IOP Conference Series: Earth and Environmental Science

PAPER • **OPEN ACCESS**

Modeling of sea surface temperature using linear models with autocorrelation Indian Ocean

To cite this article: M Miftahuddin and Y Ilhamsyah 2018 *IOP Conf. Ser.: Earth Environ. Sci.* **176** 012038

Modeling of sea surface temperature using linear models with autocorrelation Indian Ocean

M Miftahuddin[1*], Y Ilhamsyah[1,2]

[1]Faculty of Mathematics and Science, Syiah Kuala University,
Jln. Teuku Nyak Arief Darussalam, Banda Aceh, Aceh, 23111, Indonesia
[2]Applied Climatology, Graduate School of Bogor Agriculture University, Bogor, Indonesia
*e-mail: miftah@unsyiah.ac.id

Abstract. Sea surface temperature (SST) is one of several indicators of climate system of the Earth. We used model SST to observe SST dataset from buoy arrays in the eastern equatorial Indian Ocean. Relationship between SST and other climate parameters can be represented in linearity approach. This approach shows that temporal variability of the SST as a dominant effect. Linear model fitting (LMF) has been examined with four treatments, with and without: covariate transformation, interaction, centering, and addition time covariate in the model. The LMF chosen as basic construction in the model with covariate interaction combination and transformation, which increases magnitude of multiple-R^2 (56.62%) and adjusted-R^2 (56.13%), i.e. 0.31% and 0.43% respectively. This shows that time covariates have a strong significance effect in the model, compared to continuous covariates. However, the model has autocorrelation, which has large Akaike Information Criterion (AIC) value then this deletion of effects can be done through the autoregressive moving average. Moreover we obtained that LMF which suitable to SST is model with AIC value 403.2987 by using three climate features include two time covariates. Furthermore, we observed that using GAM model fitting showed an increase in explained deviance to 65.90%, a significant decrease in AIC from 678.24 to 634.99 and significant increase in adjusted-R^2 from 51.20% to 64.40% by using sixteen climate features include two times covariates without interaction and transformation.

1. Introduction

Sea surface temperature (SST) is one of many important indicators that indicates/marks the climate system of the Earth. The SST data are useful to early detection climate change and global warming. In fact, the ocean is the particular region of the earth where 70.9% areas and 29.1% is mainland. The largest ocean is the Pacific Ocean (165,759,239.06 km^2) and the second largest is the Atlantic Ocean (106,448,511.33 km^2), and third largest is the Indian Ocean (72,519,667.09 km^2). SST data is essential indicator to recognize climate variability in the earth [1-3], such as El Nino and La Nina phenomena in the Pacific Ocean that affects on dry and wet seasons in Indonesian and its adjacent region.

There are several reasons we modeled the SST data in the eastern equatorial Indian Ocean. First, the observed SST data show complex structure like missing values that represented as gaps which vary among buoys array. Second, in addition Aceh province is located in the western most Indonesia region and it directly connect with Indian Ocean on southern and western waterbody. Third, in 2012 Magnus et all proposed a climate model to investigate the effects of solar radiation and the greenhouse effect on global warming [4]. Their analysis is based on the data from land stations only and does not consider the relationship between sea and land dataset. Hence, the objectives our study, the SST data is used to reveal the relationship among variables in both by using linear and generalized additive model (GAM). Further, firstly, we reviewed the basic construction regarding linear model fitting without and with covariate transformation. Secondly, we used through linear combination of covariates interaction and without/with transformation in the linear model. Thirdly, we applied the second steps with centering approach. Fourth, we developed model by using linear model with consider autocorrelation aspect. Finally, we developed model by using GAM model with the number of extended climate features.

2. Material and Methods

2.1. Materials and data

In this study we used one of moored buoy arrays in the Indian Ocean at 1.5^0N,90^0E, operated by JAMSTEC Japan as part of the Tropical Atmosphere Ocean (TAO) program [5]. The buoy collect real-time SST daily data at 1 m depth from 1 January 2006 to 8 June 2012. The SST dataset consist of 2,066 daily observations with the response variable SST (^{0}C) from 00.00 - 12.00 pm GMT [6].

To show essential time covariates in the linear model, we used *f*(.) function as represented by model with three continuous covariates and two time covariates, i.e. month and year. We assume that three continuous covariates which correspond linearly with SST, i.e. the relationship will be the same for all levels of the time covariate and without interaction between the covariates, as follows [6],

$$SST_i = \beta_0 + \beta_1 Temperature_i + \beta_2 Humidity_i + \beta_3 Rainfall_i + \eta_k Month_i + \gamma_l Year_i + \varepsilon_i \quad (1)$$

for $k = 1, \ldots, 12$; $l = 1, 2, \ldots, 6$, and $i = 1,\ldots,n$, where η and γ are vector parameter of time covariate for month and year respectively. We construct unrestricted model where the model has seasonal effect η_m and annual effect γ_l are restricted to 0. We evaluate several models with different order include time covariate in the model fitting.

Previously, modeled SST data is as linear combination of three parameters of climate features and two time covariates [2]. These parameters are air temperature (^{0}C), relative humidity (%) covariates. They have mean with the same time records on 07.00 WIB, 13.00 WIB and 18.00 WIB, whereas rainfall (mm) during 3-hours period, seasons and annual factors. We extend to SST dataset to buoy position at 4^0N; 90^0E, depth 1 m, on period 2011 to 2015 with 16 climatic parameters, such as wind speed, subsurface temperature, shortwave radiation, salinity, humadity, precipitation, dynamic height, density, air temperature, isoterm, conductivity, zonal and merid current velocity, and time covariates (month and year). The complexity of the linking between the SST and the parameters becomes challenging in construction a dynamic model. Treatments with respect to the model constructed are without and with covariate transformation, without and with covariate interaction, and mixed both mentioned treatment.

2.2. Linear and Interaction Model Fitting

In preliminary modelling, daily SST data are used with linear in parameters assumption that denoted as (x_i, y_i), for $i = 1, 2,\ldots,n$ where x_i as covariates and y_i as response variables. The relationship between the random variables X and Y can be written in the matrix form as [6,7]:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad (2)$$

where $\boldsymbol{Y}$ represents a vector of observations $(y_1,\ldots, y_n)^T \epsilon\ R^n$, $\boldsymbol{\beta}$ represents a vector of parameters $(\beta_0,\beta_1,\ldots,\beta_p)^T \epsilon\ R^{p+1}$, $\boldsymbol{X} \epsilon\ R^{nx(p+1)}$ represents a matrix with rows n and columns $p + 1$ of a set p covariates $X_0, X_1,\ldots, X_p$ of length n including an *intercept* and *errors* ε are assumed *independent and identically distributed*, i.e. normal random variable $\boldsymbol{\varepsilon} \sim N(\mathbf{0_n}, \sigma^2\mathbf{I_n})$, where $\boldsymbol{\varepsilon}$ represents a column vector $(\varepsilon_1,\ldots,\varepsilon_n)^T$ of error and $\mathbf{I}_n$ as the identity matrix. In formulation, for the covariates we can describe as quantitative values, qualitative, transformation, interaction among covariates, and various data types.

Let random variable Y then the probability density function (pdf) of a continuous (or the probability mass function (pmf) for Y is discrete) is referred as a probability distribution and denoted as $f(y; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ represents the parameters of the distribution. The conditional expectation describes the linear or functional relationship of parameter in model,

$$E[Y|X] = \textstyle\sum_{j=0} \beta_j X_j \quad (3a)$$

or

$$\mu_i = E[Y|X_i] = f(X_i),\ i = 1,\ldots, n \quad (3b)$$

Assumption *additivity* with interaction model:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \quad (4)$$

where μ is constant, α_i and β_j are main effects, and $(\alpha\beta)_{ij}$ is interaction effect.

2.3. Generalized Additive Models Fitting

In a generalized additive models, suppose denoted Y is conditionally independent of the covariates x given the additive covariates [6,7,8]

$$Y = \beta + \sum_{j=1} f_j(xj)$$

where $f_j(xj)$ for some functions fj. Let Y and $\boldsymbol{X}$ are random variables that representing response (*output*) and covariate (*input*), respectively then conditional relationship between both the variables can be written as

$$Y = E[Y|X_1, \ldots, X_p] + \varepsilon, \quad (4a)$$

$$\text{where } E[Y|X_1, \ldots, X_p] = \beta_0 + \sum_{j=1}^{p} f_j(X_j). \quad (4b)$$

Thus, model aditive is defined by the following:

$$Y_i = \beta_0 + \sum_{j=1}^{p} f_j(X_{ij}) + \varepsilon_i, \quad i = 1, \ldots, n, \quad (5)$$

where the β_0 is an intercept, f_j is model types, such as linear, nonlinear, smooth function, spatial, interaction, etc, that combining covariate effects. Expectation, variance and covariance ε errors independently of X_j,

$$E[\varepsilon_i] = 0, \text{ and } \mathrm{var}(\varepsilon_i) = \sigma^2, \mathrm{cov}(\varepsilon) = \sigma^2 I_n. \quad (6)$$

GAM model is an extension of the Linear Models (LM) and Generalized Linear Models (GLM) through a link function $g(.)$, with assume that variable response follows several exponential family distributions. In general GAM structure is given as follows:

$$g(\mu) = g(E(Y|X_1, X_2, \ldots, X_p)), \quad (7)$$

In other words, from equation (5) is,

$$f^*(\mathbf{X}) = \beta_0 + \sum_{j=1}^{p} f_j(\mathbf{X_j}). \quad (8)$$

where f^* is response expectation by estimating additive function. From equation (5) we assume $\varepsilon_i \sim N(0, \sigma^2)$, [7]. In GAM model, [7] used loess smoothers or smoothing splines functions, whereas GAM via mgcv package by Simon Wood [9] and the model for large data sets [10]. By using spline regression method, we used knots to minimalize penalize as in the following:

$$\| Y - X\beta \|^2 + \lambda \int f''(x)^2 dx \quad (9)$$

where given λ or through cross validation to select optimal penalty.

3. Results and discussion

3.1. Linear model fitting of the SST dataset

In this section we apply our methodology to SST dataset on several treatments in model fitting.
In preliminary, the linear model is developed by without and with covariate transformation, as follows

M1: $SST = \beta_0 + \beta_1 Temp + \eta_k\, Month + \gamma_l Year + \beta_2 Humd + \beta_3 Rain + \boldsymbol{\varepsilon}$

By analysis of its variance, it is revealed that humidity and rainfall have insignificant effects as shown in table 1. Temperature and time covariates (Month and Year) have p-value <2e-16. It means that three covariates have significant effects to sea surface temperature compared with humidity and rainfall effects by using M1 model. Through linier model (M1) reveals that air temperature around ocean given affect to sea surface temperature condition on the certain month and year, as seen in figures 1 and 2.

Table 1. ANOVA of the M1 model.

Source	df	SS	MS	F-value	Pr(>F)
Temperature	1	83.331	83.331	581.2248	<2e-16 ***
Month	11	236.830	21.530	150.1698	<2e-16 ***
Year	6	57.463	9.577	66.7999	<2e-16 ***
Humidity	1	0.183	0.183	1.2750	0.2590
Rainfall	1	0.002	0.002	0.0132	0.9087
Residuals	2045	293.193	0.143		

By the parameters estimation of the M1 model showed that time effects on September and October have insignificant, whereas for 2011 and 2012 are insignificant effects as well. The model fitting reaches 56.31% (R^2 value). The specific time effects on the SST data can be showed by M1 model as seen in figure 1. Season effect occurs on monthly unit, exclude on September and October. Probably on the both month occurs time transition of the change season for the sea surface temperature phenomenon in the investigated period.

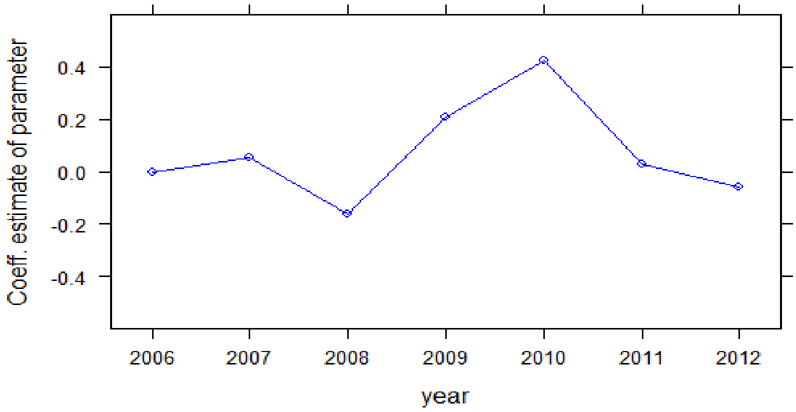

Figure 1. Annual effect of the M1 model.

In figure 1, the annual effect shows slightly increase of annual variation that happened between 2006-2007, rapid decreasing that occurred between 2007-2008, and rapid increasing between 2008-2010. Hereinafter, it decreased sharply to 2011 and slowly decreased between 2011-2012. In general, between 2007-2008 the SST data fitting by using M1 model showed non-positive effect. Likewise, it is between 2011-2012. However, between 2006-2007 and between 2008-2011, it gives positive effect.

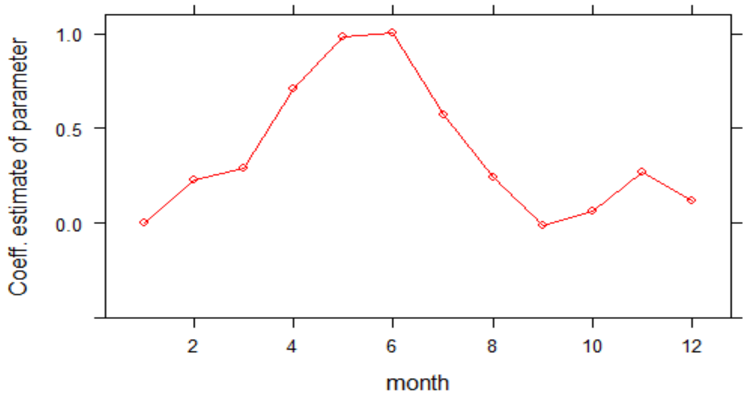

Figure 2. Seasonal effect of the M1 model.

The seasonal effect showed that increase trend that happened from January to June and a decrease trend from June to September (figure 2). Furthermore, a second feeble increase trend appeared again until November and subsequent decrease until December. A peak of the highest month effects occurs in June and lowest in September (figure 2).

Whereas the seasonal effect between January and February occurred rapid increase, and slightly increase from February to March. Further it occurs rapid increase return from March to May. The slightly increase the seasonal effect also occurs again from May to June. However, it rapid decrease happens from June to September. The slightly increase the seasonal effect go on September to October and rapidly increase go on October to November. Finally, it decreasing again occurs from November to December.

3.2. Linear model with treatment of the SST dataset

We know that generalized additive models (GAMs) are an extension function of linear models (LM) and generalized linear models (GLMs). Before our discussed regarding GAMs models fitting for the SST dataset, in this section we explore linear models fitting by using several treatments. Each treatment can be constructed as a M model and symbolized as M1, M2, M3, …, M15 models. The main treatments are transformation, interaction, and centering.

3.2.1. Linear model fitting with covariate transformation. We constructed M2 model as seen in M1 model with covariate transformation. It is used Rain = log(RAIN+0.01). By analysis of variance process it revealed that humidity and rainfall does not significant effects. The transformation rainfall covariate did not change fitting magnitude on model in the multiple R^2. However, it changed on adjusted R^2 i.e., 0.01%. This is caused by continuous covariate effect which is very small, compared to time covariate effect in the linear model fitting (see M1 and M2 models).

3.2.2. Linear model fitting with covariate interaction and without transformation. Effect of interaction modeling on the continuous covariate as in M1 and M2 models is developed to the M3 model. Result of ANOVA analysis for this model showed that humidity, rainfall and all interactions of continuous covariates did not have significant effect. Interaction effect in the model without transformation provided the multiple R^2 increasing on 0.06%. Although magnitude of adjusted R^2 is stable on 55.88% (see R^2 values pre and post interaction in the M2 and M3 models).

3.2.3. Linear model fitting with covariate interaction and transformation. In this section, we can be expressed again M3 model as M4 model with rainfall transformation. From ANOVA table knows that humidity, rainfall and interaction of continuous covariate does have insignificant effect, exclude interaction between humidity and rainfall covariates gives strong significant effect. It is also for temperature, Month and Year covariates. By parameter estimation of the M4 model shows the combination effect between interaction and transformation in the M4 model gives changing multiple R^2 and adjusted R^2 magnitudes on 0.31% and 0.43% respectively (see R^2 value on M1 and M4 models).

3.2.4. Linear model fitting with covariate interaction and without transformation and centering. Analog M3 model, we construct M5 model without centering. By analysis of variance as seen in the ANOVA table shows that humidity, rainfall, and all interaction of does have insignificant effect of treatments without transformation and centering. The centering effect in the interaction model does not changes R^2 value. This is caused by there is domination effect of time covariate bigger than continuous covariate effect (see R^2 value in the M3 and M5 models).

3.2.5. Linear model fitting with covariate interaction, transformation and centering. In this section, analog the construction of model M5 we developed the M6 model with transformation. Through ANOVA table knows that humidity, rainfall and several interaction of continuous covariate does not have significant effect, exclude interaction between humidity and rainfall covariate gives strong significant effect. It shows that treatment by interaction with transformation more effect than only with centering without transformation. Treatment for covariate interaction, transformation of continuous covariate and centering in the M6 model interaction gives effect with respect to R^2 values, multiple-R^2 and adjusted-R^2 respectively on 0.18% and 0.12% (see R^2 values on the M1 and M6).

3.2.6. Rainfall transformation and interaction of time covariates. The construction M7 model can be written, as follows SST = (Temp + Humd + Rain)*Month + (Temp + Humd + Rain)*Year. Furthermore, interaction between continuous covariate and factor covariate is combined in the linear model, where treatments with and without transformation. From this treatment knows that an increase R^2 and adjusted R^2 values, from 56.62% (M4) to 60.26% (M7) and from 56.13% (M4) to 58.85% (M7). However, by this treatment combination effect causes changing in annual affect patterns and monthly effects, so that this interaction combination can be ignored in the model fitting. Although with and without transformation as in M7 and M8 models provide significance between several factor interactions of time covariates and continuous covariates. It can be obtained by ANOVA table. It shows that significant effect, except for humidity and interaction between rainfall and year as time covariates.

3.2.7. Without rainfall transformation and with interaction of time covariates. The construction M8 model can be written as analog with the M7 model without transformation. In this treatment knows also that an increase R^2 and adjusted R^2 values, from 56.37% (M5) to 60.20% (M8) and from 55.88% (M5) to 58.78% (M8). Based on results of interaction combination effect between continuous covariate and factor covariate then further investigation with respect to autocorrelation in the model fitting.

Table 2. Summary of linear model fitting for SST data.

Model	Residual SE	df	Multiple R^2	Adjusted R^2	F-statistic	AIC
M1	0.3786	2045	56.31%	55.88%	131.8 (df=20 & 2045)	1864.176
M2	0.3786	2045	56.31%	55.89%	131.8 (df=20 & 2045)	1872.767
M3	0.3786	2042	56.37%	55.88%	114.7 (df=23 & 2042)	1876.128
M4	0.3776	2042	56.62%	56.13%	115.9 (df=23 & 2042)	1864.176
M5	0.3786	2042	56.37%	55.88%	114.7 (df=23 & 2042)	1876.128
M6	0.3781	2042	56.49%	56.00%	115.3 (df=23 & 2042)	1870.220
M7	0.3657	1994	60.26%	58.85%	42.59 (df=71 & 1994)	1779.054
M8	0.3660	1994	60.20%	58.78%	42.48 (df=71 & 1994)	1782.316

Table 2 shows that the M1-M2 models without interaction and the M3-M6 models with interaction have a similar residual standard error (SE) values, except for the M4 model. The model (M1- M8) has t he same p-value < 2.2e-16. However, all models mentioned has large AIC so that it needs fitting impr ovement.

Residual on within and between groups for time covariate become concerns in linear model fitting by autocorrelation structure. The M9-10 models are constructed through a correlation ARMA(p = 1, q =1) M11 model with ARMA(p = 2, q =2) produces AIC negative values (as seen in table 3).For this reason then we needs upgrading fit performance with respect to model fitting with the autocorrelation construction by using AR(1) model, in the M12 and M13 models without and with transformation respectively.

Table 3. AIC of Model Structure Constructed.

Model	Structure	AIC
M9	SST = Temp + Month + Year + Humd + Rain	-1357.3650
M10	SST= Temp + Month + Year + Humd + Temp*Humd + Temp*Rain + Rain + Humd*Rain	-1313.4470
M11	Analog M10 with transformation	-1318.3260
M12	Analog M10 with transformation	-755.2367
M13	Analog M10 with transformation	-771.0071
M14	Analog M10 with transformation	556.4355
M15	Analog M10 with transformation	403.2987

The result as in table 3 gives negative point of half AIC values on previous models. Further by using autocorrelation ARMA approach with ~1| Month formulation implemented on the M14 and M15 models gives AIC positive value. In this study, correlation matrix of the model is not shown. Through generalized least squares (GLS) fit by using Restricted Maximum Likelihood to construct model is as follows:

Table 4. GLS fit by REML for Constructed Model.

Model	AIC	BIC	Loglik	Parameter Est.	df (total)	Residual
M12	-755.2367	-609.0729	403.6184	0.88001	2066	2042
M13	-771.0071	-624.8433	411.5035	0.87993	2066	2042
M14	556.4355	702.5993	-252.2178	0.71406	2066	2042
M15	403.2987	549.4625	-175.6493	0.78685	2066	2042

Table 4 shows similar parameter estimation with different Akaike Information Criteria (AIC), Baye sian Information Criteria (BIC) and Logic values of the M12 to M15 models. All models have the same of total degrees of freedom (df) and residual. Based on ANOVA table, then the M15 model has smalles t value of residual SE compared with other model constructed.

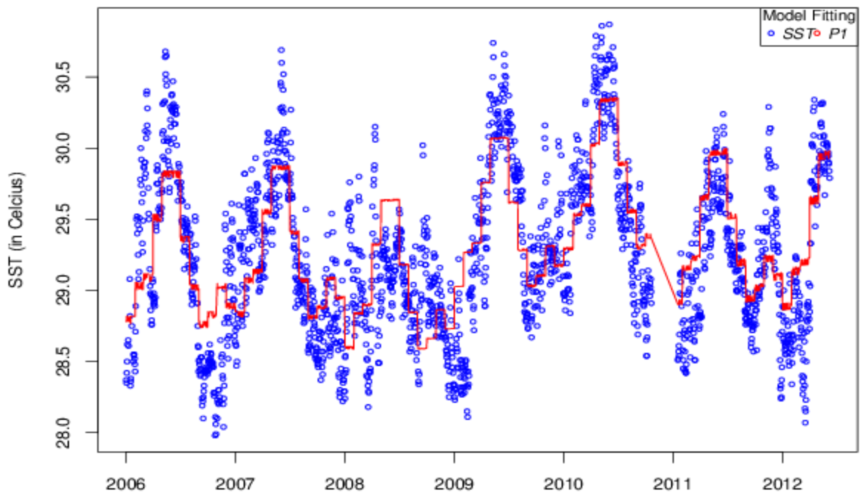

Figure 3. The M15 linear model fitting of the SST data.

Results of ANOVA analysis of M12 and M13 models with coefficients correlation in the Rho = 0.5 – 0.9 are discussed, as follows. The significant effect on all parameters, except for humidity, rainfall and interaction between continuous covariate terms in both models. Based on parameter estimation shows that rainfall transformation given effect pretty significant with respect to changing magnitude of parameter values, such as intercept value changes from 26.89 become 28.181. Although several magnitude of parameters shows similar value or category. If we observed the SST data fitting between the M12 and M13 models indicates that the M13 model is smoother than the M12 model.

Using the M14 and M15 models with the correlation coefficient in the Rho= 0.5-0.9 obtained ANO VA table shows that the significance effect for all parameters, except for humidity, rainfall and interact ion between continuous covariates in the M14 and M15 models. Further based on parameter estimation is known that the transformation gives a significant effect on the change of parameter values such as intercept value from 30.134 to 31.501. Although some parameters show quantities that it can be categ orized as similar or not statistically significant change as in the M14 and M15 models. figure 3 shows t hat the M15 model fitting smoother than M14 model. It has AIC value smaller than M14 model. The M 15 model fitting more appropriate compared in the M4 model due to it have autocorrelation effect.

3.3. Generalized additive models fitting of the SST dataset
By using generalized additive model (GAM) model, we have extended to use 16 climate features, where 14 continuous covariates and 2 times covariates. We obtained several significanse parameters as seen table 5 with AIC value is 634.99, GCV is 0.10, and the adjusted R-squared is 64,4%. In this model, we used wind speed and conductivity with large missing values, as seen in table 5.

In model fitting, we separates between continous covariates and time covariates to find time covariates contribution in the model. We obtain several parametric coeficients and smoothing parameter approximation for AIC and GCV is 665.82 and 0.10 respectively with adjusted R-squared value is 63,1%. Thus it shows that this approach (separate and compound) has the smallest AIC value and the largest R^2 adj. Both models show that time effects have a strong significance and the largest contribution in applied GAM models to the SST data.

Table 5. Analysis of variance for GAM model.

Predictor	edf	Red.df	F-value	P-value
Wind speed	4.951	6.094	4.371	0.000
Subsurface temperature	2.979	2.999	61.932	<2e-16
Shortwave radiation	1.896	2.381	0.639	0.555
Salinity	1.800	2.221	0.688	0.518
Humidity	1.000	1.000	12.112	0.000
Prec. rain	4.370	5.318	2.508	0.026
Dynamic height	2.850	2.978	21.432	2.21e-13
Density	2.845	2.976	6.774	0.000
Air temperature	1.956	2.215	3.882	0.017
Isoterm	1.000	1.000	0.035	0.852
Conductivity (SSS)	2.757	2.938	6.459	0.000
Current velocity	2.293	2.563	2.435	0.073
Zonal	1.000	1.000	30.385	4.36e-08
Merid	3.734	4.765	0.722	0.601
Year	8.000	8.000	97.955	<2e-16
Month	6.908	7.000	40.567	<2e-16

4. Conclusion
Linear model fitting of the SST data with interaction covariates in a combination and transformation gives the highest R^2 value among treatments given to the model i.e., with and without interaction, with and without transformation, with and without covariate centering, and a combination among given treatments. Although the effect value of interaction covariates in combination and transformation on R^2 value small classified (under 0.5%) gives magnitudes changing in multiple R^2 and adjusted R^2 on 0.31% and 0.43% respectively. This proof shows that time covariate does have strong significant effect in the linear model fitting for SST dataset.

In addition, for one buoy investigation shows that year effect does have significance highest peak happen on 2010 and lowest on 2008. Whereas for month effect shows increase trend occurs from January to June and decrease trend from June to September. Furthermore, trend continuous increase again till November and decrease on December. Thus, the highest peak of month effect occurred in June and the lowest in September. To obtain the optimal value of multiple R^2 and adjusted R^2 and the smallest AIC value, then the autocorrelation is a major concern in the fitting treatment to overcome the complexity of the SST dataset. Time effects have a strong significance and the largest contribution in the linear and GAM models in modeled the SST dataset, for with and without the number of climate features extended and different periods.

References

[1] Schott F A, Xie S P and McCreary J P 2009 Indian ocean circulation and climate variability *Reviews of Geophysics* **47** 1–46

[2] North G R and Stevens M J 1998 Detecting climate signals in the surface temperature record *Journal of climate* **11**(4) 563–577

[3] Deser C, Alexander M A, Xie S P and Phillips A S 2010 Sea surface temperature variability: Patterns and mechanisms *The Annual Review of Marine Science* **2** 115–143

[4] Magnus J R, Melenberg B and Muris C 2011 Global warming and local dimming: the statistical evidence *Journal of the American Statistical Association* **106**(494) 452–464

[5] McPhaden M J, Ando K, Bourlès B, Freitag H P, Lumpkin R, Masumoto Y, Murty V S N, Nobre P, Ravichandran M, Vialard J, Vousden D and Yu W 2010 The global tropical moored buoy array *Proceedings of Ocean Obs*. **9** 668-682

[6] Miftahuddin M 2016 Fundamental fitting of the SST data using linear regression models *IEEE* 128-133

[7] Hastie T, Tibshirani R and Friedman J 2008 *The Elements of Statistical Learning* (Stanford, California: Springer)

[8] James G, Witten D, Hastie T and Tibshirani R 2013 *An Introduction to Statistical Learning* (Stanford, California: Springer)

[9] Wood SN 2009 mgcv. R Package Version 1.6-0. (Available from http://CRAN.R-project.org/package=mgcv)

[10] Wood S N, Goude Y and Shaw S 2015 Generalized additive models for large data sets *Applied Statistics* **64** 139-15