**Journal of Natural Sciences and Mathematics Research**
*J. Nat. Scien. & Math. Res.* Vol. 9 No. 1 (2023) 60-68
ISSN: 2460-4453 (Online); 2614-6487 (Printed)
Available online at http://journal.walisongo.ac.id/index.php/jnsmr

# Prediction of seawater salinity based on comparison of truncated spline estimators, Fourier Series and Kernel

**Faisol[1*], Sarmiatul Hasanah[2], M. Fariz Fadillah M.[3], Ira Yudistira[4], Tony Yulianto[5]**
[1,2,4,5] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Islam Madura, Pamekasan, East Java, Indonesia
[3] Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia

* Corresponding author's e-mail: faisolmunif@gmail.com

**ABSTRACT**
Production, being defined as the degree of saltiness or excessive salt content in water. The salt contributing to salinity comprises various ions dissolved in water, including common table salt (NaCl). The higher the concentration of NaCl, the higher the quality of the resultant salt. Indonesia faces the challenge of salt importation, encompassing both consumable and industrial salt, due to the insufficient quality of domestically produced salt, which often falls short of meeting the SNI criteria. Consequently, predicting seawater salinity becomes essential for informing subsequent measures or policies aimed at enhancing salt quality in Indonesia, particularly in the Madura region. This study employs a nonparametric regression curve estimator with a truncated spline estimator approach, Fourier series, and kernel to examine and address these salinity-related concerns. From the comparison results, the best model in predicting seawater salinity is the estimator of the Fourier series base sine cosine with an oscillation parameter (k) of 2 with a GCV value of 5.017987 and MSE and a coefficient of determination of 0.06299933 and 94.64373%. So that the prediction results obtained in this study are close to accurate with MAPE values of 0.07225208%, MSE of 0.0001441417 and coefficient of determination of 99.99%.
**Keywords:** *Fourier Series, Salt, Kernel, Salinity, Truncated Spline*

## Introduction

Indonesia as an archipelagic country is a coastal and marine area that has a variety of very large biological and non-biological resources. The ocean, which is 70% of the country's total area, holds a lot of potential that can be exploited. One of them is commodity salt. Its position as an archipelagic country with a very wide sea causes each region to have the potential to produce salt (Tahta, 2015).

Madura is one of the largest salt producers in Indonesia so that Madura is identical as the Salt Island (Hadi & Ahied, 2017). The area of salt-making land in Madura is 15,347 ha (Aini, 2020). However, in reality, the wealth of the land area has not been able to meet domestic salt needs, especially industrial needs because the quality of the salt produced has not been able to meet the limit requirements that have been regulated in government regulations (Hanik & Mutmainah, 2020).

The quality of salt depends on the content of NaCl salt (Arwiyah, Zainuri, & Efendy, 2015). The quality of consumption salt according to the Indonesian National Standard (SNI) is a minimum of 94.7% NaCl which is included in the good quality range (Zainuri, 2016). Traditionally produced salt often exhibits subpar quality, necessitating reprocessing for both consumable and industrial applications. The inferior quality of salt can be attributed to various factors, with salinity being one of the significant contributors (Assadad & Utomo, 2011).

Salinity refers to the concentration of dissolved salts in water, describing the salt content in a given water source. This includes various ions, including table salt (NaCl). Predicting salinity values is crucial to enhancing the quality of domestically produced salt (Aini, 2020).

Nonparametric regression techniques employing a truncated spline estimator, Fourier series, and kernel can be utilized for predicting seawater salinity.

Nonparametric regression is an approach to regression where the curve's shape is not predetermined. Various estimators, such as splines, kernels, Fourier series, and wavelets, are employed in nonparametric regression methods (Adrianingsih, Dani, & Ainurrochmah, 2020). Spline regression is a regression analysis capable of estimating data that lacks a specific pattern, relying on its ability to identify and derive estimated data from the formed pattern (Fadhillah & Suparti, 2016). The Fourier series is a trigonometric polynomial function characterized by its high flexibility, exhibiting sine and cosine functions in its curve (Adrianingsih, Dani, & Ainurrochmah, 2020). Conversely, the kernel estimator is a technique for approximating the unknown density function by employing the kernel function (Jannah, 2020).

The objective of this study is to assess seawater salinity prediction results through a comparison of the truncated spline estimator, Fourier series, and kernel methods.

## Methods

*Descriptive Statistics*

There are two variables used in this study, namely salinity (y) and seawater temperature (x) . The data for the two variables are primary data taken directly from the coast in Tlesah Village, Kec. Tlanakan Kab. Pamekasan every two days as much as 76 data from October 1, 2021 to February 28, 2022. Descriptive analysis of the data can be seen in Table 1.

Table 2. Descriptive Statistics of Seawater Salinity Data

| Data | Min | Max | Mean | Standard Deviation |
|---|---|---|---|---|
| Y | 23.2 | 32.8 | 28.39 | 2.29 |
| X | 27.1 | 37 | 32.26 | 1.94 |

Referring to Table 2, the seawater salinity data collected from October 2021 to February 2022 reveals a minimum salinity value (y) of 23.2, a maximum value of 32.8, with an average of 28.39342105 and a standard deviation of 2.28988125. Additionally, the sea water temperature data (x) shows a minimum of 27.1, a maximum of 37, an average of 32.25657895, and a standard deviation of 1.944656647.

*Nonparametric Regression Model Estimator*

*Spline truncated*

In truncated spline regression, the identification of the optimal knot point is crucial. Attaining the optimal knot point ensures the derivation of the best truncated spline model. The selection of the optimal knot point involves examining the minimum value of the Generalized Cross Validation (GCV). Once the knot points are chosen, the parameter estimates necessary for predicting seawater salinity can be derived.

*Fourier Series*

Nonparametric regression approach with Fourier series estimator has oscillation parameter($k$). The GCV method is used to determine the $k$optimal value. Optimal selection $k$is done by looking at the minimum GCV value. From $k$those selected, a Fourier series model is obtained which can be used to estimate seawater salinity.

*Kernels*

Nonparametric regression approach with gaussian kernel estimator has *bandwidth parameter* ($h$).The GCV method is used to determine the value of$h$ optimal. Optimal selection $h$is done by looking at the minimum GCV value. From $h$those selected, a Fourier series model is obtained which can be used to estimate seawater salinity.

**Results and Discussions**

*Truncated Spline Estimator For Seawater Salinity Estimation*

In the context of truncated spline regression, identifying the optimal knot point holds significant importance. Obtaining the optimal knot point leads to the derivation of the best truncated spline model. The selection of the optimal knot point involves assessing the Generalized Cross Validation (GCV) value, aiming for the minimum. The knot points formed are presented in Table 3.

Table 3. Linear *Spline* Regression Knot Points

| Number of Knots | Selected Knot Point | GCV |
|---|---|---|
| **1** | $\boldsymbol{k_1 = 34}$ | **4.578829** |
| 2 | $k_1 = 33$ | 4.588081 |
| | $k_2 = 34$ | |
| 3 | $k_1 = 33$ | |
| | $k_2 = 34$ | 4.594714 |
| | $k_3 = 35$ | |

Based on Table 3, it can be seen that 1 number of knots produces the best model from the number of other knots so that the selected knot point is 34 with a minimum GCV value of 4.578829. The selected knot points and the estimated parameters formed are presented in Table 4.

Table 4. Selected Knot Points and Estimation of Linear *Spline Regression Parameters*

| Number of Knots | Selected Knot Point | Variable | Parameter Estimation |
|---|---|---|---|
| 1 | $k_1 = 34$ | $x_1$ | $\hat{\beta}_0$= 50.41936 |
| | | | $\hat{\beta}_1$= -0.683172 |
| | | | $\hat{\beta}_2$= 0.5251897 |

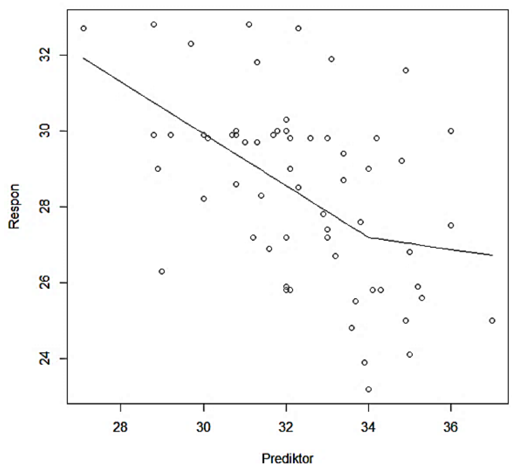


Figure 1. Selected Plot of *Truncated Spline* Model

From Table 4 it can be seen that the knots formed are 1 knot. Meanwhile, the number of parameter estimates formed is 3 including *intercepts.* Based on the GCV values in Table 4, the best model can be obtained by using the *spline* line a r at the 34th knot point. In addition, the *spline* line a r model at the 34th knot also has a good model, which has $R^2$0.2551549 and an MSE value of 4.074939.

In Figure 1 it can be seen that the *spline model* used to model seawater salinity is a linear *spline* at the 34th knot point so that the model formed.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 (x_1 - k_{1.1})_+$$

$$\hat{y} = 50{,}41936 - 0{,}683172x_1 + 0{,}5251897(x_1 - 34)_+$$

$$\hat{y} = 50{,}41936 - 0{,}683172x_1 + 0{,}5251897x_1 - 17{,}8564498$$

$$\hat{y} = 32{,}5629102 - 0{,}1579823x_1$$

The interpretation of the best *spline model* is that if it is $x_1$increased by one unit, it $y$will decrease by 0.1579823. This means that if the temperature increases it is likely to cause a decrease in seawater salinity by 0.1579823.

*Fourier Series Estimator for Salinity Estimation of Seawater*

The nonparametric regression approach utilizing the Fourier series estimator involves an oscillation parameter (k). The GCV method is applied to ascertain the optimal value for k. The results of the optimal GCV calculation can be seen in Table 5.

Table 5. GCV Value with Sinus Cosine Basis

| $k$ | GCV . value |
|---|---|
| 1 | 5,214231 |
| **2** | **5.017987** |
| 3 | 5,337114 |
| 4 | 5,601613 |
| 5 | 5.646689 |
| 6 | 5.831032 |
| 7 | 5.910873 |
| 8 | 6.386375 |
| 9 | 6.661501 |
| 10 | 6.949918 |

Based on Table 5, it can be seen that the minimum GCV value in the estimator of the Fourier sine cosine series is 5.017987 with $k$ optimally 2.

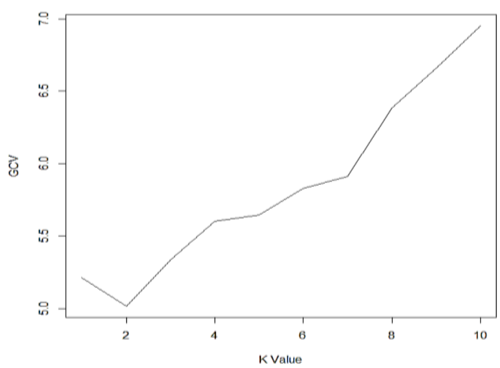


Figure 2. Changes in GCV Value of the Fourier Series Basis Sinus Cosine

In Figure 2 it can be seen that there is a decrease in the GCV value from $k = 1$towards $k = 2$the next, there is an increase in the GCV value from $k = 3$towards $k = 10$, thus for the estimation result of $k$ 2, we get a good Fourier series based on sine cosine with a GCV value of 5.017987, MSE of 0.06299933, and $R^2$of 0.9464373. So that the model formed is as follows:

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{\lambda} \left[\hat{a}_j \cos(2\pi j x_i) + \hat{b}_j \sin(2\pi j x_i)\right]$$

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{2} \left[\hat{a}_j \cos(2\pi j x_i) + \hat{b}_j \sin(2\pi j x_i)\right]$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{a}_1 \cos(2\pi x_i) + \hat{a}_2 \cos(4\pi x_i) + \hat{b}_1 \sin(2\pi x_i) + \hat{b}_2 \sin(4\pi x_i)$$

$$= 28.43385 + 0.01864553 \cos(2\pi x_i) + 0.30955448 \cos(4\pi x_i) + 1.2055340 \sin(2\pi x_i) + 0.9169853 \sin(4\pi x_i)$$

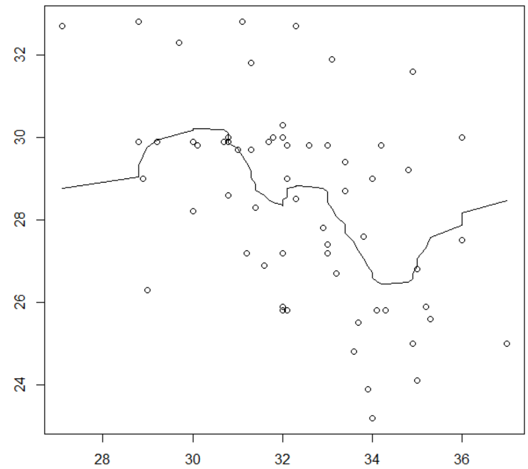


Figure 3. Plot of the Fourier Series Estimator Basis Sine Cosine

Based on Figure 3 the Fourier series stimator is a good model, with an oscillation $(k)$value of 2 , a minimum GCV of 5.017987, an MSE of 0.06299933 and $R^2$0.9464373 .

*Gaussian Kernel Estimator For Seawater Salinity Estimation*

Nonparametric regression approach with gaussian kernel estimator has *bandwidth parameter* $(h)$.The GCV method is used to determine the $h$optimal value. The results of the optimal GCV calculation can be seen in Table 6.

Table 6. Gaussian Kernel Estimator GCV Value

| ***H*** | **GCV . value** |
|---|---|
| 1.20 | 4.589479 |
| 1.21 | 4.589167 |
| 1.22 | 4.588920 |
| 1.23 | 4.588736 |
| 1.24 | 4.588616 |
| **1.25** | **4.588556** |
| 1.26 | 4.588557 |
| 1,2 7 | 4.588618 |
| 1.28 | 4.588738 |
| 1.29 | 4.588915 |
| 1.30 | 4.589149 |

Based on Table 6 , it can be seen that the minimum GCV value is 4.588556 with optimal h obtained is 1.25.

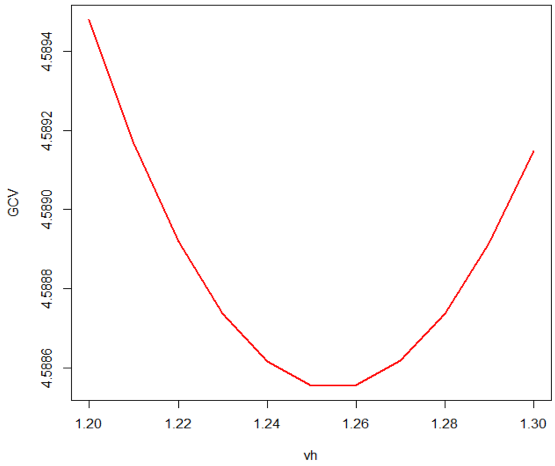

Figure 4. GCV Value with Gaussian Kernel Estimator

In Figure 4 it can be seen that there is a very significant decrease in the GCV value from $h =$ 1,20towards $h = $ 1,25the next, there is an increase in the GCV value from $h = $ 1,26leading $h =$ 1,30to thus for the estimation results from$h$ of 1.25 obtained a good gaussian kernel model with a GCV value of 4.588556 , MSE of 0.415, and $R^2$of 0.7592 so that the model formed is as follows:

$$\hat{y}_i = \hat{m}_h(x_i)$$

$$\hat{m}_h(x_i) = \frac{1}{n}\sum_{i=1}^{n} K_h\,(x - X_i)$$

$$= \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x - X_i}{h}\right)^2\right)$$

$$= \frac{1}{65(1,25)}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x - X_i}{1.25}\right)^2\right)$$

$$= \frac{1}{81,25}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x - X_i}{1.25}\right)^2\right)$$

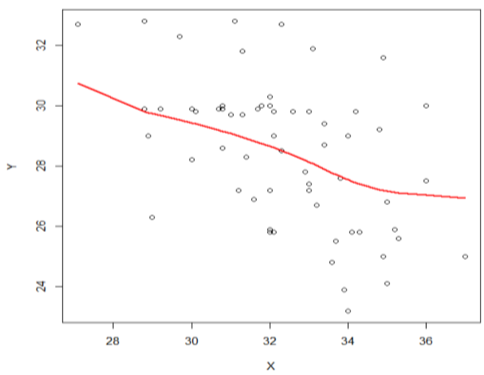

Figure 5. Plot of Gaussian Kernel Estimator

Based on Figure 5 the gaussian kernel stimator is a good model, with an optimal h value of 1.25, a minimum GCV of 4.588556 , an MSE of 0.415 and $R^2$of 0.7592 .

*Best Estimator Comparison for Seawater Salinity Estimation*

Nonparametric regression approach with *truncated spline estimator* has optimum knot k parameter, Fourier series estimator base sine cosine has oscillation parameter and $(k)$gaussian kernel estimator has *bandwidth parameter* $(h)$. The following is a comparison table between the *truncated spline estimator*, the Fourier sine cosine series and the gaussian kernel as shown in Table 7.

Table 7. Comparison of *Truncated Spline* Estimators, Fourier Series and Kernel

| Estimator Type | Optimal Quantity Size | GCV | MSE | $R^2$ |
|---|---|---|---|---|
| Spline Truncated | $k = 34$ | 4,579 | 4.075 | 0.255 |
| Fourier Series | $k = 2$ | 5.018 | 0.063 | 0.946 |
| Kernels | $h = 1.25$ | 4,588 | 0.415 | 0.759 |

Based on Table 7, it can be seen that among the *truncated splines* , the Fourier series base sine cosine and the gaussian kernel, the minimum GCV value was chosen while still paying attention to the minimum and $R^2$high MSE values in the selection of the model goodness criteria. So that the selected model is a Fourier series based on sine and cosine with an oscillation parameter $(k)$of 2 with a GCV value of 5.017987 and MSE and $R^2$0.06299933 and 0.9464373 so that the model used to predict seawater salinity is a Fourier series based on a sine cosine.

*Marine Salinity Prediction Using Selected Models*

Based on the comparison results, the best model chosen is the Fourier series estimator based on sine cosine with an oscillation parameter $(k)$ of 2. The prediction results based on the model can be seen in Table 8.

Table 8. Prediction Results of Seawater Salinity

| Data *Out Sample* | | $\hat{y}$ |
|---|---|---|
| $x$ | $y$ | |
| 33 | 30 | 28.26656 |
| 30 | 28.2 | 28.27048 |
| 33 | 30.3 | 28.27428 |
| 30.8 | 23.8 | 28.27797 |
| 32.9 | 30 | 28.28155 |
| 33.2 | 26.7 | 28.28503 |
| 32.9 | 27.8 | 28.28842 |
| 31.6 | 26.9 | 28.29171 |
| 32 | 28 | 28.29491 |
| 29 | 29.8 | 28.29803 |
| 30 | 28.2 | 28.30106 |

In Table 8 it can be seen the value of the *out sample data* and the predicted results of seawater salinity. The prediction results have a MAPE value of 0.07225208%, an MSE of 0.0001441417 and a coefficient of determination of 99.99%. The comparison plot of the *out sample data* for seawater salinity prediction can be seen in Figure 6.

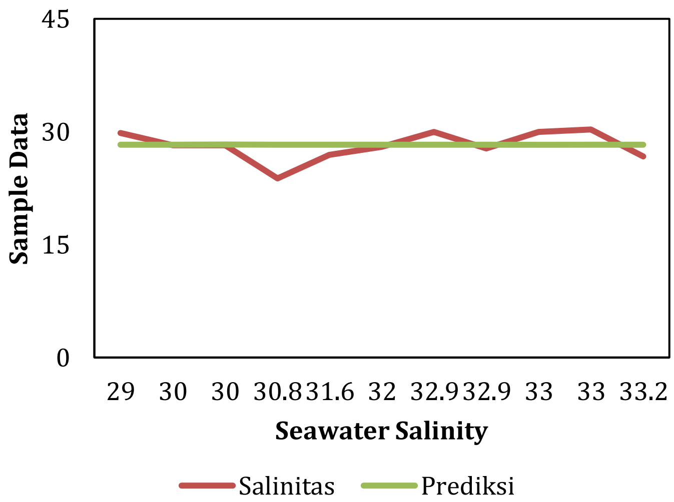


Figure 6. Comparison of *Out Sample* Data Values and Prediction of Seawater Salinity

Based on Figure 6, it can be seen that the salinity prediction results are close to the actual *out sample data value* .

## Conclusion

Based on the previous discussion, it can be concluded that the comparative analysis using the *truncated spline estimator*, Fourier series and gaussian kernel, the best model is obtained by using the Fourier series estimator on the basis of sine cosine with an oscillation parameter ($k$)of 2, GCV value of 4.578829, MSE of 0.06299933 and the coefficient of determination is 94.64373 %. So that the prediction results are very accurate with a MAPE value of 0.07225208%, an MSE of 0.0001441417 and a coefficient of determination of 99.99%.

## Acknowledgments

The research is supported by mathematics department, mathematics and science faculty and LP2M Universitas Islam Madura who has provided support in the form of assistance to conduct a research

## References

Adiraga, Y. (2013). *Analisis Dampak Perubahan Curah Hujan, Luas Tambak Garam, dan Jumlah Petani Garam Terhadap Produksi Usaha Garam Rakyat di Kecamatan Juwana Kabupaten Pati (Periode 2003-2012).* Semarang: Universitas Diponegoro Semarang.

Adrianingsih, N. Y., Dani, A. T., & Ainurrochmah, A. (2020). Pemodelan Dengan Pendekatan Deret Fourier Pada Kasus Tingkat Pengangguran Terbuka Di Nusa Tenggara Timur. *Prosiding Seminar Edusainstech*, 400-407.

Aini, C. K. (2020). *Prediksi Salinitas Menggunakan Fuzzy Sugeno.* Pamekasan: Universitas Islam Madura.

Arwiyah, Zainuri, M., & Efendy, M. (2015). Studi Kandungan NaCl di Dalam Air Baku dan Garam yang Dihasilkan Serta Produktivitas Lahan Garam Menggunakan Media Meja Garam yang Berbeda. *Jurnal Kelautan, 8*(1), 1-9.

Assadad, L., & Utomo, B. S. (2011). Pemanfaatan Garam Dalam Industri Pengolahan Produk Perikanan. *Squalen*, 26-37.

Bella, A., Putri, E. R., & Mandang, I. (2021). Rancang Bangun Sistem Monitoring Suhu dan Salinitas pada Air Laut. *Progressive Physics Journal, 2*(1), 38-48.

Fadhillah, K. N., & Suparti. (2016). Pemodelan Regresi Spline Truncated untuk Data Longitudinal. *Jurnal Gaussian*, 447-454.

Faisol, , Ukhrowi, P., Mardianto, M. F., Yudistira, I., & Kuzairi, . (2022). Comparison Of Salinity And Seawater Temperature Predictions Using VAR And Biresponse Fourier Series Estimator. *BAREKENG: Journal of Mathematics and Its Application*, 1465–1476.

Faisol, , Yulianto, T., Arsyiah, Sugiono, Basuki, A., & Zainuddin, M. A. (2022). Seawater salinity modeling using bivariate probit. *Journal of Physics: Conference Series*, 1742-6596.

Faisol, , Yulianto, T., Yaqin, M., Basuki, A., & Zainuddin, M. A. (2022). Prediction of seawater salinity using truncated spline regression method. *AIP Conference Proceeding*, 2641.

Hadi, W. P., & Ahied, M. (2017). Kajian Etnosains Madura dalam Proses Produksi Garam sebagai Media Pembelajaran IPA Terpadu. *Rekayasa*, 79-86.

Hanik, U., & Mutmainah. (2020). Analisis Kinerja Dan Kebutuhan Petani Garam Di Kabupaten Pamekasan Sebagai Dasar Pengembangan Desain Model Social Learning. *Sosek KP*, 237-249.

Jannah, H. (2020). *Estimasi Jumlah Pemohon Paspor di Kabupaten Pamekasan Berdasarkan Estimator Kernel dalam Regresi Nonparametrik.* Pamekasan: Universitas Islam Madura.

Kustianingsih, R., Mardianto, M. F., Ardhani, B. A., Kuzairi, , Thohari, A., Andriawan, R., & Yulianto, T. (2021). Fourier series estimator in semiparametric regression to predict criminal rate in Indonesia. *AIP Conference Proceeding*, 2329.

Mardianto, M. F., Ariyanto, R. A., Andriawan, R., & Husada, D. A. (2021). Contribution Analysis of “Suroboyo Bus” in Waste Management Based on Two Form of Complete Fourier Series Estimator. *Jurnal Matematika Mantik, 7*(1), 86-95.

Melinda, N., & Suryono. (2018). Rancang Bangun Sistem Wireless Sensor Salinitas Model Kapasitif. *Youngster Physics Journal, 7*(2), 76-84.

Pabiban, D. (2016). Rancang Bangun Sistem Distilasi Surya Tipe Parabolic Untuk Menurunkan Kadar Salinitas Air Laut. *Jurnal Ilmiah Flash*, 131-141.

Riskiyah, S. (2017). *Penerapan Regresi Nonparametrik Spline Truncated untuk Mengetahui Faktor-Faktor yang Mempengaruhi Hasil Tangkapan Ikan Nelayan di Kabupaten Pamekasan.* Universitas Islam Madura.

Tahta, A. (2015). *Rancang Bangun Sistem Informasi Geografis Pemetaan Produksi Garam Untuk Ketersediaan Garam Berbasis Web Di Kabupaten Sumenep.* Universitas Pesantren Tinggi Darul Ulum.

Takezawa, K. (2006). *Introduction To Nonparametric Regression.* new jersey: john wiley and sons.

Walid, M., & Darmawan, A. K. (2018). Sistem Cerdas Pendugaan Salinitas Air Laut Berdasarkan Citra Landsat Menggunakan Metode Adaptive Neuro Fuzzy Inference System ( ANFIS ). *Jurnal Buana Informatika, 9*(1), 1-10.

Yulianti, A. (2015). *Prototype Alat Pengolahan Air Laut Menjadi Air Minum (Pengaruh Variasi Packing Filter Terhadap Kualitas Air Dengan Analisa DO, Salinitas, dan Konduktivitas).* Sriwijaya: Politeknik Negeri Sriwijaya.

Yulianto, T., Faisol, , Zahroh, F., Suryanti, S., & Tafrikan, M. (2021). Forecasting the recovery of COVID-19 patients in East Java using the Fuzzy time series Cheng method. *Journal Of Natural Sciences And Mathematics Research*, 44-50.

Zahrah, M. (2020). *Prediksi Tingkat Hunian Hotel di Kabupaten Pamekasan Berdasarkan estimator deret Fourier dengan dan Tanpa Tren.* Pamekasan: Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Madura.

Zainuri, M. (2016). Hubungan Kandungan Natrium Chlorida (NaCl) dan Magnesium (Mg) Dari Garam Rakyat di Pulau Madura. *Prosiding Seminar Nasional Kelautan*, 167-172.