COMPARISON OF SALINITY AND SEAWATER TEMPERATURE PREDICTIONS USING VAR AND BIRESPONSE FOURIER SERIES ESTIMATOR

By Faisol, dkk.

BAREKENG: Jurnal Ilmu Matema 8 a dan Terapan
March 2022 Volume xx Issue xx Page xxx–xxx
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengxxxxxxxxxxxx

Comparison of Salinity and Seawater Temperature Predictions Using VAR and Biresponse Fourier Series Estimator

Commented [a1]: Mohon untuk delete 1 Spasi
Commented [F2R1]: Sudah
Commented [F3R1]:

Faisol[1*], Putri Ukhrowi[2], M. Fariz Fadillah M.[3], Ira Yudistira[4], Kuzairi[5]

[1,2,4] Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Islam Madura, Komplek PP. Miftahul Ulum Bettet, Pamekasan, 69317
[3] Department of Mathematics, Faculty of Science and Technology, Universias Airlangga, Jl. Dr. Ir. H. Soekarno, Surabaya, East Java 60115
Corresponding author e-mail: [1] faisol.munif@gmail.com*

Abstract. Salinity is the concentration of dissolved salts in water. The salt in question is a variety of ions dissolved in water, including table salt (NaCl). Salinity and seawater temperature are one of the factors that affect salt production. The higher the NaCl content, the better the quality of the salt. Currently, people's salt production is still unable to meet the needs of national salt, especially industrial salt, because most of the quality of people's salt still does not meet the SNI criteria for industrial salt. Thus, it is necessary to predict the salinity and temperature of seawater to help determine the next steps or policies in improving the quality of people's salt. Predictions of salinity and seawater temperature were carried out by applying the Vector Autoregressive (VAR) Analysis method and nonparametric Fourier series regression with primary data of salinity and seawater temperature on the coast of Tlesah Tlanakan Beach, Pamekasan. The best model chosen is the model that has the smallest error size and the highest accuracy measure. The best models are nonparametric regression of the Fourier series of sine and cosine bases with the predicted result obtaining a MAPE value is 0.00496 and coefficient of determination is 100%.

Keywords: *Biresponse Fourier Series, Salt, Salinity, temperature, VAR.*

Article info:
Submitted: date, month, year *Accepted: date, month, year*

How to cite this article:
First author, second author, etc, "TITLE OF ARTICLE", *BAREKENG: J. Il. Mat. & Ter.*, vol. xx, iss. xx, pp. xxx-xxx, Month, Year.

1. INTRODUCTION

Madura is one of the regions of Indonesia with the largest salt production so, Madura is called the Salt Island. The area of salt land in Madura is 15.347 ha which is located in several districts, namely in Sumenep, Pamekasan and Sampang. One of the factors affecting salt production is salinity and seawater temperature. Salinity is the content of salts dissolved in water. The salinity of the waters describes the salt content in a body of water. The salt in question is a variety of ions dissolved in water including table salt (NaCl) [1]. Pond salt is divided into 3, namely, the first quality (KW1) is salt with a NaCl content between 95%-98%, the second quality (KW2) is salt with a NaCl content between 90%-95%, and the third quality (KW3) is salt with a NaCl content of less than 90% [2]. Based on SNI 01-3556-2000 the minimum level of NaCl in salt consumption is 94.7% [3]. So far, the amount of people's salt production that is included in the KW1 category has only reached 31.04% of the NaCl content of salt produced domestically only ranging from 81%-96% while for industrial needs salt with NaCl quality reaches the same or more than 97% [4].

Based on data from the Ministry of Industry in 2018, the national industry salt needs in 2018 were around 3.7 million tons. However, industry salt production in Indonesia is only 1.9 million tons every year, so Indonesia has to import around 1.8 million tons every year to meet the needs of industrial salt in Indonesia. Therefore, predicting salinity and seawater temperature is considered necessary to help determine the next step or policy in improving the quality of local pond salt so that it can meet industrial salt needs and reduce salt imports. [2].

The method that can be used to predict the salinity and temperature of seawater simultaneously is a Vector Autoregressive (VAR) and biresponse Fourier series estimator. Vector Autoregressive (VAR) is one of the analysis methods multivariate time series in the form of simultaneous equations, that is, variables used are interconnected with each other [5]. Previous research using VAR modeling for time series data is about the forecasting Covid-19 in West Java Province using the VAR model by Yuriska [6] with the results of obtaining a Mean Absolute Percentage Error (MAPE) value of 4.7%. Another study using the VAR method is a study by Rajab [7] namely, Forecasting COVID-19 Vector Autoregression-Based Model with the results interpolating predictions to forecast the cumulative number of cases, obtained MAPE of 0.0017% for UAE, 0.002% for Saudi Arabia , and 0.024% for Kuwait.

The other method that can be used to predict salinity and seawater temperature is the biresponse nonparametric regression based on Fourier series estimator. Research using the application of biresponse nonparametric regression based on Fourier series estimator such as Utami and Nur [8] applied for modelling to on High Water Level (HWL) data in Semarang City, based on the results determination of the optimal k by the GCV method obtained $k = 276$ with the maximum results of HWL data or it can be said that the maximum tide occurred on November 21, 2016 with R^2 of 94% and MSE of 10.31.

Based on the description, no one has compared the two methods to predict salinity and seawater temperature. Therefore, the researcher will conduct research on the prediction of salinity and seawater temperature using VAR analysis and Fourier series nonparametric regression. Comparison measures used are Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), and coefficient of determination (R^2), and the estimator to be chosen is the estimator with the minimum MSE and MAPE values and the maximum determination coefficient .

This study aims to determine how the results of prediction of salinity and seawater temperature using the best model selected from the comparison of VAR analysis and nonparametric Fourier series regression as an illustration to make it easier for interested parties in planning policies.

2. RESEARCH METHODS

2.1 Vector Autoregressive (VAR)

Vector Autoregressive (VAR) is a method that does not distinguish between the dependent variable and the independent variable. The dependent variable is a variable whose value is determined in the model. One of the assumptions that must be met in conducting VAR analysis is that between variables must be correlated.

VAR is a system of equations that shows each variable as a linear function of the constant and the lag (past) value of the variable itself and the *lag value* of other variables in the system of equations. VAR has a model for lag p and n variables can be formulated as follows [9]:

$$Y_t = b_0 + b_1 Y_{t-1} + \cdots + b_p Y_{t-p} + \varepsilon_t$$

with :

Y_t : response data for time t

Y_{t-i} : response data for time $t-i$

b_0 : intercept vector not $n \times 1$(constant)

p : long lag VAR

t : observ [10]ed period

ε_t : residual for observation to$t$

The steps for forecasting modeling using the VAR method are:

1. Stationary test

In time series analysis, the formation of a time series analysis model is determined with the assumption that the data is in a stationary state. Stationarity means that in the data there are no drastic changes or fluctuations in the data around a constant mean value and does not depend on the time and variance of these fluctuations. The visual form of a plot of time series data is often sufficient to ensure that the data is stationary or not. One of the unit root tests can be done with the Augmented Dickey Fuller (ADF test). The ADF test is a stationary test by determining whether the time series contains a unit root. The ADF test was introduced by Dickey and Fuller in 1979 with a simple model

$$\Delta Y_t = b_0 + \gamma Y_{t-1} + \varepsilon_1$$

with $\gamma = b_1 - 1$and $\Delta Y_t = Y_t - Y_{t-1}$with Y_t is the data at time t. The hypothesis used is

$H_0 : \gamma = 0$(contains unit root or is not stationary)

$H_1 : \gamma \neq 0$(does not contain unit or stationary roots)

Hypothesis testing is carried out using $-\tau$ defined statistics by the following formula.

$$\tau = \frac{\hat{\gamma}}{se(\hat{\gamma})}$$

with $\hat{\gamma}$is the least squares estimate of $\gamma$and $se(\hat{\gamma})$is the standard error of $\hat{\gamma}$, with the critical area of this test is to reject H_0if the ADF statistic value is or $\tau$is greater than the absolute critical value of the statistical distribution , t namely $\left|t_{\frac{\alpha}{2};df=n-n_p}\right|$, where $n$is the number of observations and $n_p$is the number of parameters. If the data is not stationary in the mean then differencing is done, where as if the data is not stationary in the variance then a transformation is performed [10].

Differencing is one of the common methods used to deal with non-stationary data. The differencing process can be carried out for several periods until the data is stationary, namely by subtracting a data from the previous one. Differencing is performed when the data is not stationary in the mean. If Y_t' it is data that has been differencing, then the differencing process is formulated with the following equation:

$$Y_t' = (1-b)^d Y_t$$

with b is a backward shift operator, which is an operator that shows a data shift back one period. Meanwhile d, it is a variable that shows the order of differencing, namely the number of differencing performed until the data is stationary.

2. VAR lag determination

Lag determination 9 sed to determine the optimal lag length. Determining the optimal lag can use several methods, namely Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), Hannan Quin Information Criterion (HQ). Optimal lag selection criteria are AIC, SC and HQ with the smallest value [11]. The Akaike Information Criterion (AIC) equation is as follows:

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

with

AIC : Akaike information criteria

k : Number of parameter estimates in the model

$\hat{L}$: The maximum value of the possible functions for the model

3. Granger causality test

Granger causality is a causal test or to see whether or not there is a unidirectional relationship or reciprocal relationship between variables [12]. General model of granger causality equation unrestricted as follows [13]:

$$\begin{cases} Y_{1t} = \sum_{i=1}^{p} a_i Y_{1(i-1)} + \sum_{i=1}^{p} \beta_i Y_{2(i-1)} + \varepsilon_{1t} \\ Y_{2t} = \sum_{i=1}^{p} \gamma_i Y_{2(i-1)} + \sum_{i=1}^{p} \tau_i Y_{1(i-1)} + \varepsilon_{2t} \end{cases}$$

Granger causality test hypothesis
H_0: There is no influence between variables
H_1: There is an influence between variables

Determination of the decision is if the p-value <α then it is H_0rejected, meaning that there is an influence between the variables studied. On the other hand, if the p-value$\geq$ α then it is H_0accepted, meaning that there is no influence between the variables studied.

4. Parameter significance test

Parameter significance test can be done by individual test (t-test), which is a test conducted to test the effect of each parameter on the model.
Hypothesis:
H_0: $b_{ij}^{(l)} = 0$(for all $i = 1,2 \ldots m\,; j = 1,2 \ldots m\,; l = 1.2 \ldots p$)
H_1: $b_{ij}^{(l)} \neq 0$(for all $i = 1,2 \ldots m\,; j = 1,2 \ldots m\,; l = 1.2 \ldots p$)
Significance level : α
Test statistics:

$$t_h = \frac{\hat{b}_{ij}^{(l)}}{SE(\hat{b}_{ij}^{(l)})}$$

Test criteria:
Reject H_0if $|t_h| > t_{\alpha/2,(n-k)}$or p- *value* < α, where $n$is the number of observations [14].

5. Model Verification

After estimating the parameters, the next step is to verify the model to see if the model is feasible or not to be used. The model is said to be suitable for use if it meets the White nose assumption. One way to test the White nose is to perform the Multivariate Portmanteau test. The hypothesis of the Portmanteau test is as follows:
H_0:$\rho_1 = \cdots = \rho_m = 0$
H_1: there is at least one $\rho \neq 0$
with ρ is the correlation matrix of the error vector and the statistical test used is

$$Q_N(m) = T^2 \sum_{t=1}^{m} \frac{1}{T-t} tr(\hat{\boldsymbol{\Gamma}}_t' \hat{\boldsymbol{\Gamma}}_0^{-1} \hat{\boldsymbol{\Gamma}}_t \hat{\boldsymbol{\Gamma}}_0^{-1}) \tag{1}$$

with :
T : number of observations
$\hat{\Gamma}_T(k) = \frac{1}{T-k+1} \sum_{t=0}^{T-k} \boldsymbol{Y}(t)\boldsymbol{Y}'(t-k)$is the element of the covariance matrix$\boldsymbol{\Gamma}^{(p)}$
$\hat{\Gamma}_T(-k)= \hat{\boldsymbol{\Gamma}}_T'(k)$fork ≥ 0

The first stage of the Pormanteau test is to calculate the Q statistical value as in equation (1). Q distribute $\mathrm{Chi-square}$ with degrees of freedom N^2m. Next is to compare the value Qwith the value $x^2{}_{N^2m}$ at the level of confidence $100(1-\alpha)\%$. If $Q < x^2{}_{N^2m}$ $(Q < \mathrm{chi-square})$ or $p - \mathrm{value} > \alpha$ then accept H_0. These results indicate that the residuals meet the white noise assumption and it can be said that the model fits the data. Vice versa, if $Q < x^2{}_{N^2m}$ or $p - \mathrm{value} < \alpha$ then reject $H_0$and it can be concluded that the model does not fit the data because the residuals do not meet the white noise assumption [15].

2.2 Biresponse Fourier Series Estimator

1. Biresponse Nonparametric Regression

Regression analysis involving two response variables and between the response variables there is a strong correlation or relationship, both logically and mathematically, is called biresponse regression. The nonparametric approach is used when the shape of the biresponse regression curve is unknown. In general, the model for biresponse nonparametric regression can be written as follows [16]:

$$\begin{cases} y_{i1} = g_1(x_{i1}) + \varepsilon_{i1} \\ y_{i2} = g_2(x_{i2}) + \varepsilon_{i2} \end{cases}$$

2. Fourier Series Estimator

Fourier series is a trigonometric polynomial function that has a high degree of flexibility. The Fourier series is a curve that shows the sine and cosine functions. By expansion into the form of a Fourier series, a periodic function can be expressed as the sum of several harmonic functions, namely functions of sine and cosine, including sinusoidal functions [17]:

Definition 2.1.

If given $g(x)$is a function that can be integrated and differentiable on the interval $[a, a+2L]$, then the representation of the Fourier series on that interval with respect to $g(x)$the trigonometric components sine and cosine is as follows:

$$g(x) = \frac{a_1}{2} + \sum_{p=1}^{\infty}(a_p \cos k^* x + b_p \sin k^* x)$$

with $k^* \approx \frac{n\pi}{L}$; $n = 1,2,3,\ldots$The Fourier coefficient is determined by the following formulation

$$a_1 = \frac{1}{L}\int_a^{a+2L} g(x)dx;\ a_p = \frac{1}{L}\int_a^{a+2L} g(x)\cos k^* x\, dx; b_p = \frac{1}{L}\int_a^{a+2L} g(x) \sin k^* x\, dx$$

3. Estimation of Biresponse Fourier Series Model

The estimator for the parameter regression curve of a biresponding nonparametric model using a Fourier series on a sine basis is

$$\hat{y}_{1i} = \hat{g}_{1i} = \frac{\hat{a}_{11}}{2} + \beta_1 t_{il} + \sum_{p=1}^{K}(\hat{a}_{k1} \sin kt_{il})$$
$$\hat{y}_{2i} = \hat{g}_{2i} = \frac{\hat{a}_{12}}{2} + \beta_2 t_{il} + \sum_{p=1}^{K}(\hat{a}_{k2} \sin kt_{il})$$

The estimator for the nonparametric biresponse parameter regression curve with the Cosine basis Fourier series approach is

$$\hat{y}_{1i} = \hat{g}_{1i} = \frac{\hat{a}_{11}}{2} + \beta_1 t_{il} + \sum_{p=1}^{K}(\hat{a}_{k1} \cos kt_{il})$$
$$\hat{y}_{2i} = \hat{g}_{2i} = \frac{\hat{a}_{12}}{2} + \beta_2 t_{il} + \sum_{p=1}^{K}(\hat{a}_{k2} \cos kt_{il})$$

The estimator for the parameter regression curve of a biresponding nonparametric model using the Fourier series approximation of the basis of sine and cosine is

$$\hat{y}_{1i} = \hat{g}_{1i} = \frac{\hat{a}_{11}}{2} + \beta_1 t_{il} + \sum_{p=1}^{K}(\hat{a}_{k1} \cos kt_{il} + \hat{b}_{k1} \sin kt_{il})$$
$$\hat{y}_{2i} = \hat{g}_{2i} = \frac{\hat{a}_{12}}{2} + \beta_2 t_{il} + \sum_{p=1}^{K}(\hat{a}_{k2} \cos kt_{il} + \hat{b}_{k2} \sin kt_{il})$$

2.3 Goodness of Fit Criteria

The indicator of the goodness of the model cn be seen from the model which has the smallest error size (MSE and MAPE) and the highest accuracy (coefficient of determination):

1. Mean Square Error (MSE) and Generalized Cross Validation (GCV)

MSE is the estimated value of the error variance. MSE is determined by the following equation:

$$MSE[k] = \frac{1}{n}\mathbf{y}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{A}(\mathrm{k})\right)^{\mathrm{T}}\mathbf{V}(\mathbf{I} - \mathbf{A}(\mathrm{k}))\mathbf{y}$$

With $\mathbf{A}$(k)is the hat matrix:

$$\boldsymbol{A}(k) = \boldsymbol{T}(k)(\boldsymbol{T}(k)^T \boldsymbol{V} \boldsymbol{T}(k))^{-1}\boldsymbol{T}(k)^T\boldsymbol{V}$$

The model is said to be good if the MSE value is minimum. Apart from being seen from the minimum MSE, the GCV indicator is also very influential for the best model. The GCV value is expressed in the following equation:

$$GCV(k) = \frac{MSE(k)}{(n^{-1}\, trace(\boldsymbol{I} - \boldsymbol{A}(k)))^2}$$

2. Coefficient of Determination (R^2)

One of the criteria used in selecting the best model is to use the coefficient of determination R^2. The coefficient of determination (R^2) is a quantity that describes the percentage of variation in the response variable that is explained by the predictor variable [18]. The Formula for the coefficient of determination is given as follows:

$$R^2 = \frac{(\hat{\mathbf{y}}-\bar{\mathbf{y}})^T(\hat{\mathbf{y}}-\bar{\mathbf{y}})}{(\mathbf{y}-\bar{\mathbf{y}})^T(\mathbf{y}-\bar{\mathbf{y}})}$$

with $\bar{y}$is a vector containing the average response data. A good model can be measured by R^2 great value [19]

3. Mean Absolute Percentage Error (MAPE)

MAPE is used to measure the error in the estimated value of the model which is expressed in the form of an average absolute percentage of residual. The MAPE calculation can be written as follows.

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \times 100\% \right|}{n}$$

with n is the amount of data or observations Y_t is the actual data and $\hat{Y}_t$ is the data forecast results. The best model is the model that has the smallest MAPE value [14]. The interpretation of the MAPE value is as follows [20].

Table 1 Interpretation of the MAPE Value

MAPE (%)	*Interpretation*
< 10	*Highly accurate prediction* (HAP)
10-20	*Good prediction* (GPR)
20-50	*Reasonable prediction* (RP)
>50	*Inaccurate prediction* (IPR)

Commented [a4]: Mohon menggunakan ukuran huruf 10 dan penulisan mohon disesuaikan dengan Templete BAREKENG.

Commented [F5R4]: Sudah diperbaiki

2.4 Data Source

The data used in this study are primary data, namely data on salinity (y_1) and sea water temperature (y_2) on the coast of Tlesah Tlanakan Pamekasan taken for 5 months (every 2 days) in October 2021-February 2022 with a total of 76 data. Data consisting of 65 data in sample (training) and 11 data out sample (testing). The in sample data used for the model formation process is data for the period 1 to 65, while the out sample data is used to evaluate the prediction results is the data period 66 to 76

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics

Analysis of the data can be seen in Table 2 below.

Table 2 Descriptive Statistics of Salinity and Seawater Temperature.

	Variance	mean	Stands. Deviation	Minimum	Maximum
Salinity	5.24	28.39	2.29	23.20	32.80
Temperature	3.78	32.26	1.94	27.10	37.00

Commented [a6]: Mohon mengikuti Templete.

Mohon penulisan , diganti dengan menggunakan "."

Based on Table 4.1, it is known that the average salinity of seawater on the coast of Tlesah Village is 28.39. The maximum value is 32.80 , while the minimum value is 32.20 with a standard deviation of 2.29 and a variance of 5.24. Meanwhile, sea water temperature has an average value of 32.26 , a maximum value of 37.00 , while minimum value 27.10 with a standard deviation of 1.94 and a variance of 3.78.

Commented [F7R6]: Sudah diperbaiki

3.2 Parametric Modeling Based on VAR Analysis

1. Data Stationarity Test

Stationarity test can be done with the ADF test. Based on the calculation of the stationary test with the ADF test using the R program, the p-value for salinity data is 0.0478 and temperature is 0.0486 (> 0.01). Therefore, it can be said that the data is not stationary. Because the data is not stationary, differencing is performed.

Unit root test on data which after differencing produces p-value less than 0.01 with the same hypothesis as the previous test. then it is concluded that the data y_1 (salinity) and y_2 (temperature) after first differencing is stationary.

2. Autoregressive Vector Lag Determination

Lag length included in this test is from 1 to 8 because the data used is daily data (2 days) for 5 months. The length of this lag is considered sufficient to describe the data for that period. The AIC value can be seen in Table 3 below:

Table 3 AIC Value Lag 1 to Lag 8

Model	*AIC*
VAR(1)	3,593
VAR(2)	**3,151**
VAR(3)	3,169
VAR(4)	3,156
VAR(5)	3,211
VAR(6)	3,216
VAR(7)	3,192
VAR(8)	3.186

In Table 2 it can be seen the results of the identification of the smallest AIC value contained in VAR (2) which is equal to 3.151 so the VAR model used is second-order VAR or VAR (2).

3. Granger Causality Test

The result of the causality test in this study is that salinity has a causal relationship to sea water temperature by looking at the probability value of 0.003 which is smaller than 0.05 (α) so that the decision is rejected H_0. Meanwhile, seawater temperature does not affect or have a causal relationship to salinity because it has a failure to reject decision H_0 because the probability value is 0.305 which is greater than 0.05

4. Estimation of Model Parameters

The VAR model used in this study is the 2nd order VAR or VAR (2). There are several model parameters that are not significant, because the p-value > 0.01 . The insignificant parameters are b_{10}, b_{12}, b_{14}, b_{20}, b_{22} and b_{24}. The results of the estimation of the parameters of the VAR (2) model on the salinity and seawater temperature data can be seen in Table 4.

Table 4 Parameter Estimation

Parameter	Estimate	*P-Value*	Significance
b_{10}	0.003	0.992	
b_{11}	-0.902	4.12×10^{-8}	***
b_{12}	-0.274	0.147	
b_{13}	-0.589	0.0001	***
b_{14}	-0.187	0.301	
b_{20}	-0.045	0.857	
b_{21}	0.217	0.046	*
b_{22}	-0.211	0.136	
b_{23}	0.380	0.0007	***
b_{24}	-0.086	0.521	

**** : significant at 0.01 . level*
*** : significant at 0.05 . level*

The model formed is a model that is estimated using the least squares method and the following equation is obtained:

$$\hat{y}_1 = 0{,}003 - 0{,}902 y_{1_{t-1}} - 0{,}274 y_{2_{t-1}} - 0{,}589 y_{1_{t-2}} - 0.187 y_{2_{t-2}}$$
$$\hat{y}_2 = -0{,}045 + 0{,}217 y_{1_{t-1}} - 0{,}211 y_{2_{t-1}} + 0{,}380 y_{1_{t-2}} - 0{,}086 y_{2_{t-2}}$$

Model has good criteria with MSE value of 5.086 and MAPE of 1.642. The R^2 model value is 0.463.

5. Model Verification

After estimating the parameters, the next step is to verify the model to see if the model is feasible or not to be used. In this stage the test used is the Pormentau test. The p-value in the Pormentau test is 0.125. This value is greater than 0.05, so it can be interpreted that the VAR (2) model meets the assumption of White noise or is suitable for predicting salinity and seawater temperature data.

3.2 Biresponse Fourier Series Estimation

1) Sine

Nonparametric regression with Fourier series estimation has an oscillation parameter ($\boldsymbol{k}$). The optimum $\boldsymbol{k}$ value is used to form the model and is determined based on the minimum GCV value. GCV values of some $\boldsymbol{k}$ values with a sine base can be seen in Table 5

Commented [a8]: .

Commented [F9R8]: Sudah diperbaiki

Commented [a10]: .

Commented [F11R10]: Sudah diperbaiki

Commented [a12]: .

Table 5 GCV Values with Sine Base

k	GCV
5	6.677×10^{2}
6	2.841×10^{2}
7	1.095×10^{2}
8	2.885×10^{1}
9	1.153×10^{21}
10	3.661×10^{-1}
11	25.470×10^{1}
12	1.096×10^{-2}
13	1.899×10^{-2}
14	3.274×10^{2}
15	8.630×10^{-2}

Based on Table 5, the minimum GCV value or optimum k is found in the 12th k, which is 1.096×10^{-2}. Based on the optimum value of the oscillation parameter (k), which is 12, an estimator model with a k of 12 was obtained. By substituting the estimated value of the parameter, it gets a nonparametric regression model as follows.

$$\hat{y}_{i1} = 24.197 + 0.111t_{i1} - 4.209\sin\ t_{i1} - 5.732\sin 2\,t_{i1} \ldots + 2.040\sin\ 12t_{i1}$$
$$\hat{y}_{i2} = 33.015 - 0.019t_{i2} - 0.244\sin\ t_{i2} + 0.539\sin 2\,t_{i2} + \cdots + 0.419\sin\ 12t_{i2}$$

This model has a goodness criterion with a GCV value of 1.096×10^{-2}, an MSE of 1.265 and a MAPE of 4.602. The value of the coefficient of determination of the model is 0.730.

2) Cosine Base

The GCV values of some k values with a cosine base can be seen in the following Table 6.

Table 6 GCV Values with Cosine Base

k	GCV
1	1.838×10^{6}
2	1.786×10^{6}
3	1.707×10^{6}
4	1.623×10^{6}
5	1.543×10^{6}
6	1.548×10^{6}
7	1.482×10^{6}
8	1.428×10^{6}
9	$\mathbf{1.359\times 10^{6}}$
10	1.442×10^{6}

Based on Table 6, the minimum GCV value or optimum k is found in the 9th k, which is $1{,}359\times 10^{6}$. Based on the optimum value of the oscillation parameter (k), namely 9, an estimator model with k of 9 was obtained. By substituting the estimated value of the parameter, it gets a nonparametric regression model as follows.

$$\hat{y}_{i1} = 30.012 - 0.042t_{i1} + 0.575\cos\ t_{i1} - 0.630\cos 2\,t_{i1} + \cdots + 0.287\ cos\ 9t_{i1}$$
$$\hat{y}_{i2} = 32.879 - 0.015t_{i2} + 0.191cos\ t_{i2} + 0.462\cos 2\ t_{i2} + \cdots - 0.515\ cos\ 9t_{i2}$$

This model has a goodness criterion with a GCV value of 1.359×10^{6}, an MSE of 0.896 and a MAPE of 4.318. The value of the coefficient of determination of the model is 0.540.

3) Sine and Cosine Bases

The GCV values of some k values with a sine and cosine base can be seen in the following Table 7.

Commented [a13]: .

Commented [F14R13]: Sudah diperbaiki

Commented [a15]: .

Commented [F16R15]: Sudah diubah

Commented [a17]: .

Commented [F18R17]: Sudah diperbaiki

Table 7 GCV Values with Sine and Cosine Bases

k	GCV
30	2.225×10^{3}
31	1.157×10^{2}
32	$\mathbf{1.759 \times 10^{-7}}$
33	2.942×10^{-5}
34	1.207×10^{3}
35	1.704×10^{-5}
36	1.757×10^{5}
37	1.042×10^{-5}
38	1.455×10^{-4}
39	2.079×10^{2}
40	1.359×10^{6}

Based on Table 7, the minimum GCV value or optimum k is found in the 32nd k, which is 1.759×10^{-7}. Based on the optimum value of the oscillation parameter (k), which is 32, an estimator model with k of 32 was obtained. By substituting the estimated value of the parameter, it gets a nonparametric regression model as follows.

$$\hat{y}_{i1} = 3.566 + 0.024t_{i1} + 3.208\cos\, t_{i1} + \cdots - 1.47\, cos\, 32t_{i1} - 4.165 \sin\, t_{i1} + \cdots - 2.331 \sin\, 32t_{i1}$$
$$\hat{y}_{i2} = 3.4307 - 6.0309t_{i2} + 1.520cos\, t_{i2} + \cdots + 0.507cos\, 32t_{i2} + 1.494 \sin\, t_{i2} + \cdots + 8.370 \sin\, 32t_{i2}$$

This model has a goodness criterion with a GCV value of 1.759375×10^{-7}, MSE of 1.191 and MAPE of 4.179. The value of the coefficient of determination of the model is 0.999.

3.1 Comparison of The Best Models For Seawater Salinity and Temperature Prediction

After obtaining the VAR model and nonparametric regression of the Fourier series of cosine and sinus bases, the next stage is to carry out the selection of the best model to be used. Model selection is carried out by looking at indicators, namely MSE, coefficient of determination and MAPE. A better model is one with the smallest MSE and MAPE values and the largest coefficient of determination. The MSE, MAPE and coefficients of determination values of the two selected models can be seen in table 8 below.

Table 8 Comparison of VAR and Fourier Series Birespon Estimates

Estimations	MSE	Coefficients of Determination	MAPE
VAR	5.086	0.463	1.642
Series Fourier Biresponse Sine Base	1.265	0.730	4.602
Series Fourier Biresponse Cosine Base	1.896	0.540	4.318
Series Fourier Biresponse Sine and Cosine Bases	**1.191**	**0.999**	**4.179**

Based on Table 8 it can be seen that between var models and nonparametric regressions of the fourier series of the base of the sine and cosine, the best model is the nonparametric regression of the birespon of the Fourier series of the base of the sinus and the cosine with oscillation parameter 32, MSE is 1.191, coefficient of determination is 0.999 and MAPE is 4.179.

Commented [a19]: .

Commented [F20R19]: Sudah diperbaiki

Commented [a21]: Mohon agar dibuat sejajar.

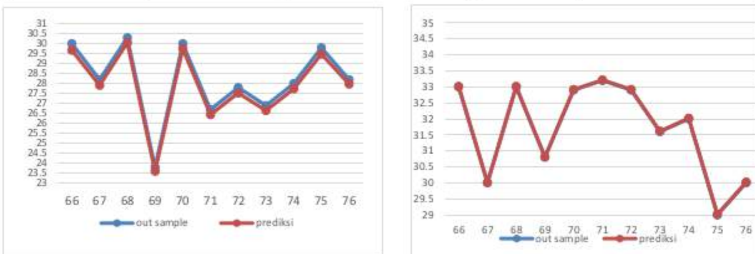
Commented [F22R21]: Sudah diperbaiki

Commented [a23]: .

Commented [F24R23]: Sudah diperbaiki

3.2 Prediction of Salinity and Seawater Temperature Using the Best Model

Based on the results of the comparison, the best model selected was the estimation of the fourier series of sine and cosine bases with oscillation parameter 32. The prediction results with the model have a MAPE value of 0.00496 and a coefficient of determination of 100%. The comparison plot of data out sample prediction of salinity and seawater temperature can be seen in Figure 1a and Figure 1b below.



(a) (b)

Figure 1 a) Comparison of out sample data values and salinity prediction results b) Comparison of out sample data values and seawater temperature predictions

Commented [a25]: Size 10

Commented [F26R25]: Sudah diperbaiki

Based on Figure 1a and 1b, it can be seen that the results of the prediction of salinity and seawater temperature using fourier series beespons estimates with a sine and cosine base are very close to the out sample data values in the period 66 to 76.

4. CONCLUSION

From this study, it can be concluded that several estimation models were obtained to predict the salinity and temperature of seawater, namely VAR of order 2 or VAR(2), nonparametric regression of the fourier series of sine bases with 12 oscillation parameters (k), cosine bases with 9 oscillation parameters (k), and sinus and cosine bases with 32 oscillation parameters (k). The best model selected was a nonparametric regression of the Fourier series of sine and cosine bases with predictive results having a MAPE value of 0.00496 and a cooefficient determination of 100%.

REFERENCES

[1] C. K. Aini, Prediksi Salinitas Menggunakan Fuzzy Sugeno, Pamekasan: Universitas Islam Madura, 2020.

[2] S. N. Putri, Y. I. Satria and N. Hendrianie, "Pra Desain Pabrik Garam Industri dari Garam Rakyat," *Jurnal Teknik ITS*, vol. 9, no. 2, pp. F151-F156, 2020.

[3] K. D. Maulana, m. m. Jamil, P. E. M. Putra, B. Rohmawati and Rahmawati, "peningkatan Kualitas Garam bledug Kuwu Melalui Proses Rekristalisasi dengan Pengika Pengotor," *Journal of Creativity Student*, vol. II, no. 1, pp. 42-46, 2017.

[4] Pusat Pengkajian Perdagangan Dalam Negeri, "Analisis Struktur Biaya Produksi Garam Rakyat," in *Laporan Akhir*, Jakarta, Badan Pengkajian dan Pengembangan Perdagangan Kementerian Perdagangan, 2019.

[5] M. R. A. Pranata, PENERAPAN STRUCTURAL SEASONAL VECTOR AUTOREGRESSIVE (SSVAR) PADA INFLOW-OUTFLOW UANG KARTAL, Malang: Universitas Brawijaya, 2018.

[6] R. Yuriska, A. A. Rohmawati and A. Aditsania, "Forecasting Jumlah Kasus Harian Covid-19Di Provinsi Jawa Barat Menggunakan Model Vector Autoregressive (VAR)," *e-Proceeding of Engineering*, vol. VIII, no. 5, pp. 11376-11387, 2021.

[7] K. Rajab, F. Kamalov and A. K. Cherukuri, "Forecasting COVID-19: Vector Autoregression-Based Model," *Arabian Journal for Science and Engineering*, no. 47, p. 6851–6860, 2022.

[8] T. W. Utami and I. M. Nur, "Aplikasi Regresi Nonparametrik Deret Fourier Pada Data High Water Level (HWL) Kota Semarang," *Statistika*, vol. 5, no. 2, pp. 57-61, 2017.

[9] L. J. Christiano and A. Cristoper, "Sims and Vektor autoregressive," *Jurnal of Ekononic*, pp. 1082-1104, 2012.

[10] R. Handayani, S. Wahyuningsih and D. Yuniarti, "Modeling Generalized Space Time Autoregressive (GSTAR) on Inflation Data In Samarinda And Balikpapan," *Jurnal Eksponensial*, vol. IX, no. 2, pp. 153-161, 2018.

[11] B. Junda and Junaidi, Ekonometrika Deret Waktu, Bogor: IPB Press, 2012.

[12] W. W. Utami, Pemodelan Ispa, Faktor Cuaca dan PM10 dengan Mengunakan Vector Autoregressive, Pekanbaru: Universitas Islam Syarif Kasim Riau, 2020.

[13] E. N. Susanti and R. Zamora, "Analisis Kausalitas Pertumbuan Ekonomi Terhadap Indeks Pembangunan Manusia Di Provinsi Kepulauan Riau," *Dimensi*, vol. VIII, no. 3, pp. 473-484, 2019.

[14] P. R. Hardani, A. Hoyyi and Sudarno, "Peramalan Laju Inflasi , Suku Bunga indonesia dan Indeks Harga Saham Gabungan Menggunkan Metode Vector Autoregressive (VAR)," *Gaussian*, vol. VI, no. 1, pp. 101-110, 2016.

[15] F. N. Hayati, Peramalan Harga Saham Jakarta Islamic Index Menggunakan Metode Vector Autoregressive, Surabaya: Institut Teknologi Sepuluh Nopember, 2016.

[16] D. R. Hardine, A. Abdullah, M. Ikbal and N. Chamidah, "Pemodelan Kadar Gula Darah dan Tekanan Darah Pada Remaja Penderita Diabetes Miletus Tipe II dengan pendekatan Regresi Nonparametrik Birespon Berdasarkan Estimator Spline," *Seminar Nasional Matematika dan Aplikasiny*, pp. 308-312, 2017.

[17] M. F. F. Mardianto, E. Tjahjono and M. Rifada, "Semiparametric Regression Based on Three Forms of Trigonometric Function in Fourier Series Estimator," in *IOP Publishing*, 2019.

[18] R. Kustianingsih, M. F. F. Mardianto, B. A. Ardhani, Kuzairi, A. Thohari, R. Andriawan and T. Yulianto, "Fourier series estimator in semiparametric," *AIP Conference Proceedings*, vol. 2329, pp. 060023-1-060023-9, 2021.

[19] A. Sholiha, Kuzairi and M. F. F. Mardianto, "Estimator Deret Fourier Dalam Regresi Nonparametrik dengan Pembobot Untuk," *Zeta – Math Journal*, vol. 4, no. 1, pp. 18-23, 2018.

[20] E. Vivas, H. Allende-Cid and S. Rodrigo, "ASystematic Review of Statistical and Machine Learning Methods for Electrical Power Forcasting with Reported MAPE Score," *entropy*, vol. XXII, no. 1412, 2020.

COMPARISON OF SALINITY AND SEAWATER TEMPERATURE PREDICTIONS USING VAR AND BIRESPONSE FOURIER SERIES ESTIMATOR

ORIGINALITY REPORT

5%
SIMILARITY INDEX

PRIMARY SOURCES

1	Kuzairi, Miswanto, I Nyoman Budiantara. "Three form fourier series estimator semiparametric regression for longitudinal data", Journal of Physics: Conference Series, 2020 Crossref	91 words — 2%
2	knepublishing.com Internet	51 words — 1%
3	repository.uph.edu Internet	31 words — 1%
4	jurnalsaintek.uinsby.ac.id Internet	23 words — < 1%
5	www.mdpi.com Internet	19 words — < 1%
6	ijisrt.com Internet	17 words — < 1%
7	ojs3.unpatti.ac.id Internet	17 words — < 1%
8	Koko Hermanto, Ryan Suarantalla, Sahdan -. "APLIKASI PROGRAM LINIER INTEGER 0-1 UNTUK	16 words — < 1%

MENYUSUN JADWAL USULAN PIKET SATPOL PP", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2020
Crossref

9	ijecm.co.uk Internet	15 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 15 WORDS