56. This means that the higher the air temperature, the lower the humidity.

Based on the relationship between response variables and predictor variables obtained that sequentially the highest closeness relationship between Y (SSTA) and X_1 (air temperature) has the highest correlation coefficient of 0.45, meaning the higher the SSTA then the higher the air temperature. Then Y and X_5 (humidity) with a correlation coefficient of -0.25, meaning the higher SSTA the lower the humidity. Y and X_4 (solar radiation) with a correlation coefficient of 0.09, meaning the higher SSTA the higher the solar radiation. Y and X_3 (wind speed) with a correlation coefficient of -0.09, meaning the higher the SSTA the lower the wind speed. Y and X_2 (precipitation) with a correlation coefficient of -0.05, meaning the higher SSTA the lower the rainfall.

2. Cross-Correlation

Cross-correlation is very well used to know the closeness of the relationship between two variables in the form of time-series data and assumed stationary against mean and variance. The collection of time-series data by applying the cross-correlation method must be discrete data that has the same time interval. Correlation values (r) range from -1 to +1, where zero indicates no relationship between the two variables. Here is the value of the cross-correlation coefficient of the SSTA variable with several other variables.

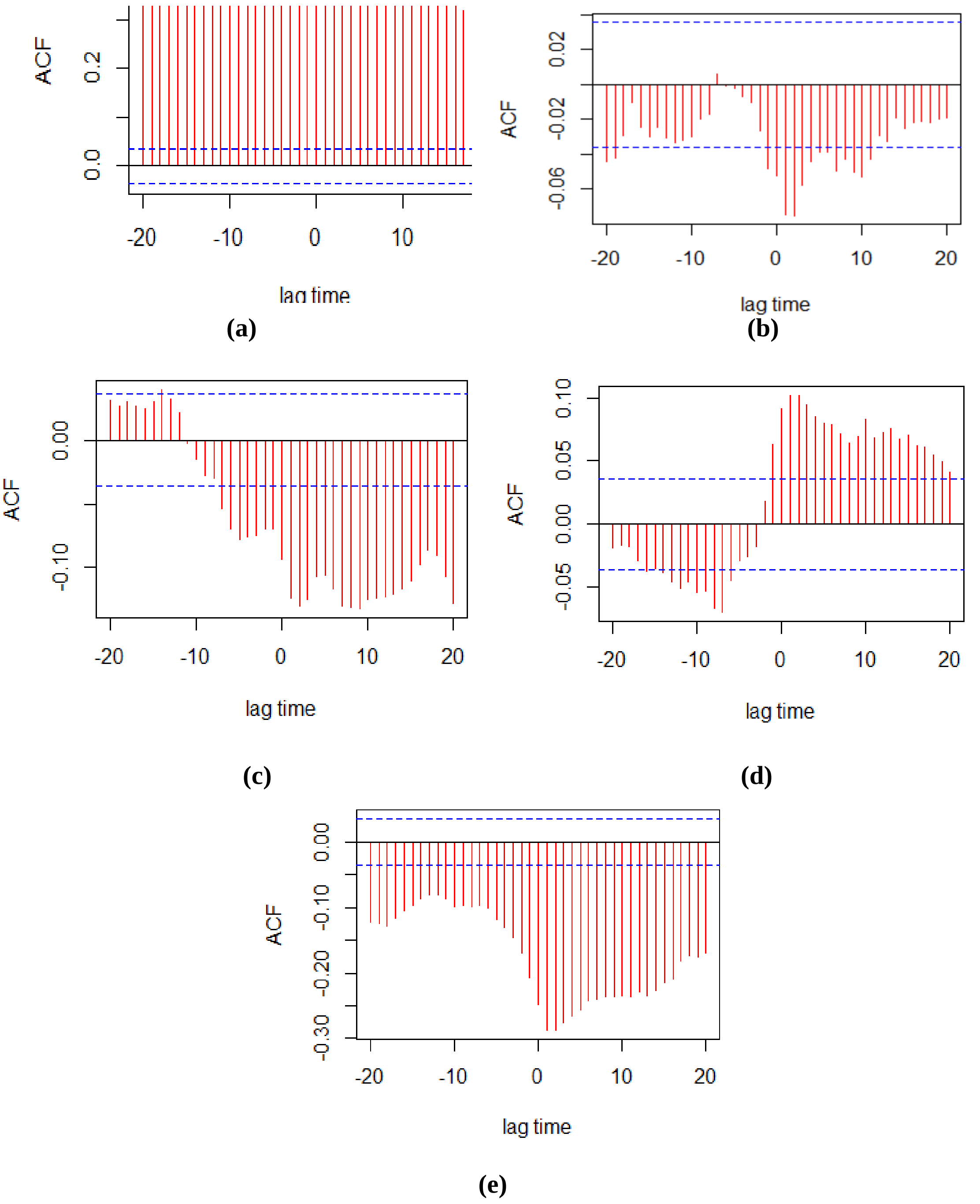

Figure 3. Cross-correlation (a) SSTA and Temperature, (b) SSTA and Air Rainfall, (c) SSTA and Wind Speed, (d) SSTA and ShortWaver Solar Radiation, (e) SSTA and Relative Humidity

Based on Figure 3, it shows the best time shift on variables Y (SSTA) and X_1 (air temperature) occurred at the highest cross-correlation of 0.469 with the 1st positive time shift. It means that the rise or decrease in air temperature would happen one day before the SSTA event. The SSTA increase occurred on April 27, 2016, so the temperature increase would occur on April 26, 2016. Similarly, the lowest SSTA drop occurred on December 6, 2010, so the drop in air temperature would happen on December 5, 2010.

The best time shifts on variables Y (SSTA) and X_2 (precipitation) occur at the highest cross-correlation of -0.075 occurring at the 2nd positive time shift. It means that the increase or decrease in rainfall would occur within two days before the SSTA event. The increase in SSTA occurred on April 27, 2016, resulting in a decrease in rainfall on April 25, 2016. Similarly, the lowest SSTA decrease occurred on December 6, 2010, so the increase in rainfall would occur on December 4, 2010.

The best time shifts on variables Y (SSTA) and X_3 (wind speed) occurred at the highest cross-correlation of -0.132 occurring at the 9th positive time shift. It means that the increase or decrease in wind speed would occur 9 days prior to the SSTA event. The SSTA upgrade occurred on April 27, 2016, resulting in a decrease in wind speeds on April 18, 2016. On the other hand, the lowest SSTA decline occurred on December 6, 2010, so the increase in wind speeds would occur on November 28, 2010.

The best time shifts on variables Y (SSTA) and X_4 (solar radiation) occur at the highest cross-correlation of 0.102 occurring at the 1st positive time shift. It means that the increase or decrease in solar radiation would occur 1 day before the SSTA event. The increase in SSTA occurred on April 27, 2016, so the increase in solar radiation will occur on April 26, 2016. Similarly, the lowest SSTA decrease occurred on December 6, 2010, so the decrease in solar radiation will occur on December 5, 2010.

The best time shifts on variables Y (SSTA) and X_5 (humidity) occur at the highest cross-correlation of -0.287 with the 1st positive time shift. It means that the increase or decrease in humidity would happen one day before the SSTA event. The SSTA increase occurred on April 27, 2016, resulting in a decrease in humidity on April 26, 2016. Similarly, the lowest SSTA decline occurred on December 6, 2010, so the increase in humidity will occur on December 5, 2010.

CONCLUSION

In study, we applied the cross-correlation method for SSTA dataset. By using Pearson correlation obtained that sequentially the highest relationship density is Y (SSTA) and X_1 (air temperature) with a correlation coefficient value of 0.45, Y and X_5 (humidity) of -0.25, Y and X_4 (solar radiation) of 0.09, Y and X_3 (wind speed) of -0.09, and the lowest closeness of the relationship is Y and X_2 (precipitation) of -0.05. Whereas the highest cross-correlation sequentially between Y (SSTA) and X_1 (air temperature) is 0.469 with the 1st positive time shift. Y and X_5 (humidity) have a cross-correlation of -0.287 with the 1st positive time shift. Y and X_3 (wind speed) have a cross-correlation of -0.132 with the 9th positive time shift. Y and X_4 (solar radiation) have a cross-correlation of 0.102 with a 1st positive time shift. Y and X_2 (precipitation) have a cross-correlation of -0.075 with a 2nd positive time shift.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Statistics, Faculty of Mathematics & Sciences, Syiah Kuala University & Directorate of Research and Community Service (or DPRM) Ristekdikti, Ministry of Education and Culture of Indonesia in Campus Intellectual Product Business Development Program. Jakarta. 2020.

REFERENCES

Derrick, Timothy R., and Thomas, Joshua M. (2004). Time Series Analysis: The Cross-Correlation Function. Innovative Analyses of Human Movement. Vol 7: 189-205.

Guilford, J.P. (1956). Fundamental Statistic in Psychology and Education. 3rd Ed. New York: McGraw-Hill Book Company, Inc.

Hashmi,M.Z., Shamseldin, A.Y., and Melville, B.W. (2009). Statistical Downscaling of Precipitation: State of the Art and Application of Bayesian Multimodel Approach for Uncertainty Asessment. Hydrol Earth Sys Sci Discuss.Vol 6: 6536-6579.

Hermawan, E., and Komalaningsih, K. (2008). Karakteristik Indian Ocean Dipole Mode di Samudera Hindia Hubungannya dengan Perilaku Curah Hujan di Kawasan Sumatera Barat Berbasis Analisis Mother. Jurnal Sains Dirgantara. 5(2): 109–129.

Miftahuddin, Andriani,R., Setiawan I., dan Mulsandi, A., 2014, Penerapan Analisis Korelasi Kanonik pada Kajian ENSO dalam Identifikasi Hubungan Fitur Iklim, Jurnal Natur Indonesia, 15, 36-44, ISSN: 1410-9379.

Nugroho, S., Akbar, S., and Vusvitasari, R. (2008). Kajian Hubungan Koefisien Korelasi Pearson (r), Spearman-rho(ρ), Kendall-tau (τ), Gamma (G), dan Somers(dyx). Journal Statistika, 41–54.

Pramudia, A., Estiningtyas, W., Susanti, E., and Sucianti. (2015). Fenomena dan Perubahan Iklim Indonesia serta Pemanfaatan Informasi Iklim untuk Kalender Tanam. Badan Penelitian dan Pengembangan Pertanian, Departemen Pertanian

Reid, C., Marshall, J. Logan, D., and Kleine, D. (2009). Terumbu Karang dan Perubahan Iklim. Terjemahan dari Coral Reefs and Climate Change, oleh Badan Penelitian dan

Pengembangaan Kelautan dan Perikanan & Kementerian Kelautan dan Perikanan Indonesia. Panca Jaya, Jakarta.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., and Wang, W. (2001). An Improved in Situ and Satellite SST Analysis for Climate. Journal of Climate. 15(2): 1-3.

Saji, N. H., Goswami, B.N., Vinayachandran, P.N., and Yamagata, T. (1999). Adipole Mode in the Tropical Indian Ocean. Journal Nature. 401:360-363.

Shumway, Robert H., and Stoffer, David S. (2011). Time Series Analysis and Its Applications with R Examples Third Edition. Springer. USA.