Mapping of Nearshore Bathymetry Based on Random Forest Machine Learning in Kemujan Island Waters

Kurniawan, A.,[1] Khakhim, N.[2*] and Wicaksono, P.[2]
[1]Coastal and Watershed Management Planning, Postgraduate Geography, Faculty of Geography Universitas Gadjah Mada, Indonesia, E-mail: agung.kurniawan.16@mail.ugm.ac.id
[2]Cartography and Remote Sensing, 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.v21i2.3933

Abstract
The availability of shallow water bathymetry data is essential for support shipping safety, marine spatial planning, and conservation. Conventional bathymetric data acquisition generally requires complex instruments and is expensive. Remote sensing through multispectral image data allows the application of satellite-derived bathymetry (SDB) to obtain shallow water bathymetry data quickly and efficiently. Through this study, we tested the influence of seasonal variations across Indonesian waters: Northwest Monsoon (NWM), TS1 (Transition Season 1), Southeast Monsoon (SEM), and TS2 (Transition Season 2) on the accuracy produced by SDB through an empirical approach using Random Forest (RF) algorithm. The RF algorithm is applied to multispectral multiscenario images, which scale to surface reflectance (SR), deglint, and band ratio. The main variable is input depth data, which automatically divided into training and testing. The entire process is carried out on a cloud computing device called Google Earth Engine (GEE), thereby reducing processing residue and significantly saving processing time. We use two approaches to assessing accuracy: quantitatively using R^2, MSE, and RMSE, and qualitatively descriptive using a 1:1 plot approach and underwater topographic profile via transect, which compares reference data with model depth. Overall, the resulting RMSE range is 0.86–1.7 m; for MSE, it is 0.84–2.9 m; and R^2 is in the range of 0.63–0.86 m. This study found that seasonal variations have a systematic effect on accuracy, where NWM produces the lowest accuracy, which is thought to be due to atmospheric factors, and SEM TS2, respectively, produces the best accuracy. Through depth distribution, the SDB model is able to show maximum performance up to a total depth of 4 m. The underwater topographic profile shows that the overall scenario can replicate the depth well. This study provides comprehensive insight into the influence of seasonal variations on the accuracy of shallow-water bathymetric mapping.

Keywords: Accuracy, Machine Learning, Random Forest, Season, Shallow Water Bathymetry

1. Introduction

The availability of bathymetric data in shallow waters plays an important role in shipping and navigation safety [1], hydrodynamics and wave simulations [2], monitoring small and outer island conservation areas [3] and biodiversity and benthic habitats [4][5] and [6]. The coastal environment is dynamic and temporary, so resources are needed to explore effective techniques for capturing topographic data on underwater environments through efficient measurements [7]. Conventionally, bathymetry measurements involve ships using acoustic instruments such as multibeam (MBES) or single-beam echosounder (SBES) [1] and [8], as well as by using airborne laser technology such as LiDAR [1][9] and [10]. However, the main challenge in obtaining shallow water bathymetry data in coastal areas using conventional techniques is the expensive budget required [3] and [5], logistics [10], difficult access to remote areas and shallow waters [11] as well as longer operational time constraints [8]. Thus, for researchers and governments in 3rd world countries, this technology is less efficient for carrying out fast mapping, where shallow water bathymetry data often plays an important role in the process of preparing marine management documents [1].

Remote sensing-based technology has developed massively in the last decade and makes it possible to complement conventional bathymetric data acquisition methods [7], because it has wide coverage and has impressive spatiotemporal resolution and is cost-effective [12].

However, the use of satellite imagery is very vulnerable to atmospheric disturbances and physical conditions of the sea surface, so selecting appropriate and consistent imagery is very important [5]. On the other hand, the phenomenon of water-leaving radiation, which is a function of radiation penetration at depth, is important and must be taken into account when using satellite imagery in water areas [5][13] and [14] especially in waters. with complex waters, variations in bottom reflection are often found due to various benthic conditions benthic [14][15] and [16]. A cost-effective, flexible and reasonable alternative for conducting shallow water bathymetry mapping is satellite image-based using the Satellite-Derived Bathymetry (SDB) method [4][8][17] and [18]. The SDB methods that are being developed are generally based on two main approaches, namely physics-based and empirical approaches, with the main difference being the existence of depth data as sample data [3] and [19].

The physics-based approach relies on physical principles that reduce radiation exponentially as the depth of the water column increases, thereby reducing reflectance from the bottom of the water [20] and [21]. Through this approach, bathymetric information is predicted through a radiative transfer model that calculates inherent optical properties (IOPs), as well as a forward modeling algorithm with the assumption that water quality tends to be constant and the bottom substrate is homogeneous [17], resulting in in situ depth data is no longer needed as analysis material [7]. Physics-based approaches are generally more challenging because there is a complex correlation between reflectance properties in waters and CDOM (Colored dissolved organic matter) which may interfere with the penetration of solar radiation. So the more general and more widely developed approach is empirically based, or also known as the statistical approach [4][17] and [19].

Empirical algorithms or in a broader perspective include semi-empirical, requiring depth data from measurements as a representation of in-situ data to calibrate the model [22], in order to obtain a statistical relationship between depth data and water reflectance in a single channel as well as the ratio [8]. The empirical algorithm basically uses the effectiveness of multiple regression and spectral log ratio regression, which is applied to the blue channel and ratio channel as channels with short wavelengths so that they can be optimal references because radiation is possible to penetrate higher even in objects with low albedo [7] [13] and [14]. Developing computing technology allows the application of more comprehensive empirical approaches involving machine learning (ML).

In empirically based approaches there is often weak linearity between bottom reflectance and sample depth [1], therefore the use of ML is intended to handle this nonlinear relationship to prevent sampling errors [1][2][22] and [23]. Through the ML method, we can optimize the selection of spectral bands for depth estimation and the accuracy of the ML model results is greatly influenced by the sample data, both quality and quantity [4]. ML algorithms that are widely used in shallow water bathymetry extraction include Random Forest (RF) which has the ability to handle images with noise or non-linear relationships [1], [23] and [24], Support Vector Machine (SVM) which is a data-based technique for maintaining sampling error within certain limits and minimizing confidence intervals by relating machine capacity to input complexity [22]. On the other hand, there are also Multi-Adaptive Regression Splines [25], neural network (NN) [26] to linear [1] and Logistic regression [27].

Several algorithms have been applied and performance tested in the waters of the study area, where [1] used Worldview-2 high resolution imagery to carry out comparisons between SVM, RF, and Linear Regression (LR) with superior results from RF which outperformed another algorithm with RMSE ± 0.8 meters, and is able to describe water depth accurately. [3] conducted a performance test study using empirical methods on a combination of the Coastal Blue, Green I, Yellow, and Red Edge channels on PlanetScope imagery and the visible channel on Sentinel-2 imagery. The test results show that the new PlanetScope channel cannot outperform the accuracy produced from Sentinel-2 imagery with the best accuracy RMSE = 0.8. [3] tested SDB performance using a physics-based approach, with RMSE accuracy results > 1 meter. At the same water location, [28][29] and [30] used PlanetScope imagery to obtain shallow water bathymetry using an empirical approach (SVR and LR and conventional machine learning algorithms) and produced the best accuracy results through RMSE up to ± 0.6.

In this study, the RF machine learning algorithm was chosen to be run to obtain shallow water bathymetry information, because in general it can represent shallow water topography with precision [1], but with improvements in tuning parameters and involving the influence of seasonal variations on the band. single and band ratio. SDB results via RF have been proven to be able to comprehensively describe the depth of shallow waters [1]. The increase in RF model input parameters based on seasonal variations is intended to determine its accuracy and significance compared to previous research in the same water area, namely: [1][3][28][29] and [30].

A comprehensive evaluation and update of seasonal parameters in the RF algorithm in shallow waters with complex ecosystems is expected to provide a real contribution to mapping coastal and near-shore areas. The image dataset used is Sentinel-2 MSI with a resolution of 10 meters. This image data was chosen because of its continuous availability and ability to record the same area every 10 days [31], and is continuously updated by the developer, so that it can provide a comprehensive representation of the quality of the data in each full season. Improvement parameters according to season is the highlight of this research, so that it can provide factual insight into SDB-based bathymetric mapping in general, and in the Kemujan water area specifically.

2. Material and Methods

2.1 Study Area

Administratively, the study area is part of the waters of Kemujan Island, Karimunjawa Islands, Jepara Regency, and is under Central Java Province. As a National Park, this island is a conservation area under the supervision of the Ministry of Environment and Forestry, Republic of Indonesia. Kemujan Island is the main and largest island of the Karimunjawa archipelago [3] (see Figure 1). This island has unique geomorphic characteristics with complex topography and has diverse varieties of coral reefs [1]. Kemujan Island in general is an area with waters rich in benthic habitat resources [32]. emujan Island has various types of reef morphology, namely Lagoon and Back Reef, Flat Reef, Reef Cut, Fore Reef, Bank/Shelf, and Escarpment [33]. In the shallow waters near the coast of Kemujan Island, sunlight can effectively penetrate the water column to reach the bottom of the waters. Significant differences in hue can reflect differences in geomorphological appearance or different benthic habitats.

2.2 Field Measurement Data

The depth data that is input into the ML model ensemble process for training and validation data is the result of measurements in the field carried out at several different times. The instrument used was the Garmin GPSmap 2108 single-beam echo sounder which was used to measure the latest data in 2021 [1]. Additional data is also used to complete training data using survey results in 2016 [34] and 2009 [35].

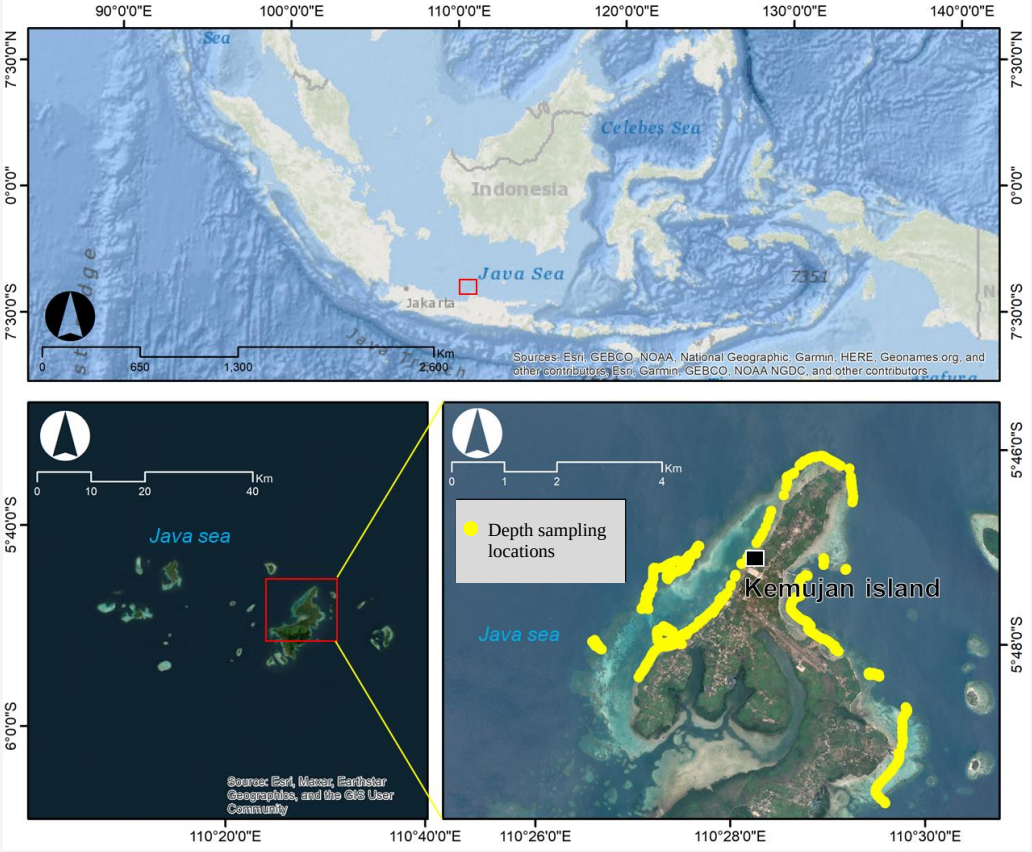

Figure 1: Kemujan Island

Even though the depth data as samples for the ML model were taken at different time periods, we assume that the relative depth has not changed because it is in the Karimunjawa National Park conservation area [1]. The depth data is then scaled against the tidal data to obtain the mean sea level (MSL). Tidal data is obtained from the tidal model developed by the Geospatial Information Agency (see: https://srgi.big.go.id/map/pasut-prediction-active). The 1199 depth sample data have been corrected with MSL values needing to be converted into raster form and adjusted to the targeted image resolution [1], in this case 10x10 m^2 in accordance with the Sentinel-2 image resolution. We divide the data from field measurements into two forms, namely training data used to build a shallow water bathymetry model and testing data used to validate the depth of the model results, but this separation process is automatically carried out by a machine with a 70:30 portion for training and testing sequentially. Apart from that, we also prepared two transects to carry out underwater topography analysis, as well as qualitatively assess the accuracy of the SDB results. See Figure 1 for detailed information regarding the distribution of sample data.

2.3 Sentinel-2 MSI Data Collection

The satellite image used is Sentinel-2 L2A MSI obtained via the Google Earth Engine (GEE) database. The Sentinel-2 Multi-Spectral Instrument (MSI) consists of two twin satellites launched in 2015 and 2017. The instrument has a field of view of 290 km, 12-bit radiometric quantization capability, and consists of thirteen spectral bands covering from light visible to shortwave infrared (SWIR) [7]. In the petabyte data storage facility at GEE, users can access almost all of MSI's Sentinel-2 L2A data collection which has been scaled to surface reflectance (SR). A description of the detailed image information that we used can be seen in Table 1. We carried out an analysis for the SDB by considering the main seasonal variations that pass through Indonesian waters. Two main seasons that pass through Indonesian waters, namely: a) the west season which originates from the northwest wind which brings moist and warm air from the Asian Continent to the Australian Continent and is also known as the northwest monsoon (NWM), and b) In June – August, on the contrary, the southeast wind blows from Australia towards Asia bringing dry air or also known as the southeast monsoon season (SEM). Between the two main seasons there is transition season I (TS1) (March – May), and transition season II (TS2) (September – November). The time period used in this research is from December 2022 – November 2023. In one season that lasts three months, the Sentinel-2 L2A MSI image with a temporal resolution of ten days, theoretically has at least ± 30 data archives that specifically record the waters of the Karimunjawa Islands [36].

Table 1: General specifications of Sentinel-2 imagery used in this study [31]

Band Number	Spectral Band	Wavelength (nm)	Spatial Resolution (m)
B1	Coastal Aerosol	443	60
B2	Blue	490	10
B3	Green	560	10
B4	Red	665	10
B5	Vegetation Red Edge	705	20
B6	Vegetation Red Edge	740	20
B7	Vegetation Red Edge	783	20
B8	Near Infrared (NIR)	842	10
B8b	Vegetation Red Edge	865	20
B9	Water Vapour	940	60
B10	SWIR-Cirrus	1,375	60
B11	SWIR	1,610	20
B12	SWIR	2,190	20

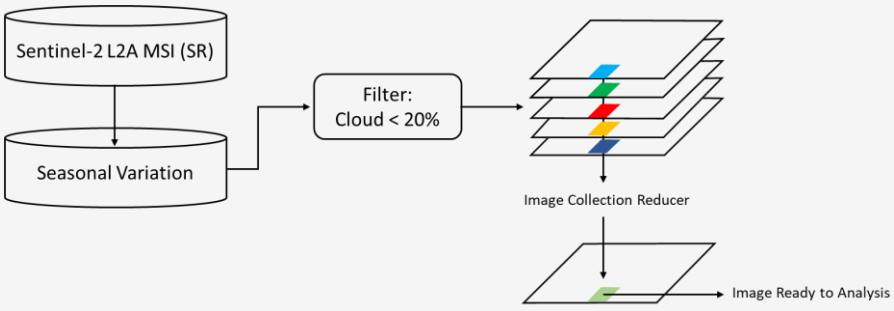

Figure 2: Sentinel-2 L2A MSI image data pre processing

Images are selected from within the collection by applying filters to select clouds ≤ 20% and "mean" for the reducer. Reducer terminology is a tool in GEE to efficiently collect data within a certain time period into one data using simple statistics (e.g., minimum, maximum, mean, median, and standard deviation). The reducer applied automatically in the earth-engine will be adjusted to all channels in the Sentinel-2 L2A MSI image. A detailed explanation regarding the acquisition of Sentinel-2 L2A MSI image data and the reducer method can be seen in the illustration in Figure 2. In GEE, a reducer is a function that aggregates data across multiple values, such as different time steps, spatial areas, or image bands, to produce a single value or a smaller set of values. Using the image, the stack of layers represents different time steps, bands, or other dimensions. The reducers process this stack to generate a new output image, where each pixel's value reflects the minimum, maximum, mean, median, or standard deviation of the original pixel values across the layers.

2.4 Sunglint Correction

Seasonal variations will directly and indirectly influence the radiometric properties of waters both vertically and horizontally. Sunglints are specular reflections of solar radiation that contact the water surface in bright conditions and the water surface tends to be wavy [37]. This phenomenon depends on physical aspects of the ocean including radiation intensity, viewing angle, and roughness of the water surface [38]. Generally, this phenomenon can be found in high resolution images [1]. The sunglint correction applied in this study uses an algorithm developed by [37] on GEE devices. Although generally this correction is applied to high resolution images, we carried out sunglint correction to avoid the potential for sunglints that are not visually clearly visible in the water appearance in the image. In addition to using suglint-corrected image data as input for the ML model ensemble, we also use images that have been scaled in SR for comparison of results.

2.5 Clear Water Computing

We selected Sentinel-2 images that have been scaled in SR on the GEE platform with the minimum clouds and are in the range of ≤ 20% in each season, so that there are four image collection packages (e.g., West Season, Transition Season I, East Season, and Season Transition II). After the cloud filter is applied, we use the "reducer" facility in GEE to produce a collection of image data using the "mean" statistical method. Based on observations made in the study area, the mean statistical value is able to produce promising visual appearances when compared with other statistical methods such as median. To separate land and water features in this study, an automated NDWI (normalized difference water index) is defined in Equation 1:

$$NDWI = \frac{Green - NIR}{Green + NIR}$$

Equation 1

The threshold value that we use as a delimiter between land pixel values and water bodies is 0.2, where values less than the threshold are defined as water and values more than the threshold are land features. At this stage, the clean water image scaled at the SR level is ready to be regressed into ML. Next, a deglint clean water image was constructed by applying the sunglint correction algorithm by [37] on the GEE device. The clean water image scenario on the SR and deglint scale is then exported for further analysis.

2.6 Random Forest Regression

Random forest or in this study will be shortened to RF is an ensemble learning method that initially produces many sets of subsamples. To improve the prediction accuracy of the model, the model samples a data set and forecasts a decision tree with randomly selected characteristics from each subset. Next, the model integrates the prediction results from each decision tree [11] and [39].

The RF method is very good for regression prediction and classification because it shows nonlinear relationships between variables [39], and has recently shown significant results in SDB [1]. Without using a logarithmic transformation, this technique can determine the relationship between directly measured water depth and spectral reflectance values [10]. his study uses the random forest regression library on the Google Earth Engine device. The main input data used as an ensemble of ML models using the algorithm in this RF study include depth data, clean water images with SR scale and deglint bands.

2.7 Accuracy Assessment

We randomly selected 20% of the total sample data to be input into the RF model, using MSE, RMSE, and R^2 as quantification methods to assess performance and accuracy. Qualitatively, we use 1:1 plots and compare the transverse profile of the bottom topography of the waters between the model data and the reference depth from measurements in the field. There are three transects as instruments for conducting qualitative analysis of the RF algorithm's ability to reconstruct the depth of shallow waters around the waters of Kemujan Island. The three transect locations are relatively similar in location to previous research [1] (see Figure 3). The existence of transects in relatively similar locations is intended to compare the results of qualitative accuracy of the results of tuning the influence of seasonal variables with previous research. Transect 1 is in the southwest part of Kemujan Island which crosses waters with complex topography, namely sandy bottom near the coastline, lagoon, reef flat, back reef, and ends at reef crast. The depth distribution in these waters ranges from 1.0 – 13 meters. Transect 2 is in the waters of the southeastern part of Kemujan Island which stretches across the basic geomorphology which is dominated by reef flats, back reefs, and reef crests. The reference depth distribution shows a range between 1.2 – 3.3 meters. Transect 3 is in the waters of the northwestern part of Kemujan Island. Geomorphologically, the location of this transect is not much different from the characteristics of the first transect. The depth distribution in this last transect is in the range 1.6 – 5.0 meters.

Recent studies involving machine learning in SDB reconstruction in the same area were conducted by [1] and [31] and studies by [32] and [36] were conducted using an empirical approach, but most of these studies were still conventional in nature using GIS-based software, while this study is slightly more sophisticated by using cloud-based tools to improve the efficiency of SDB processing. The detailed workflow of this study can be seen in Figure 4. This study uses Sentinel-2 imagery processed seasonally for nearshore bathymetry mapping with RFR. The process begins by applying NDWI to separate water from land, producing a clear water image with spectral reflectance (SR) in blue, green, and red bands as predictors. Simultaneously, sunglint correction was applied, generating a deglint clear water image with corrected blue, green, and red bands as additional predictors. The depth data is split into 70% for training and 30% for testing. The RFR model is then used to predict water depth, followed by accuracy assessment and performance analysis to evaluate model precision.

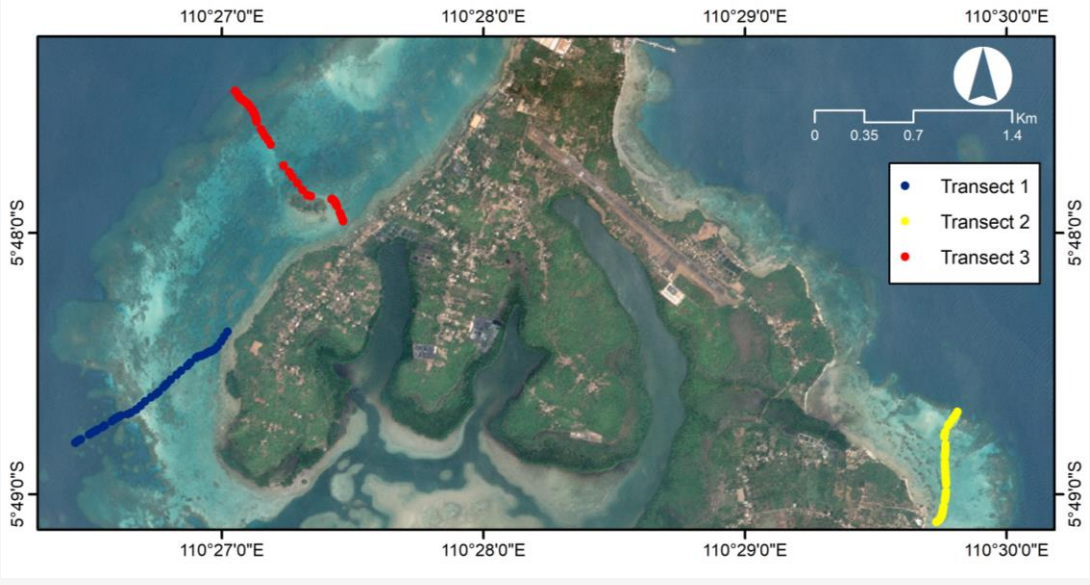

Figure 3: Transect locations for conducting descriptive accuracy analysis

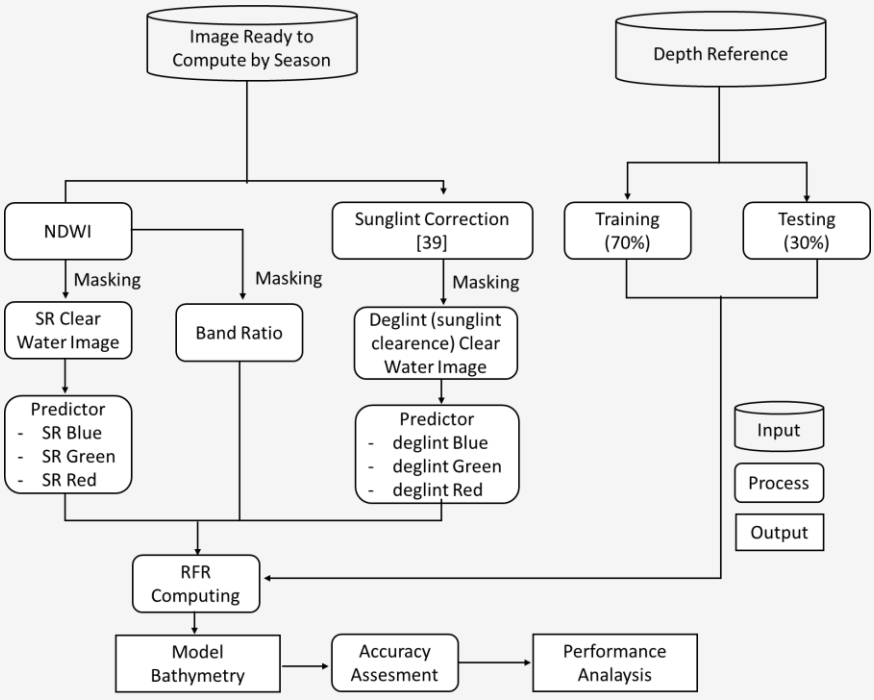

Figure 4: Nearshore bathymetry accuracy study workflow

3. Results

3.1 Spatial Distribution of Bathymetric Model

Based on the results of the RF model on SR imagery, the bathymetry of shallow waters can be well illustrated which is reflected in the basic geomorphological class of the waters around the waters of Kemujan Island (see Figure 5). Lagoons and reef slopes, which have topographical features that tend to be deeper, can be depicted with precision, although in each season the area of the lagoon that is depicted is not always the same (Point A in Figure 5). Nearshore water areas, reef flats, reef crests and intertidal areas can be described as shallower than other areas in the surrounding waters. In general, there were no abnormal bathymetric model results based on the geomorphic class. Additionally, inconsistencies can be seen in the geomorphic reconstruction of the shallow lagoon in the western part of Kemujan Island in Transition Season II (Area B in Figure 5). However, in the west season and transition I, where Indonesian waters are dominated by high precipitation, there is quite massive cloud cover, which reduces the ability of the RF model to generate shallow water bathymetric information. Due to massive disturbances from the concentration of cloud cover in the West season and Transition I season, the eastern and central water areas of Kemujan Island cannot be depicted properly, as well as the intertidal areas (Area C in Figure 5). This cloud disturbance also has an impact on the low accuracy obtained during both seasons. The maximum depth difference that can be reconstructed is 2.6 meters.

In the Sentinel-2 image which undergoes a sunglint correction process and the result is referred to as a deglint image, shallow water bathymetry can be replicated by the model accurately, where the geomorphic appearance of the lagoon and reef flat is consistently depicted similarly throughout all seasons, as well as the fore reef area and intertidal zone (Point A in Figure 6). The maximum depth difference resulting from shallow water bathymetric reconstruction each season is 2.3 meters. The influence of cloud disturbances in the western season around the intertidal zone can be reduced (Area B Figure 6) but in the eastern part it cannot be overcome significantly (Area C Figure 6). However, the SDB reconstruction results using deglint image input show that there was a miscalculation in the sunglint correction process which resulted in land residue (Area D Figure 6). Land residues that are not successfully removed in the correction process participate in the RF model ensemble process and become invalid values. The results of SDB reconstruction using the RF algorithm in the final scenario are using band ratios in the calculation process. The bands that are rationed are blue and green, where these two bands tend to be able to significantly capture solar radiation in the water column.

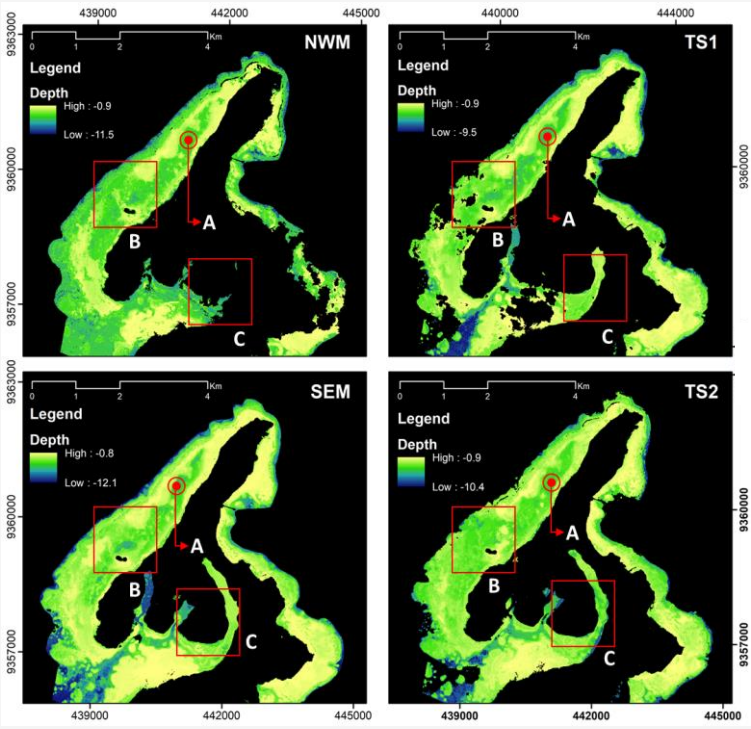

Figure 5: Shallow water bathymetry models resulting from RF algorithm on SR imagery. [A] Lagoons [B] Inconsistencies in replicating the geomorphic features of reef flats [C] Cloud interference which reduces the algorithm's ability to depict intertidal areas

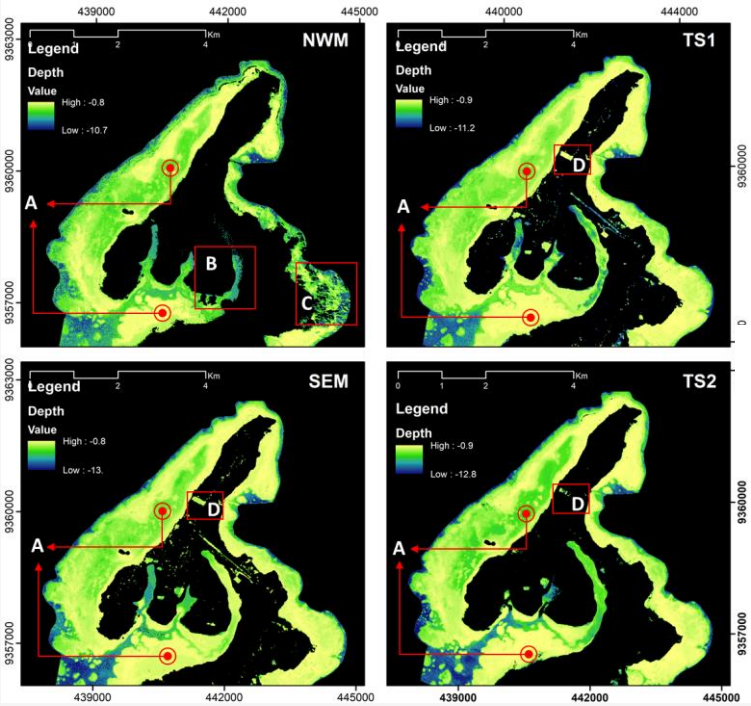

Figure 6: Shallow water bathymetry models resulting from the RF algorithm on deglint images [A] well-delineated lagoon, reef flat and intertidal zone areas [B] disturbances in the intertidal zone that have been successfully reduced [C] Cloud disturbances in the west monsoon in the eastern area in the West Monsoon, [D] remaining sunglint correction residues be an invalid result

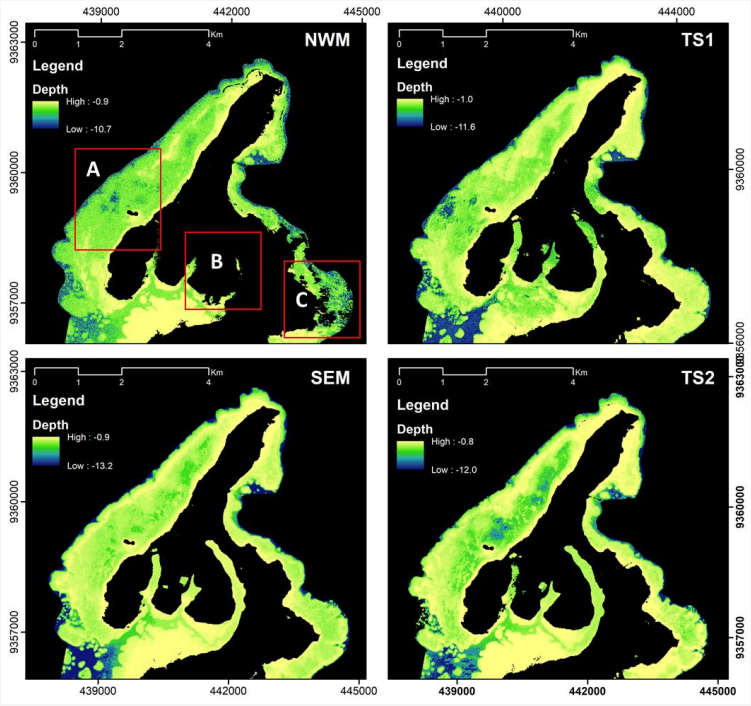

Figure 7: Shallow water bathymetry models resulting from the RF algorithm on ratio images (blue/green) [A] Excessive appearance of the lagoon based on the depth of the model results [B] Cloud disturbance to the intertidal zone [C] Cloud disturbance to the waters of the eastern part of Kemujan Island

As a result, in transition season 1 (TS1) to TS2, spatially the bathymetry model is well distributed, but the NWM shows that the depth of the geomorphic features of the lagoon tends to be more excessive than in other seasons (Area A in Figure 7). The intertidal and eastern regions (Area B and C in Figure) in the west monsoon cannot be illustrated perfectly due to the influence of massive cloud cover disturbances. The maximum depth difference resulting from shallow water bathymetry reconstruction each season is 2.5 meters.

3.2 Seasonal Accuracy Comparison

A summary of accuracy tests based on R^2, MSE, and RMSE can be seen in Table 2. In direct proportion to the spatial distribution of the resulting shallow water bathymetry model, NWM produces the lowest accuracy for the MSE, RMSE, and R^2 parameters in all scenarios. In the RF model with SR image input, the best accuracy for MSE and RMSE is found in TS2 with values of 0.84 and 0.92 respectively. However, the best R^2 is found in SEM. Apart from that, the RF bathymetry model with deglint image input shows a pattern similar to the previous scenario, where in TS2 the best accuracy for MSE and RMSE is obtained with values of 0.92 and 0.96 respectively. Apart from that, the best R^2 that can be obtained in the band deglint scenario is in SEM.

The RF model scenario involving band ratio (blue/green) shows significant results, where the best quantification values for R^2, MSE, and RMSE are found in SEM. Sequentially the accuracy values are 0.87, 0.74, and 0.82. Overall, the band ratio in SEM produces the best accuracy of all scenarios run to test machine learning models using the RF algorithm.

3.3 References Plot by 1:1

A 1:1 plot was performed to provide an illustration of the RF model's ability to replicate shallow water bathymetry based on the same data points used to measure R^2 MSE, and RMSE accuracy. Overall, the comparison plot between the depth reference and the model in all scenarios can be seen in Figures 8-10. Based on the model results for all scenarios and seasons, the bathymetry model produced at SEM and TS-2 for all scenarios produces the best accuracy between the data reference depth with the results of SDB reconstruction via the RF algorithm. Accuracy will be considered better if the data set is close to the fit line (solid black). Overall, it can be seen that the best accuracy based on 1:1 plot analysis that can be produced is in the depth range of 1 to 4 meters, and accuracy begins to decrease after a depth of 4 meters and fails to replicate the maximum depth.

Table 2: Quantitative accuracy test results

Model Input	Season	R^2	MSE	RMSE
SR Sentinel-2	North West Monsoon	0.71	1.88	1.37
	Transition 1 (TS1)	0.77	0.86	0.92
	South East Monsoon	0.86*	0.96	0.98
	Transition 2 (TS2)	0.74	0.84*	0.92*
Deglint Sentinel-2	North West Monsoon	0.63	1.94	1.39
	Transition 1 (TS1)	0.77	1.27	1.12
	South East Monsoon	0.86*	0.99	0.99
	Transition 2 (TS2)	0.85	0.92*	0.96*
Band Ratio	North West Monsoon	0.52	2.90	1.70
	Transition 1 (TS1)	0.74	1.39	1.18
	South East Monsoon	0.87*	0.74*	0.86*
	Transition 2 (TS2)	0.79	0.85	0.92

** indicates the best accuracy based on the lowest MSE, RMSE and highest R^2*

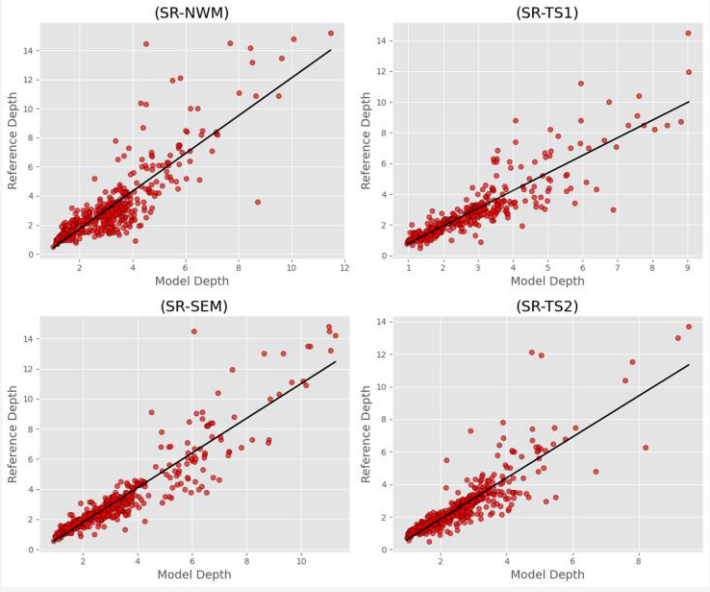

Figure 8: Accuracy based on a 1:1 plot between reference data and the bathymetry model via the RF algorithm in the ML SR image scenario

3.4 Underwater Topography Profile

We analyzed depth data in the field using a shallow water bathymetry model reconstructed by the RF algorithm. Comparative data was prepared in three transects in the Southeast, Southwest and Northwest sections. We compared field data with the best SDB model in each scenario, namely transition season II (TS2) for the SR and deglint band scenarios, while SEM for the scenario using the ratio band. Overall, in the three transects, the depth profile resulting from the SDB model is able to completely replicate the water depth using the existing depth reference. At relatively shallow depths, namely 0-4 meters, the bathymetry can be replicated well, but after the depth reaches >4 meters, there is a significant difference in depth even though descriptively the bathymetry pattern tends to be similar. In transect 1, it can be identified that the SDB model with deglint image input is able to replicate the bathymetric depth of shallow water with precision, so that from the initial depth point near the coast to the end of the transect it appears to adhere to the reference data. Apart from that, the SDB model on images with band ratio input gives good results up to a depth of 4 meters, but after going deeper than that the accuracy starts to decrease and tends to give underestimated depth values. Meanwhile, the image with SR image input failed to replicate the depth at the 33rd depth point (see Figure 11).

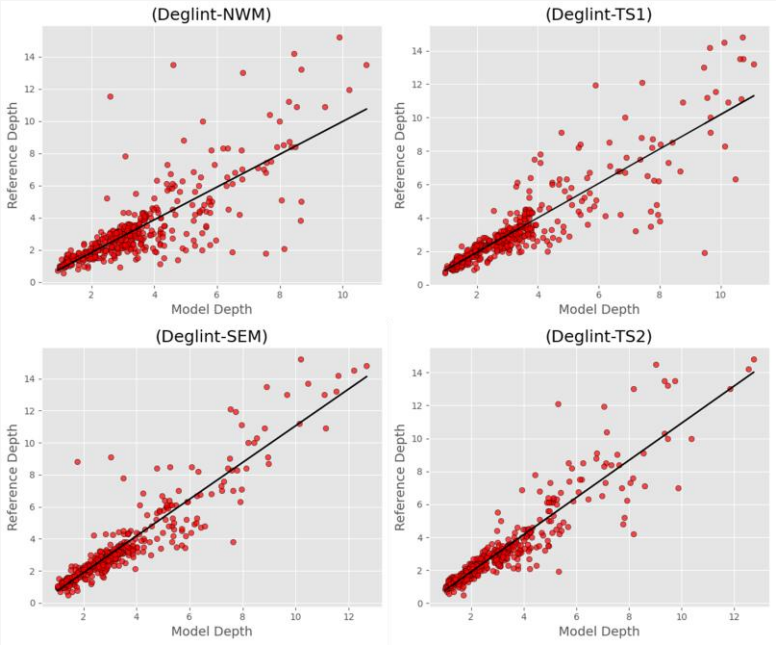

Figure 9: Accuracy based on a 1:1 plot between reference data and the bathymetry model via the RF algorithm in the ML image deglint scenario

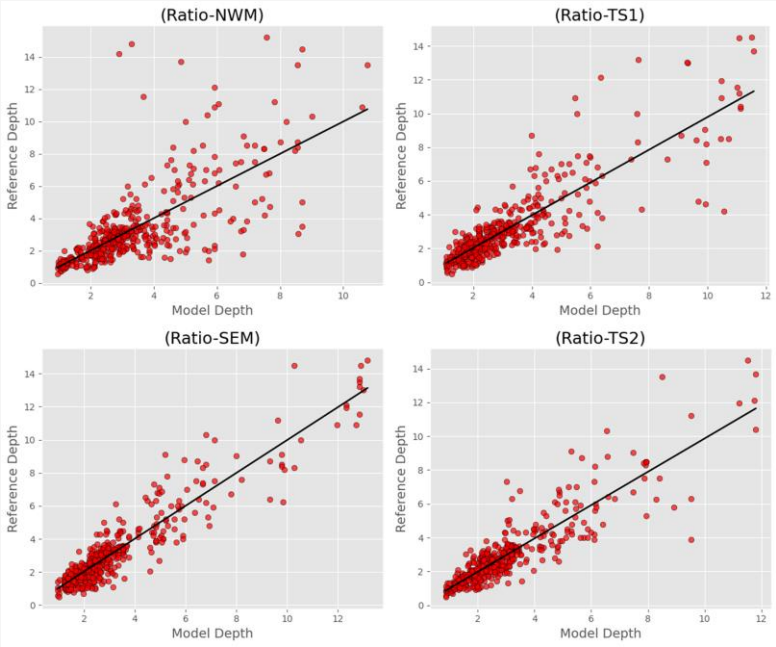

Figure 10: Accuracy based on a 1:1 plot between reference data and the bathymetry model via the RF algorithm in the ML band ratio image scenario

Transect 2 is in a water area that tends to slope with the main geomorphic appearance being reef flat and dominated by a sandy substrate. In the image, the appearance of the second transect also tends to be bright. As a result, the SDB model with SR and deglint image input is able to describe depth precisely, but in the ratio band the bathymetry results tend to overestimate along the transect to a depth of 6 meters, but beyond that depth the SDB model with glance ratio band input provides significant performance (see Figure 12).

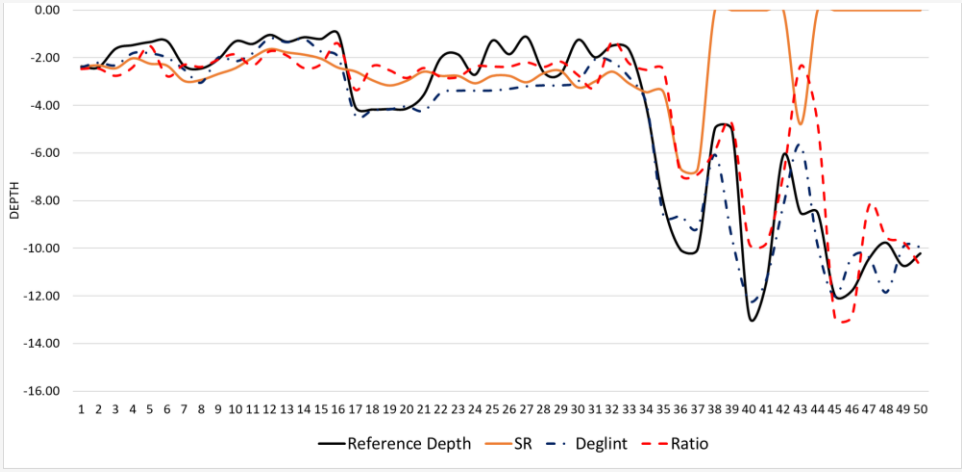

Figure 11: Underwater topographic profile on transect 1

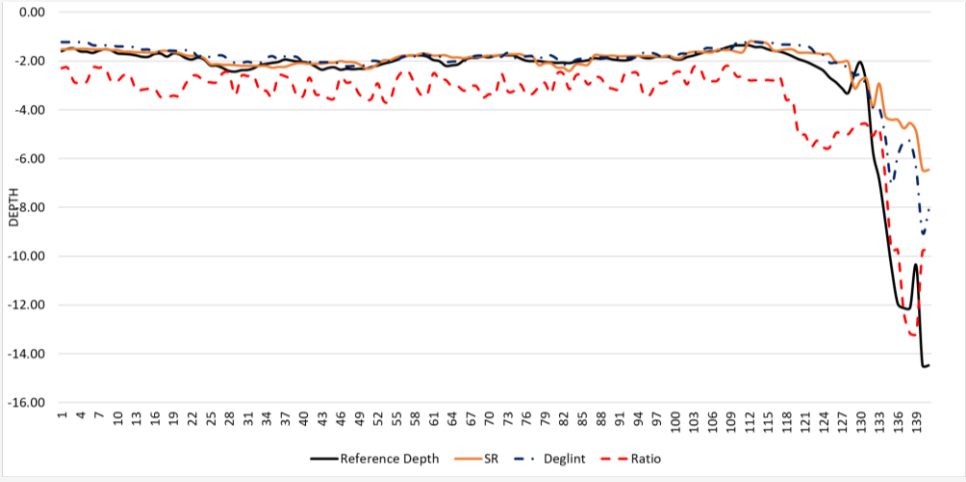

Figure 12: Underwater topographic profile on transect 2

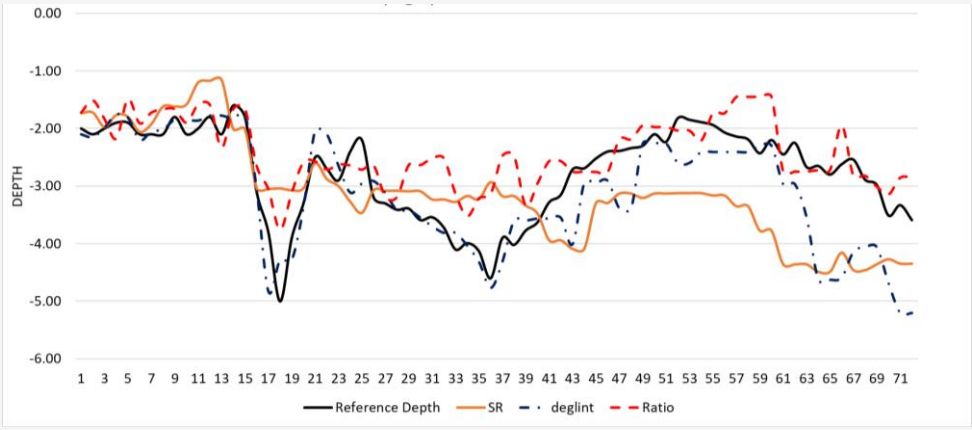

Figure 13: Underwater topographic profile on transect 3

Transect 3 is the most complex location with geomorphic features in the form of shallow reef near the beach, lagoon, back reef, reef crest and fore reef with various benthic varieties such as macroalgae, coral and seagrass [1] and [40]. The transect results of the bathymetry model at the third location tend to be varied, but it is clear that the bathymetry model with input deglint image data provides significant performance and is able to replicate near perfect topography up to point 37, the rest of the time the deglint model gives results that tend to overestimate (see Figure 13).

4. Discussion

This study places the influence of seasonal variations as the main indicator in comparison to test the accuracy of shallow water bathymetry models using the SDB method. The approach used is statistical (empirical) based with machine learning (ML) via the Random Forest (RF) algorithm. The seasons that cross Indonesian waters are divided into four main seasons, namely NWM, TS1, SEM, and TS2. This study was conducted to assess the influence of seasonal variations on the accuracy resulting from the SDB model with an RF algorithm that considers input data scenarios in the scale of surface reflectance (SR), image deglint, and input image band ratio. Assessment of accuracy is carried out using quantitative and qualitative approaches. We put R^2 RMSE and MSE to provide an explanation of the performance of the SDB model in a quantitative approach. Another unique consideration, as stated by [1] is that accuracy must also be assessed from the model's ability to replicate the bottom topography of the waters as reflected in the underwater topography profile and 1:1 plot which shows the depth distribution and errors.

In general, based on the performance shown by R^2, RMSE and MSE, the SDB model with band ratio input gives the best results with RMSE = 0.86 m, and this value almost matches the best performance produced by [1] which have RMSE = 0.82 m using remote sensing imagery with much higher resolution (Worldview-2, Spatial resolution = 2 m). However, the R^2 performance produced in this study far exceeds previous research, namely R^2 = 0.87 compared to previous research, R^2 = 0.71 in the same regression model and study location. The use of the RF algorithm in this study was able to significantly outperform the use of a physics-based approach by previous research [30] with RMSE = 1.1 m, as well as the use of PlanetScope imagery using new emerged bands [3] yang memperoleh akurasi RMSE = 1.5 – 1.9 m, and [41] with RMSE = 1.01 m.

This study considers the fact that high R2 values are not always directly proportional to quantitative accuracy based on RMSE, and this is reflected through the SDB model with input data deglint to SEM season. Qualitatively, in the 1:1 plot, this study agrees with [1] that only very shallow waters can be replicated well by SDB, especially the RF algorithm (approximately 0 – 4 m). However, significant differences can be seen in the resulting underwater topographic profiles, where the profiles of transects 1 and 3 are more complex in terms of morphology, while transect 2 produces relatively similar profiles.

This study considers that atmospheric influences, especially massive cloud concentrations in the NWM season, result in a significant decrease in accuracy, where increasing the use of band deglint and ratio does not provide a comprehensive effect, although it can visually reduce the effect of atmospheric disturbances in the intertidal zone. The distribution of the bathymetry model is accurate as long as the depth does not exceed 4 meters, deeper than that there are many model errors in replicating shallow water bathymetry (see plot 1:1, Figures 8-10). Furthermore, this study highlights the influence of seasonal variations (NWM, TS1, SEM, TS2) on the accuracy of the SDB model in the waters of Kemujan Island. Apart from that, the multiple scenarios used (SR, deglint, and ratio) did not provide significant differences in accuracy results. The results of the study show that season has an influence on accuracy, so that in other applications it is hoped that the use of SDB in obtaining depth information can take into account the influence of season, especially in more strategic applications such as: marine spatial planning and shipping navigation. The cloud computing device that we use provides superior processing and storage time efficiency, so that processing residue does not require a specific storage device.

5. Conclusion

This study shows that seasonal variations can have an influence on the resulting accuracy. The RF algorithm, which is generally able to overcome the non-linear relationship between depth and reflectance value, is then applied to the image on the SR, deglint and ratio scales. In the study time period, namely December 2022 – November 2023, SEM and TS2 which are close to each other produce higher accuracy than NWM in the multiple scenarios applied. The NWM variation is proven to have systematically low accuracy due to atmospheric disturbances, especially cloud concentration, whereas the SEM variation shows its superiority

based on the results of quantitative accuracy calculations. The maximum depth distribution that can be described by the RF model is 0-4 meters, below which the model generally fails to comprehensively replicate shallow water bathymetry. In this case, the underwater topographic profile is significant for assessing the accuracy of bathymetry from a descriptive qualitative perspective. This research found that season (NWM, TS1, SEM, TS2) significantly influences accuracy compared to multi-scenarios (SR, deglint, ratio) in Kemujan Island waters. Cloud processing enables efficient RF algorithms to run without leaving residue, as well as saving processing time and storage capacity. For recommendation, further research needs to improve accuracy by adding the number of samples to make it more comprehensive, or conducting an in-depth analysis of the image radiometry system to avoid bias in time series imagery.

Acknowledgement

The researchers extend their gratitude to Universitas Gadjah Mada for the support provided through the Final Assignment Recognition (Rekognisi Tugas Akhir/RTA) grant, as outlined in the 2024 RTA Program announcement letter: 4971/UN1.P1/PT.01.01/2024, and the assignment letter: 5286/UN1.P1/PT.01.03/2024. This support has played a significant role in advancing the research and contributing to the development of scientific knowledge.

References

[1] 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.

[2] Sagawa, T., Yamashita, Y., Okumura, T. and Yamanokuchi, T., (2019). Shallow Water Bathymetry Derived by Machine Learning and Multitemporal Satellite Images. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 8222–8225. https://doi.org/10.1109/IGARSS.2019.8899043.

[3] Khakhim, N., Kurniawan, A., Wicaksono, P. and Hasrul, A., (2024). 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.

[4] 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.

[5] Eugenio, F., Marcello, J., Mederos-Barrera, A. and Marques, F., (2022). High-Resolution Satellite Bathymetry Mapping: Regression and Machine Learning-Based Approaches. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60. https://doi.org/10.1109/TGRS.2021.3135462.

[6] Ma, Y., Xu, N., Liu, Z., Yang, B., Yang, F., Wang, X. H. and Li, S., (2020). Satellite-Derived Bathymetry using the ICESat-2 Lidar and Sentinel-2 Imagery Datasets. *Remote Sensing of Environment*, Vol. 25. https://doi.org/10.1016/j.rse.2020.112047.

[7] Kurniawan, A., Khakhim, N., and Wicaksono, P. (2024). Seasonal Variation Influence to Water Image Properties to Retrieve Nearshore Bathymetry Based on Cloud Machine Learning Approach. *International Journal of Geoinformatics*, Vol. 20(7), 77–92. https://doi.org/10.52939/ijg.v20i7.3407.

[8] Zhu, J., Yin, F., Qin, J., Qi, J., Ren, Z., Hu, P., Zhang, J., Zhang, X. and Wang, R., (2022). Shallow Water Bathymetry Retrieval by Optical Remote Sensing Based on Depth-Invariant Index and Location Features. *Canadian Journal of Remote Sensing*, Vol. 48(4), 534–550. https://doi.org/10.1080/07038992.2022.2104235.

[9] Tonina, D., McKean, J. A., Benjankar, R. M., Wright, C. W., Goode, J. R., Chen, Q., Reeder, W. J., Carmichael, R. A. and Edmondson, M. R., (2019). Mapping River Bathymetries: Evaluating Topobathymetric LiDAR Survey. *Earth Surface Processes and Landforms*, Vol. 44(2), 507–520. https://doi.org/10.1002/esp.4513.

[10] Kasvi, E., Salmela, J., Lotsari, E., Kumpula, T. and Lane, S. N., (2019). Comparison of Remote Sensing Based Approaches for Mapping Bathymetry of Shallow, Clear Water Rivers. *Geomorphology*, Vol. 333, 180–197. https://doi.org/10.1016/j.geomorph.2019.02.017.

[11] Wu, Z., Mao, Z., Shen, W., Yuan, D., Zhang, X. and Huang, H., (2022). Satellite-Derived Bathymetry Based on Machine Learning Models and an Updated Quasi-Analytical Algorithm Approach. *Optics Express*, Vol. 30(10). https://doi.org/10.1364/oe.456094.

[12] Niroumand-Jadidi, M., Bovolo, F. and Bruzzone, L., (2020). SMART-SDB: Sample-Specific Multiple Band Ratio Technique for Satellite-Derived Bathymetry. *Remote Sensing of Environment*, Vol. 251. https://doi.org/10.1016/j.rse.2020.112091.

[13] 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.

[14] Stumpf, R. P., Holderied, K. and Sinclair, M., (2003). Determination of Water Depth with High-Resolution Satellite Imagery. *Limnology and Oceanography*, Vol. 48(1), 547–556. https://doi.org/10.4319/lo.2003.48.1_part_2.0547.

[15] Xu, N., Ma, X., Ma, Y., Zhao, P., Yang, J. and Wang, X. H., (2021). Deriving Highly Accurate Shallow Water Bathymetry from Sentinel-2 and ICESat-2 Datasets by a Multitemporal Stacking Method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, 6677–6685. https://doi.org/10.1109/JSTARS.2021.3090792.

[16] Liu, Y., Zhou, Y. and Yang, X., (2024). Bathymetry Derivation and Slope-Assisted Benthic Mapping Using Optical Satellite Imagery in Combination with ICESat-2. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 127. https://doi.org/10.1016/j.jag.2024.103700.

[17] Amrari, S., Bourassin, E., Andréfouët, S., Soulard, B., Lemonnier, H. and Le Gendre, R., (2021). Shallow Water Bathymetry Retrieval Using a Band-Optimization Iterative Approach: Application to New Caledonia Coral Reef Lagoons Using Sentinel-2 Data. *Remote Sensing*, Vol. 13(20). https://doi.org/10.3390/rs13204108.

[18] 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.

[19] Wan, J. and Ma, Y., (2021). Shallow Water Bathymetry Mapping of Xinji Island Based on Multispectral Satellite Image using Deep Learning. *Journal of the Indian Society of Remote Sensing*, Vol. 49(9), 2019–2032. https://doi.org/10.1007/s12524-020-01255-9.

[20] 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.

[21] Chu, S., Cheng, L., Cheng, J., Zhang, X, and Liu, J., (2023). Comparison of Six Empirical Methods for Multispectral Satellite-derived Bathymetry. *Marine Geodesy*, Vol. 46(2), 149–174. https://doi.org/10.1080/01490419.2022.2132327.

[22] Misra, A., Vojinovic, Z., Ramakrishnan, B., Luijendijk, A. and Ranasinghe, R., (2018). Shallow Water Bathymetry Mapping Using Support Vector Machine (SVM) Technique and Multispectral Imagery. *International Journal of Remote Sensing*, Vol. 39(13), 4431-4450. https://doi.org/10.1080/01431161.2017.1421796.

[23] Susa, T., (2022). Satellite Derived Bathymetry with Sentinel-2 Imagery: Comparing Traditional Techniques with Advanced Methods and Machine Learning Ensemble Models. *Marine Geodesy*, Vol. 45(5), 435-461. https://doi.org/10.1080/01490419.2022.2064572.

[24] Manessa, M. D. M., Kanno, A., Sekine, M., Haidar, M., Yamamoto, K., Imai, T. and Higuchi, T., (2016). Satellite-Derived Bathymetry Using Random Forest Algorithm and Worldview-2 Imagery. *Geoplanning: Journal of Geomatics and Planning*, Vol. 3(2), 117. https://doi.org/10.14710/geoplanning.3.2.117-126.

[25] 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.

[26] Abdul Gafoor, F., Al-Shehhi, M. R., Cho, C. S. and Ghedira, H., (2022). Gradient Boosting and Linear Regression for Estimating Coastal Bathymetry Based on Sentinel-2 Images. *Remote Sensing*, Vol. 14(19). https://doi.org/10.3390/rs14195037.

[27] Bué, I., Catalão, J. and Semedo, Á., (2020). Intertidal Bathymetry Extraction With Multispectral Images: A Logistic Regression Approach. *Remote Sensing*, Vol. 12(8). https://doi.org/10.3390/RS12081311.

[28] Khakhim, N., Kurniawan, A., Wicaksono, P. and Hasrul, A., (2024). 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.

[29] Wulandari, S. A. and Wicaksono, P., (2021). Bathymetry Mapping using PlanetScope Imagery on Kemujan Island, Karimunjawa, Indonesia. *IOP Conference Series: Earth and Environmental Science*, Vol. 686(1). https://doi.org/10.1088/1755-1315/686/1/012032.

[30] Sesama, A. S., Setiawan, K. T. and Julzarika, A., (2021). Bathymetric Extraction Using Planetscope Imagery (Case Study: Kemujan Island, Central Java). *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, Vol. 17(2). https://doi.org/10.30536/j.ijreses.2020.v17.a3445.

[31] European Space Agency. (2015). *Sentinel-2 User Handbook*. [Online]. Available: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook. [Accessed: Dec. 25, 2024].

[32] Wicaksono, P., (2016). Improving the Accuracy of Multispectral-Based Benthic Habitats Mapping Using Image Rotations: The Application of Principle Component Analysis and Independent Component analysis. *European Journal of Remote Sensing*, Vol. 49, 433–463. https://doi.org/10.5721/EuJRS20164924.

[33] Wicaksono, P. and Lazuardi, W., (2018). Assessment of Planetscope Images for Benthic Habitat and Seagrass Species Mapping in a Complex Optically Shallow Water Environment. *International Journal of Remote Sensing*, Vol. 39(17), 5739-5765. https://doi.org/10.1080/01431161.2018.1506951.

[34] Rahman, W. and Wicaksono, P., (2019). Aplikasi Citra WorldView-2 untuk Pemetaan Batimetri di Pulau Kemujan Taman Nasional Karimunjawa. [WorldView-2 Imagery Application for Bathymetric Mapping on Kemujan Island, Karimunjawa National Park] *Jurnal Penginderaan Jauh Indonesia*, Vol. 1(1), 333–338. http://jurnal.mapin.or.id/index.php/jpji/issue/view/1.

[35] Wicaksono, P., (2010). *Integrated Water Column Correction for Improving Satellite-Based Benthic Habitat Mapping*. Master's Thesis, Faculty of Geography, Universitas Gadjah Mada. https://etd.repository.ugm.ac.id/home/detail_pencarian/45791.

[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] Hedley, J. D., Harborne, A. R. and Mumby, P. J., (2005). Simple and Robust Removal of Sun Glint for Mapping Shallow-Water Benthos. *International Journal of Remote Sensing*, Vol. 26(10), 2107–2112. https://doi.org/10.1080/01431160500034086.

[38] Wang, M., Hu, Q., Zhu, X., Lu, Y., Jiao, J., Zhou, J., Ju, W., Chen, Z., Li, C., Huang, Y. and Hong, Q., (2022). Correction of Multi-Scale Sunglint Reflections from the Water Surface in Airborne High-Spatial Resolution Optical Images. *Optics Express*, Vol. 30(25). https://doi.org/10.1364/oe.478887.

[39] 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.444557.

[40] Wicaksono, P., (2015). Perbandingan Akurasi Metode Band Tunggal Dan Band Rasio Untuk Pemetaan Batimetri Pada Laut Dangkal Optis. *Simposium Nasional Sains Geoinformasi*, Vol. 4. https://doi.org/10.13140/RG.2.1.1340.3286.