

# Seasonal Variation Influence to Water Image Properties to Retrieve Nearshore Bathymetry Based on Cloud Machine Learning Approach

Kurniawan, A.,<sup>1</sup> Khakhim, N.<sup>2\*</sup> and Wicaksono, P.<sup>2</sup>

<sup>1</sup>Coastal and Watershed Management Planning, Postgraduate Geography, Faculty of Geography, Universitas Gadjah Mada, Indonesia, E-mail: agung.kurniawan.16@mail.ugm.ac.id

<sup>2</sup>Department of Geography Information Science, Faculty of Geography, Universitas Gadjah Mada, Indonesia E-mail: nurulk@ugm.ac.id,\* prama.wicaksono@ugm.ac.id

\*Corresponding Author

DOI: <https://doi.org/10.52939/ijg.v20i7.3407>

## Abstract

*This research aims to develop a computational framework for shallow water bathymetry reconstruction using machine learning-based Satellite-derived bathymetry (SDB) running on cloud computing. The RF and LR algorithms were tested for performance by considering the influence of seasonal variations. Both algorithms were trained using bathymetric data from hydrographic surveys, converted to the number of test and validation samples which determine the number independently. The accuracy test considering quantitative aspects through RMSE, MAE and  $R^2$ , as well as qualitative aspects using cross-sectional transects of underwater topography and 1:1 plot. The complex bottom topography and supported by various benthic varieties causes differences in the water reflectance of in each season, it is necessary to analyze their influence on the machine learning algorithm in SDB. Overall, the best RMSE, MAE, and  $R^2$  were produced by the RF algorithm in transition season II with values of 0.34 m, 0.21 m, 0.944 respectively. For the LR algorithm, the best performance is shown in the east season with respective accuracies of 0.60 m, 0.46 m, 0.83. Through cross-sections of underwater topography, SDB algorithm can represent accurately in various geomorphological bottom variations, such as lagoons and reef flats. The LR algorithm is not yet able to optimally reconstruct shallow water bathymetry because outlier values in the accuracy test by 1:1 plot. In general, the RF and LR algorithms show high accuracy results at depths of up to 2 meters, and accuracy tends to decrease at depths > 3 meters. Through this study we found a relationship between the low reflectance of waters in the west season, which is correlated with the low performance of the SDB RF and LR algorithms. This study provides a cloud computing framework for the SDB reconstruction, efficiently in time and storage facilities without leaving any residue. The impressive archive facilities also enable multi-season analysis.*

**Keywords:** Cloud Computing, Machine Learning, Seasonal Variations, Shallow Water Bathymetry, Tested Performance

## 1. Introduction

Remote sensing-based computing technology makes it possible to carry out monitoring activities in remote areas and shallow waters effectively and efficiently [1][2] and [3]. This technique makes it possible to obtain estimates of shallow water bathymetry in certain areas using the Satellite-derived bathymetry (SDB) method [4] by utilizing optical multispectral remote sensing images [5] and [6], as well as radar images, which can be real aperture radar (RAR) or synthetic aperture radar (SAR) [7]. SDB is a series of techniques for reconstructing depth based on remote

sensing sensors [8]. Nowadays, multispectral optical remote sensing imagery has become the most popular alternative for use in rapid mapping of shallow water bathymetry [9] and [10] with the main consideration being the low cost aspect [3], fast processing time [2], and the flexibility of algorithm modification and optimization available [11] and [12]. Multispectral imagery is able to describe depth estimates through reflectance schemes from shallow waters [6] which reflects electromagnetic energy radiation in the visible spectrum and is recorded by sensors [13].

However, this energy reflection has limitations due to a gradual decrease in energy intensity due to Inherent Optical Properties (IOP) and turbidity in the water column [13] and [14]. So, when reconstructing depth models, the optical properties of water are generally taken into account, which are described by [12], including the spectral characteristics of suspended solids and dissolved substances as well as basic reflectance, such as the concentration of chlorophyll-a (Chl-a), the diffuse attenuation coefficient of the water body, detritus concentration, spectral shape, absorption, and backscatter coefficients.

Common approaches known for reconstructing water depth using the SDB method on optical multispectral images include statistical and physics-based ones. Statistical-based approaches are classified into empirical methods, while Physics-Based Approaches are classified into semi-empirical and analytical methods (semi-analytical and quasi-analytical) [13]. The main difference in these two approaches lies in the role of the presence of depth samples [1]. The statistical approach requires in-situ data as a variable to carry out "training" [15], while the Physics-Based Approach is concerned with and emphasizes the passage of electromagnetic waves and their attenuation in the atmosphere and water [16]. A consideration that needs to be noted is the precision aspect, and the accuracy is relatively lower than traditional survey methods (Echo-Sounder or Side scan sonar) with depth coverage that only reaches optically shallow waters that require clear visibility [3]. However, SDB-based water depth reconstruction allows researchers to obtain depth data quickly, cheaply, and efficiently over large areas [3][17] and [18].

Optimization of empirical methods in statistical approaches has increased rapidly with the development of machine learning-based computing. The SDB method in reconstructing depth utilizes linearity between reflectance values and in-situ depth samples [3], but there are conditions where this linearity cannot be met, so statistical approaches such as machine learning (ML) or Neural Network (NN) are used to reduce the sampling error rate to overcome the non-linear relationship between reflectance and in-situ depth data [3][19][20][21] and [22]. ML algorithms such as convolutional neural networks (CNN) have been used to reconstruct bathymetry on coasts and reservoirs using different bands simultaneously [21] and [23]. Other approaches such as random forest (RF) are also compatible with high resolution imagery to obtain detailed results in coastal areas [24] and [25].

In coastal and near-shore areas, reflectance values can show different intensities even in the same area due to differences in seasons (for example: wet or dry

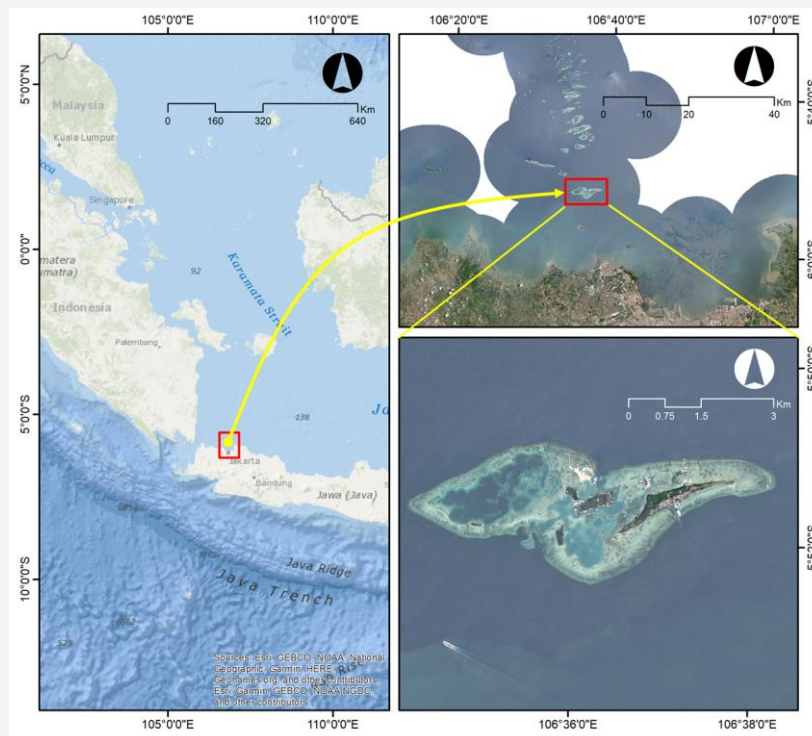
season) [26] and [27]. According to [28] in the southwest monsoon (SWM) period which is synonymous with rain and wetness, it brings material content from upstream to estuaries and areas near the coast and influences colored dissolved organic matter (CDOM), but in other seasons the influence mainly influenced by microbial activity, anthropogenic sources, high temperature and radiation. We believe that there is a need for an in-depth study of the influence of season on the quality of the reflectance value produced by each recorded image, and its correlation with the resulting SDB.

The ML approach was chosen to reconstruct SDB for seasonal variations in shallow waters to deal with water complexity due to coral reef cover and benthic variations resulting in differences in energy reflectance [3] and [29]. This study aims to evaluate the effect of seasonal variations on the quality of image reflectance for optically reconstructing bathymetry in shallow waters using machine learning techniques with complex topographic variations. This evaluation plays an important role in selecting multispectral optical images that are adapted to the current season, so that this research can complement previous research [1][3][19][20][25] and [30]. ML algorithms applied in current SDB studies include random forest (RF), convolutional neural networks (CNN), and support vector machine (SVM), which are optically capable of providing good accuracy [3][19] and [30]. Through studies [3] RF is able to produce products with better accuracy than SVR and linear regression.

## 2. Material and Methods

### 2.1 Study Area

This study was conducted in the waters of Pari Island, which is part of the Thousand Islands chain, DK Jakarta Province. Based on absolute location, this inhabited island is located at 106°34'4.28" E – 106°38'33.87" and 5°50'36.69" S – 5°52'32.5" S. This island has a land area of around 41.32 Ha [31], which is administratively part of the Seribu Islands Regency, Special Region of Jakarta Province (see Figure 1). The shallow waters on Pari Island consist of various types of coral reefs that are in healthy condition [32]. Coastal geomorphology in shallow waters tends to show high complexity, where the shallow water area of Pari Island is divided into two significant areas. In the eastern part the geomorphological zone is dominated by inner reef flat, outer reef flat, reef crest and reef slope, while in the eastern part it is dominated by inner reef flat, deep lagoon, shallow lagoon, reef crest and reef slope, and in general the reef crest completely surrounds shallow water zone on Pari Island [33].



**Figure 1:** Location of Pari Island as a study area

Benthic habitat classes discovered by [33] and [34] are live coral, live coral + rubble, rare seagrass + sand, dense seagrass, sand + rubble, sand, sand + rare seagrass, pavement/rock, and rubble. Live coral dominates 17% of all benthic classes, and sand dominates 36% of all classes.

## 2.2 Image Data Collection and Correction

The image data used in this research is Sentinel-2 L2A which has been scaled to the Bottom of Atmosphere (BOA), and has been geometrically corrected with 12 multispectral image channels. The spatial resolution of images varies, from 10 meters (channels B2, B3, B4 and B8) to 20 meters (channels B5, B6, B7, B11 and B12), and 30 meters (channels B1, B9 and B10) [35]. The channels used are B2 (blue, 490 nm), B3 (green, 560 nm), B4 (red, 665 nm) and B5 (VNIR, 705 nm). The sentinel-2 L2A image used is filtered based on the season in Indonesian waters. These seasons are winter monsoon or Northwest monsoon (NWM) in the range of December – February, southeast monsoon (SEM) in June – August, and between the two main seasons there is transition season I (March – May)), and transition season II (September – November) [36]. Images are cloud computed using Google Earth Engine (GEE) platform.

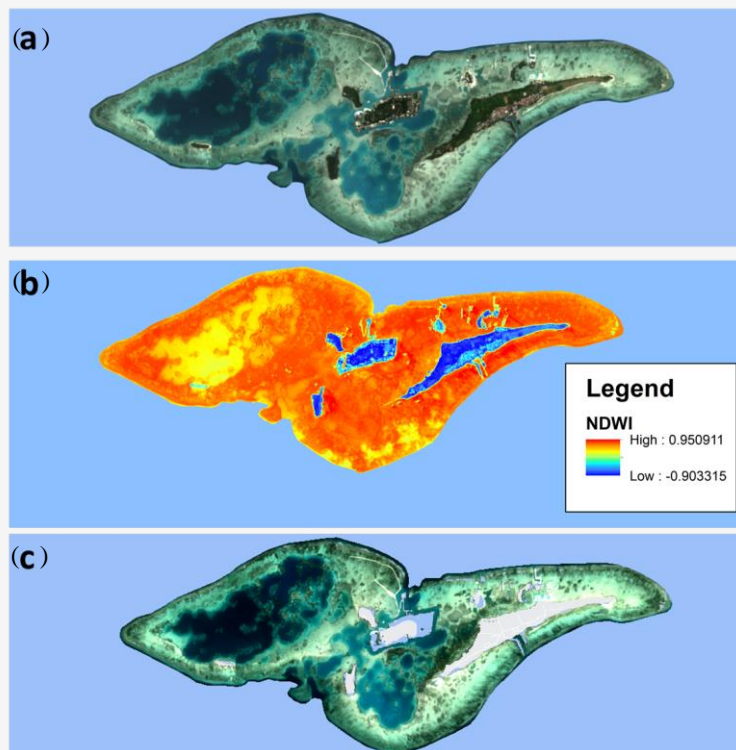
The selected image is a collection of Sentinel-2 L2A image data on the GEE database with minimum

cloud cover and turbidity. The image is then subjected to a process of reducing the number of images by calculating the "mean" of all pixel values in the entire image collection in the same season. We also present and analyze the spectral signature on sentinel-2 L2A images in each season with similar water bottom material objects, using Semi Automatic Classification Plugin in QGIS. Separation of land, deep water and shallow water objects is carried out using the normalized difference water index (NDWI) transformation approach (see Equation 1). Only positive NDWI values are identified as shallow water [17] and used as input to perform cropping on the image. The equation of NDWI is defined in equation 1.

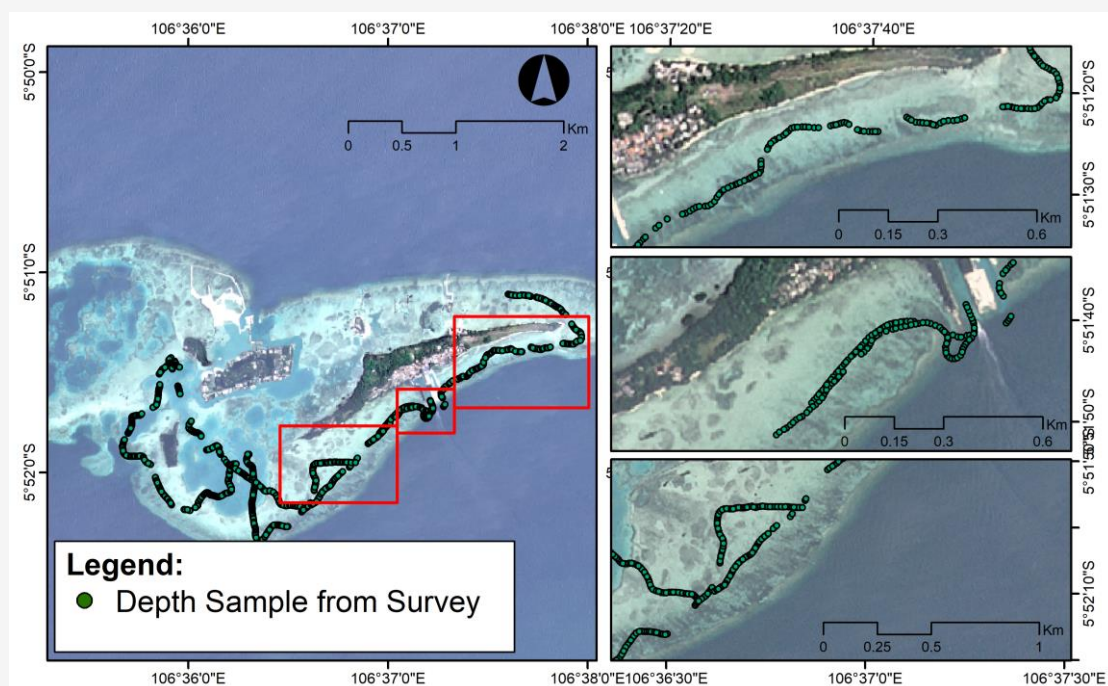
$$NDWI = \frac{\rho(\text{Green}) - \rho(\text{NIR})}{\rho(\text{Green}) + \rho(\text{NIR})}$$

Equation 1

Where  $\rho(\text{green})$  is the green band spectral channel (B3 in the Sentinel 2 L2A image),  $\rho(\text{NIR})$  is the Near Infrared band spectral channel (NIR/B4). Figure 2 is a sentinel-2 L2A image that has been cropped to obtain only shallow water objects. However, work domain adjustments are still made to obtain maximum shallow water objects.



**Figure 2:** (a) Multispectral Sentinel 2 L2A image data in visible spectrum display, (b) NDWI transformation results and (c) Image that has been masked based on NDWI transformation



**Figure 3:** Spatial distribution of bathymetry data for training and validation during the computing process

### 2.3 Data Collection using Hydrographic Survey

The distribution of sample depth data used in machine learning operations in this research can be

seen in Figure 3. Depth data was obtained through a hydrographic survey using a Teledyne CV 100 single beam echosounder device equipped with a Trimble



R8S GNSS system. The survey was carried out on October 15 2023 with clear weather conditions and solar radiation was visually visible penetrating the shallow water column to the maximum (see Figure 4).

The depth data from the hydrographic survey is then carried out in a division process to select data that will become samples for training needs and validation depth data to assess the accuracy of the regression results. The division of the depth data into sample data and validation are completely automated in the GEE software. The program code is set to divide  $\pm 465$  total depth data into sample data and training data with portions of 80:20 and iterates in the computing process 100 times. In general, bathymetric data from hydrographic survey measurements is in the data range from 0.5 to 7 meters above sea level. The depth data is then converted into a raster to be uploaded to GEE and used as a sample for the regression model.

#### 2.4 Random Forest Regression

This study involves a random forest (RF) algorithm from machine learning to reconstruct bathymetric information in shallow waters in the study area. RF is an algorithm that involves decision trees in carrying out ensemble classification or regression methods [3] and [37]. RF which carries the concept of ensemble decision trees is suitable for use on various types of data, where irrelevant predictors can be ignored, and linear and non-linear mechanisms can be handled well for easy interpretation [38] and [39]. RF was chosen to be run in this study by considering its ability to select an optimal decision tree, so that high accuracy can be obtained by dealing with outliers and noise data [3]. RF's unique capabilities are capable of

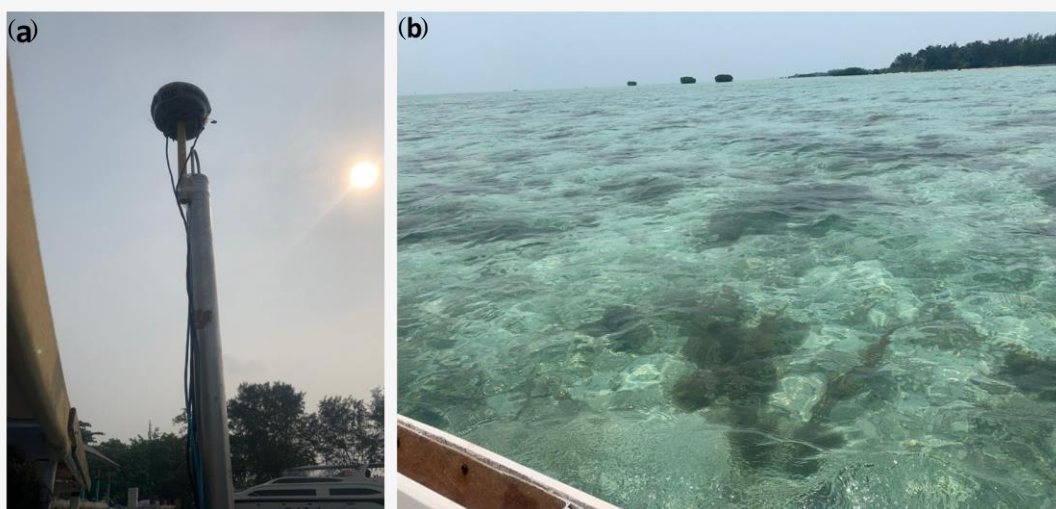
handling data in large quantities and dimensions, effectively handling multicollinearity, and normally distributed data is not a necessity. This method does not depend on the assumption of linearity between predictor variables and response variables, so it is relatively strong against collinearity compared to other regression methods, such as linear regression [38]. In general, in GEE the arguments that must be met to run RF regression are *numberOfTrees*, *variablesPerSplit*, *minLeafPopulation*, *bagFraction*, *maxNodes*, and *seed*.

#### 2.5 Linear Regression

Linear regression/transformation is based on the concept of attenuation of light radiation which occurs exponentially with a positive correlation with depth [14]. The use of multiple channels in the imagery is recommended as the use of a single band in the LR is not feasible because variations in benthic cover reflectance can lead to inaccurate depth predictions. For example, at the same depth, pixels with bright sand will appear shallower than pixels with seagrass which has lower reflectance, thereby causing errors in depth estimation [3]. GEE can apply linear regression through the functions: a) *linearFit()*, b) *linearRegression()*, c) *robustLinearRegression()*, and d) *ridgeRegression()*.

#### 2.6 Accuracy Assessment

Accuracy tests were carried out to measure the ability of the machine learning algorithm applied to reconstruct the bathymetry of shallow waters around Pari Island. The accuracy test is related to the quality of shallow water bathymetric maps produced using validation data which is proportionally determined along with training data in the GEE program code.



**Figure 4:** (a) Weather conditions on the day the bathymetry data was announced, and (b) Visual conditions at the bottom of shallow water which shows maximum solar radiation penetrating the water column

Accuracy tests were systematically applied to shallow water bathymetric reconstruction results representing each season (four bathymetric reconstruction maps). This study uses a mathematical and visual approach in interpreting the accuracy of SDB results, namely a) RMSE and b) MAE which represent mathematical methods and c) comparison of topographic profiles with reference depth, and d)  $R^2$  includes a 1:1 plot between validation data measured through hydrographic survey with shallow water bathymetric reconstruction results via SDB using machine learning methods. RMSE and MAE are able to provide absolute quantification of results regarding the difference between real values in the field and SDB data, thus producing a range of instability which is denoted in  $\pm$ . Further qualitative analysis is needed to spatially describe the distribution of shallow water bathymetry reconstruction results using a comparison approach of topographic profiles with reference depths. Comparative analysis of topographic profiles can provide a clear picture of the capacity of the bathymetry model against validation data in each season.  $R^2$  include Plot 1:1 is useful as a tool for detecting inaccurate depth estimates, either too low or too high [3]. The RMSE and MAE are defined in equations 2 and 3, respectively.

$$RMSE = \sqrt{\frac{\sum (Y_{image} - Y_{field})^2}{N}}$$

Equation 2

$$MAE = \frac{\sum |Y_{image} - Y_{field}|}{N}$$

Equation 3

Where  $Y_{image}$  is bathymetric data from SDB reconstruction,  $Y_{field}$  is bathymetric data from measurements as validation data, and N is the size of the dataset [12]. The research flow diagram can be seen in Figure 5. Based on the flowchart in Figure 5, the entire process was carried out on GEE with the Sentinel-2 image dataset which was filtered based on season and cloud cover. Images that have undergone a filtering process are masked using NDWI which separates water and non-water objects. Then, we simultaneously uploaded the depth dataset as material for regression into the form of training and validation depth. Depth data that has been converted to raster via topo to raster processing, and has been uploaded via assets to GEE is then divided automatically in the code editor into training and validation data.

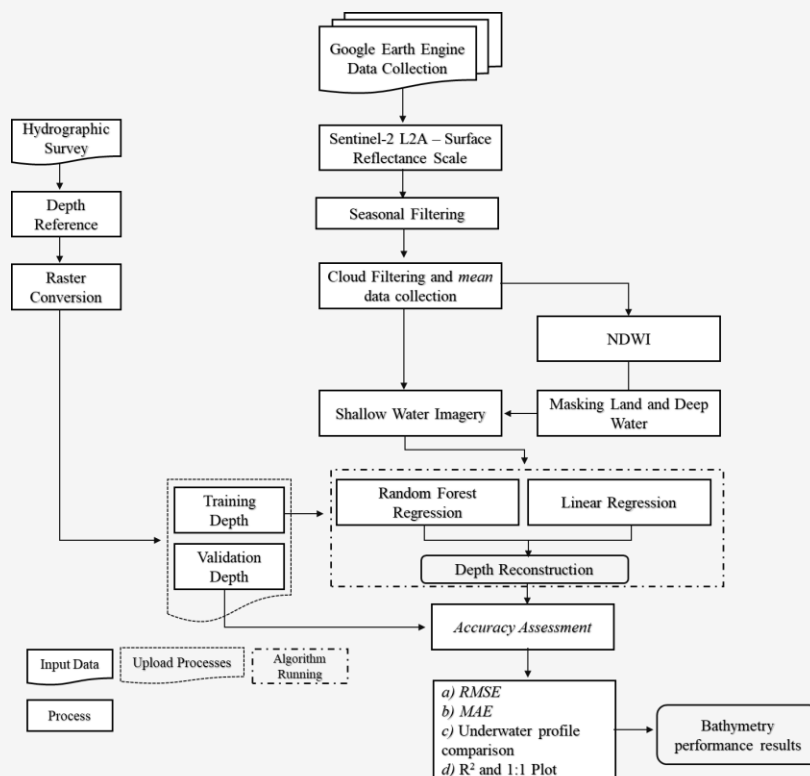


Figure 5: Research flow diagram

The training data is then operated into RF and LR regression to perform depth reconstruction. The results of the depth reconstruction via SDB are then tested for accuracy to carry out performance analysis.

### 3. Results

#### 3.1 Results

##### 3.1.1 Spectral signature of coastal waters in different seasons

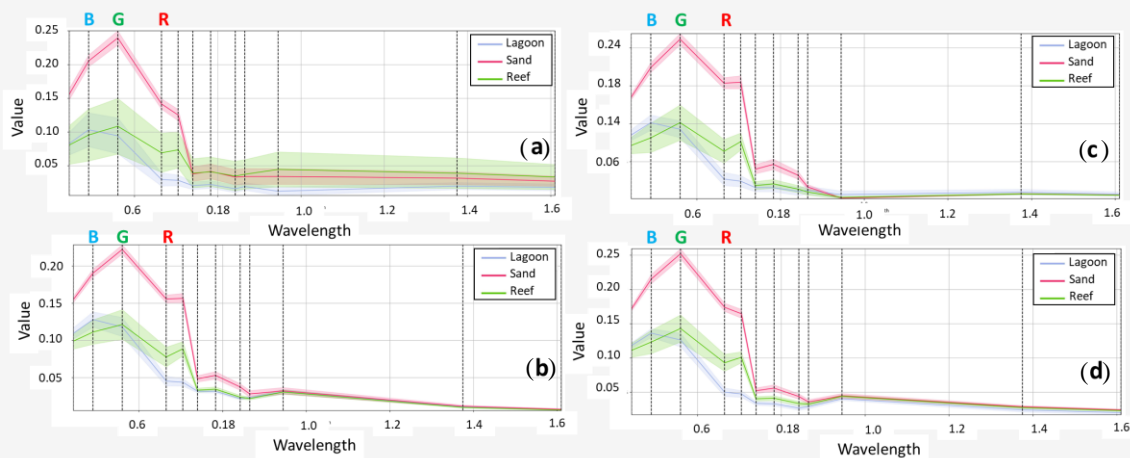
This study was carried out from the end of 2022 to the end of 2023, or in detail divided into four seasons, namely: December 2022 – February 2023 (West Season), March 2023 - May 2023 (Transition Season I), June 2023 – August 2023 (Eastern Season), and September 2023 – November 2023 (Transitional Season II). We analyzed Spectral Signatures from a variety of geomorphological zones and possible benthic habitats. This analysis aims to describe the influence of seasonal variations on the spectral response emitted by waters. Spectral response samples were taken in the lagoon zone, sandy reef flats, and bottom waters with coral cover. We used the QGIS tool with the Semi-Automatic Classification (SCP) plugin [40] to extract the spectral reflectance values of the three dominant objects in the waters. Figure 6 is a spectral reflectance curve extracted from four images representing seasonal variations.

The three objects observed showed a similar pattern, where the reflectance emitted by the lagoon object was significantly at the highest value in the blue wavelength channel (B2). The reflectance of bottom waters dominated by coral reefs and sand is significantly responded to by the green wavelength channel (B3). The spectral response of the three objects slopes gradually when entering the channel with the red wavelength (B4). Water bottom objects

with sandy substrates have the highest spectral reflectance of all objects observed. We carried out detailed identification of the reflections from the three objects by narrowing the analyzed spectral channels (Figure 7). We found that the lowest reflectance signal from the three objects as a whole occurs in the west season. The lagoon and reef substratum get the highest reflectance in the transition season I (March – May 2023), while the sand substratum gets the highest reflectance in the east season (June – August 2023). This study found the influence of seasonal variations on the reflectance level of waters with different bottom substrates. Although the variations shown are not significant, this empirical evidence can be used as an important basis for filtering data used by other researchers in the future in coastal and shallow water applications.

##### 3.1.2 Bathymetry spatial distribution

The results of shallow water bathymetry reconstruction can be seen in Figure 8 for the spatial distribution of bathymetry using the RF algorithm, and Figure 9 shows the results of bathymetry reconstruction using the LR algorithm, in various seasons. The depth values produced by RF in various seasonal variations do not show significant differences. In transition season 2, the model reconstructs the bathymetry slightly deeper than in other seasons. Deeper water areas such as lagoons are well reconstructed. Reef flat areas are the dominant geomorphological type surrounding islands and lagoons, and can be reconstructed well. The results of shallow water bathymetric reconstruction throughout the season consistently show a similar distribution, where in areas of water with fore reef geomorphology show a deeper sloping pattern than waters on the reef flat.



**Figure 6:** Spectral response of Lagoon objects (blue), sandy substrate (red), and substrate with coral reefs (green) during seasonal variations: (a) West Season, (b) Transition Season I, (c) East Season and (d) Transition Season II

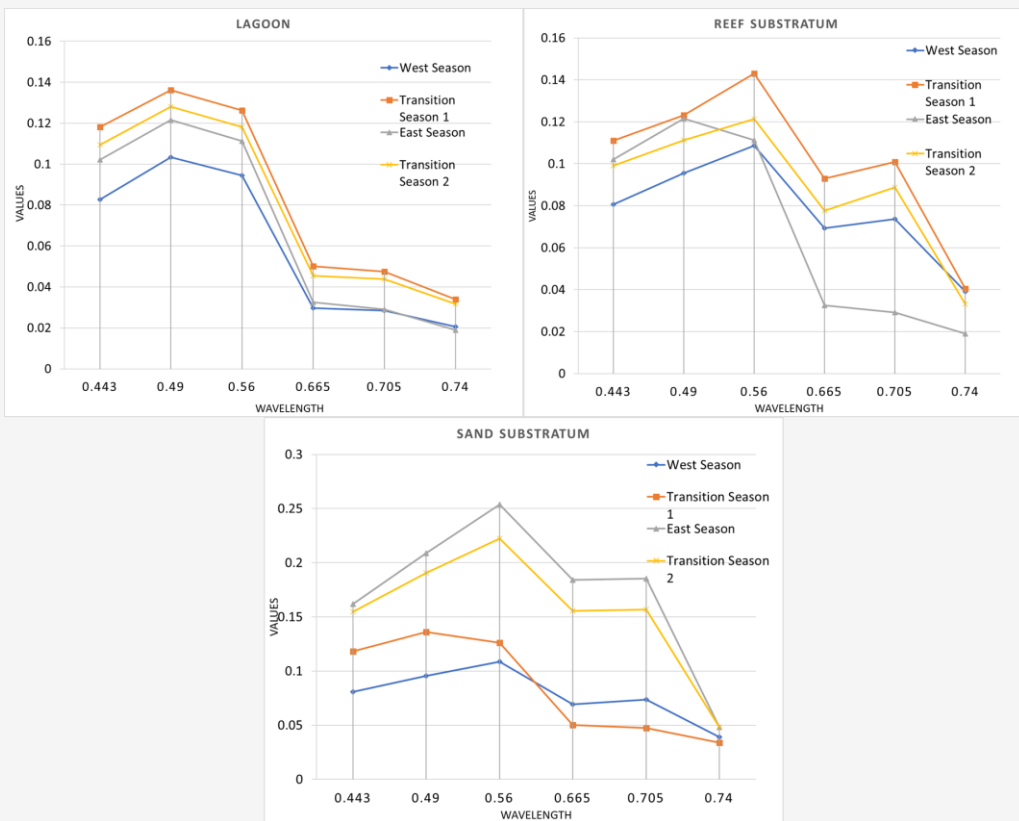


Figure 7: Detailed reflectance during seasonal variations

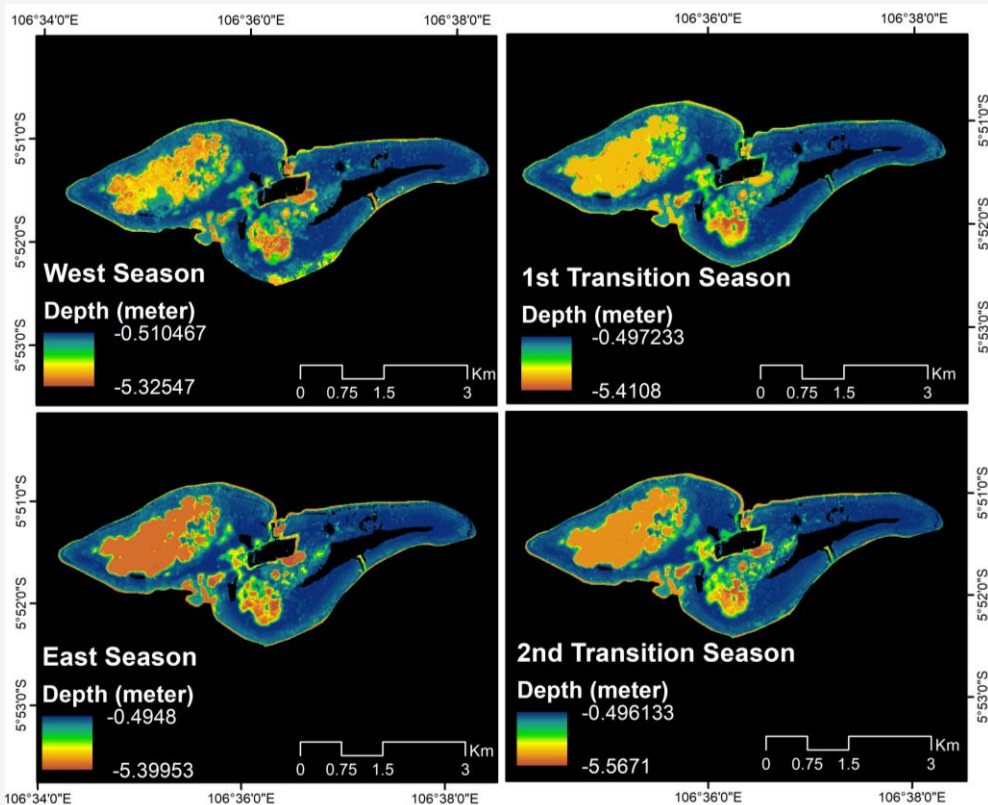
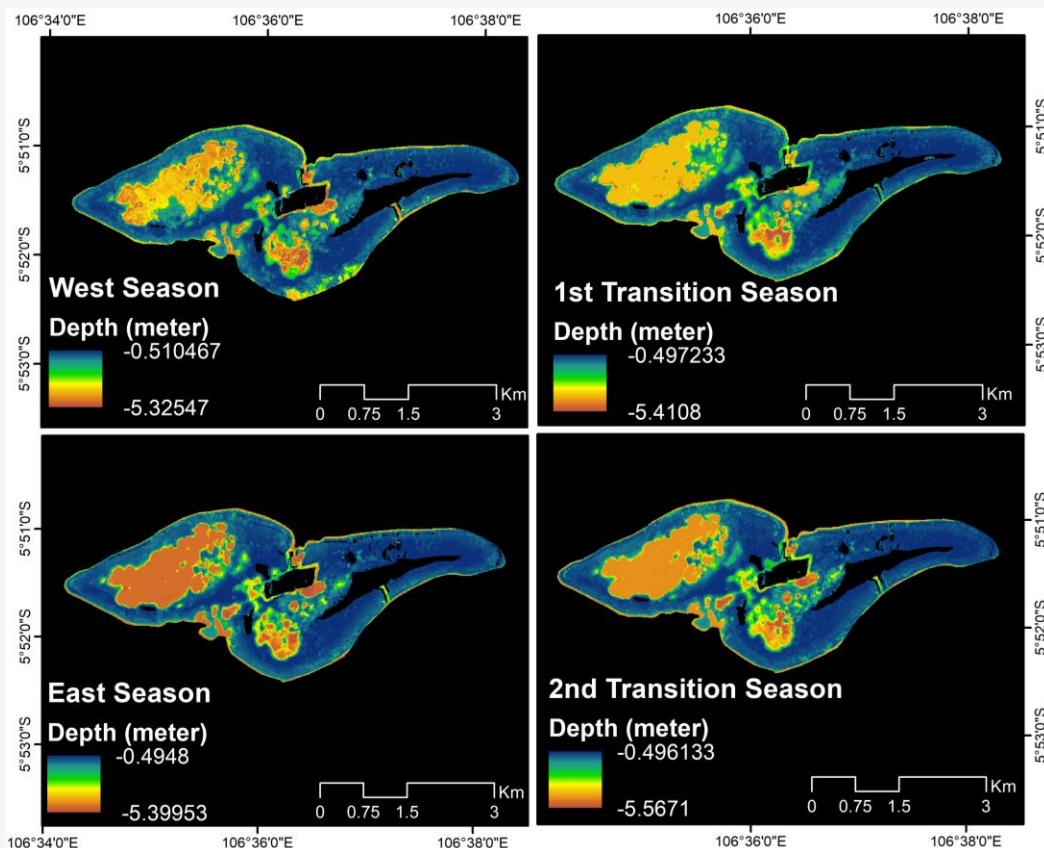


Figure 8: Results of shallow water bathymetric reconstruction in various seasons using the RF algorithm





**Figure 9:** Results of shallow water bathymetric reconstruction in various seasons using the LR algorithm

**Table 1:** Summary of model variables and accuracy test quantification of the RF algorithm in various seasonal variations

Seasonal Input	Depth Sample		Variable Importance	Accuracy Assessment		
	Training	Testing		RMSE	MAE	R <sup>2</sup>
West Season	346	96	B2 B3 B4	0.43	0.28	0.905
1 <sup>st</sup> Transition Season	354	88	B2 B3 B4	0.39	0.22	0.928
East season	358	84	B2 B3 B4	0.39	0.22	0.938
2 <sup>nd</sup> Transition Season	353	89	B2 B3 B4	0.34	0.21	0.944

This depth distribution significantly limits and surrounds the shallow water domain around Study Island. Different results are shown in shallow water bathymetric reconstruction using machine learning methods via the LR algorithm, where there are significant differences in depth results in each season. The resulting maximum depth difference reached 1.9 meters. The LR algorithm produces water areas with "invalid" results, where the water areas are systematically reconstructed by the LR algorithm as land pixel values, and are outside the range of values provided by the sample/training data.

### 3.1.3 RMSE and MAE evaluation

RMSE and MAE and R<sup>2</sup> are applied to all shallow water bathymetric reconstruction results in each seasonal variation. In the RF and LR algorithms, the portion of depth data used for training and testing (validation) is in the ratio range of 80:20 of the total depth data that is input to the model. The system created will automatically calculate sample requirements, so there will be differences in the number of samples for training and testing in each season. Tables 1 and 2 provide a summary of accuracy test results (RMSE, MAE, and R<sup>2</sup>) on machine learning methods using RF and LR algorithms.

The range of RMSE and MAE values in the reconstruction results using the RF algorithm does not show a significant difference, but in the RF algorithm the range of RMSE differences between seasons has a distance that needs to be considered. The difference between the highest and lowest RMSE - MAE values in the RF algorithm is 0.09 and 0.07 respectively, while in the LR algorithm the resulting difference is 0.16 and 0.1 respectively for RMSE and MAE. The best performance was produced by the RF algorithm in the transition season II variation (September – November 2023).

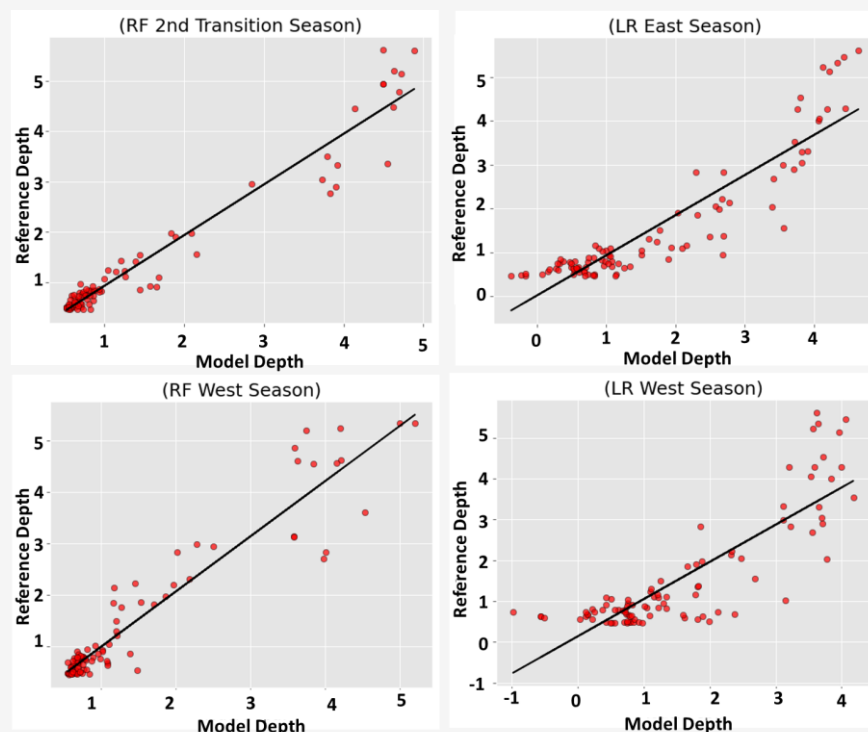
### 3.1.4 Depth analysis of 1:1 plot

Figure 10 depicts a 1:1 plot of the algorithms with the highest and lowest performance. RF produces the highest and lowest performance in the transition season II and west season respectively, while in the

LR algorithm the east season shows the highest performance, and the west season has the lowest performance. The overall plot shows that at depths < 1 meter there is a strong relationship between the SDB results and the reference depth. Through  $R^2$  calculations in the RF algorithm, the two seasons were able to show significant performance with  $R^2$  values > 0.9, while in the LR algorithm, there was a difference of 0.12 between SDB with the highest and lowest performance, which were 0.83 and 0.71 respectively. In general, depths below 2 meters contribute to significant accuracy, but accuracy tends to decrease after a depth of 3 m, and outlier data looks increasingly massive. Specifically, in the LR algorithm, there are reconstructions that are considered "invalid" because they produce data above normal figures (water level = 0 m), so they are interpreted as land.

**Table 2:** Summary of model variables and accuracy test quantification of the LR algorithm in various seasonal variations

Seasonal Input	Depth Sample		Variable Importance	Accuracy Assessment		
	Training	Testing		RMSE	MAE	$R^2$
West season				0.76	0.56	0.71
1 <sup>st</sup> Transition Season	371	94	Band Ratio	0.65	0.51	0.79
East Season				0.60	0.46	0.83
2 <sup>nd</sup> Transition Season				0.62	0.47	0.82

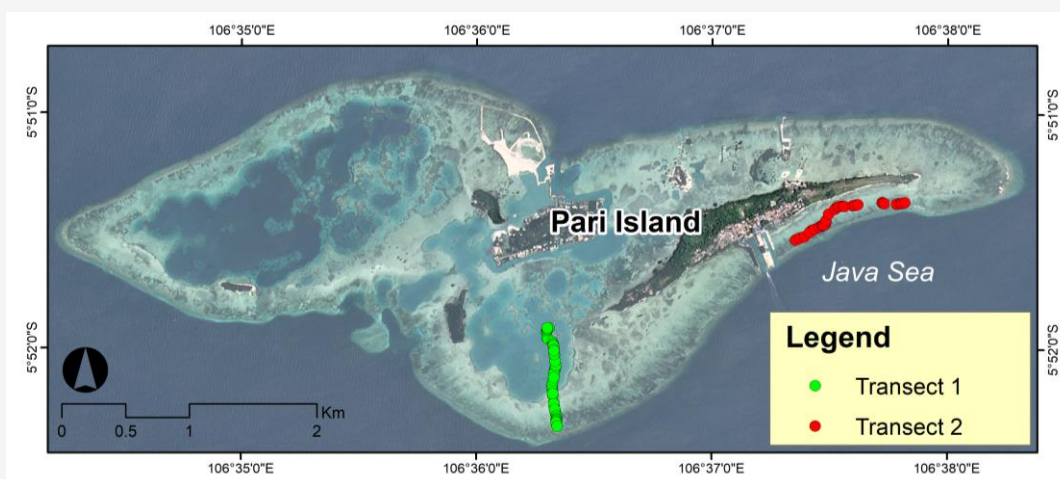


**Figure 10:** 1:1 plot results of bathymetric reconstruction results using the SDB method with the best performance represented by the RF 2nd transition season and LR east season, while the lowest performance is represented by the RF west season and LR west season

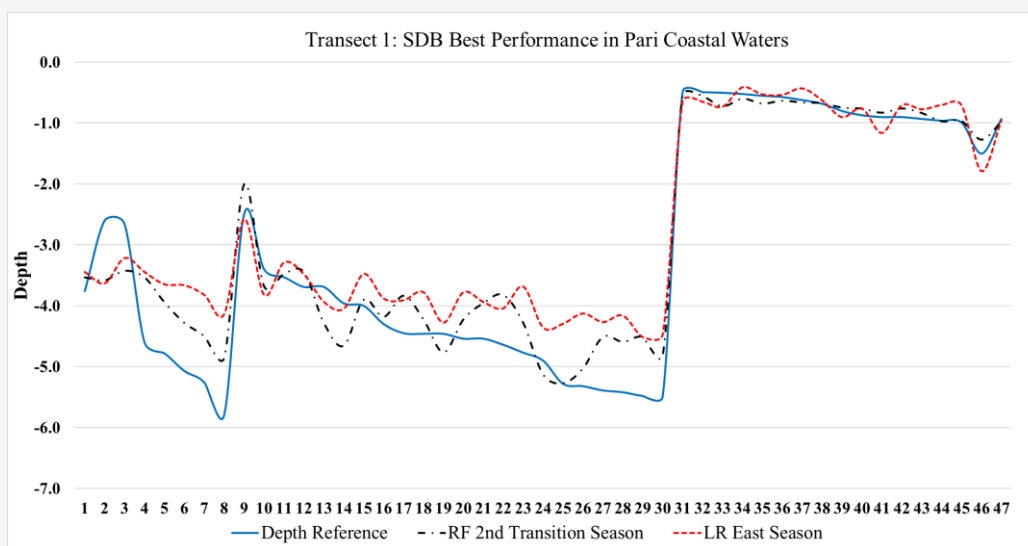
### 3.1.5 Underwater topographic profile analysis

We complete the process of analyzing the results of bathymetric reconstruction using SDB through machine learning (RF and LR) with analysis of the bottom topography of the waters. Our analysis was carried out using data from hydrographic measurements that were not used as sample data for SDB reconstruction. The comparison is carried out by comparing real measurement data in the field with the highest and lowest performance results from each RF and LR algorithm. The transect location can be seen in Figure 11. In transect 1, both the RF and LR algorithms with the best and lowest performance were able to show a pattern similar to the reference depth (measurements in the field), except for the SDB reconstruction results in the west monsoon LR algorithm which were not able to represent shallow

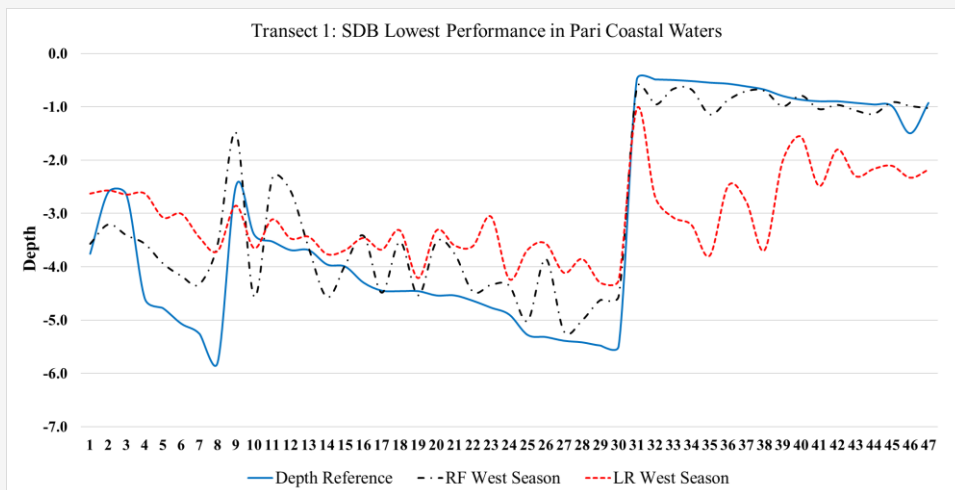
areas (see Figure 12 red dash line). The lagoon to the north (start of the transect) is consistently described as a deeper area than the waters to the south which tend to be shallower (see Figure 13). West Monsoon shows the lowest performance for both RF and LR algorithms. Overall, a non-uniform pattern is shown by the SDB reconstruction results from the reference depth measurements, although the resulting depth differences tend to be small. The bathymetry resulting from reconstruction via SDB at illustration points 1 – 26 is random so it is difficult to capture depth variations and is unable to illustrate depth patterns well, but the results tend to be uniform starting to be shown by illustration points > 26 (see Figures 14 and 15).



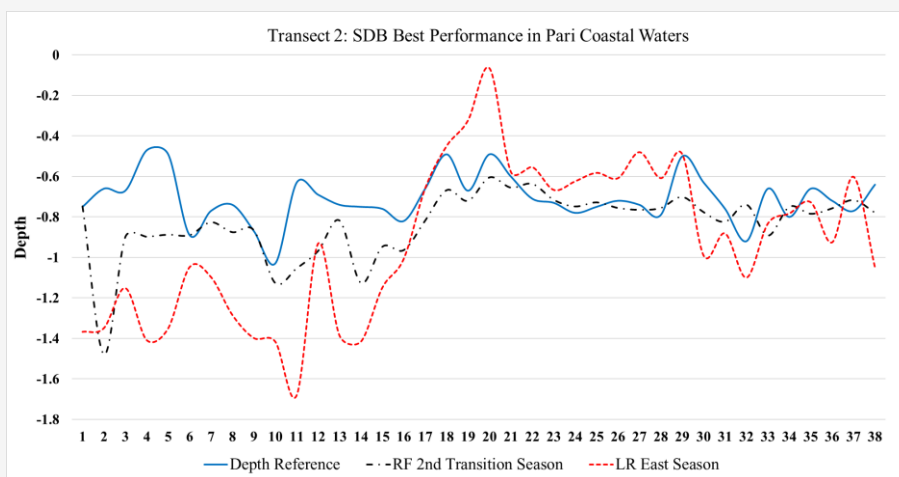
**Figure 11:** Location of transect 1 (green color) which is in the western part of Pari Island waters by cutting through the lagoon area, and location of transect 2 (red color) which is in a relatively uniform reef flat area



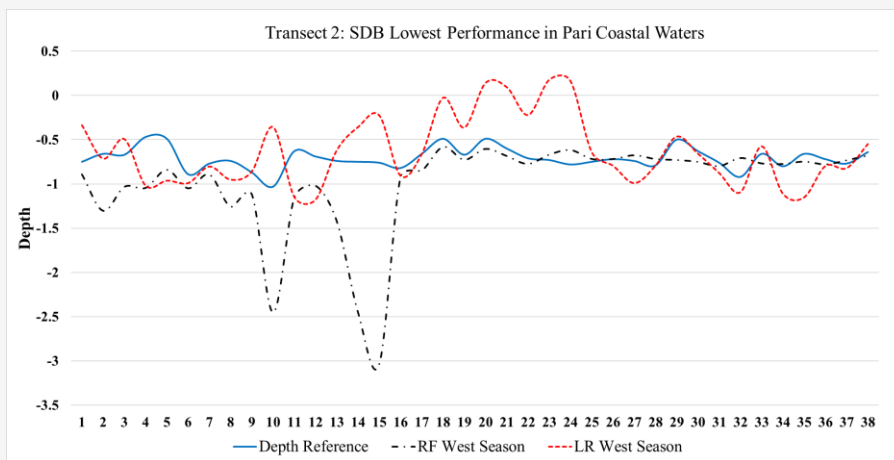
**Figure 12:** Transect 1 which depicts the underwater profile of the SDB with the best performance, illustrated from the north (lagoon) to the south (reef flat)



**Figure 13:** Transect 1 which depicts the underwater profile of the SDB with the lowest performance, illustrated from the north (lagoon) to the south (reef flat)



**Figure 14:** Transect 2 which depicts the underwater profile of the SDB with the best performance, illustrated from West to East which is in the reef flat zone



**Figure 15:** Transect 2 which depicts the underwater profile of the SDB with the lowest performance, illustrated from West to East which is in the reef flat zone



### 3.2 Discussion

This research aims to reconstruct shallow water bathymetry using the SDB method by considering the influence of seasonal variation, namely the west season, transition season I, east season and transition season II. Performance testing was carried out on the RF and LR algorithms on sentinel 2 images, the computation of which was carried out in the cloud on the GEE platform. Based on performance tests that are quantitatively based by looking at the RMSE, MAE and  $R^2$  values, the RF algorithm is able to provide significant results in each season by consistently producing minimal differences in accuracy between seasons. Qualitatively, the algorithm's ability to reconstruct bathymetry can be seen in the results of cross-sections of bottom water topography, in this case the model with the best performance is able to describe depth variations more accurately. The SDB model with the best performance is reflected in its ability to describe depth variations well to meet qualitative aspects, and is supported by adequate quantitative aspects. This is in accordance with the statement of [3] which states that accuracy analysis through cross-sectional identification is a more effective way to assess the accuracy of SDB results, because it provides an in-depth evaluation of the algorithm's ability to replicate seabed morphology in the real world.

In areas with geomorphological and benthic complexity in transect 1, the RF algorithm shows significance in replicating the shape of the water bed and quantitatively shows good results through RMSE and MAE. Based on the results obtained, the RF algorithm produces RMSE and MAE ranges from 0.34 – 0.43 m, and 0.28 – 0.21. The LR algorithm quantitatively produces RMSE and MAE ranges of 0.76 – 0.60 m, and 0.46 – 0.56 m. In general, the performance produced in this study outperforms previous research by [41] with the best  $R^2$  of 0.84, and matches the latest study by [42] with  $R^2 > 0.9$  at the same study location. The significance provided by the RF algorithm is the ability to reconstruct the morphology of the bottom waters in the study area even though there is interference from benthic commodities in these waters, and become eminence of this study. Differences in interpretation of results still appear in the form of differences between reference depth and SDB results, but these differences tend to be insignificant.

This research highlights seasonal variations and their influence on the reflectance ability of waters on different substratum. In the west season the three pilot samples (sand, reef and lagoon) were at the lowest reflectance levels, this corresponds to the accuracy produced in the west season being the lowest for both the RF and LR algorithms.

In the western season, various outlier data appeared (see Figure 10) which resulted in low quantitative accuracy test results. The superiority of RF as stated by [3] is that it is able to overcome variations in benthic cover in carrying out bathymetric reconstruction with SDB. We believe the level of significance and accuracy of SDB results of various machine learning algorithms can be continuously improved by considering seasonal variations with optimal reflectance evenly distributed, as well as more samples covering different benthic varieties and geomorphologies. The emergence of SDB reconstruction results with the lowest performance in the west monsoon and in line with the low level of reflectance in various benthic and morphological variations makes it a research alternative that needs to be explored in the future.

Additionally, our research is processed in cloud-based computing devices (GEE), making it possible to cut processing time and save storage facilities without leaving residue during the processing process. The data archive facility in peta byte [43] allows for multi-season SDB analysis which may not be possible easily using desktop-based computing devices. However, the main limitation is the image data that can be used, which is limited to images with medium resolution (namely: Sentinel 2 and Landsat series). Bathymetric reconstruction of shallow waters through SDB is a rational alternative, considering the difficulty of conducting surveys in the field with complex instruments, and the expensive [1][3][8] and [44]. The consequence that must be accepted is the fact that there is a difference between the real depth in the field and the SDB reconstruction results [3] which are recorded in the form of RMSE, MAE and other accuracy test methods.

### 4. Conclusion

This study conducts performance tests for reconstructing shallow water bathymetry using the SDB method using machine learning on RF and LR algorithms, taking into account the influence of seasonal variations. In general, the RF algorithm shows significant results with impressive accuracy test results over all seasonal variations, whether seen through quantitative aspects in the form of RMSE, MAE and  $R^2$  tests or qualitatively demonstrated through the ability to replicate seabed morphology based on visualization of cross-sectional bottom topography. waters. Apart from that, the LR algorithm can still provide good results, even though in several seasons it still shows outlier results through quantitative-based accuracy tests. The RF algorithm has the advantage of reconstructing the bathymetry of shallow waters, by overcoming the influence of reflectance differences due to seasonal variations, as

well as the influence of benthic varieties and bottom water geomorphology. In addition, our research found that the west monsoon with minimal water reflectance produces SDB reconstruction with the weakest performance of all scenarios for RF and LR. The cloud-based computing facilities used make it possible to save time and storage capacity without leaving processing residue. Further research that can be carried out based on the advanced topic of this study is to consider the influence of turbidity, temperature and salinity of waters which have an influence on CDOM to see their influence on the resulting accuracy of the SDB.

### Acknowledgment

Researchers thanks to Universitas Gadjah Mada for the support provided through the Final Assignment Recognition (Rekognisi Tugas Akhir/RTA) grant via the 2024 RTA Program announcement letter: 4971/UN1.P1/PT.01.01/2024. Additionally, author give gratitude to LPPM Indonesian Naval Postgraduate Military Service School also known as STTAL (Jakarta) for assistance with hydrographic survey instruments.

### References

- [1] Khakhim, N., Kurniawan, A., Wicaksono, P. and Hasrul, A., (2024a). Assessment of Empirical Near-Shore Bathymetry Model Using New Emerged PlanetScope Instrument and Sentinel-2 Data in Coastal Shallow Waters. *International Journal of Geoinformatics*, Vol. 20(2), 95–105. <https://doi.org/10.52939/ijg.v20i2.3071>.
- [2] Khakhim, N., Kurniawan, A., Wicaksono, P. and Hasrul, A., (2024b). Rapid Bathymetry Mapping Based on Shallow Water Cloud Computing in Small Bay Waters: Pilot Project in Pacitan-Indonesia. *Journal of Environmental Management and Tourism*, Vol. XV(1(73)), 41–51. [https://doi.org/https://doi.org/10.14505/jemt.v15.1\(73\).04](https://doi.org/https://doi.org/10.14505/jemt.v15.1(73).04).
- [3] Wicaksono, P., Djody Harahap, S. and Hendriana, R., (2024). Satellite-derived Bathymetry from WorldView-2 Based on Linear and Machine Learning Regression in the Optically Complex Shallow Water of the Coral Reef Ecosystem of Kemujan Island. *Remote Sensing Applications: Society and Environment*, Vol. 33. <https://doi.org/10.1016/j.rsase.2023.101085>.
- [4] Gülher, E. and Alganci, U., (2023a). Satellite-Derived Bathymetry Mapping on Horseshoe Island, Antarctic Peninsula, with Open-Source Satellite Images: Evaluation of Atmospheric Correction Methods and Empirical Models. *Remote Sensing*, Vol. 15(10). <https://doi.org/10.3390/rs15102568>.
- [5] Wang, Y., Chen, Y., Feng, Y., Dong, Z. and Liu, X., (2023). Multispectral Satellite-Derived Bathymetry Based on Sparse Prior Measured Data. *Marine Geodesy*, Vol. 46(5), 426–440. <https://doi.org/10.1080/01490419.2023.2213840>.
- [6] Lyzenga, D. R., Malinas, N. P. and Tanis, F. J., (2006). Multispectral Bathymetry Using a Simple Physically Based Algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44(8), 2251–2259. <https://doi.org/10.1109/TGRS.2006.872909>.
- [7] Mudiyansele, S. D., Wilkinson, B. and Abd-Elrahman, A., (2024). Automated High-Resolution Bathymetry from Sentinel-1 SAR Images in Deeper Nearshore Coastal Waters in Eastern Florida. *Remote Sensing*, Vol. 16(1). <https://doi.org/10.3390/rs16010001>.
- [8] Westley, K., (2021). Satellite-derived Bathymetry for Maritime Archaeology: Testing its Effectiveness at Two Ancient Harbours in the Eastern Mediterranean. *Journal of Archaeological Science: Reports*, Vol. 38. <https://doi.org/10.1016/j.jasrep.2021.103030>.
- [9] Laporte, J., Dolou, H., Avis, J. and Arino, O., (2023). Thirty Years of Satellite Derived Bathymetry – The Charting Tool that Hydrographers Can No Longer Ignore. *The International Hydrographic Review*, Vol. 29(1), 170–184. <https://doi.org/10.58440/ihr-29-a20>.
- [10] Gülher, E. and Alganci, U., (2023b). Satellite-Derived Bathymetry in Shallow Waters: Evaluation of Gokturk-1 Satellite and a Novel Approach. *Remote Sensing*, Vol. 15(21). <https://doi.org/10.3390/rs15215220>.
- [11] Sukmono, A., Aji, S., Amarrohman, F. J., Bashit, N. and Saputra, L. R., (2022). The Extraction of Near-Shore Bathymetry using Sentinel-2A Satellite Imagery: Algorithms and their Modifications. *TEM Journal*, Vol. 11(1), 150–158. <https://doi.org/10.18421/TEM111-17>
- [12] Ji, X., Ma, Y., Zhang, J., Xu, W. and Wang, Y., (2023). A Sub-Bottom Type Adaption-Based Empirical Approach for Coastal Bathymetry Mapping Using Multispectral Satellite Imagery. *Remote Sensing*, Vol. 15(14). <https://doi.org/10.3390/rs15143570>.

- [13] Ashphaq, M., Srivastava, P. K. and Mitra, D., (2021). Review of Near-Shore Satellite Derived Bathymetry: Classification and Account of Five Decades of Coastal Bathymetry Research. *Journal of Ocean Engineering and Science*, Vol. 6(4), 340–359. <https://doi.org/10.1016/j.joes.2021.02.006>.
- [14] Stumpf, R. P., Holderied, K. and Sinclair, M., (2003). Determination of Water Depth with High-Resolution Satellite Imagery Over Variable Bottom Types. *Limnology and Oceanography*, Vol. 48(1), 547–556. [https://doi.org/10.4319/lo.2003.48.1\\_part\\_2.0547](https://doi.org/10.4319/lo.2003.48.1_part_2.0547).
- [15] Lee, Z., Carder, K. L., Mobley, C. D., Steward, R. G. and Patch, J. S., (1999). Hyperspectral Remote Sensing for Shallow Waters: 2 Deriving Bottom Depths and Water Properties by Optimization. *Applied Optics*, Vol. 38(18). <https://doi.org/10.1364/ao.38.003831>.
- [16] Duplančić Leder, T., Baučić, M., Leder, N. and Gilić, F., (2023). Optical Satellite-Derived Bathymetry: An Overview and WoS and Scopus Bibliometric Analysis. *Remote Sensing*, Vol. 15(5). <https://doi.org/10.3390/rs15051294>.
- [17] Li, J., Knapp, D. E., Lyons, M., Roelfsema, C., Phinn, S., Schill, S. R. and Asner, G. P., (2021). Automated Global Shallowwater Bathymetry Mapping Using Google Earth Engine. *Remote Sensing*, Vol. 13(8). <https://doi.org/10.3390/rs13081469>.
- [18] Duan, Z., Chu, S., Cheng, L., Ji, C., Li, M. and Shen, W., (2022). Satellite-derived Bathymetry using Landsat-8 and Sentinel-2A Images: Assessment of Atmospheric Correction Algorithms and Depth Derivation Models in Shallow Waters. *Optics Express*, Vol. 30(3). <https://doi.org/10.1364/oe.4444557>.
- [19] Knudby, A. and Richardson, G., (2023). Incorporation of Neighborhood Information Improves Performance of SDB models. *Remote Sensing Applications: Society and Environment*, Vol. 32. <https://doi.org/10.1016/j.rsase.2023.101033>.
- [20] Zhou, W., Tang, Y., Jing, W., Li, Y., Yang, J., Deng, Y. and Zhang, Y., (2023). A Comparison of Machine Learning and Empirical Approaches for Deriving Bathymetry from Multispectral Imagery. *Remote Sensing*, Vol. 15(2), 1–17. <https://doi.org/10.3390/rs15020393>.
- [21] Liu, S., Wang, L., Liu, H., Su, H., Li, X. and Zheng, W., (2018). Deriving Bathymetry from Optical Images with a Localized Neural Network Algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56(9), 5334–5342. <https://doi.org/10.1109/TGRS.2018.2814012>.
- [22] Mateo-Pérez, V., Corral-Bobadilla, M., Ortega-Fernández, F. and Rodríguez-Montequín, V., (2021). Determination of Water Depth in Ports Using Satellite Data Based on Machine Learning Algorithms. *Energies*, Vol. 14(9). <https://doi.org/10.3390/en14092486>.
- [23] Gupta, G. K., Bhat, R. and Balan, M. S., (2023). Satellite-Derived Bathymetry of an Inland Reservoir in India. *TechRxiv*. [https://www.techrxiv.org/articles/preprint/Satellite-Derived\\_Bathymetry\\_of\\_an\\_Inland\\_Reservoir\\_in\\_India/22210603](https://www.techrxiv.org/articles/preprint/Satellite-Derived_Bathymetry_of_an_Inland_Reservoir_in_India/22210603).
- [24] Çelik, O. İ., Büyüksalih, G. and Gazioğlu, C., (2023). Improving the Accuracy of Satellite-Derived Bathymetry Using Multi-Layer Perceptron and Random Forest Regression Methods: A Case Study of Tavşan Island. *Journal of Marine Science and Engineering*, Vol. 11(11). <https://doi.org/10.3390/jmse11112090>.
- [25] Khomsin, Mukhtashor, Suntoyo, and Pratomo, D. (2023). Dredging Volume Analysis Using Bathymetric Multifrequency. *International Journal of Geoinformatics*, Vol. 19(4), 1–12. <https://doi.org/10.52939/ijg.v19i4.2623>.
- [26] Nababan, B., Wirapramana, A. A. G. and Arhatin, R. E., (2013). Spektral Remote Sensing Reflektansi Permukaan Air Laut. *Jurnal Ilmu Dan Teknologi Kelautan Tropis*, Vol. 5(1), 69–84.
- [27] Yang, C., Ye, H. and Tang, S., (2020). Seasonal Variability of Diffuse Attenuation Coefficient in the Pearl River Estuary from Long-Term Remote Sensing Imagery. *Remote Sensing*, Vol. 12(14). <https://doi.org/10.3390/rs12142269>.
- [28] Bhuyan, M., Jayaram, C., Menon, N. N. and Joseph, K. A., (2020). Satellite-Based Study of Seasonal Variability in Water Quality Parameters in a Tropical Estuary along the Southwest Coast of India. *Journal of the Indian Society of Remote Sensing*, Vol. 48(9), 1265–1276. <https://doi.org/10.1007/s12524-020-01153-0>.
- [29] Sagawa, T., Yamashita, Y., Okumura, T. and Yamanokuchi, T., (2019). Satellite Derived Bathymetry Using Machine Learning and Multi-Temporal Satellite Images. *Remote Sensing*, Vol. 11(10). <https://doi.org/10.3390/rs11101155>.

- [30] Darmanin, G., Gauci, A., Deidun, A., Galone, L. and D'Amico, S., (2023). Satellite-Derived Bathymetry for Selected Shallow Maltese Coastal Zones. *Applied Sciences (Switzerland)*, Vol. 13(9). <https://doi.org/10.3390/app13095238>.
- [31] Wulan, D. R., Sintawardani, N., Marganingrum, D., Triyono, T., Barid, V. B., Santoso, H. and Yulianto, E., (2023). Water Sources, Consumption, and Water-Related Sanitation on Pari Island, Indonesia: A Mixed-Focus Group Discussion and Survey Study. *Aqua Water Infrastructure, Ecosystems and Society*, Vol. 72(8), 1359–1372. <https://doi.org/10.2166/aqua.2023.137>.
- [32] Samadi, S., Munandar, A., Ayuni, V. T. and Amelia, R. N., (2023). Coral Density Level in the Context of Ecological Resilience and Its Effect on Sustainable Tourism in Pari Island, DKI Jakarta. *IOP Conference Series: Earth and Environmental Science*, Vol. 1190(1). <https://doi.org/10.1088/1755-1315/1190/1/012004>.
- [33] Anggoro, A., Siregar, V. P. and Agus, S. B., (2018). Klasifikasi Multiskala Untuk Pemetaan Zona Geomorfologi Dan Habitat Bentik Menggunakan Metode Obia Di Pulau Pari (Multiscale Classification for Geomorphic Zone and Benthic Habitats Mapping Using Obia Method in Pari Island). *Jurnal Penginderaan Jauh Dan Pengolahan Data Citra Digital*, Vol. 14(2). <https://doi.org/10.30536/j.pjpdcd.1017.v14.a2622>.
- [34] Anggoro, A., Siregar, V. P. and Agus, S. B., (2016). The Effect of Sunlight on Benthic Habitats Mapping in Pari Island Using Worldview-2 Imagery. *Procedia Environmental Sciences*, Vol. 33, 487–495. <https://doi.org/10.1016/j.proenv.2016.03.101>.
- [35] ESA. (2015). Sentinel-2 User Handbook. In *ESA Standard Document Date* (Issue 1). European Space Agency. <https://doi.org/10.1021/ie51400a018>.
- [36] Purwanto, Sugianto, D. N., Zainuri, M., Permatasari, G., Atmodjo, W., Rochaddi, B., Ismanto, A., Wetchayont, P. and Wirasatriya, A., (2021). Seasonal Variability of Waves Within the Indonesian Seas and its Relation with the Monsoon Wind. *Ilmu Kelautan: Indonesian Journal of Marine Sciences*, Vol. 26(3), 189–196. <https://doi.org/10.14710/ik.ijms.26.3.189-196>.
- [37] Zhang, C., (2015). Applying Data Fusion Techniques for Benthic Habitat Mapping and Monitoring in a Coral Reef Ecosystem. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 104, 213–223. <https://doi.org/10.1016/j.isprsjprs.2014.06.005>.
- [38] Islam, K. I., Elias, E., Carroll, K. C. and Brown, C., (2023). Exploring Random Forest Machine Learning and Remote Sensing Data for Streamflow Prediction: An Alternative Approach to a Process-Based Hydrologic Modeling in a Snowmelt-Driven Watershed. *Remote Sensing*, Vol. 15(16). <https://doi.org/10.3390/rs15163999>.
- [39] Svetnik, V., Liaw, A., Tong, C., Christopher Culberson, J., Sheridan, R. P. and Feuston, B. P., (2003). Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences*, Vol. 43(6), 1947–1958. <https://doi.org/10.1021/ci034160g>.
- [40] Congedo, L., (2021). Semi-Automatic Classification Plugin: A Python Tool for the Download and Processing of Remote Sensing Images in QGIS. *Journal of Open Source Software*, Vol. 6(64). <https://doi.org/10.21105/joss.03172>.
- [41] Wahyuningrum, P. I., Jaya, I. and Simbolon, D., (2008). Algoritma Untuk Estimasi Kedalaman Perairan Dangkal Menggunakan Data LANDSAT-7 ETM + (Studi Kasus: Perairan Gugus Pulau Pari, Kepulauan Seribu, Jakarta) Algorithm to Estimate Shallow Water Depth by using Landsat-7 ETM + Data (Case Study: Pari Island, Seri. *Buletin PSP*, Vol. XVII(3), 333–340.
- [42] Sanova, A. S. S., Saputro, S. and Helmi, M., (2018). Shallow Water Depths Mapping Using ALOS – AVNIR Satellite Imagery at Pari Islands, Seribu Archipelago, Jakarta. *Jurnal Laut Khatulistiwa*, Vol. 1, 1–6.
- [43] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R., (2017). Remote Sensing of Environment Google Earth Engine: Planetary-scale Geospatial Analysis for Everyone. *Remote Sensing of Environment*, Vol. 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>.
- [44] Lumban-Gaol, Y. A., Ohoi, K. A. and Peters, R. Y., (2021). Satellite-derived Bathymetry Using Convolutional Neural Networks and Multispectral Sentinel-2 Images, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 43(B3-2021), 201–207. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2021-201-2021>.