P-ISSN: 2798-1134 | E-ISSN: 2797-3549 | DOI: 10.52562/injoes.2025.1738

Indonesian Journal of Earth Sciences

http://journal.moripublishing.com/index.php/injoes

| Research Article |

Sea Surface Salinity (SSS) Prediction Using Landsat 8 OLI Image Data in The Bangka Strait Waters with Five Prediction Model Combinations

Khoirun Nisa[1]*, Gentio Harsono[1], Sukendra Martha[1], Dangan Waluyo[1]

Sensing Technology Study Program, Faculty of Engineering and Technology, Indonesian Defense University, Bogor 16810, Indonesia

**Correspondence: khoirun.samsunga11@gmail.com*

Received: 02 August 2025 / Accepted: 08 September 2025 / Published: 09 September 2025

Abstract: Salinity is the most important parameter for controlling the biological components of ecosystems, seas, and estuaries, which also control the components that make up an ecosystem. Conventional water quality monitoring is considered inaccurate and inefficient in terms of energy and time. Therefore, research is needed to predict sea surface salinity as a type of water quality monitoring using remote sensing reflectance or Remote Sensing Reflectance (RRS) from Landsat imagery. The Landsat image data used is level 2 Surface Reflectance (SR), which is ready to use without additional processing by the user, whereas previous research required corrections to the image data to obtain Surface Reflectance image data. This study aims to determine the performance of the prediction model produced by using five combinations of Landsat image bands. The data used are Landsat 8 OLI image data (recording date 05 August 2024) downloaded from the USGS website and in situ salinity data in the Bangka Strait sea (09 March 2025), as many as 5 samples that can be used. The obtained data were processed using multiple linear regression analysis with Rrs as the independent variable and in situ salinity as the dependent variable. The salinity prediction model consisted of five band combinations. The analysis produced R^2 values for each model combination of 0.8166287408, 0.935603228, 0.820745745, 0.869209652, and 0.574027060. The RMSE validity tests for combination 1, combination 2, combination 3, combination 4, and combination 5 were 2.41327, 1.43012, 2.38602, 2.03811, and 3.67817. Then for the NMAE value, namely 10.10152205%, 5.32713015%, 9.58011308%, 8.8868031%, and 14.51012574%. The combination rankings that have the best prediction performance are combination 2, combination 4, combination 3, combination 1, and combination 5. So the best model in predicting seawater salinity is the combination of the 2 prediction models, with its constituent band components being band 1, band 2, and band 4.

Keywords: Combination, Landsat 8 OLI image, Multiple Linear Regression, Rrs, Salinity

INTRODUCTION

Among the relevant physical, chemical, and biological parameters of water, water quality is the primary measure of the sustainability of estuarine aquatic ecosystems. Salinity level is one of the criteria used to classify these environments (Lew et al., 2022). Salinity is the level of salinity or the number of grams of salt dissolved per liter of solution (Pasaribu et al., 2024). Therefore, salinity can be considered a water quality parameter that also controls the components of an ecosystem.

To monitor water quality, this monitoring can be done conventionally or by utilizing technology. According to Wulandari et al. (2024), conventional water quality monitoring or observation means directly observing water quality at the water's location. However, this conventional method is still considered inaccurate and inefficient in terms of energy and time. Therefore, more efficient and effective methods are needed for water quality monitoring.

According to Arafah et al. (2015), the technology that can provide information on water quality conditions is remote sensing technology. Several water quality parameters, such as suspended solids (SS) and chlorophyll-a (Chl-a), can be detected remotely because the spectral properties of reflected light are altered by optically active components (Sood & Zhu, 2024). Remote sensing techniques are more cost- and time-effective than traditional methods for observing and measuring water quality parameters in a body of water (Abdelmalik, 2018). Research by Octaviana et al. (2020) explains that remote sensing has been widely used as an alternative for monitoring water quality. This technology is less time-consuming or expensive, and water quality can be monitored regularly and effectively. Remote sensing solution

techniques can be used to collect data about the Earth's surface over a constantly changing spatial and temporal range.

Parameters such as pH, conductivity, TDS, salinity, and turbidity are measured to determine water quality (Rusli et al., 2023). Water salinity can be analyzed using remote sensing technology (Nafizah et al., 2016). In their research, Muhsi et al. (2022) used Remote Sensing Reflectance (RRS) as a method for predicting seawater surface salinity. The data used were Landsat 8 OLI (Operational Land Imager) imagery, with five bands: Band 1 (433 nm-453 nm), Band 2 (450 nm-515 nm), Band 3 (525 nm-600 nm), Band 4 (630 nm-680 nm), and Band 5 (845 nm-885 nm).

A study by Ansari & Akhoondzadeh (2020) used radiometric and atmospheric corrections to generate Surface Reflectance (SR) imagery. This study used Landsat 9 OLI level 2 SR imagery data. Landsat Collection 2 science products have been recognized as Analysis Ready Data (ARD) by the Committee on Earth Observation Satellites (CEOS). The satellite data has been organized in a way that allows for direct analysis with minimal user effort to obtain this certification. This also allows for collaboration with the data over time (USGS, 2024). Therefore, in this study, no correction was carried out to obtain the Surface Reflectance image as was done in previous studies. The problems arising from conventional water quality monitoring processes gave rise to the idea of conducting research in order to predict seawater surface salinity using remote sensing reflectance (Rrs) from Landsat 8 OLI imagery and how the prediction model performance is generated using five combinations of seawater salinity prediction models from Landsat 8 OLI imagery bands. Five combinations of three bands were taken from the five bands in the Landsat 8 OLI imagery as used in the study conducted by Muhsi et al. (2022), namely band 1 (Coastal/Aerosol), band 2 (Blue), band 3 (Green), band 4 (Red), and band 5 (Near Infrared). Previous studies used all five bands. Therefore, researchers will use three of them for one model combination and analyze the five combinations of salinity prediction models, then compare them to find the model with the best band combination. The Bangka Strait is located between Bangka Island and the east coast of South Sumatra Island. Large rivers on the east coast of South Sumatra, especially in Banyuasin Regency and Ogan Komering Ilir Regency, which flow along the Bangka Strait, carry large amounts of freshwater into the Bangka Strait. The dynamics and quality of the waters will be influenced by the mixture of freshwater masses from the upstream rivers with seawater masses (Surbakti et al., 2022).

The combination of Remote Sensing Reflectance (RRS) from Landsat 8 OLI image bands downloaded from https://earthexplorer.usgs.gov/, along with in-situ salinity data, was analyzed to develop a salinity prediction model using a multiple linear regression algorithm. The results were then compared based on the coefficient of determination (R^2), the Root Mean Squares Error (RMSE), and the Normalized Mean Absolute Error (NMAE).

METHOD

This research uses quantitative methods. According to Creswell in Berlianti, cited in Jauza & Albina (2025), quantitative research focuses on measurements, numbers, and statistical analysis. The research flow is explained in Figure 1. This research was conducted in the Bangka Strait. The Bangka Strait is a body of water located between Bangka Island and Sumatra Island (Rahma et al., 2024). The population in this study was the seawater of the Bangka Strait, and the sample (target) of the research was seawater salinity.

Data Collection

Seawater samples from the Bangka Strait were taken at 11 locations. These 11 sampling locations were then sorted based on salinity measurement results and the position of the downloaded image. Field (in situ) salinity measurements were made using a salinity refractometer, with measurement units in ppt (parts per thousand). The downloaded image data was Landsat 8 OLI Level 2 Surface Reflectance (SR) imagery. Sharaf El Din et al. developed an algorithm using Landsat 8 surface reflectance data to calculate turbidity and total suspended solids (Jin et al., 2023). Through the website https://earthexplorer.usgs.gov/, Landsat 8 OLI image data with recording and processing dates as described in Table 1 below:

Table 1. Landsat 8 OLI Image Data

Name	Recording Date	Processing Date
LC08_L2SP_124061_20240805_20240808_02_T1	2024-08-05	2024-08-08

Considering that the image used is a level 2 surface reflectance image, no further processing, including filtering of cloud cover, was performed. Based on the metadata file information of the Landsat 8 OLI image used, the overall cloud cover percentage is 5%, and the cloud cover for land is 10.82%. This

is in accordance with the statement in the study (Li et al., 2018) regarding the difference in acquisition probability in Landsat observation cloud coverage; generally, previous studies have used a threshold of less than or equal to 30% CC (Cloud Cover).

Sea Surface Salinity (SSS) Prediction Equation

Salinity prediction calculations using the RRS method use the following bands in Landsat 8 OLI imagery: Band 1 (433 nm-453 nm), Band 2 (450 nm-515 nm), Band 3 (525 nm-600 nm), Band 4 (630 nm-680 nm), and Band 5 (845 nm-885 nm) (Muhsi et al., 2022). An evaluation of the consistency of Landsat-8 data for high-spatiotemporal inland and coastal water quality monitoring shows that Landsat-8, including the use of Band 1, can be consistently used quantitatively and qualitatively for monitoring coastal, inland, and nearshore waters (Hafeez et al., 2022). Research conducted by Quynh et al. (2024) evaluated surface water salinity indicators from Landsat-8 OLI imagery using machine learning, where the results of the analysis using the BMA (Bayesian Model Averaging) technique showed that the band 2 (B2) variable from Landsat 8 imagery was ranked and the variable weight was better than band 5 (B5). Then for band 3, or the green band, research by Mamun et al. (2024) showed that the ratio of red/blue bands and green bands in Sentinel-2 MSI and Landsat 8 OLI had the highest predictive performance of SD (Secci Depth). Secchi depth interacted significantly with maximum specific conductance, where the interaction occurred slightly but significantly increased the negative effect of Secchi depth; this indicates that the impact of warming is amplified in high salinity (Loewen & Jackson, 2024). Bands 3 and 4 were found to be the most suitable spectral data for estimating chlorophyll-a (Chl-a) (Mohan et al., 2025). Sea surface salinity (SSS) has a moderate and negative relationship with chlorophyll-a (Chl-a)(Hidayat et al., 2023). This means that if the concentration of chlorophyll-a in water increases, salinity decreases, and vice versa. The Landsat 8 OLI Near-Infrared (NIR) reflectance band values combined with other bands were used in the EC calculation formula, resulting in R^2 of 0.894 and 0.960, respectively (Abdul Wahid & Arunbabu, 2022). The highest correlation was found in EC (Electrical Conductivity) and EVI (Enhanced Vegetation Index) measurements, with $R^2 = 0.736$ (Tran et al., 2020).

Field salinity is the dependent variable and Rrs is the independent variable. Rrs (Remote Sensing Reflectance) is defined as the ratio of the outgoing water radiation (Lw) to the incoming sky radiation (Ed) (Mobley et al., 2016). According to Novoa et al. (2017) the Rrs equation can be expressed as follows.

$$Rrs = \frac{\rho}{\pi} \quad (1)$$

Rrs is the remote sensing reflectance value, ρ is the water reflectance value in a dimensionless space. In determining PI (π), Archimedes (287–212 BC) found an estimated value of 3.14159 (Dewangan, 2023). The Rrs value obtained together with the in-situ salinity data were then analyzed to obtain a salinity prediction model. In this study, there was a limitation in the number of in situ salinity samples, of the 11 sample points taken, only 5 points could be used for analysis for reasons that will be explained in the next section. The added value of multiple linear regression is that the method and resulting equations are simple and easy to understand (Dimitriadou & Nikolakopoulos, 2022). The multiple linear regression prediction equation according to Sahbeni (2021):

$$Y = A_0 + A_1 . x_1 + A_2 . x_2 + A_3 . x_3 + \ldots A_n . x_n \quad (2)$$

where: Y is dependent variable; A_0 is intercept; A_i is coefficient of variable i; and X_i is explanatory variable.

Prediction Model Performance

Three different methods can be used to test hypotheses in multiple linear regression analysis: the individual parameter significance test (t), the simultaneous significance test (F), and the coefficient of determination (R^2) test (Fitri et al., 2024). The classification of coefficients of determination is presented in Table 2.

Table 2. Classification of Statistical Index (R^2)

R^2 Value Range	Criteria
$R^2 \leq 0.25$	Deviate
$0.25 < R^2 \leq 0.50$	Bad
$0.50 < R^2 \leq 0.60$	Moderate
$0.60 < R^2 \leq 0.75$	Good
$0.75 < R^2 \leq 1.00$	Very Good

Source: Fitri et al. (2024)

The hypothesis H_0 is that Rrs has no significant effect on salinity, and Ha is that Rrs has a significant effect on salinity. Validity tests can be carried out to determine the performance of the salinity prediction model. Validity tests can use RMSE (Root Mean Square Error). The RMSE equation can be represented by Equation (3) (Hodson, 2022) as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{1}^{n}(yi - \hat{y}i)^2} \quad (3)$$

n is the number of data, yi is the observed value in the validation data set (actual value), and yi is the predicted value (Li, 2017). In addition to using RMSE to assess the performance of the prediction model, the NMAE (Normalized Mean Absolute Error) validity test is also used. The NMAE number is calculated using the following formula:

$$\text{NMAE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|Yi - Y'i|}{\sqrt{Yi \cdot Y'i}} \quad (4)$$

Yi is the actual value, and Y'i is the predicted value in time period i, where n is the test data set (Zhao et al., 2022).

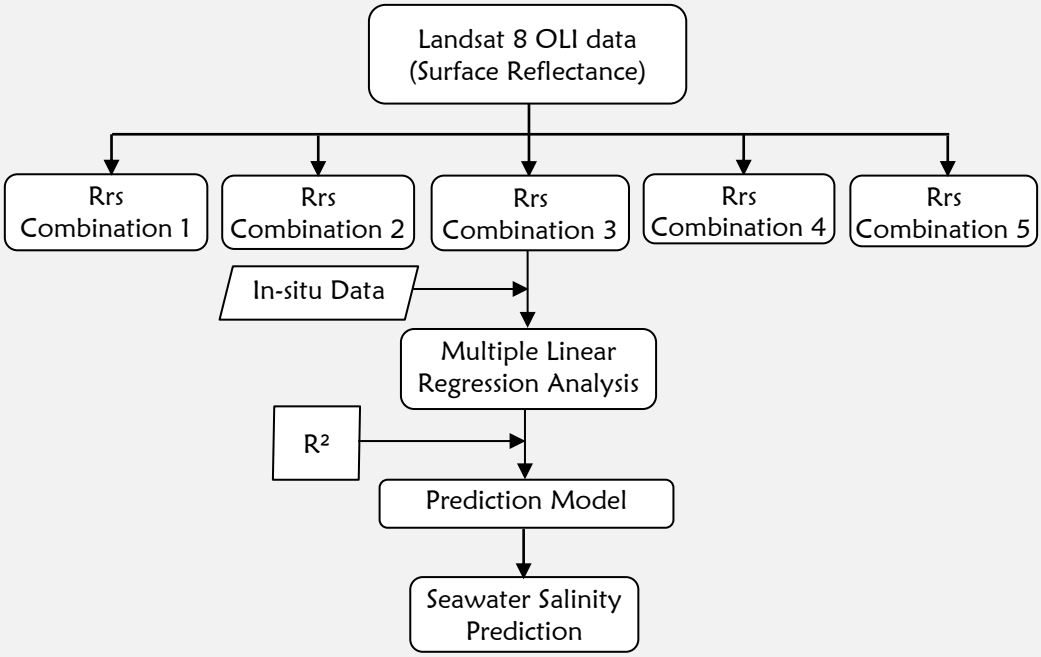


Figure 1. Research Design.

RESULTS & DISCUSSION

In-situ Data and Landsat 8 OLI Imagery

In-situ salinity was measured at 11 points. Due to signal interference during the recording of the water sampling coordinates, 10 seawater samples were taken to measure salinity using a salinity refractometer. The unit of salinity measurement using this instrument is parts per thousand (ppt). The summary of the field salinity measurements in the Bangka Strait waters is as following Table 3.

The results of field salinity measurements as shown in Table 3, STA 10 with coordinates 2°10'37"S and 104°57'47"E (located in the ocean), the salinity measurement results are 0. This deviation is likely due to an error in taking seawater samples at that point. Therefore, the seawater salinity at the STA 10 coordinate point is not used in the next process. So, in the calculation of Rrs (Remote Sensing Reflectance), 9 salinity sample points are used. Field salinity data, as shown in Table 4, as many as nine points are analyzed using linear regression analysis.

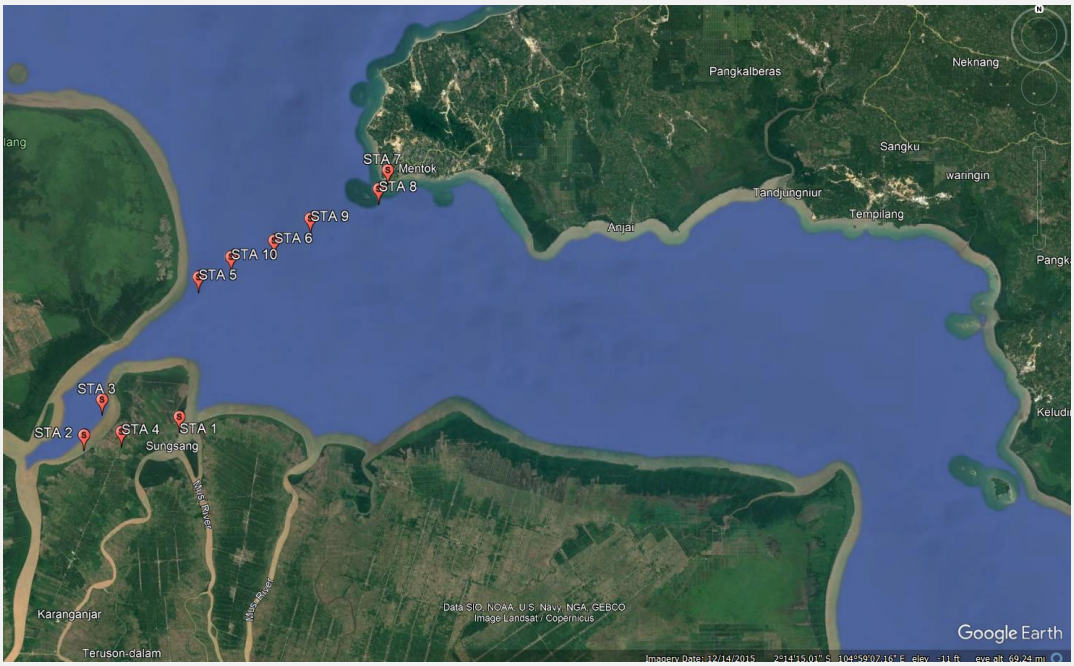


Figure 2. Coordinate Points of Salinity Sampling Stations in the Study Area.

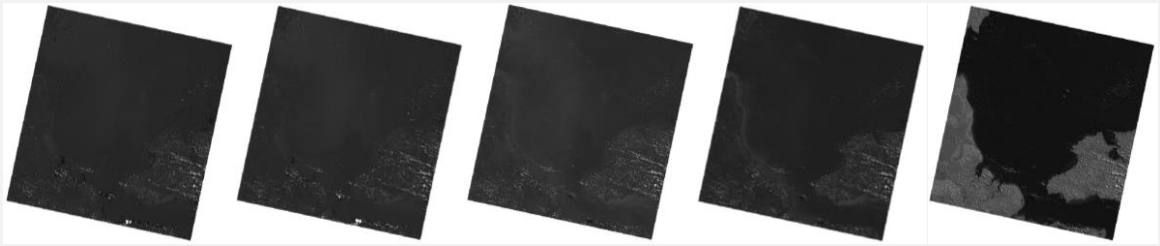
Table 3. Salinity of Bangka Strait Seawater Salinity (Coordinate Selection)

STA	Latitude	Longitude	Time Capture	Salinity (ppt)
1	2°20'58"S	104°54'25"E	11:32:56	5
2	2°22'9"S	104°48'13"E	13:01:54	14
3	2°19'51"S	104°49'24"E	14:17:16	15
4	2°21'56"S	104°50'39"E	15:24:11	7
5	2°11'56"S	104°55'40"E	16:07:41	15
6	2°9'34"S	105°0'34"E	16:32:52	20
7	2°5'1"S	105°7'56"E	18:29:21	30
8	2°6'14"S	105°7'20"E	19:32:31	25
9	2°8'9"S	105°2'55"E	20:15:21	16
10	2°10'37"S	104°57'47"E	21:08:09	0

Table 4. Salinity of Bangka Strait Seawater Salinity (Salinity Selection)

STA	Latitude	Longitude	Latitude	Longitude	Salinity (ppt)
1	2°20'58"S	104°54'25"E	-2.34944	104.90694	5
2	2°22'9"S	104°48'13"E	-2.36917	104.80361	14
3	2°19'51"S	104°49'24"E	-2.33083	104.82333	15
4	2°21'56"S	104°50'39"E	-2.36556	104.84417	7
5	2°11'56"S	104°55'40"E	-2.19889	104.92778	15
6	2°9'34"S	105°0'34"E	-2.15944	105.00944	20
7	2°5'1"S	105°7'56"E	-2.08361	105.13222	30
8	2°6'14"S	105°7'20"E	-2.10389	105.12222	25
9	2°8'9"S	105°2'55"E	-2.13583	105.04861	16

Calculation of Remote Sensing Reflectance (Rrs) of Landsat 8 OLI imagery using the QGIS version 3.40 Bratislava Long Time Release Raster Calculator tool. The Rrs calculation uses a raster calculator, namely the image band divided by 10000 then divided by the π value because the scale factor for Landsat imagery is 0.00001 (https://usgs.gov/). According to Table 1, the data used is Landsat 8 OLI imagery for recording on August 5, 2024. The calculated Rrs image in geotiff (raster) format is as follows.



(a) (b) (c) (d) (e)

Figure 3. Rrs display in geotiff format of GQIS Rrs calculation results: (a) Rrs band 1; (b) Rrs band 2; (c) Rrs band 3; (d) Rrs band 4; (e) Rrs band 5: (a) Rrs band 1; (b) Rrs band 2; (c) Rrs band 3; (d) Rrs band 4; Rrs band 5.

The Rrs raster extraction was performed to obtain the Rrs values for the field salinity sample coordinates. Four empty sampling points were marked with NULL. Therefore, only five sampling coordinate points were used for analysis using multiple linear regression analysis. The Rrs and in situ salinity levels were analyzed to obtain several parameters that form the elements of the salinity prediction equation. Microsoft Excel version 2021 with Data Analysis Regression as a tool used to help perform analysis.

Table 5. Recapitulation of Rrs extraction results for each image band against sampling points

Latitude	Longitude	Salinity	Rrs band 1	Rrs band 2	Rrs band 3	Rrs band 4	Rrs band 5
-2.19900	104.92800	15	0.237937	0.257322	0.302777	0.3009622	0.241279
-2.15900	105.00900	20	0.242139	0.262988	0.288357	0.2451943	0.233258
-2.08400	105.13200	30	0.331997	0.344507	0.373569	0.3847097	0.559844
-2.10400	105.12200	25	0.223040	0.248346	0.279317	0.2301382	0.219602
-2.13600	105.04900	16	0.232685	0.246754	0.258181	0.2310613	0.229215

Data Analysis Results

Multiple linear regression analysis using five combinations of Landsat 8 OLI image bands for each band's RRS values is shown in Table 6 below.

Table 6. Rrs Calculation Band Combination

Combination	Band Composition
1	Band 1, Band 2, Band 3
2	Band 1, Band 2, Band 4
3	Band 1, Band 2, Band 5
4	Band 2, Band 3, Band 4
5	Band 2, Band 3, Band 5

Table 7. Recapitulation of Data Analysis Results

Parameter	Combination 1	Combination 2	Combination 3	Combination 4	Combination 5
Constant (α)	0.94438	15.12139	96.01550	- 42.19642	- 6.87629
β band 1	971.26747	407.44901	- 466.66195		
β band 2	-1656.23108	- 816.42411	5.427086	115.92997	134.86625
β band 3	839.12975			297.11385	- 32.65655
β band 4		480.26739		-206.16677	
			162.89213		4.06643
R^2	0.816628741	0.935603228	0.820745745	0.869209652	0.574027060

Several parameters obtained from the analysis results are then entered into equation (2), namely the multiple linear regression equation, namely the salinity prediction equation. Calculations of seawater salinity predictions are grouped based on combinations of image bands, namely combination 1, combination 2, combination 3, combination 4, and combination 5.

Table 8. Salinity Prediction Recapitulation

In-situ Salinity (ppt)	Salinity Prediction				
	Combination 1	Combination 2	Combination 3	Combination 4	Combination 5
15	15.6281	14.4654	14.3724	15.5456	18.9213
20	16.7495	22.183	22.288	23.4157	20.1237
30	29.5728	30.0291	29.8769	29.4204	29.6632
25	23.9436	22.7888	20.942	22.1364	18.3886
16	20.1061	16.5337	18.5207	15.4819	18.9033

Based on the salinity prediction results, it is known that the salinity predictions of combination 1, combination 2, combination 3, and combination 4 are close to in situ salinity. This is supported by the R^2 generated by each combination in the analysis. The prediction models of combination 1, combination 2, combination 3, and combination 4 are considered very good in predicting salinity because the average R^2 value is above 0.75. Meanwhile, the prediction model of combination 5 shows a significant difference from in situ salinity because this combination has the smallest R^2 number, namely 0.574027059677348. Based on the statistical index classification in Table 2, the prediction model of combination 5 is in the medium or moderate performance in salinity prediction because the R^2 value is in the range of $0.50 < R^2 \leq 0.60$. Table 9 compiles the results of the hypothesis test:

Table 9. Hypothesis Test Results

Hypothesis Testing Parameters	Combination 1	Combination 2	Combination 3	Combination 4	Combination 5
R^2	0.81662874	0.93560323	0.82074575	0.86920965	0.57402706
Sig. F	0.52807055	0.31960178	0.52250108	0.45022224	0.76750089
P-value (α)	0.97921554	0.58727863	0.62068992	0.36031530	0.97035222
P-value (Rrs band 1)	0.43592000	0.21946066	0.74492661		
P-value (Rrs band 2)	0.53578856	0.24471550	0.99016083	0.66863653	0.91225763
P-value (Rrs band 3)	0.58117000			0.49214084	0.94218788
P-value (Rrs band 4)		0.31071186		0.37381092	
P-value (Rrs band 5)			0.57195507		0.98822125

Simultaneous significance test (F test): namely, the significant F value in Combination 1, Combination 2, Combination 3, Combination 4, and Combination 5 is 0.528070550170780, 0.319601776451194, 0.522501076065491, 0.450222243534862, and 0.767500894038141, which means the model is not feasible (Sig > α (5%)). Likewise, the value produced in the individual parameter significance test (t-test), each variable Rrs band 1, Rrs band 2, Rrs band 3, Rrs band 4, and Rrs band 5, produces a p-value > 0.05; this indicates that there is no partial influence between the independent variable and the dependent variable, and the multiple linear regression model is not suitable for measuring independent variables and dependent variables interacting with each other. However, in the study by Andrade (2019), it explains that if the P-value > 0.5 is obtained, it does not reject H0, but it can be interpreted in this study that it cannot be concluded that Rrs has no significant effect on salinity, meaning Rrs has no effect on salinity at all. Thus, based on the data, there is an effect of Rrs on salinity, but this effect is not large or consistent enough to meet the statistical requirements, namely a P-value ≤ 0.05. This occurs because the amount of in situ data is limited, as Muff et al. (2022) suggest a range for translating the P-value into evidence language is more appropriate than the significance test, as in Figure 4. Meanwhile, the coefficient of determination (R^2) test in each prediction model of Combination 1, Combination 2, Combination 3, Combination 4, and Combination 5 produces a number close to one; in other words, almost all the data needed to predict the variation of the dependent variable is provided by the independent variables. To find out how much of the percentage change (variance) in the dependent variable (Y) can be explained by the independent variable (X) in the population as a whole, not just in

the sample, the adjusted R^2 is used (Karch, 2020). The calculation of adjusted R^2 (Ohtani & Tanizaki, 2004) is as follows :

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}(1 - R^2) \quad (5)$$

Description:
R^2: Adjusted R^2
n: Number of data
k: Number of independent variables (predictors)
R^2: Coefficient of determination

Table 10. Adjusted coefficient of determination ($\bar{R}^2$) per combination

Combination	n	k	R^2	$\bar{R}^2$
1	5	3	0.816629	0.633257
2	5	3	0.935603	0.871206
3	5	3	0.820746	0.641491
4	5	3	0.86921	0.738419
5	5	3	0.574027	0.148054

The coefficient of determination used in regressions using more than two independent variables is the adjusted R^2. The adjusted coefficient of determination, or adjusted R-squared, may have a negative value and is always smaller than R-squared (Enggal et al., 2019). Table 10 shows something in line with the previous statement, so in this study the adjusted R^2 was used. Interpretation of the $\bar{R}^2$ value refers to Zaka & Sutopo (2017). The adjusted coefficient of determination ($\bar{R}^2$) in combination 2 is 0.871206, meaning that the independent variable (Rrs) is able to explain the variation that occurs in the dependent variable (salinity) by 87.1206%, while the remaining 12.8794% is by other variables outside the variables used in this study.

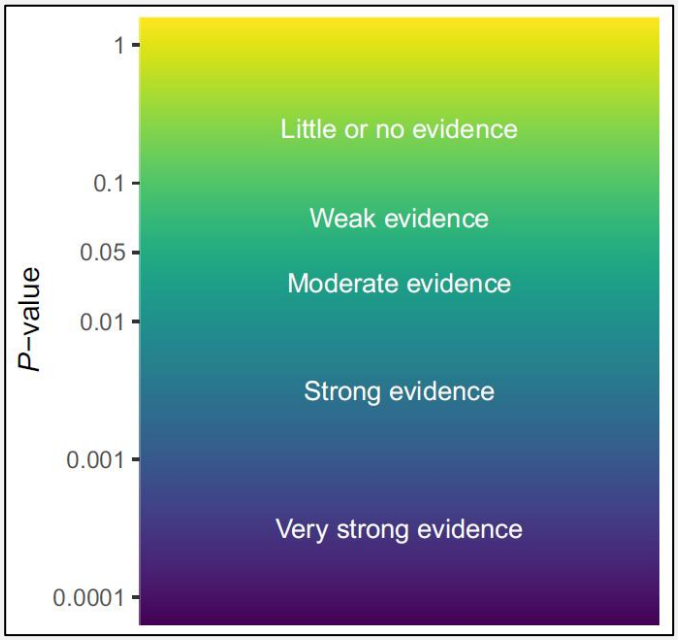


Figure 4. Recommended intervals for translating estimated P-values into the language of evidence
Source: (Muff et al., 2022).

The calculation of RMSE and NMAE as a validity test uses equations (2) and (3). Then together with R^2 and $\bar{R}^2$are presented in the following Table 11:

Table 11. Performance Metrics per Combination

Combination	R^2	$\overline{R}^2$	RMSE	NMAE (%)
1	0.816628741	0.633257	2.41327	10.10152205
2	0.935603228	0.871206	1.43012	5.32713015
3	0.820745745	0.641491	2.38602	9.58011308
4	0.869209652	0.738419	2.03811	8.88680310
5	0.574027060	0.148054	3.67817	14.51012574

The numbers resulting from the RMSE validity test for the five prediction model combinations, namely combination 1, combination 2, combination 3, combination 4, and combination 5, are 2.41327, 1.43012, 2.38602, 2.03811, and 3.67817. The NMAE values are 10.10152205%, 5.32713015%, 9.58011308%, 8.8868031%, and 14.51012574%. The RMSE and NMAE charts in Figure 5 provide an overview of the relationship between the five combinations.

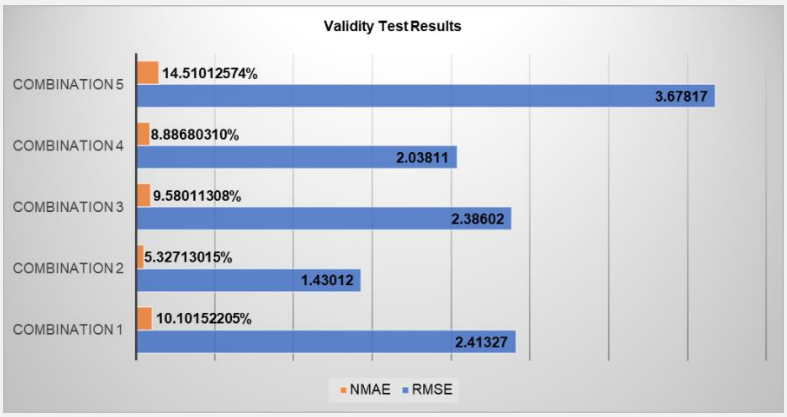


Figure 5. Validity Test Results Chart.

From the validity test results chart, it can be interpreted that the RMSE and NMAE figures, which are the results of testing the five prediction model combinations, have a directly proportional relationship. This means that the higher the RMSE number, the higher the NMAE number will also be. Conversely, the lower the RMSE number, the lower the NMAE number. The lower the number resulting from the validity test, the better the performance of a model. However, the coefficient of determination (R^2) states the opposite; the higher the R^2 value, the better the performance of a model. If depicted through a chart as shown in Figure 4, the R^2 number will be inversely proportional to the RMSE and NMAE numbers. Combination 2, as the best model in salinity prediction, produces the lowest RMSE and NMAE, namely 1.43012 and 5.32713015%, and produces the highest R^2 number, namely 0.935603228.

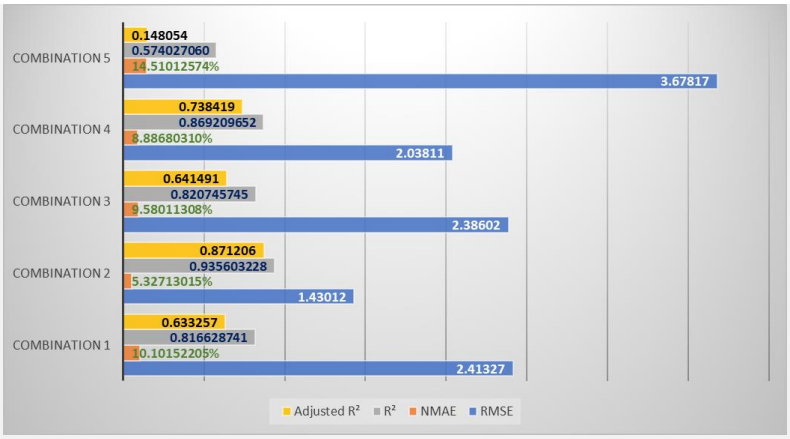


Figure 6. Chart of Validity Test Results, Determination Coefficient (R^2) and Adjusted Determination Coefficient ($\overline{R}^2$).

A study by Muhsi et al. (2022) found an NMAE value of 0.47% and an RMSE value of 0.17, indicating that the developed algorithm model is adequate for estimating seawater surface salinity. The average NMAE and RMSE in this study differ significantly from the previously mentioned studies. The lowest RMSE and NMAE obtained in this study were 1.43012 and 5.32713015%, respectively. This shows the difference in RMSE values between the two studies of 1.26012, indicating that the RMSE of this study is very large, making the prediction model inadequate for salinity prediction. Because this study only used 5 in-situ data sets, while the previous study used around 20 data sets, this is supported by Hoffman (2021), where, of the 6 tables listing RMSE, 4 tables show the highest RMSE values obtained from a small number of observational data. However, the average R^2 value of this study for combination 1 to combination 4 is 0.8, which is in accordance with the range of R^2 values from studies predicting oil palm yields using Landsat 7 satellite data where neural networks with wider, basic, and deeper topologies produce the highest overall prediction accuracy, namely 0.79, 0.77, and 0.85 (Ang et al., 2025). Then the $\overline{R}^2$ produced for the four models is still in the range of 0.633257 to 0.871206 with good to very good criteria. Based on a series of data processing that has been done, the results of the validity test (RMSE and NMAE) and the coefficient of determination (R^2) and $\overline{R}^2$, as well as the components of the bands in each combination of salinity prediction models, can be summarized as in Table 12.

Table 12. Summary of Validity Test Results and Determination Coefficient (R^2)

Combination	Band Combination	RMSE	NMAE (%)	R^2	$\overline{R}^2$
1	Band 1, Band 2, Band 3	2.41327	10.10152205	0.816628741	0.633257
2	Band 1, Band 2, Band 4	1.43012	5.32713015	0.935603228	0.871206
3	Band 1, Band 2, Band 5	2.38602	9.58011308	0.820745745	0.641491
4	Band 2, Band 3, Band 4	2.03811	8.88680310	0.869209652	0.738419
5	Band 2, Band 3, Band 5	3.67817	14.51012574	0.574027060	0.148054

Judging from the ranking of the best model combinations for predicting salinity, the top two are combinations 2 and 4. These two prediction model combinations contain band 2 (blue band) and band 4 (red band). Therefore, these two bands, band 2 (blue band) and band 4 (red band), are the best for salinity prediction.

Table 13. Summary of Salinity Prediction Model Combination Rankings

Rank	Model	Band Combination	RMSE	NMAE (%)	R^2	$\overline{R}^2$
1	Combination 2	Band 1, Band 2, Band 4	1.43012	5.32713015	0.935603228	0.633257
2	Combination 4	Band 2, Band 3, Band 4	2.03811	8.88680310	0.869209652	0.871206
3	Combination 3	Band 1, Band 2, Band 5	2.38602	9.58011308	0.820745745	0.641491
4	Combination 1	Band 1, Band 2, Band 3	2.41327	10.10152205	0.816628741	0.738419
5	Combination 5	Band 2, Band 3, Band 5	3.67817	14.51012574	0.574027060	0.148054

From the data analysis, especially after conducting several tests on the prediction model, both hypothesis testing and validity testing, this research can be used as a pilot study with limited observation data, so that the general application of the results in other regions must be carefully considered.

CONCLUSION

The processing and analysis of image data and field salinity data in modeling salinity prediction provides information that the best combination of prediction models in predicting salinity is the combination prediction model 2, which consists of components of band 1, band 2, and band 4. Then the second-best ranking in salinity prediction is combination 4, consisting of bands 2, 3, and 4. Thus, it is known that band 2 and band 4 are the best bands in salinity prediction. Therefore, it is highly recommended to include these two bands in the salinity prediction model. By using sensing technology, the salinity prediction model will be easier to monitor water salinity levels. This will also help in seawater management, such as planning desalination installations to use seawater as a raw water source. Despite the results of the hypothesis test (F test and t test) and validity test (RMSE and NMAE), which do not meet statistical requirements and do not meet previous research standards, the prediction model shows the potential to predict salinity provided that observation data is added.

ADVANCED RESEARCH

Further research is needed to determine other combinations of Landsat 8 image bands in predicting salinity, the need to use non-linear learning models, and the need to enrich the amount of in

situ data in order to obtain the best combination of salinity prediction models and improve model performance.

ACKNOWLEDGMENT

Thank you to all parties who have played a role in completing this article.

REFERENCES

Abdelmalik, K. W. (2018). Role of statistical remote sensing for Inland water quality parameters prediction. *Egyptian Journal of Remote Sensing and Space Science*, *21*(2), 193–200. https://doi.org/10.1016/j.ejrs.2016.12.002

Abdul Wahid, A., & Arunbabu, E. (2022). Forecasting water quality using seasonal ARIMA model by integrating in-situ measurements and remote sensing techniques in Krishnagiri reservoir, India. *Water Practice and Technology*, *17*(5), 1230–1252. https://doi.org/10.2166/wpt.2022.046

Almeida, R. A., Pereira, S. B., & Pinto, D. B. F. (2018). Calibration and validation of the SWAT hydrological model for the Mucuri river basin. *Engenharia Agricola*, *38*(1), 55–63. https://doi.org/10.1590/1809-4430eng.agric.v38n1p55-63/2018

Andrade, C. (2019). The P value and statistical significance: Misunderstandings, explanations, challenges, and alternatives. *Indian Journal of Psychological Medicine*, *41*(3), 210–215. https://doi.org/10.4103/IJPSYM.IJPSYM_193_19

Ang, Y., Shafri, H. Z. M., Lee, Y. P., Bakar, S. A., Abidin, H., Hashim, S. J., Samad, M. N., Che'ya, N. N., Hassan, M. R., Lim, H. S., Abdullah, R., Yusup, Y., Muhammad, S. A., Yin, T. S., & Gibril, M. B. A. (2025). Block-scale Oil Palm Yield Prediction Using Machine Learning Approaches Based on Landsat and MODIS Satellite Data. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *45*(1), 90–107. https://doi.org/10.37934/araset.45.1.90107

Ansari, M., & Akhoondzadeh, M. (2020). Mapping water salinity using Landsat-8 OLI satellite images (Case study: Karun basin located in Iran). *Advances in Space Research*, *65*(5), 1490-1502. https://doi.org/10.1016/j.asr.2019.12.007

Arafah, F., Taufik, M., & Jaelani, L. M. (2015). Analisis Parameter Kualitas Air Laut di Perairan Kabupaten Sumenep untuk Pembuatan Peta Sebaran Potensi Ikan Pelagis (Studi Kasus: Total Suspended Solid (TSS)). *Seminar Nasional Aplikasi Teknologi Prasarana Wilayah*, 21–30. http://dx.doi.org/10.13140/RG.2.1.4871.8801

Dewangan, L. (2023). The Exact Measurement of Pi. *International Journal for Research in Applied Science and Engineering Technology*, *11*(8), 2217–2233. https://doi.org/10.22214/ijraset.2023.55555

Dimitriadou, S., & Nikolakopoulos, K. G. (2022). Multiple Linear Regression Models with Limited Data for the Prediction of Reference Evapotranspiration of the Peloponnese, Greece. *Hydrology*, *9*(7). https://doi.org/10.3390/hydrology9070124

Enggal, T. W., Bukhori, M., & Sudaryanti, D. (2019). Analisa Bauran Pemasaran Yang Mempengaruhi Keputusan Pembelian Baju di Beberapa Departement Store di Kota Malang. *Jurnal Ilmiah Bisnis Dan Ekonomi Asia*, *13*(2), 61–70. https://doi.org/10.32812/jibeka.v13i2.116

Fitri, N. L., Pangaribuan, D., & Yuniati, T. (2024). Pengaruh Free Cash Flow, Investment Opportunity Set, Dan Struktur Modal Terhadap Kebijakan Dividen Pada Perusahaan Food and Beverage. *SENTRI: Jurnal Riset Ilmiah*, *3*(2), 740–760. https://doi.org/10.55681/sentri.v3i2.2324

Hafeez, S., Wong, M. S., Abbas, S., & Asim, M. (2022). Evaluating Landsat-8 and Sentinel-2 Data Consistency for High Spatiotemporal Inland and Coastal Water Quality Monitoring. *Remote Sensing*, *14*(13). https://doi.org/10.3390/rs14133155

Hidayat, M. N., Wafdan, R., Ramli, M., Muchlisin, Z. A., & Rizal, S. (2023). Relationship between chlorophyll-a, sea surface temperature and sea surface salinity. *Global Journal of Environmental Science and Management*, *9*(3), 389–402. https://doi.org/10.22035/gjesm.2023.03.03

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, *15*(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Hoffman, S. (2021). Estimation of prediction error in regression air quality models. *Energies*, *14*(21). https://doi.org/10.3390/en14217387

Jauza, N. A., & Albina, M. (2025). *Model dan Pendekatan Penelitian Kuantitatif : Kajian Filosofis , Metodologis , dan Aplikatif*. 104–111. https://doi.org/10.61104/qb.v2i1.280

Jin, H., Fang, S., & Chen, C. (2023). Mapping of the Spatial Scope and Water Quality of Surface Water Based on the Google Earth Engine Cloud Platform and Landsat Time Series. *Remote Sensing*, *15*(20), 1–21. https://doi.org/10.3390/rs15204986

Karch, J. (2020). Improving on adjusted R-squared. *Collabra: Psychology*, *6*(1), 45. https://doi.org/10.1525/collabra.343

Lew, S., Glińska-Lewczuk, K., Burandt, P., Kulesza, K., Kobus, S., & Obolewski, K. (2022). Salinity as a Determinant Structuring Microbial Communities in Coastal Lakes. *International Journal of Environmental Research and Public Health*, *19*(8). https://doi.org/10.3390/ijerph19084592

Li, J. (2017). Assessing the accuracy of predictive models for numerical data: Not r nor r2, why not? Then what? *PLoS ONE*, *12*(8), 1–16. https://doi.org/10.1371/journal.pone.0183250

Li, P., Feng, Z., & Xiao, C. (2018). *Acquisition probability differences in cloud coverage of the available Landsat observations over mainland Southeast Asia from 1986 to 2015. 8947*. https://doi.org/10.1080/17538947.2017.1327619

Loewen, C. J. G., & Jackson, D. A. (2024). Salinization, warming, and loss of water clarity inhibit vertical mixing of small urban ponds. *Limnology And Oceanography Letters*, *9*(2), 155–164. https://doi.org/10.1002/lol2.10367

Mamun, M., Hasan, M., & An, K. G. (2024). Advancing reservoirs water quality parameters estimation using Sentinel-2 and Landsat-8 satellite data with machine learning approaches. *Ecological Informatics*, *81*(April), 102608. https://doi.org/10.1016/j.ecoinf.2024.102608

Mobley, C. D., Werdell, J., Franz, B., Ahmad, Z., & Bailey, S. (2016). *Atmospheric Correction for Satellite Ocean Color Radiometry: A Tutorial and Documentation of the Algorithms Used by the NASA Ocean Biology Processing Group*. http://www. NASA/TM–2016-217551. Document ID 20160011399

Mohan, S., Kumar, B., & Nejadhashemi, A. P. (2025). Integration of Machine Learning and Remote Sensing for Water Quality Monitoring and Prediction: A Review. *Sustainability (Switzerland)*, *17*(3). https://doi.org/10.3390/su17030998

Muff, S., Nilsen, E. B., O'Hara, R. B., & Nater, C. R. (2022). Rewriting results sections in the language of evidence. *Trends in Ecology and Evolution*, *37*(3), 203–210. https://doi.org/10.1016/j.tree.2021.10.009

Muhsi, M., Sukojo, B. M., Taufik, M., Aji, P., & Jaelani, L. M. (2022). Estimation of Sea Surface Salinity Concentration from Landsat 8 OLI Data in The Strait of Madura, Indonesia. *Forum Geografi*, *36*(2), 149–159. https://doi.org/10.23917/forgeo.v36i2.19941

Nafizah, N., Jaelani, L. M., & Winarso, G. (2016). Evaluasi Algoritma Wouthuyzen dan Son untuk Pendugaan Sea Surface Salinity (SSS) (Studi Kasus: Perairan Utara Pamekasan). *Jurnal Teknik ITS*, *5*(2), 176-180.

Novoa, S., Doxaran, D., Ody, A., Vanhellemont, Q., Lafon, V., Lubac, B., & Gernez, P. (2017). Atmospheric corrections and multi-conditional algorithm for multi-sensor remote sensing of suspended particulate matter in low-to-high turbidity levels coastal waters. *Remote Sensing*, *9*(1), 61. https://doi.org/10.3390/rs9010061

Octaviana, A., Prasetyo, Y., & Amarrohman, F. J. (2020). Analisis perubahan nilai total suspended solid tahun 2016 Dan 2019 menggunakan citra sentinel 2a (Studi Kasus : Banjir Kanal Timur, Semarang). *Jurnal Geodesi Undip*, *9*(2), 167–176. https://doi.org/10.14710/pilars.%25v.%25i.%25Y.%25p

Ohtani, K., & Tanizaki, H. (2004). Exact Distributions of R^2 and Adjusted R^2 in a Linear Regression Model with Multivariate *t* Error Terms. *Journal of the Japan Statistical Society*, *34*(1), 101–109. https://doi.org/10.14490/jjss.34.101

Pasaribu, R. P., Djari, A. A., Rahman, A., & Handayani, R. (2024). Penentuan Karakteristik Salinitas Menggunakan Conductivity Temperature Depth (Ctd) Di Perairan Pulau Sumba. *Jurnal Kelautan Nasional*, *19*(2), 129. https://doi.org/10.15578/jkn.v19i2.13283

Quynh, B. D., Ha, H., Truong, T. X., & Luu, C. (2024). Geoinformatics for Spatial-Infrastructure Development in Earth and Allied Sciences. In *Proceedings of GIS-IDEAS 2023* (Vol. 411). Springer Nature. https://doi.org/10.1007/978-3-031-71000-1

Rahma, A. A., Adrianto, D., & Malik, K. (2024). Pemodelan Numerik Arus Pasang Surut 2D Menggunakan Software Mike 21 (Studi Kasus Selat Bangka). *Jurnal Hidrografi Indonesia*, *4*(2), 87–94. https://doi.org/10.62703/jhi.v4i2.36

Rusli, H. A. R., Berd, I., Hutasoit, L. M., & Tanjung, D. A. (2023). The Tests Of pH, Conductivity, Total Dissolved Solids, Salinity And Turbidity in the Batang Arau River Surrounding, Padang City-West Sumatra. *BIOLINK (Jurnal Biologi Lingkungan Industri Kesehatan), 10*(1), 25-34. https://doi.org/10.31289/biolink.v10i1.9642

Sahbeni, G. (2021). Soil salinity mapping using Landsat 8 OLI data and regression modeling in the Great Hungarian Plain. *SN Applied Sciences*, *3*(5), 1–13. https://doi.org/10.1007/s42452-021-04587-4

Sood, R., & Zhu, K. (2024). Improving Water Quality Time-Series Prediction in Hong Kong using Sentinel-2 MSI Data and Google Earth Engine Cloud Computing. *arXiv preprint arXiv:2408.14010*. http://dx.doi.org/10.48550/arXiv.2408.14010

Surbakti, H., Nurjaya, I. W., Bengen, D. G., & Prartono, T. (2022). Kontribusi Massa Air Tawar dari Estuari Banyuasin ke Perairan Selat Bangka pada Musim Peralihan II. *Positron*, *12*(1), 29. https://doi.org/10.26418/positron.v12i1.53035

Tran, T. V., Tran, D. X., Pham, H. V., Truong, T. V., Trinh, H. P., Nguyen, D. Q., Nguyen, B. A., & Nguyen, H. C. (2020). Exploring spatial relationship between electrical conductivity and spectral salinity indices in the Mekong delta. *Journal of Environmental Science and Management*, *23*(1), 39–49. https://doi.org/10.47125/jesam/2020_1/05

U.S. Geological Survey. (2024). *Landsat-8 imagery of Bangka Strait area* [satellite image data]. EarthExplorer. Retrieved on February 27, 2025, from https://earthexplorer.usgs.gov

Wulandari, S. A., Sucipto, A., Rosyady, A. F., Ardana, M. D. R., Cahyono, O. D. P., & Khomarudin, A. N. (2024). Rancang Bangun Sistem Monitoring Kualitas Air Untuk Mendeteksi Keadaan Tidak Normal atau Penyakit Pada Tambak Ikan Mujaer Menggunakan Fuzzy Logic Mamdani Berbasis Mobile. *Technologica*, *3*(1), 42–54. https://doi.org/10.55043/technologica.v3i1.153

Zaka, A. R., & Sutopo, S. (2017). Analisis Faktor-Faktor Yang Mempengaruhi Kepuasan Pelanggan Pada LBB Antologi Semarang. *Diponegoro Journal of Management*, *6*(3), 33-45.

Zhao, L., Li, Z., & Qu, L. (2022). Forecasting of Beijing PM2.5 with a hybrid ARIMA model based on integrated AIC and improved GS fixed-order methods and seasonal decomposition. *Heliyon*, *8*(12), e12239. https://doi.org/10.1016/j.heliyon.2022.e12239