Using Machine Learning (Multiple Regression Model) to Conduct Satellite Derived Bathymetry with Sentinel-2 in Sattahip Bay, Thailand.

สัณหณัฐ อิสมาแอล[1]

Sanhanat Itsama-ael[1]

Received: April 26, 2024

Revised: May 27, 2024

Accepted: May 28, 2024

Abstract

The aim of this study is to utilize machine learning (Multiple Regression Model) to perform Satellite Derived Bathymetry (SDB) with six image bands, namely blue, green, red, near-infrared, NDWI, and NDVI, from Sentinel-2 in Sattahip Bay, Thailand. The results demonstrated that the multiple regression model achieved higher accuracy (48.27%) compared to the ratio algorithm approach with the same areas and datasets. However, the depth values from the multiple regression model cannot meet the standard set by the International Hydrographic Organization (IHO). Nonetheless, they can serve as an initial identifying underwater objects, enhancing the safety of hydrographic surveyors, and reducing survey time.

Keywords: Machine Learning, Satellite Derived Bathymetry, Multiple Regression Model, GIS, SDB

[1]ประจำ กพ.ทร.

Attached to Naval Personnel Department

E-Mail: sanhanat@buffalo.edu, mee_sanhanat52@hotmail.com

1. Introduction

Bathymetric information is crucial for mariners and shoreline residents alike. For seafarers, this data ensures the safety of navigation at sea, aiding in the development of effective sailing plans to reach destinations safely. Additionally, coastal topography along the shorelines assists in determining location, preventing sailors from getting lost. Moreover, bathymetric information is necessary for preserving marine ecosystems. For instance, mangrove forests, acting as nurseries for marine life, provide protection against storm surges. Coral reefs also play a vital role, serving as homes and shelters for various marine species. These ecosystems, in turn, contribute to the economy, especially in maritime shipping, a key mode of transportation for import and export products. Accurate and adequate hydrographic information is essential for safe navigation of ships at sea. Additionally, coastal areas, including beaches, islands, and coral reefs, contribute to the tourism industry. Therefore, bathymetric information is undeniably important.

Bathymetric surveys are conducted on a large scale, requiring significant budgets and time when using SONAR and GNSS equipment on survey ships or when surveyors walk along shorelines collecting positions with GNSS. Furthermore, some areas are inaccessible due to being dangerous for survey ships and teams to reach, posing obstacles to obtaining bathymetric and shoreline information.

Nowadays, there is a method called satellite derived bathymetry (SDB) that is used by many countries globally because satellite images from reliable platforms such as Landsat 8 and Sentinel-2 are freely accessible. In terms of satellite derived bathymetry algorithm, Stumpf et al (2003) introduced the algorithm using fundamental of the ratio transform or empirical method, which is popular and accurate [1]. However, there is another SDB method that uses machine learning approaches that are able to give high accuracy such as multiple regression model.

With the advancement in remote sensing, the challenges above can be overcome by some remote sensing techniques because remote sensing is able to obtain large scales of information without in contact those survey areas. Therefore, the objective of this project is to conduct SDB with the machine learning approach in Sattahip Bay, Thailand, using reference bathymetric data from Hydrographic Department of Royal Thai Navy (HDRTN) and satellite images from Sentinel-2, and validate the results whether they can meet International Hydrographic Organization (IHO) standard or not.

2. Literature Review

SDB is a remote sensing technique that uses satellite images to measure depth values. This technique has been developed since 1978 with a standard algorithm employing a linear transform [2]. The equation is shown in Equations (1) and (2)

$$Z = a_0 + a_i X_i + a_j X_j \tag{1}$$

$$X_i = \ln[R_w(\lambda_i) - R_\infty(\lambda_i)] \tag{2}$$

where z is the depth; a_0, a_i, and a_j are coefficients determined from multiple regression [2]. In terms of X_i and X_j, they are derived from Equation (2), where R_w is the reflectance of the water and R_∞ is the water column reflectance if the water is optically deep [2].

As can be seen from the standard algorithm, there are five variables that require tuning that are a_0, a_i, a_j, R_w, *and* R_∞ , making this equation more complicated and less effective in areas with a seafloor that has an extremely low albedo [1]. The main reason why the linear transform algorithm struggles in low albedo areas is that if Ad (bottom albedo) is lesser than R_∞ in Equation (3)

$$R_w = (A_d - R_\infty)\exp(-gz) + R_\infty \tag{3}$$

this causes R_w to be a negative value. Thus, the Equation (2) will fail, because the inside logarithm cannot be negative. Consequently, a simpler and more accurate algorithm was developed by using the ratio transform, which required only two bands, blue and green, as shown in Equation (4) [1].

$$Z = m_1 \frac{\ln(band\ blue)}{\ln(band\ green)} + m_0 \tag{4}$$

It is evident that this equation is the same as the linear regression equation y = ax + b. The ratio transform utilizes the fundamental concept that blue and green bands can penetrate through shallow water, but the green band decays faster than the blue band, resulting in an algorithm that increases with depth.

The ratio algorithm was applied to IKONOS images with spatial resolution of 4 meters to compare the results with the linear transform algorithm. The result showed that the ratio algorithm can obtain depth values in area up to 25 meters with clear water, while the linear transform algorithm only distinguishes depth up to 15 meters. Moreover, the ratio transform can be used in extremely low albedo areas, whereas the linear transform algorithm cannot, given a spatial resolution of four meters or more [1].

In addition, the ratio algorithm was compared with the regression kriging algorithm using satellite images from Landsat 8, RapidEye, and Pleiades. The result showed all sources of satellite images provide satisfactory results, but Landsat 8 demonstrated more accuracy with linear regression than regression kriging. Meanwhile, the results with linear regression than regression kriging from Pleiades showed high accuracy. Overall, regression kriging performed better than linear regression [3].

In terms of the machine learning approach, it utilizes the relationships between the satellite images and bathymetric depths from the training data, and the training model is used to derive the rest of the depth values in the study area [4]. There are many machine learning models that have been used for SDB such as random forest regressors, Convolution Neural Networks, Support Vector Machine, U-Net, etc. [4]. However, a multiple regression model was used in this study. In multiple regression, there is one dependent variable and several independent variables instead of one independent variable in linear regression as shown in the Equation (5) [5].

$$y' = a + b_1x_1 + b_2x_2 + \cdots + b_kx_k \tag{5}$$

Where y′ is the dependent variable in this case, which is depth values, x_1, x_2, ..., x_k are independent variable that are image bands for SDB, b_1, b_2, ..., b_k are coefficients for each independent variables, and a is intercept.

From the previous studies above, the ratio algorithm has been used with various kinds of satellite images because it is less complex and highly accurate as compared to other methods. However, it is interesting to use the machine learning approach for SDB with Sentinel-2 since the multiple number of image bands can be included in the multiple regression model, resulting in more accurate and flexible predictions. Therefore, the multiple regression model was utilized with Sentinel-2 in this research.

3. Study Area and Data

The Sattahip Bay, Thailand, was selected as the study area because its proper topographic and hydrographic characteristics. As depicted on the Thai nautical chart no. 115 [6] in Figure 1, there are various depth areas and dangerous objects (e.g., underwater rocks, wrecks, and shallow areas), making it particularly useful for testing whether SDB can detect them or not. Furthermore, all required data, especially bathymetric data, can be obtained in this area.

Firstly, Thai nautical chart no. 115 [6] was used as background as well as topographic and hydrographic reference. Secondly, bathymetric data in point cloud format in the study area from HDRTN was used as the reference depth to compute SDB and validate the results [7]. Next, satellite images from Sentinel-2 in 2023 were used, as shown in Table 1. Finally, ArcGIS Pro, Microsoft Office Excel, and Python were used to conduct SDB, multiple regression model and compute statistics.

Table 1 Satellite Bands Used.

Spectrum	Sentinel 2
Blue	Band 2
Green	Band 3
Red	Band 4
NIR	Band 8

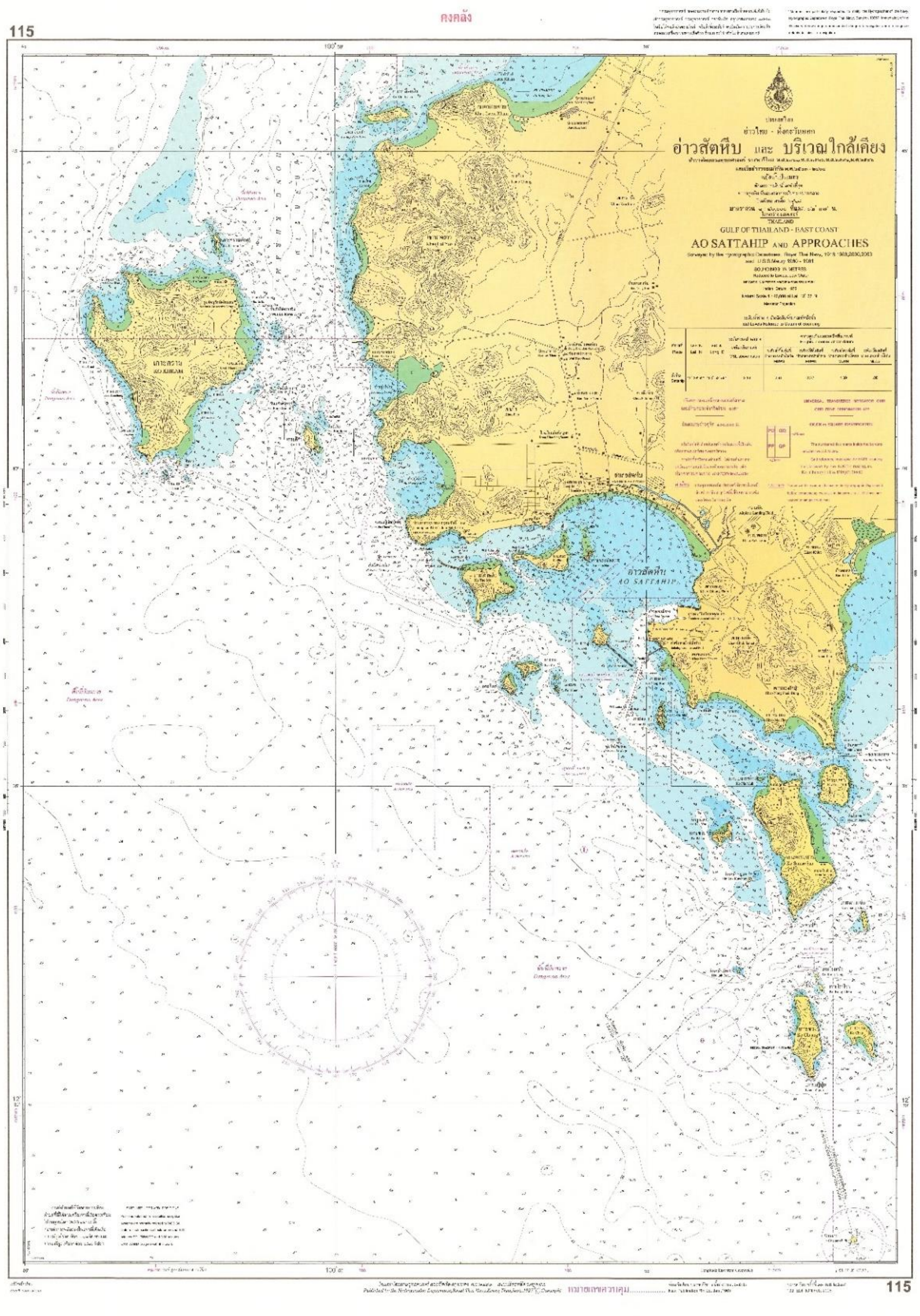

Figure 1 Thai Nautical Chart No.115, Sattahip Bay [6]

4. Methods

According to chapter 11 of the IHO-IOC GEBCO Cook Book, "LANDSAT 8 Satellite-Derived Bathymetry" demonstrates how to conduct SDB using Landsat 8 [8], which was used as reference for the methodology in this project. However, some steps were adjusted to suit this project, such as using a machine learning approach instead of the empirical method, utilizing a Normalized Difference Water Index (NDWI) and an NDWI threshold to remove the land area. Additionally, the images from Sentinel-2 were used in this project due to its higher spatial resolution, which is 10-meter spatial resolution. The SDB method with machine learning approaches can be divided into five steps below.

4.1 Download and Filter Data

As discussed in the previous section, Sentinel-2 bands blue, green, red and NIR were downloaded from European Space Agency (ESA), as well as bathymetric data in point cloud format in the study area were provided by HDRTN. Afterward, each pixel value was converted into a floating – point representation for finer resolution. Then, all bands were filtered with the low pass filter tool to obtain smoother images that mitigate outlines in images.

4.2 Remove Land Areas

Since water areas are only required for the SDB algorithm, the water areas were extracted using the Normalized Difference Water Index (NDWI) and NDWI threshold, as shown in Equation (6) and (7).

$$NDWI = \frac{(Green-IR)}{(Green+IR)} \tag{6}$$

$$T = (mean_w + mean_L)/2 \tag{7}$$

where $Green$ is green band, IR is infrared band, T is NDWI threshold, $mean_w$ is mean of NDWI value on the water, and $mean_L$ is mean of NDWI value on the land. Firstly, an NDWI surface was generated using Equation (6), as depicted in Figure 2. Then, a sample area was taken using the coastline on the nautical chart as a reference for computing the NDWI threshold to obtain means of water and land areas, as shown in Figure 2. After that, the NDWI threshold was computed based on Equation (7). Lastly, the land areas of each image band were removed, as presented in Figure 2.

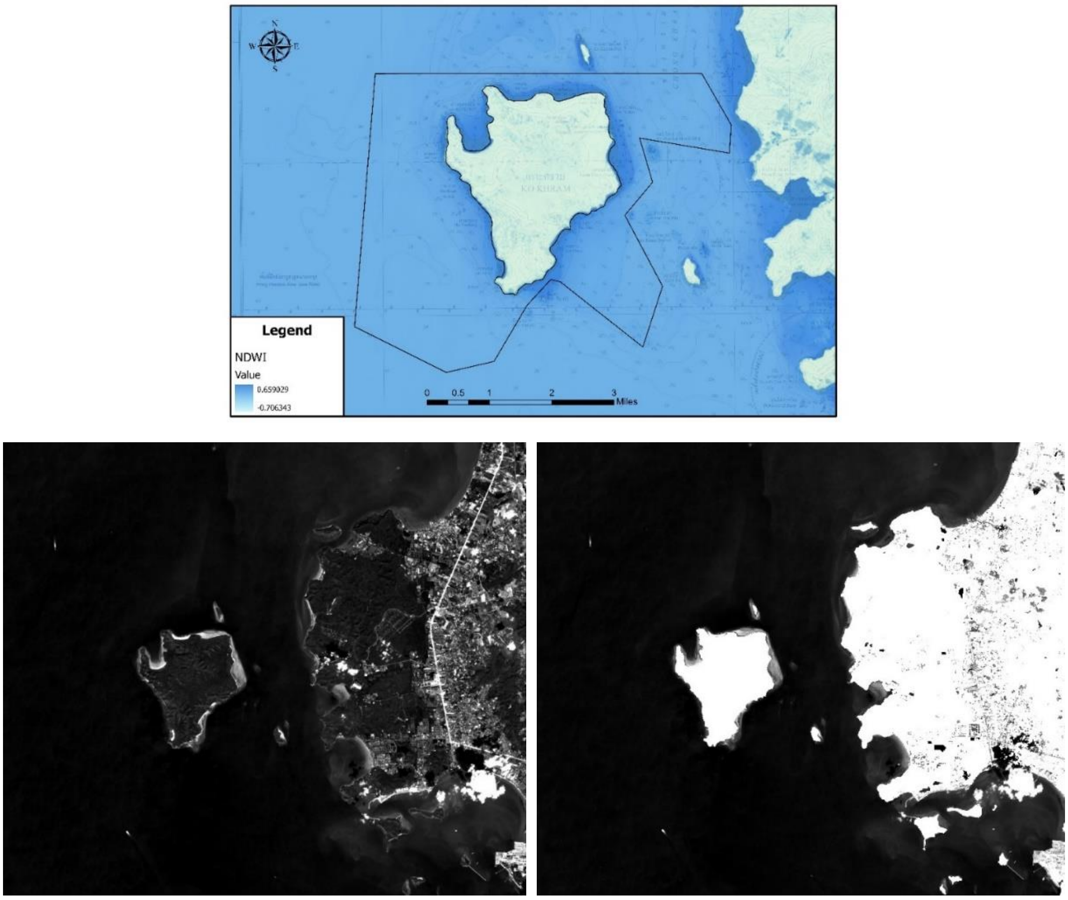

Figure 2 The Sample Area for the NWDI Threshold (Top), before Remove Land Areas (Bottom Left), after Remove Land Areas (Bottom Right)

4.3 Generate the Multiple Regression Model

As multiple regression models require several independent variables, six image bands were introduced to the model: blue, green, red, near infrared (NIR), NDWI, and NDVI. In the case of NDVI, red and NIR bands were applied with Equation (8) to obtain the NDVI layer.

$$NDVI = \frac{(NIR-Red)}{(NIR+red)} \quad (8)$$

Then, bathymetric information from HDRTN was used to be the dependent variable, and the area between the mainland and Ko Khram Island was chosen as the training area, as shown in Figure 3. This area was selected due to its various water depth ranges from 0 to 25 meters, which is the maximum depth that the ratio algorithm could survey based on the previous studies. Hence, 772 samples were selected in this area to build the model. Nevertheless, samples deeper than 10 meters were removed from the model because, after 10 meters, the model began to saturate based on the ratio algorithm. This implies that based on the atmospheric and environmental conditions at that time the images were taken, the blue and green lights could penetrate through the water to a depth of 10 meters. Consequently, a ten-meter depth was set as the maximum depth prediction in this study.

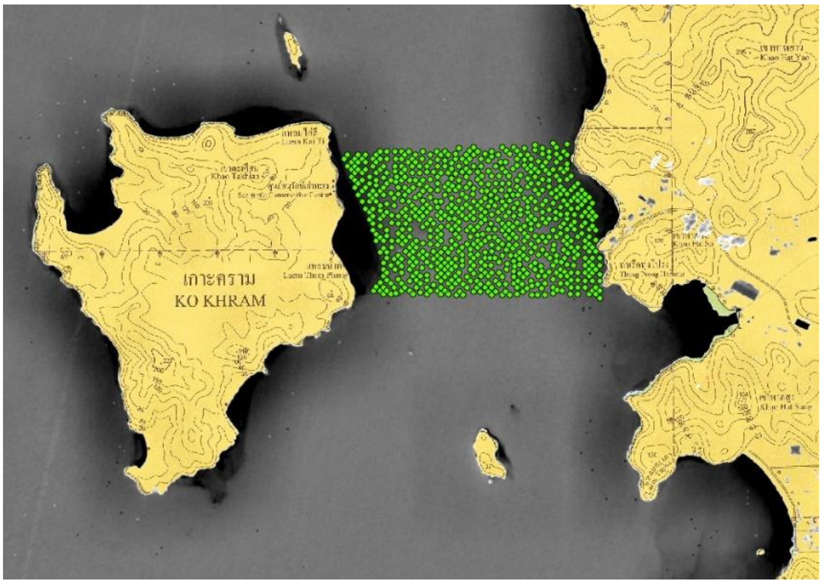

Figure 3 The Training Area and 772 Samples for Building the Model [6].

After all independent variables were prepared, each layer was extracted based on the bathymetric information points in the training area, resulting in a spreadsheet in CSV format containing independent and dependent variables. After that, those variables were fed into the machine learning model, which was a multiple regression model using Python programming language, based on the Equation (9), to obtain all coefficients (b_1, b_2, ..., b_6) and intercept (a).

$$SDB\ Depth = a + b_1 Blue + b_2 Green + b_3 Red + b_4 NIR + b_5 NDWI + b_6 NDVI \quad (9)$$

4.4 Apply the Linear Formulas

After obtaining the model as depicted in Equation (9) from section 4.3, the results were obtained by applying these models back in ArcGIS Pro using the “Raster Calculator” tool. The results are the surface layers representing depth values, as shown in Figure 4.

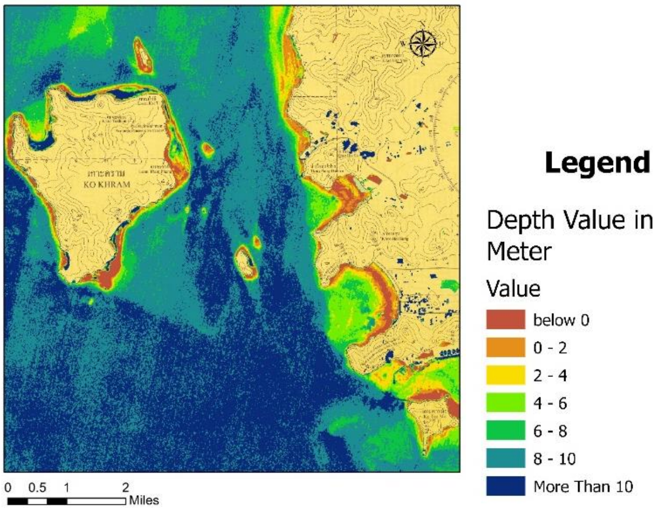

Figure 4 The Final Result from Sentinel-2 [6]

4.5 Validation

To validate the result, testing sets were selected from the areas where depth values ranged from 0 to 10 meters because the model was generated to predict depth less than 10 meters. There were 341 depth values in the testing sets chosen throughout the study area, as presented in Figure 5.

Then, “Extract Value to Point” tool in ArcGIS Pro was utilized to extract a set of paired values between reference depth and SDB depth values. This set of pair values was exported in CSV format for validation in Python.

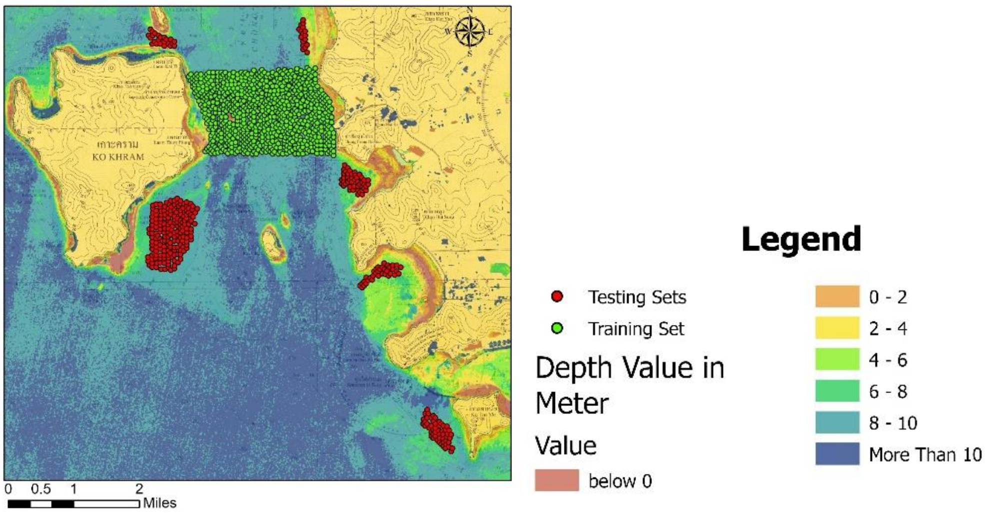

Figure 5 Training and Testing Sets [6]

Python was used to compute RMSE and a 95% confidence interval overall and for each range of depth value so that the trend of accuracy can be analyzed. Firstly, “Pandas” and “Math” packages were imported because Pandas is a powerful package for data manipulation, and “Math” is useful for complex computation. Then, the set of paired values between reference depth and SDB depth values was imported into Python to be a data frame using the Pandas package. After that, the data frame was converted into a dictionary. Lastly, the dictionary was input into the functions that compute RMSE and a 95% confidence interval overall and for each depth range such as 2 to 3 meters, 3 to 4 meters, 4 to 5 meters, and so on.

5. Results and Discussion

After the raw images were filtered with a low-pass filter, they became smoother, mitigating noises and outlines as compared to the result without the filter. Regarding the NDWI threshold, the means of NDWI values for water and land areas were 0.084 and -0.508, respectively, as shown in Table 2. Hence, the NDWI threshold as computed using Equation (7) was -0.212. As seen from the result in Figure 6, areas that experience low tide effects such as cliffs or high-slope shorelines, showed a higher accuracy of coastline delineation compared to beached areas or low-slope shorelines. This discrepancy may be due to tide effects since the shoreline in the nautical chart represents the coastline during the highest tide, particularly during spring tide. However, the images were not adjusted for tide elevation due to limitation in available information. Therefore, considering tide effects for coastline detection is an area of improvement for future research.

Table 2 Statistic Parameters of Water and Land Area Samples for Computing NDWI Threshold

Types	Min	Max	Mean	STD	Median
Water	-0.527	0.659	0.084	0.103	0.047
Land	-0.650	0.611	-0.508	0.151	-0.545

Figure 6 The Image after Remove Land in a High Slope Region (Left) [6], The Image after Remove Land in a Low Slope Region (Right) [6]

After applying the SDB algorithm, the training set of 722 samples was initially used to generate the multiple regression model. However, upon ploting 772 samples, it was found that the model trend saturated after a depth of 10 meters. Consequently, samples with depths greater than 10 meters were removed, leaving 143 samples for generating the model. This suggests that, based on environmantal conditions such as transperancy, atmosphere, and light attenuation at that time, the blue and green bands could penatrate through the water only up to 10 meters. Sattahip bay, Thailand, is located near four rivers that are Mae Klong, Tha Chin, Chao Phraya, and Bang Pakong, resuling in high sediment loads from the land, leading to low transperency in the study area. The result of mutiple regression model is presented in Figure 4 and Equation (10) with R^2 = 0.841.

$$SDB\ Depth = 7.44221 + 0.03396 Blue - 0.0002 Green - 0.03986 Red + 0.010112 NIR - 130.79153 NDWI - 139.35844 NDVI \quad (10)$$

To validate the results, testing sets were chosen throughout the study area, as shown in Figure 5, considering a variety of water depth ranging from 0 to 10 meters in the testing areas based on the models that can predict depth between 0 and 10 meters. Thus, these areas were selected. After selecting these 341 samples, standard deviations, and Root Mean Square Error (RMSE) were computed, along with 95 % confidence intervals. Finally, the 95 % confidence intervals were compared with the Standard of Hydrographic Survey. The results indicate that the multiple regression model with Sentinel-2 cannot meet the Standard of

Hydrographic Survey because the 95 % confidence intervals from the results exceed the maximum 95 % confidence intervals allowed by the standard, as shown in Table 3.

Table 3 The Sentinel 2 Validation Results in Overall with Multiple Regression Model.

Observation		IHO Standard with 10-meter depth			
RMSE	95% Con	Order 2	Order 1b	Order 1a	Special Order
0.994	1.948	1.026	0.517	0.517	0.261

Additionally, 95 % confidence intervals by depth ranges were computed, and all depth ranges failed to meet the standard. The result demonstrated that high discrepancies occurred in depths below five meters, but the accuracy improved in the depths ranging from five to nine meters, as shown in Figure 7. However, the 95 % confidence interval from the multiple regression model gave higher accuracy than the linear regression model from the empirical method at 48.27 %, as shown in Table 4

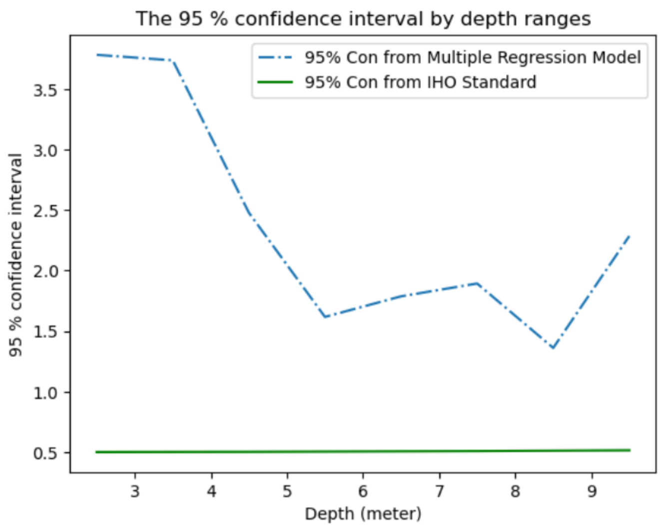

Figure 7 The 95 % confidence interval by depth ranges

Table 4 The Sentinel 2 Validation Results in Overall with Linear Regression Model.

Observation		IHO Standard with 10-meter depth			
RMSE	95% Con	RMSE	95% Con	RMSE	95% Con
1.921	3.766	1.921	3.766	1.921	3.766

Although the results cannot meet the Standard of Hydrographic Survey and cannot be used to create a nautical chart, there are some advantages that can be useful for hydrographic survey. Firstly, SDB can assist hydrographers in initially identifying underwater objects, as most underwater rocks can be detected by SDB, as presented in Figure 8. This capability can reduce the time and cost required to search for these objects from scratch, ultimately enhancing the safety of hydrographers, survey equipment, and survey ships.

In terms of “Search and Rescue” (SAR) for a sinking ship or plane crash, searching for underwater objects using survey ships can be a lengthy process, especially in large areas. Therefore, SDB can be used as an initial search to identify the most possible spots for a sinking ship or plane crash.

In remote areas that are not accessible by survey ships, SDB can serve as a tool for hydrographic surveying. Hydrographic offices can utilize SDB in such areas, but it should be noted on the map that the survey was conducted using SDB. While having some water values on the map is preferable for mariners compared to leaving it blank, clear communication about the survey method is crucial.

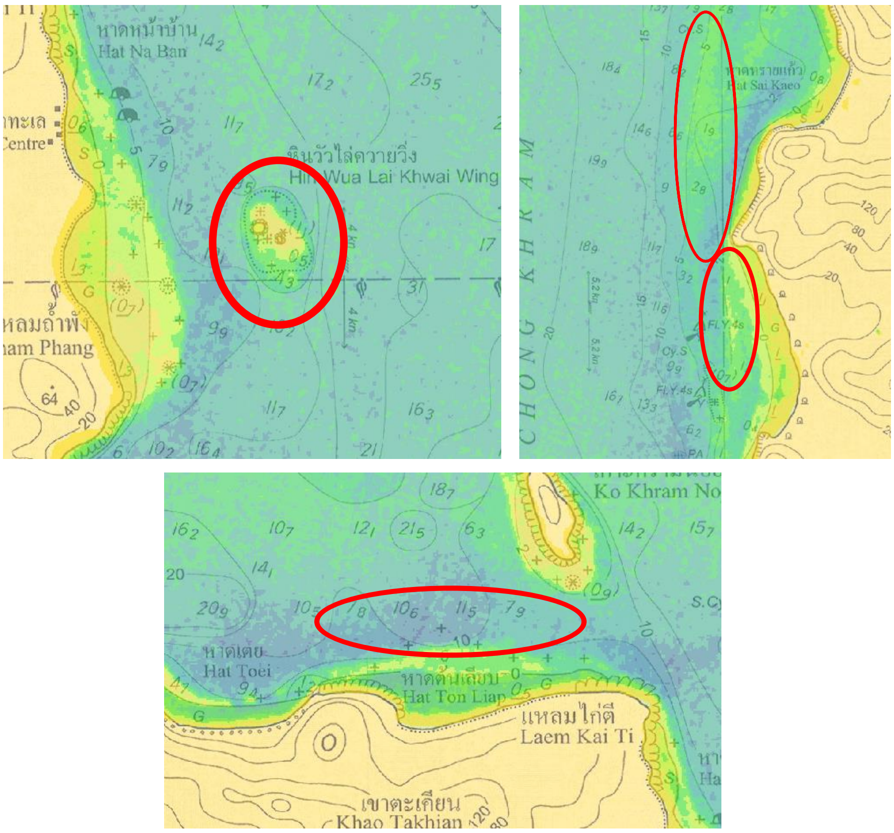

Figure 8 Underwater Rocks Detected by SDB [6].

6. Conclusions

Bathymetric information is fundamental for various activities such as water navigation, ecosystem preservation, maritime shipping, and the tourist industry. Inaccurate and insufficient bathymetric information can have significant effects on these activities, particularly for coastal countries that rely on these activities for economic purposes. Therefore, Bathymetric information is crucial for a country's prosperity.

However, on-site hydrographic survey methods require substantial budgets, and they are time consuming. Consequently, this is an area where remote sensing can bridge the gap by using satellite derived bathymetry (SDB) to obtain bathymetric information. Hence, the objective of this study is to apply the SDB technique with multiple regression model and Sentinel-2 images to measure depth values in Sattahip Bay, Thailand, and validate the results against the IHO standard.

From the results, it is observed that multiple regression model with Sentinel-2 images cannot meet the standard. However, they can be initially used to pre-identify underwater objects such as underwater rocks and wrecks, allowing hydrographers to save time in searching for them from scratch. In addition, SDB is useful for mapping unreachable areas and search and rescue (SAR) missions, such as sinking ships and plane crashes. In terms of accuracy, the multiple regression model preformed higher accuracy than the ratio algorithm, which utilizes linear regression model at 48.27%.

For further study, it is recommended that water transparency in the study area should be considered in the result because transparency plays a significant role in SDB [8]. Additionally, other hydrographic survey standards, such as United Kingdom Hydrographic Office (UKHO) should be used for comparison with SDB results. Lastly, to enhance accuracy, tide information should be taken into account in the NDWI threshold and when subtracting the final depth values.

Reference

[1] Stumpf RP, Holderied K, Sinclair M. Determination of water depth with high-resolution satellite imagery over variable bottom types. Limnol Oceanogr [Internet]. 2003 [cited 2024 Jan 23];48(1):547-56. Available from: https://doi.org/10.4319/lo.2003.48.1_part _2.0547

[2] Lyzenga D. Remote sensing of bottom reflectance and water attenuation parameters in shallow water using aircraft and LANDSAT data. Int J Remote Sens [Internet]. 1981 Jan [cited 2024 Jan 27];2(1):71-82. Available from: https://www.researchgate.net/publication/233151938

[3] Cahalane C, Magee A, Monteys X, Casal G, Hanafin J, Harris P. A comparison of Landsat 8, RapidEye and Pleiades products for improving empirical predictions of satellite-derived bathymetry. Remote Sens Environ [Internet]. 2019 [cited 2024 Jan 30];233:111414. Available from: https://doi.org/10.1016/j.rse.2019.111414

[4] International Hydrographic Organization. International Hydrographic Organization: Guidance to Satellite-Derived Bathymetry: B-13 Edition 1.0.0 [Internet]. Monaco: International Hydrographic Organization; 2024 [cited 2024 Feb 1]. Available from: https://iho.int/uploads/user/pubs/bathy/B_13_Ed100_032024.pdf

[5] Bluman AG. ELEMENTARY STATISTICS: A STEP BY STEP APPROACH [Internet]. 9th ed. New York: McGraw-Hill; 2014 [cited 2024 Feb 2]. Available from: https://students.aiu.edu/submissions/profiles/resources/onlineBook/w3Q8y4_Elementary%20Statistics_%20Step-By-Step%20Approach%209th%20ed%20(2014).pdf

[6] Hydrographic Department of Royal Thai Navy. Thailand 1: 40000 Nautical Chart Number 115 Ao Sattahip. Bangkok: Hydrographic Department of Royal Thai Navy; 2006.

[7] Hydrographic Department of Royal Thai Navy. Thailand Bathymetric Data: Number 115. Bangkok: Hydrographic Department of Royal Thai Navy; 2006.

[8] International Hydrographic Organization. The IHO-IOC GEBCO Cook Book: IHO Publication B-11: IOC Manuals and Guides 63 [Internet]. Monaco: International Hydrographic Organization; 2019 [cited 2024 Feb 5]. Available from: https://www.star.nesdis.noaa.gov/socd/lsa/GEBCO_Cookbook/documents/CookBook_20191031.pdf

[9] International Hydrographic Organization. International Hydrographic Organization: Standard for Hydrographic Surveys: S-44 Edition 6.0.0 [Internet]. Monaco: International Hydrographic Organization; 2020 [cited 2024 Feb 12]. Available from: https://repository.oceanbestpractices.org/bitstream/handle/11329/388.2/S-44_Edition_6.0.0_EN.pdf?sequence=5&isAllowed=y