PAPER • **OPEN ACCESS**

Seagrass Habitat Suitability Models using Multibeam Echosounder Data and Multiple Machine Learning Techniques

To cite this article: M A H Muhamad and R Che Hasan 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1064** 012049

Seagrass Habitat Suitability Models using Multibeam Echosounder Data and Multiple Machine Learning Techniques

M A H Muhamad[1] and R Che Hasan[1]

[1]Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

E-mail: rozaimi.kl@utm.my

Abstract. Seagrass beds are important habitats in the marine environment by providing food and shelter to dugongs and sea turtles. Protection and conservation plans require detail spatial distribution of these habitats such as habitat suitability maps. In this study, machine learning techniques were tested by using Multibeam Echo Sounder System (MBES) and ground truth datasets to produce seagrass habitat suitability models at Redang Marine Park. Five bathymetric predictors and seven backscatter predictors from MBES data were used to representing topography features and sediment types in the study area. Three machine learning algorithms; Maximum Entropy (MaxEnt), Random Forests (RF), and Support Vector Machine (SVM) were tested. The results revealed that MaxEnt and RF models achieved the highest accuracy (93% and 91%, respectively) with SVM produced the lowest (67%). Depth was identified as the most significant predictor for all three models. The contributions of backscatter predictors were more central for SVM model. High accuracy models showed that suitable habitat for seagrass is distributed around shallow water areas (<20 m) and between fringing reef habitats. The findings highlight that acoustic data and machine learning are capable to predict how seagrass beds are spatially distributed which provide important information for managing marine resources.

Keyword: Habitat Suitability Model, Maximum Entropy, Random Forest, Support Vector Machine, Bathymetric, Backscatter, Multibeam Echosounder, Redang Marine Park

1. Introduction

Around the globe, anthropogenic activities and natural disasters have become great risks as they adversely affect seagrasses and their ecological functions [1-4]. The distribution loss in Malaysia is also increasing at an alarming rate [5, 6]. Despite there being no effects from their existing distribution, studies predict a significant reduction in their coverage area, which will have a detrimental effect on the surrounding ecosystem. [7, 8]. Hence, monitoring the seagrass habitat using advanced technologies is very important for improving protection of this important resource [9-11]. This is particularly vital for Malaysia's marine protected areas, with little attention and a lack of information to conserve and preserve the seagrass habitat [12-14]. There are about 13 species of seagrass, and *Halophila decipiens*, *Halodule pinifolia*, and *Halophila minor* are among the seagrass species that are found in Redang Marine Park (RMP), Terengganu, Malaysia [15]. Their habitats in Malaysia are also expected to face serious threats as anthropogenic activities and natural disasters continue to happen [5]. As a result, the most effective and suggested method is to conserve and preserve seagrass habitats early by mapping their habitats [3, 16-18]. Creating a map of the present habitat and simulating a suitable habitat for seagrasses has made a significant contribution to marine park managers and decision-makers in making decisions and subsequently actions [2, 17, 19-21]. However, for monitoring their distribution and managing their natural habitats, a reliable, cost-effective, efficient, and precise mapping technique is required [22, 23]. In this regard, among the most widely used modelling and prediction techniques is Habitat Suitability Modelling (HSM) [2, 17, 19-21].

HSM has been used by marine scientists and researchers to predict the distribution of marine habitats [12, 24-26]. HSMs, on the other hand, have shown varying levels of model performance, and studies have yet to identify a single optimum model for several taxa and seafloor features [17]. Choosing and implementing HSM requires careful consideration as the usage of numerous modeling algorithms may have an impact on model performance, contribution of predictors, and habitat prediction [21, 27-29]. As a result, several studies used more than one HSM [29, 30]. Besides, previously published studies advocate the use of presence-absence modelling techniques, whereas some strongly encourage the use of modelling algorithms that only take into presence-only [31]. The reason behind is that comparisons between machine learning algorithms and regression model algorithms has been widely conducted [32, 33], while other researchers compared machine learning algorithms [34-36] and ensemble HSMs developed by others instead of relying on a HSM [37]. However, the application of HSM is also affected by the selection of predictors [36, 38-40], spatial resolution of predictors [28, 41-43], and the size of window analysis to derive predictors [12, 44].

As a result, the use of comparable approaches to several HSMs has enable some of the first organized efforts by marine researchers and scientists to identify suitable algorithms for mapping marine habitats. The possible options in this study are the use of multibeam echosounder (MBES) data. Due to the high spatial resolution of MBES data and extensive coverage area, it provides a significant contribution to the monitoring of seagrass habitat distributions [12, 45]. In previous studies, high spatial resolution MBES data allowed scientists or researchers to derive numerous spatially continuous datasets (i.e., MBES predictors) [46-50]. The predictive value of these variables is better than that of MBES data (i.e., bathymetric map and backscatter mosaic) as they provide the detailed information on seabed topography and substrata [51, 52]. Among the various predictors, the majority of previous research typically used slope, curvature, eastness, and northness derived from bathymetric map and Gray Level Co-occurrence Matrix (GLCM) texture features [53-57] and Angular Range Analysis (ARA) parameters derived from backscatter mosaic [58-60]. Though bathymetric map has higher importance in modelling seagrass habitats [12, 61-64], other backscatter predictors may also contribute significantly to improve the model [62, 65]. For instance, a significant correlation between the seagrass habitat distribution and fine sandy substrates exists [66]. Thus, assessing the role of MBES predictors is critical since they may reflect topographical and substratum characteristics of the study area.

Several prior studies have evaluated HSM's efficacy in mapping and modeling seagrass ecosystems. Hu, Zhang, Chen, Liu, Ye, Jiang, Zheng, Du and Chen [17] compared the accuracy of two different seagrass habitat prediction models and found that using an ensemble model be significantly better than using a single model. In addition, in order to make a more accurate model, Araújo and New [67] suggested that an ensemble model was used rather than just one model. Likewise, Effrosynidis, Arampatzis and Sylaios [68] and Stankovic, Kaewsrikhaw, Rattanachot and Prathep [69] analyzed the performance of several machine learning algorithms and concluded that the results using the Random Forest (RF) algorithm were excellent compared to those from the other algorithms. In addition, Downie, von Numers and Boström [21] made a comparison of the performances of two machine learning algorithms (i.e., Maximum Entropy and Generalised Additive Model) in the prediction of seagrass and suggested an ensemble approach is able to produce a best-performing model, rather than one algorithm. Furthermore, Valle, van Katwijk, de Jong, Bouma, Schipper, Chust, Benito, Garmendia and Borja [70] made a comparison of the performances of five machine learning models (i.e., Maximum Entropy and Generalised Additive Model) and three regression-based models in the prediction of seagrass and suggested Boosted Regression Trees (BRT) and Random Forest (RF), which were the most effective modeling algorithms. Both remote sensing and non-remote sensing datasets have been used in these studies (e.g., Hu, Zhang, Chen, Liu, Ye, Jiang, Zheng, Du and Chen [17]). Downie, von Numers and Boström [21] used EOS Terra-MODIS images and topographical data (e.g., depth and sandy shores) to produce two machine learning models, that predict seagrass in the Archipelago Sea, SW Finland. Other study used satellite imagery and ground-truth survey data to model seagrass habitat distribution [17]. In addition, Effrosynidis, Arampatzis and Sylaios [68] used satellite imagery, digital elevation models (DTMs), and bathymetric survey data to produce multiple seagrass prediction models. However, comparable studies of SDM utilizing high-resolution images of MBES data, in particular for predicting seagrass habitat, are lacking.

As a result, this study will address the following research gaps and needs: (1) assessing the potential of the bathymetric map, backscatter mosaic, and bathymetric and backscatter predictors for modelling predicted seagrass habitat distribution in the RMP area is highly essential; and (2) identifying a reliable method to model the predicted distribution of seagrass habitat is critical.

2. Materials and methods

2.1. Study area

The study area encompasses the whole region of the Redang archipelago in Terengganu, Malaysia, which encompasses an area of about 182.55 km^2. The Redang archipelago is located between latitudes $5^{\circ}55'$ and $5^{\circ}40'$ N and about longitudes $102^{\circ}55'$ and $103^{\circ}10'$ E. Under the Marine Park Malaysia Act, the Redang archipelago was officially designated as a marine park, titled as the Redang Marine Park (RMP). Order 1994 (Fisheries Act in 1995). The largest island of this archipelago is Redang, which is surrounded by Pinang, Lima, Ekor Tebu, Kerengga Kecil, Kerengga Besar, Paku Besar, Paku Kecil, and Ling, the eight tiny islands. Redang, Pinang, Lima, and Ekor Tebu are among the islands that have been designated as RMP areas. As a whole, the RMP area is inhabited by two categories of marine ecosystems, which are coral reefs [71] and seagrasses [15]. Three seagrass species, which are *Halophila decipiens, Halodule pinifolia,* and *Halophila minor*, were discovered in the vicinity of the RMP, out of a total of 13 species that live along Peninsular Malaysia's coasts [15]. Chagar Hutang, the Redang River estuary, Pasir Panjang, and Teluk Dalam are habitats for most of these species (Figure 1). These species are often present in subtidal areas with a water depth of 2.5 to 24 meters [15] and are mainly found on seafloor covered by silt and substrate [15].

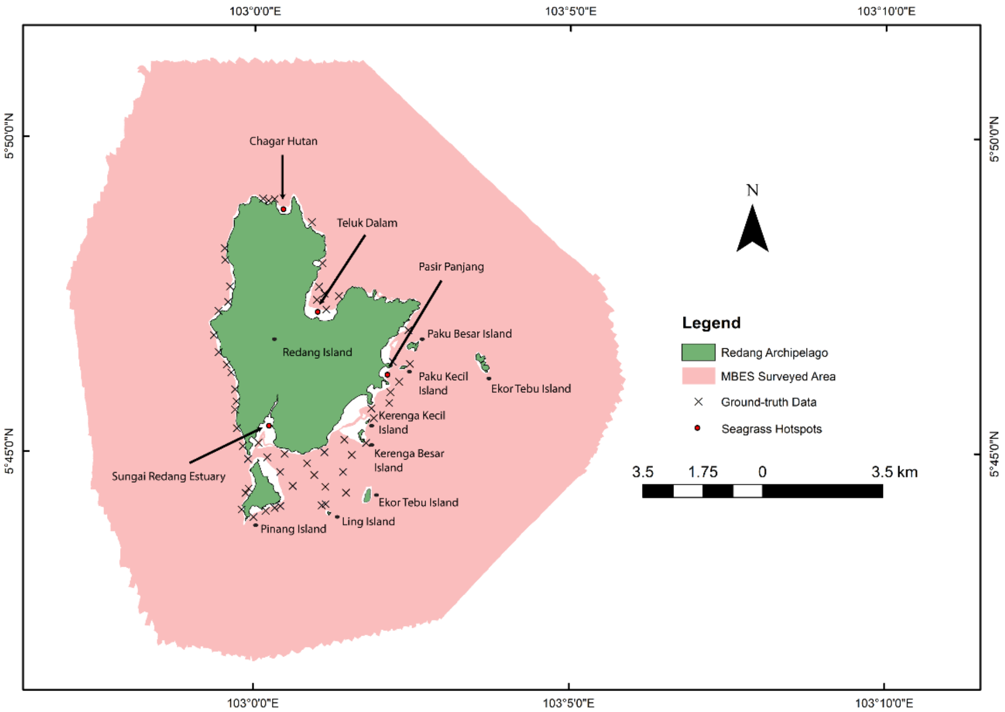

Figure 1. The study area is located at Redang Marine Park (RMP). RMP consists of Redang, the main island, and eight smaller islands, including Pinang, Lima, Ekor Tebu, Kerengga Kecil, Kerengga Besar, Paku Besar, and Paku Kecil, as well as Ling, Pinang, Lima, and Ekor Tebu.

2.2. Data acquisition

2.2.1. Multibeam Echosounder (MBES) Survey. The Multibeam Echosounder (MBES) data that was used came from MBES survey using a Kongsberg EM2040C MBES system mounted on the port side. From April 6th to 24th, 2019, this survey was conducted at the RMP area (Figure 1). The Seafloor Information System (SIS) by Kongsberg Maritime was used for data acquisition. In a high-density equidistant mode, MBES ran at 300 kHz, using a range of various ping rates and pulse lengths that were autonomously adjusted according to water depth (400 beams per ping). During data acquisition, sound velocity profiles were acquired using the Valeport Monitor Sound Velocity Profiler. During data acquisition, at regular intervals of ten minutes, tides were collected using a TideMaster Portable Tide Gauge. The position of the vessel was determined using VERIPOS LD3 and differential global positioning system (DGPS) mode, which is based on GPS/GLONASS corrections obtained through radio transmission from the satellite positioning service by Fugro Marine Star. The vessel's motion was logged (e.g., heave, pitch, roll, yaw, and heading) using Kongsberg MRU 5 and set aside for data processing using a precise motion sensor system.

2.2.2. Ground-truth survey: Underwater imagery sampling. An action camera, the GoPro Hero 4, was installed on a steel frame and was used to collect ground-truth data. The frame was then lowered into the water to the seabed, with each observation taking place at a maximum depth of about 30 metres. The targeted seagrass habitat spots were carefully selected to capture all seagrass distributions based on previous research [72]. Video and picture confirmation of the presence of seagrass at each drop site were obtained from footage. The AtlasLink H10 Smart Antenna was used to obtain surface location data. The DGPS mode, which acquired the GNSS correction service (precision of 50 cm), enhanced the position accuracy of each drop point. A total of 91 underwater imagery samples were observed, and seagrass occurrence was identified and classified for each sample.

2.3. Data processing

2.3.1. Bathymetric Maps and Backscatter Mosaics. The following methods were carried out in CARIS HIPS & SIPS version 10.4 to process raw MBES bathymetry data: a) filtering the positioning; b) filtering the motion (e.g., heading, heave, pitch, roll); c) tidal correction; d) sound speed correction; e) and removing noise. Meanwhile, Fledermaus Geocoder Toolbox (FMGT) 7.4.4 was used to process the raw MBES backscatter data. A total of 400 beams were used, with 0° and 90° starting and cut-off beam angles. Beam average backscatter was used to produce a backscatter mosaic image. The processing parameters were kept as close as possible to the FMGT default settings to maintain continuity between the surveys of the backscatter mosaics, even though this may have been to the detriment of the subjective nature of the mosaics. The bathymetry map and backscatter mosaic were projected into the UTM Zone 48 N (WGS 1984) coordinate system and gridded at 1 m.

2.3.2. Bathymetric and backscatter predictors. Geomorphometry for Ecology tool in TASSE Toolbox v.1.1 is used to derive bathymetric predictors from bathymetric map [73], including slope, curvature [74], eastness, and northness [25, 75-77]. Meanwhile, the intensity level (dB) from the backscatter mosaic 32bit was rendered to grayscale values (i.e., backscatter mosaic 8bit). Apart from these backscatter mosaics, angular range analysis (ARA) parameters (i.e., Phi and characterization) and gray-level co-occurrence matrix (GLCM) texture features (i.e., homogeneity, entropy, and correlation) were derived to characterise the sediment composition and textures of the seafloor. The following table lists the bathymetric predictors and backscatter predictors. All bathymetric and backscatter predictors were gridded at 1 m and projected using UTM Zone 48 N (WGS 1984).

2.3.3. Seagrass occurrence data. For seagrass occurrence data, the presence or absence of seagrass habitat is determined by referring to the underwater imagery sampling data. The process of determining the occurrence of seagrass habitat by analysing the image (i.e., underwater imagery sample). The images found the presence of seagrass habitat classified as seagrass presence point. Meanwhile, the images found the absence of seagrass habitat classified as a point of absence of seagrass absence point. From 91 samples, only 56 were used for seagrass habitat suitability modeling. For modeling using MaxEnt, seagrass with presence-only data was used with 9 points of seagrass occurrence data. Meanwhile, for presence-absence data, 56 points were used for RF and SVM approaches. 75% of the points were chosen at random for model training and 25% for model testing [12, 78, 79].

2.3.4. *Pearson's correlation coefficient.* Using Pearson's correlation coefficient, the relationship between all predictors were initially determined by removing highly correlated predictors [80] and excluded during the model fitting process [29, 81]. Predictors with low correlation coefficient values, i.e., r greater than 0.5, were considered to have low correlation [82]. The predictor was retained for model fitting when the correlation coefficient values were less than 0.5 and vice versa. Pearson's correlation coefficient was calculated and a visualization of a correlation matrix was drawn using the corrplot function in the corrplot r package in the RStudio 1.4.1106 [83].

2.3.5. Seagrass habitat suitability modeling (SHSM): MaxEnt, RF, and SVM. This study assessed three machine learning algorithms to predict seagrass habitat within the RMP area: Maximum Entropy (MaxEnt), Random Forest (RF), and Support Vector Machine (SVM). In this study, SHSM was carried out using the MaxEnt machine learning algorithm in MaxEnt v3.4.1 software [84], whereas the RF and SVM machine learning algorithms were carried out using sdm, which is a r package in the RStudio 1.4.1106. Three stages were involved in generating the SHSM, including data input, model fitting and evaluation, and model prediction. In order to model the predicted seagrass habitat, seagrass occurrence data was used as species occurrence data and all selected predictors were used as input variables (i.e., bathymetric maps, backscatter mosaics, bathymetric predictors, and backscatter predictors). For seagrass occurrence data, 75% was used to train the SHSM, while the other 25% was used to validate the SHSM for all model algorithms. For model fitting and evaluation, all default settings (e.g., regularized multiplier, maximum number of background points, maximum iterations, and coverage threshold) were used for all models derived using MaxEnt algorithms since these settings have been shown to produce good results [85-87]. For RF and SVM, all default settings were tested on this study. All models used 20 replicates and the bootstrap replicate method to produce SHSM. For model prediction, each cell in the models represented the seagrass habitat suitability index (SHSI). The SHSI scale varied from 0 to 1, with 0 indicating low suitability for seagrass habitat and 1 for high suitability for seagrass habitat.

2.3.6. Model performance. Each algorithm extracted Receiver Operating Characteristic (ROC) curves, and the model performance was accessed through the use of Area Under the Curve (AUC) [88]. The AUC is a statistical test that quantifies predictive model performance over threshold ranges by using both presence and absence data. Phillips, Anderson and Schapire [84] method was used in this study. This method calculates the ROC AUC using randomly generated pseudo-absences rather than actual absences. The AUC was calculated based on the specificity and sensitivity of the predictive model. The specificity and sensitivity values reflected the accuracy with which suitable or less suitable seagrass habitats were classified. The mean AUC values of twenty replicates were produced in this study. AUC values greater than 0.9 were considered excellent, those between 0.8 and 0.9 were considered very good, those between 0.7 and 0.8 were considered satisfactory, and values less than 0.7 indicated a lack of discriminative ability [89, 90].

2.3.7. Predictor importance. Furthermore, the predictor importance analysis indicated the most contributed MBES predictor [91]. While the Maxent, RF, and SVM models is being trained, they keep track of which MBES predictors are contributing for model fitting. Each step of the Maxent, RF, and SVM machine learning algorithm increases the gain of the models by modifying the coefficient for a single feature. All machine learning algorithms assign the gain increase to the MBES predictors on which the feature is dependent. At the end of the training process, convert to percentages and calculate the percent contribution [84, 92].

3. Result and analysis

3.1. Correlation of predictors

This study found weak linear relations (correlation coefficient (r) is below 0.5) among many bathymetry predictors but not among those generated from backscatter data. Five predictors were shown to have weak linear relationships (bathymetry, slope, eastness, northness, and GLCM entropy). It was discovered that the remaining predictors have strong linear relationships with the extracted predictors. The following figure illustrates a visualization of a correlation matrix (i.e., a correlogram) of predictors with weak and strong linear relations (2).

	Bathy	Slp	East	North	Curv	Bs32	Bs8b	Corr	Ent	Hom	Char	Phi
Bathy	**1.00**											
Slp	-0.15	**1.00**										
East	-0.38	0.21	**1.00**									
North	0.00	-0.39	-0.14	**1.00**								
Curv	**-0.57**	**0.83**	0.37	0.03	**1.00**							
Bs32	**-0.50**	0.36	**0.51**	-0.40	0.46	**1.00**						
Bs8	**-0.50**	0.36	**0.51**	-0.40	0.46	1.00	**1.00**					
Corr	**-0.58**	-0.39	-0.41	0.24	-0.07	-0.17	0.17	**1.00**				
Ent	0.24	0.25	0.23	-0.23	-0.02	0.21	0.21	**-0.58**	**1.00**			
Hom	**-0.71**	0.04	0.03	0.00	0.30	0.32	0.32	**0.68**	-0.47	**1.00**		
Char	0.40	-0.26	-0.46	**0.59**	-0.22	**-0.88**	**-0.88**	0.13	-0.24	-0.21	**1.00**	
Phi	0.44	-0.25	-0.48	**0.55**	-0.24	**-0.88**	**-0.88**	0.08	-0.20	-0.24	**1.00**	**1.00**

Figure 2. A correlation matrix for MBES predictors showed correlation coefficient (*r*). Bathy=Bathymetric Map; Slp=Slope; East=Eastness; North=Northness; Curv=Curvature; Bs32=Backscatter 32bit; Bs8=Backscatter 8bit; Corr=GLCM Correlation; Ent=GLCM Entropy; Hom=GLCM Homogeneity; Char=ARA Characterisation; and Phi=ARA Phi.

3.2. Model performance

For MaxEnt, RF, and SVM, the AUC values for the train models were ranged from 98% to 100. For test models, the MaxEnt algorithm showed a slight decrease in AUC values, which is 93%. While the test RF models also showed the performance of this model, the AUC value has decreased slightly, which is 93%. On the other hand, test SVM models showed a significant decrease in AUC values when they obtained 84% of AUC values. The model's overall performance was assessed based on train AUC and test AUC values, which indicate excellent performance for MaxEnt and RF models [93]. Meanwhile, only the test SVM model showed good model performance [93] (Table 1).

Table 1. Training and test AUC values for MaxEnt, RF, and SVM models.

MaxEnt		**RF**		**SVM**	
Train AUC (%)	Test AUC (%)	Train AUC (%)	Test AUC (%)	Train AUC (%)	Test AUC (%)
98	93	100	91	100	84

3.3. Predictor importance

The results showed the contribution rate and importance of MBES predictors to models developed using three machine learning algorithms (Figure 3). The bathymetric map is the most influential variable for all models, ranging from 48.7% to 85.3%. The ranking of the percent contributions of other predictors varies for each model. For example, the percent contribution of eastness and northness scored 1.2% and 6.8%, respectively, when making the MaxEnt model. While RF models showed that the percent contributions of these predictors (i.e., eastness and northness) were increased. The percent contribution of eastness and northness has increased to 30.7% and 40.3%, respectively. While SVM models showed that the percent contributions of eastness showed a slight increase when compared to the percentages from the RF model, vice versa, the percent contribution of northness showed a slight decrease when compared to the percentages from the RF model. For slope, the production of the MaxEnt model is less dependent on this predictor, for which the percentage contribution is very low, i.e., 2.4%. Meanwhile, RF and SVM models showed an increase of not more than 29.5% when compared to the percentage contribution from the MaxEnt model. While the percent contribution of GLCM entropy also showed a minor contribution to the MaxEnt model's development, i.e., 4.3%. While the increment of percent contributions of GLCM entropy for RF and SVM models also showed a similar trend to the slope, which is 8.3% and 24.0%.

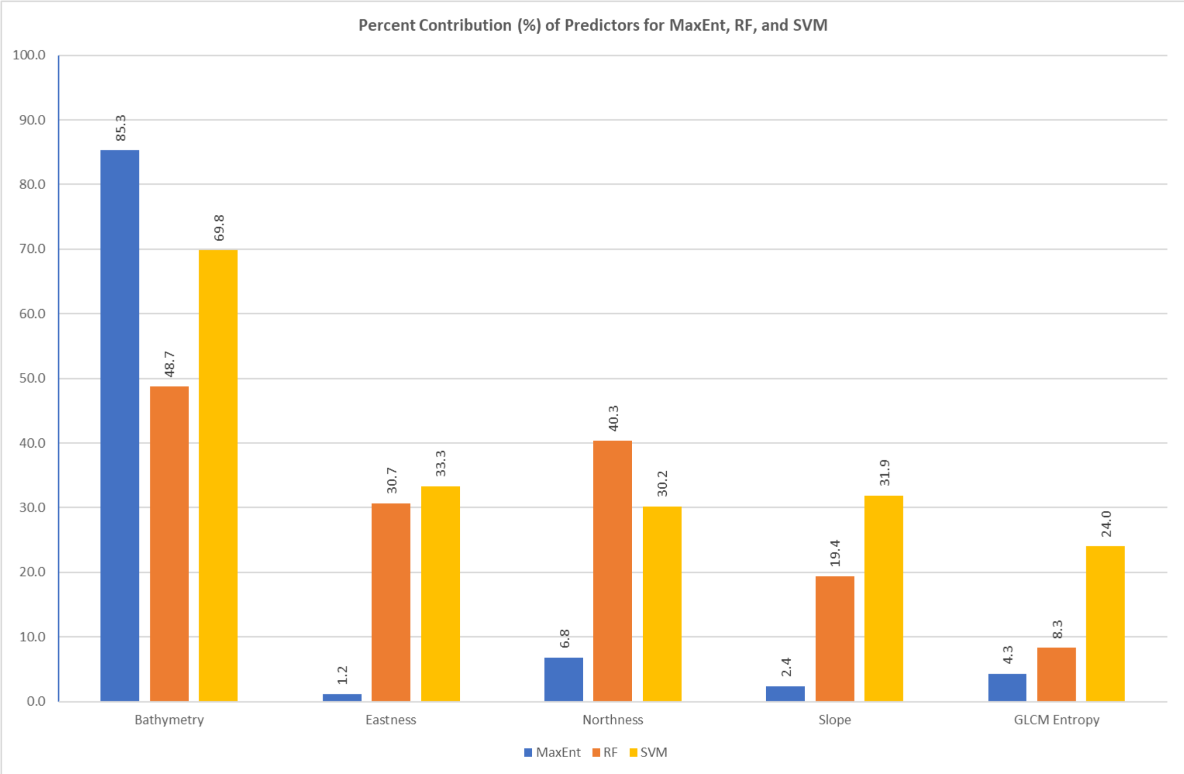

Figure 3. Percent of contributions of the MBES predictors based on MaxEnt, RF, and SVM models.

3.4. Seagrass habitat suitability model

The Maximum Entropy (MaxEnt) machine learning algorithms was applied to create a Seagrass Habitat Suitability Model (SHSM). The result from **Figure 4** showed the MaxEnt model. The habitat suitability index (SHSI) for this model is ranged from 0 to 1. showed the MaxEnt model. The habitat suitability index (SHSI) for this model ranges from 0 to 1. The MaxEnt model illustrated that the distribution of high SHSI is concentrated along the shoreline of the Redang archipelago. This result indicated that the seagrass habitats are distributed across the shallow water areas in the study area. Furthermore, seagrass habitats are expected to be suitable for living between the islands of Redang, Paku Besar, Kerengga Besar, Ekor Tebu, and Pinang. In this area, seagrass habitat was expected to be found in shallow water areas (i.e., less than 20 m) between bordering reefs. This model also indicated that seagrass is not suitable for inhabiting deep water areas (i.e., more than 20 m).

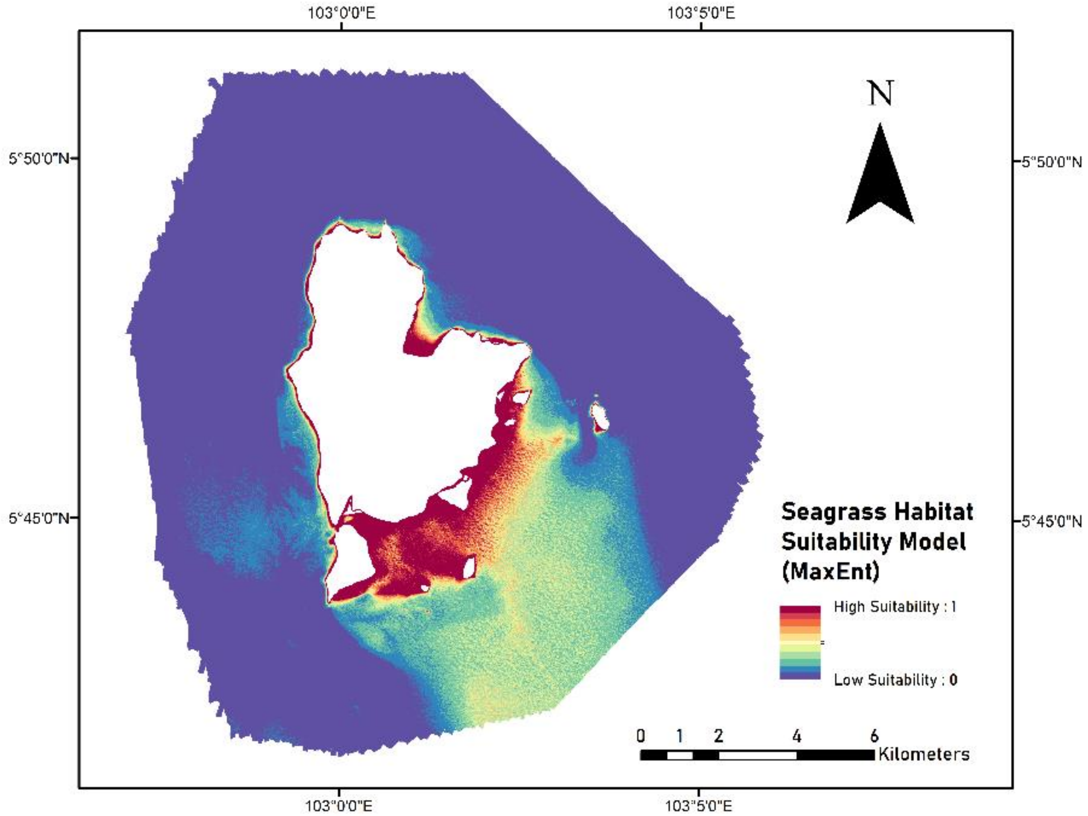

Figure 4. The prediction model made by the MaxEnt algorithm on the suitability of the seagrass habitat over the MBES surveyed area.

The Random Forest (RF) machine learning algorithm was applied to create a Seagrass Habitat Suitability Model (SHSM). The result from Figure 5 showed the RF model produced a seagrass habitat suitability index (SHSI) that ranged from 0 to 0.3. However, the distribution of SHSI range values for this model is smaller compared to the model generated using the MaxEnt algorithm. This model yielded similar findings when the distribution of SHSI and spatial pattern of this model seemed to be like that of the MaxEnt model. In general, this model illustrated that the highest SHSI was concentrated along the shoreline of the Redang archipelago. Furthermore, study findings predicted that seagrass habitats are more concentrated in the areas between the Redang, Kerengga Besar, Kerengga Kechil, Ekor Tebu, and Pinang islands. Furthermore, patchy isolated pixels were spotted throughout the RMP area and mostly found in the deep-water area that is not suitable for seagrass habitat. Like the prediction of the MaxEnt model that illustrated seagrass habitats were predicted to be suitable in the shallow water areas (i.e., less than 20 m). There are several constraints produced by this model, such as the SHSI values for this model have a small range, which means low variability in spatial distribution. In addition, SHSI values in this model are not homogeneous, there is a drastic change in SHSI values between the adjacent cells that leads to false interpretation.

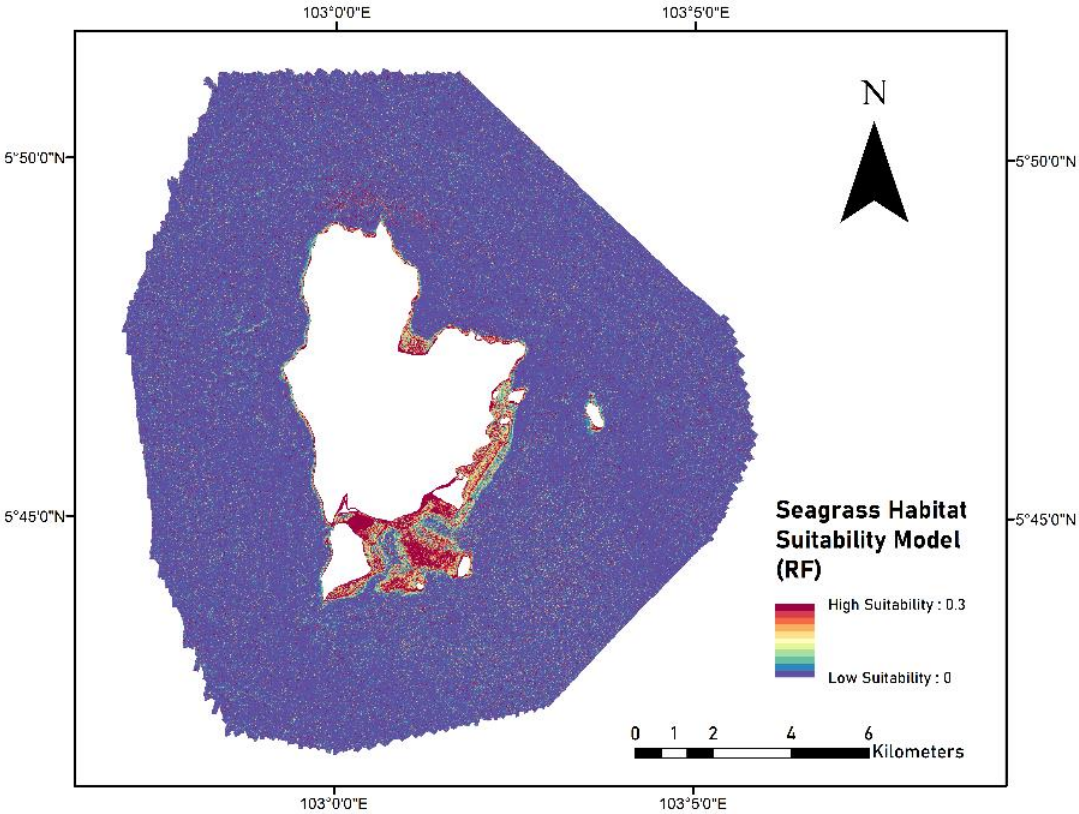

Figure 5. The prediction model made by the RF algorithm on the suitability of the seagrass habitat over the MBES surveyed area.

The Support Vector Machine (SVM) machine learning algorithm was applied to create a Seagrass Habitat Suitability Model (SHSM). The result from **Figure 6** showed the SVM model produced a seagrass habitat suitability index (SHSI) for this model that ranged from 0 to 0.07. The distribution of SHSI range values for the SVM model is smaller compared to the model generated using the RF and MaxEnt models. The SVM model produced contradictory findings when the distribution of SHSI and the spatial pattern of this model looked different from the MaxEnt and RF models. In general, this model illustrated that the highest SHSI was distributed patchily along the shoreline of the Redang archipelago. Furthermore, study findings predict that seagrass habitats are scattered in between the Redang, Kerengga Besar, Kerengga Kechil, Ekor Tebu, and Pinang islands. Furthermore, patchy isolated pixels were also spotted throughout the RMP. Most of these pixels are distributed in the deep-water area that is not suitable for seagrass habitat. Similar to the predictions of the RF and MaxEnt models, this model is also able to predict suitable habitats in shallow water areas (i.e., less than 20 m). This model generates several constraints. For example, the SHSI values for this model have a small range, implying low spatial distribution variability. Additionally, the SHSI values in this model are not homogeneous; there is a significant difference in SHSI values between adjacent cells, resulting in incorrect interpretation.

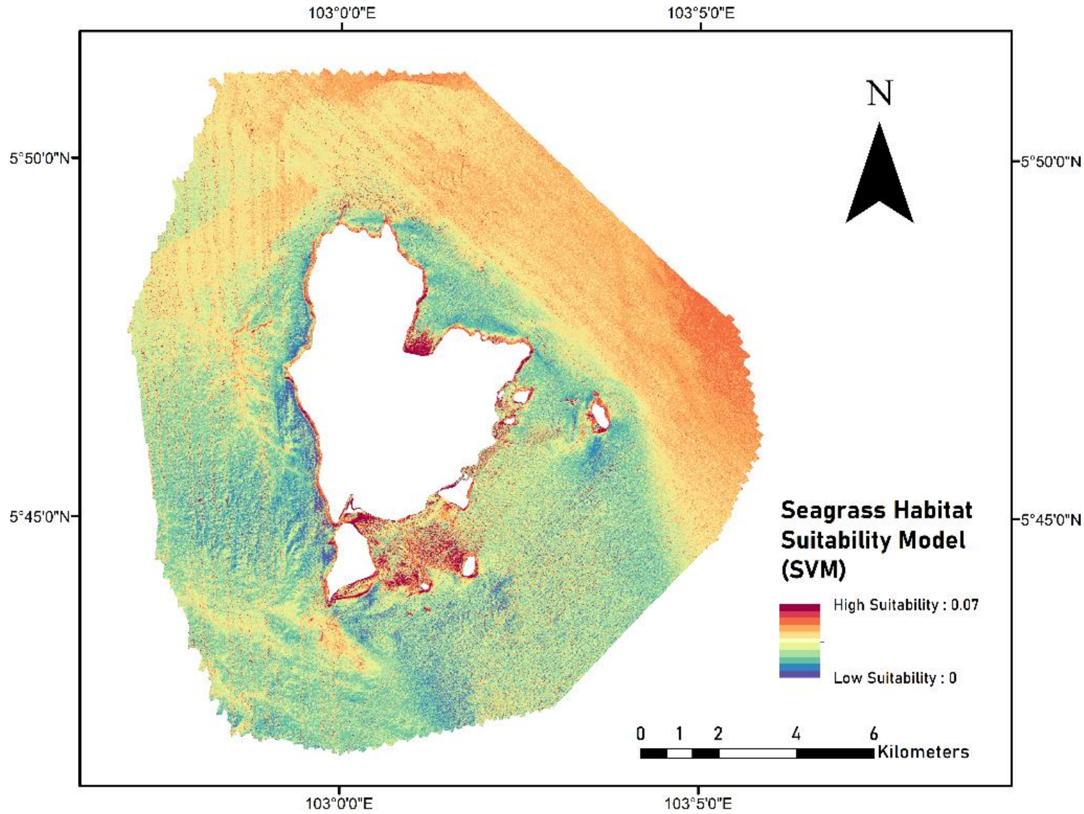

Figure 6. The prediction model made by the SVM algorithm on the suitability of the seagrass habitat over the MBES surveyed area.

4. Discussion

The performance of the models varied according to the model type utilized [29]. In summary, the findings indicate that two models, MaxEnt and RF, perform exceptionally well in terms of prediction. Meanwhile, results showed that SVM models perform well in terms of prediction. Research findings indicated that seagrass habitat distribution in the RMP region was successfully predicted using proposed machine learning techniques. While both RF and SVM require presence-absence data, MaxEnt requires only presence-only data, but all prediction models performed the best and could be utilized individually. Despite the fact that RF has been implicated in a number of prior studies [17, 68, 69], and SVM [94] has done effectively, the current study findings showed that MaxEnt may still be utilized effectively as a machine learning modeling strategy that uses presence-only data without significantly reducing model performance. However, there is a large difference in the range of SHSI values between the models. The results indicated that MaxEnt's SHSI values were slightly larger than those predicted by RF and SVM. This finding contradicts prior studies indicating that MaxEnt consistently predicted a narrow range of SHSI values, whereas RF consistently predicted a wide range of SHSI values [17, 37] and SVM [94] are predicted to have a broader range of SHSI values. Additionally, MaxEnt models predict the growth of suitable seagrass habitats throughout the entire shoreline and around the Redang archipelago. Additionally, MaxEnt models imply that appropriate seagrass habitats are dominant in shallow water locations, especially at the southern region of the study area. The distribution of predicted seagrass habitats corresponds to the distribution of ground-truth data. (i.e., seagrass presence data) and earlier study [15]. This varied geographical distribution pattern and the MBES factors implicated have significant implications for conservation and preservation efforts in the study area.

5. Conclusion

The current study demonstrated the practicality and utility of combining MBES predictors with seagrass occurrence data and HSM approaches for seagrass habitat modelling at RMP. Due to the utilisation of data with a high spatial resolution, a thorough and comprehensive SHSM was created to determine the most appropriate seagrass habitat. With varied spatial resolution data, the dataset with a high resolution provided a more consistent spatial pattern of SHSMs, compared to the dataset with a low resolution, despite identical accuracy. Additionally, this study indicated that regardless of the predictor importance analysis, the bathymetric map is the most important predictor for predicting seagrass compared to the other MBES predictors. By comparison, the various machine learning algorithms applied in this study demonstrated the importance of selecting appropriate algorithms for producing high-quality SHSM.

References

[1] Waycott M, Duarte C M, Carruthers T J, Orth R J, Dennison W C, Olyarnik S, Calladine A, Fourqurean J W, Heck K L and Hughes A R 2009 Accelerating loss of seagrasses across the globe threatens coastal ecosystems *Proceedings of the national academy of sciences* **106** 12377-81

[2] Short F T, Polidoro B, Livingstone S R, Carpenter K E, Bandeira S, Bujang J S, Calumpong H P, Carruthers T J, Coles R G and Dennison W C 2011 Extinction risk assessment of the world's seagrass species *Biological Conservation* **144** 1961-71

[3] McKenzie L J, Nordlund L M, Jones B L, Cullen-Unsworth L C, Roelfsema C and Unsworth R K F 2020 The global distribution of seagrass meadows *Environmental Research Letters* **15** 074041

[4] Hossain M S, Bujang J S, Zakaria M H and Hashim M 2016 Marine and human habitat mapping for the Coral Triangle Initiative region of Sabah using Landsat and Google Earth imagery *Marine Policy* **72** 176-91

[5] Bujang J S, Zakaria M H and Short F 2016 pp 1-9

[6] Zakaria M and Bujang J 2011 Disturbances in seagrass ecosystem in Malaysia *Seagrasses: Resource Status and Trends in Indonesia, Japan, Malaysia, Thailand and Vietnam* 67-78

[7] Hossain M, Bujang J, Zakaria M and Hashim M 2015 Application of Landsat images to seagrass areal cover change analysis for Lawas, Terengganu and Kelantan of Malaysia *Continental Shelf Research* **110** 124-48

[8] Hossain M S, Hashim M, Bujang J S, Zakaria M H and Muslim A M 2019 Assessment of the impact of coastal reclamation activities on seagrass meadows in Sungai Pulai estuary, Malaysia, using Landsat data (1994–2017) *International Journal of Remote Sensing* **40** 3571-605

[9] Hossain M S and Hashim M 2019 Potential of Earth Observation (EO) technologies for seagrass ecosystem service assessments *International Journal of Applied Earth Observation and Geoinformation* **77** 15-29

[10] Yang B, Hawthorne T L, Hessing-Lewis M, Duffy E J, Reshitnyk L Y, Feinman M and Searson H 2020 Developing an Introductory UAV/Drone Mapping Training Program for Seagrass Monitoring and Research *Drones* **4** 70

[11] Mohamed H, Nadaoka K and Nakamura T 2020 Semiautomated Mapping of Benthic Habitats and Seagrass Species Using a Convolutional Neural Network Framework in Shallow Water Environments *Remote Sensing* **12** 4002

[12] Muhamad M A H, Che Hasan R, Md Said N and Ooi J L-S 2021 Seagrass habitat suitability model for Redang Marine Park using multibeam echosounder data: Testing different spatial resolutions and analysis window sizes *PLOS ONE* **16** e0257761

[13] Mustajap F, Saleh E, Madin J and Hamid S A 2015 Marine habitat mapping at Labuan Marine Park, Federal Territory of Labuan, Malaysia *Ocean Science Journal* **50** 291-8

[14] Rohani S, Mohd Jani J, Chan A, Ismail M, Bahrin B and Jamaludin S 2019 Pulau Sibu Scientific Expedition: Connecting the Land and The Sea For Biodiversity Management Of A Marine Park Island *Malayan Nature Journal* **7** 277-84
[15] Zakaria M H, Bujang J S and Razak F 2003 Occurrence and Morphological Description of Seagrasses from Pulau Redang, Terengganu, Malaysia *Jurnal Teknologi* **38**
[16] Wabnitz C C, Andréfouët S, Torres-Pulliza D, Müller-Karger F E and Kramer P A 2008 Regional-scale seagrass habitat mapping in the Wider Caribbean region using Landsat sensors: Applications to conservation and ecology *Remote Sensing of Environment* **112** 3455-67
[17] Hu W, Zhang D, Chen B, Liu X, Ye X, Jiang Q, Zheng X, Du J and Chen S 2021 Mapping the seagrass conservation and restoration priorities: Coupling habitat suitability and anthropogenic pressures *Ecological Indicators* **129** 107960
[18] Andréfouet S, Derville S, Buttin J, Dirberg G, Wabnitz C C C, Garrigue C and Payri C E 2021 Nation-wide hierarchical and spatially-explicit framework to characterize seagrass meadows in New-Caledonia, and its potential application to the Indo-Pacific *Marine Pollution Bulletin* **173** 113036
[19] Bittner R, Roesler E and Barnes M 2020 Using species distribution models to guide seagrass management *Estuarine, Coastal and Shelf Science* **240** 106790
[20] Jayathilake D R M and Costello M J 2018 A modelled global distribution of the seagrass biome *Biological Conservation* **226** 120-6
[21] Downie A-L, von Numers M and Boström C 2013 Influence of model selection on the predicted distribution of the seagrass Zostera marina *Estuarine, Coastal and Shelf Science* **121-122** 8-19
[22] Beca-Carretero P, Varela S and Stengel D B 2020 A novel method combining species distribution models, remote sensing, and field surveys for detecting and mapping subtidal seagrass meadows *Aquatic Conservation: Marine and Freshwater Ecosystems* **30** 1098-110
[23] Hossain M S, Bujang J S, Zakaria M H and Hashim M 2015 The application of remote sensing to seagrass ecosystems: an overview and future research prospects *International Journal of Remote Sensing* **36** 61-114
[24] Downie A-L, Piechaud N, Howell K, Barrio Froján C, Sacau M and Kenny A 2021 Reconstructing baselines: use of habitat suitability modelling to predict pre-fishing condition of a Vulnerable Marine Ecosystem *ICES Journal of Marine Science* **78** 2784-96
[25] Anderson O F, Guinotte J M, Rowden A A, Tracey D M, Mackay K A and Clark M R 2016 Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand *Deep Sea Research Part I: Oceanographic Research Papers* **115** 265-92
[26] Rengstorf A M, Mohn C, Brown C, Wisz M S and Grehan A J 2014 Predicting the distribution of deep-sea vulnerable marine ecosystems using high-resolution data: Considerations and novel approaches *Deep Sea Research Part I: Oceanographic Research Papers* **93** 72-82
[27] Lindegarth M, Bergström U, Mattila J, Olenin S, Ollikainen M, Downie A-L, Sundblad G, Bučas M, Gullström M, Snickars M, von Numers M, Svensson J R and Kosenius A-K 2014 Testing the potential for predictive modeling and mapping and extending its use as a tool for evaluating management scenarios and economic valuation in the Baltic Sea (PREHAB) *Ambio* **43** 82-93
[28] Bowden D A, Anderson O F, Rowden A A, Stephenson F and Clark M R 2021 Assessing Habitat Suitability Models for the Deep Sea: Is Our Ability to Predict the Distributions of Seafloor Fauna Improving? *Frontiers in Marine Science* **8**
[29] Rowden A A, Anderson O F, Georgian S E, Bowden D A, Clark M R, Pallentin A and Miller A 2017 High-Resolution Habitat Suitability Models for the Conservation and Management of Vulnerable Marine Ecosystems on the Louisville Seamount Chain, South Pacific Ocean *Frontiers in Marine Science* **4**
[30] Anderson O F, Guinotte J M, Rowden A A, Clark M R, Mormede S, Davies A J and Bowden D A 2016 Field validation of habitat suitability models for vulnerable marine ecosystems in the

South Pacific Ocean: Implications for the use of broad-scale models in fisheries management *Ocean & Coastal Management* **120** 110-26
[31] Elith J and Leathwick J R 2009 Species Distribution Models: Ecological Explanation and Prediction Across Space and Time *Annual Review of Ecology, Evolution, and Systematics* **40** 677-97
[32] Benito B M, Cayuela L and Albuquerque F S 2013 The impact of modelling choices in the predictive performance of richness maps derived from species-distribution models: guidelines to build better diversity models *Methods in Ecology and Evolution* **4** 327-35
[33] Früh L, Kampen H, Kerkow A, Schaub G A, Walther D and Wieland R 2018 Modelling the potential distribution of an invasive mosquito species: comparative evaluation of four machine learning methods and their combinations *Ecological Modelling* **388** 136-44
[34] Hu Z, Hu J, Hu H and Zhou Y 2020 Predictive habitat suitability modeling of deep-sea framework-forming scleractinian corals in the Gulf of Mexico *Science of The Total Environment* **742** 140562
[35] da Silveira C B L, Strenzel G M R, Maida M, Gaspar A L B and Ferreira B P 2021 Coral Reef Mapping with Remote Sensing and Machine Learning: A Nurture and Nature Analysis in Marine Protected Areas *Remote Sensing* **13** 2907
[36] Rahimian Boogar A, Salehi H, Pourghasemi H R and Blaschke T 2019 Predicting Habitat Suitability and Conserving Juniperus spp. Habitat Using SVM and Maximum Entropy Machine Learning Techniques *Water* **11** 2049
[37] Georgian S E, Anderson O F and Rowden A A 2019 Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean *Fisheries Research* **211** 256-74
[38] Santos Rolando O and Lirman D 2012 Using habitat suitability models to predict changes in seagrass distribution caused by water management practices1 *Canadian Journal of Fisheries and Aquatic Sciences* **69** 1380-8
[39] Heide T, Peeters E T H M, Hermus D C R, Katwijk M M v, Roelofs J G M and Smolders A J P 2009 Predicting habitat suitability in temperate seagrass ecosystems *Limnology and Oceanography* **54** 2018-24
[40] Bittner R E, Roesler E L and Barnes M A 2020 Using species distribution models to guide seagrass management *Estuarine, Coastal and Shelf Science* **240** 106790
[41] Metz M, Rocchini D and Neteler M 2014 Surface Temperatures at the Continental Scale: Tracking Changes with Remote Sensing at Unprecedented Detail *Remote Sensing* **6** 3822-40
[42] Cisneros-Araujo P, Goicolea T, Mateo-Sánchez M C, García-Viñás J I, Marchamalo M, Mercier A and Gastón A 2021 The Role of Remote Sensing Data in Habitat Suitability and Connectivity Modeling: Insights from the Cantabrian Brown Bear *Remote Sensing* **13** 1138
[43] Miyamoto M, Kiyota M, Murase H, Nakamura T and Hayashibara T 2017 Effects of Bathymetric Grid-Cell Sizes on Habitat Suitability Analysis of Cold-water Gorgonian Corals on Seamounts *Marine Geodesy* **40** 205-23
[44] Sun X, Long Z and Jia J 2021 A multi-scale Maxent approach to model habitat suitability for the giant pandas in the Qionglai mountain, China *Global Ecology and Conservation* **30** e01766
[45] Gumusay M U, Bakirman T, Tuney Kizilkaya I and Aykut N O 2019 A review of seagrass detection, mapping and monitoring applications using acoustic systems *European Journal of Remote Sensing* **52** 1-29
[46] Zakariya R, Abdullah M A, Hasan R C and Khalil I 2018 *Engineering Applications for New Materials and Technologies,* ed A Öchsner (Cham: Springer International Publishing) pp 579-91
[47] Koop L, Snellen M and Simons D G 2021 An Object-Based Image Analysis Approach Using Bathymetry and Bathymetric Derivatives to Classify the Seafloor *Geosciences* **11** 45
[48] Brown C J, Beaudoin J, Brissette M and Gazzola V 2019 Multispectral multibeam echo sounder backscatter as a tool for improved seafloor characterization *Geosciences* **9** 126

[49] Gaida T C, Tengku Ali T A, Snellen M, Amiri-Simkooei A, Van Dijk T A and Simons D G 2018 A multispectral Bayesian classification method for increased acoustic discrimination of seabed sediments using multi-frequency multibeam backscatter data *Geosciences* **8** 455
[50] Rattray A, Ierodiaconou D and Womersley T 2015 Wave exposure as a predictor of benthic habitat distribution on high energy temperate reefs *Frontiers in Marine Science* **2**
[51] Wölfl A-C, Snaith H, Amirebrahimi S, Devey C W, Dorschel B, Ferrini V, Huvenne V A, Jakobsson M, Jencks J and Johnston G 2019 Seafloor mapping–the challenge of a truly global ocean bathymetry *Frontiers in Marine Science* **6** 283
[52] Ierodiaconou D, Schimel A C G, Kennedy D, Monk J, Gaylard G, Young M, Diesing M and Rattray A 2018 Combining pixel and object based image analysis of ultra-high resolution multibeam bathymetry and backscatter for habitat mapping in shallow marine waters *Marine Geophysical Research* **39** 271-88
[53] Masetti G and Calder B 2012 Remote identification of a shipwreck site from MBES backscatter *Journal of Environmental Management* **111** 44-52
[54] Zhao T, Montereale Gavazzi G, Lazendić S, Zhao Y and Pižurica A 2021 Acoustic Seafloor Classification Using the Weyl Transform of Multibeam Echosounder Backscatter Mosaic *Remote Sensing* **13** 1760
[55] Janowski L, Madricardo F, Fogarin S, Kruss A, Molinaroli E, Kubowicz-Grajewska A and Tegowski J 2020 Spatial and Temporal Changes of Tidal Inlet Using Object-Based Image Analysis of Multibeam Echosounder Measurements: A Case from the Lagoon of Venice, Italy *Remote Sensing* **12** 2117
[56] Parnum I M and Gavrilov A N 2011 High-frequency multibeam echo-sounder measurements of seafloor backscatter in shallow water: Part 2–Mosaic production, analysis and classification *Underwater Technology* **30** 13-26
[57] Wang M, Wu Z, Yang F, Ma Y, Wang X H and Zhao D 2018 Multifeature extraction and seafloor classification combining LiDAR and MBES data around Yuanzhi Island in the South China Sea *Sensors* **18** 3828
[58] Fonseca L, Brown C, Calder B, Mayer L and Rzhanov Y 2009 Angular range analysis of acoustic themes from Stanton Banks Ireland: A link between visual interpretation and multibeam echosounder angular signatures *Applied Acoustics* **70** 1298-304
[59] Santos R, Rodrigues A and Quartau R 2018 Acoustic remote characterization of seabed sediments using the Angular Range Analysis technique: The inlet channel of Tagus River estuary (Portugal) *Marine Geology* **400** 60-75
[60] Fakiris E, Dimas X, Georgiou N, Christodoulou D, Rzhanov Y and Papatheodorou G 2017 Integrating dual frequency side-scan sonar data and multibeam backscatter, angular response and bathymetry, for benthic habitat mapping in the Laganas Gulf MPA, Zakinthos Isl., Greece *The Journal of the Acoustical Society of America* **141** 3949-
[61] Muhamad M A H and Che Hasan R 2019 Seagrass Habitat Suitability Map at Merambong Shoal, Johor: A Preliminary Study Using Multibeam Echosounder and Maxent Modelling *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*
[62] Bakirman T and Gumusay M U 2020 A novel GIS-MCDA-based spatial habitat suitability model for Posidonia oceanica in the Mediterranean *Environmental monitoring and assessment* **192** 1-13
[63] Holon F, Marre G, Parravicini V, Mouquet N, Bockel T, Descamp P, Tribot A-S, Boissery P and Deter J 2018 A predictive model based on multiple coastal anthropogenic pressures explains the degradation status of a marine ecosystem: Implications for management and conservation *Biological Conservation* **222** 125-35
[64] Zucchetta M, Venier C, Taji M A, Mangin A and Pastres R 2016 Modelling the spatial distribution of the seagrass Posidonia oceanica along the North African coast: Implications for the assessment of Good Environmental Status *Ecological Indicators* **61** 1011-23

[65] Dartnell P and Gardner J V 2004 Predicting seafloor facies from multibeam bathymetry and backscatter data *Photogrammetric Engineering & Remote Sensing* **70** 1081-91
[66] Ahmed N, Atzberger C and Zewdie W 2021 Species Distribution Modelling performance and its implication for Sentinel-2-based prediction of invasive Prosopis juliflora in lower Awash River basin, Ethiopia *Ecological Processes* **10** 18
[67] Araújo M B and New M 2007 Ensemble forecasting of species distributions *Trends in Ecology & Evolution* **22** 42-7
[68] Effrosynidis D, Arampatzis A and Sylaios G 2018 Seagrass detection in the mediterranean: A supervised learning approach *Ecological Informatics* **48** 158-70
[69] Stankovic M, Kaewsrikhaw R, Rattanachot E and Prathep A 2019 Modeling of suitable habitat for small-scale seagrass restoration in tropical ecosystems *Estuarine, Coastal and Shelf Science* **231** 106465
[70] Valle M, van Katwijk M M, de Jong D J, Bouma T J, Schipper A M, Chust G, Benito B M, Garmendia J M and Borja Á 2013 Comparing the performance of species distribution models of Zostera marina: Implications for conservation *Journal of Sea Research* **83** 56-64
[71] Sumber B T L d P 2011 Laporan 2011. (Malaysia: Bahagian Taman Laut & Pengurusan Sumber)
[72] Zakaria M H, Bujang J S and Abdul Razak F R 2003 Occurrence and Morphological Description of Seagrasses from Pulau Redang, Terengganu, Malaysia *Jurnal Teknologi* **38**
[73] Lecours V 2017 *TASSE (Terrain Attribute Selection for Spatial Ecology) Toolbox v. 1.1*
[74] Ribó M, Macdonald H, Watson S J, Hillman J R, Strachan L J, Thrush S F, Mountjoy J J, Hadfield M G and Lamarche G 2021 Predicting habitat suitability of filter-feeder communities in a shallow marine environment, New Zealand *Marine Environmental Research* **163** 105218
[75] Dolan M 2012 Calculation of slope angle from bathymetry data using GIS - effects of computation algorithms, data resolution and analysis scale *NGU Report*
[76] Wilson M F, O'Connell B, Brown C, Guinan J C and Grehan A J 2007 Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope *Marine Geodesy* **30** 3-35
[77] Monk J, Versace V, Bellgrove A, Harvey E, Rattray A, Laurenson L and Quinn G 2010 Habitat suitability for marine fishes using presence-only modelling and multibeam sonar *Marine Ecology Progress Series* **420** 157-74
[78] Wang B, Xu Y and Ran J 2017 *Predicting suitable habitat of the Chinese monal (Lophophorus lhuysii) using ecological niche modeling in the Qionglai Mountains, China* vol 5
[79] Briscoe D K, Hiatt S, Lewison R and Hines E 2014 Modeling habitat and bycatch risk for dugongs in Sabah, Malaysia *Endangered Species Research* **24** 237-47
[80] Gaida T C, Snellen M, van Dijk T A G P and Simons D G 2019 Geostatistical modelling of multibeam backscatter for full-coverage seabed sediment maps *Hydrobiologia* **845** 55-79
[81] Trzcinska K, Janowski L, Nowak J, Rucinska-Zjadacz M, Kruss A, von Deimling J S, Pocwiardowski P and Tegowski J 2020 Spectral features of dual-frequency multibeam echosounder data for benthic habitat mapping *Marine Geology* **427** 106239
[82] Zuur A, Ieno E N, Walker N, Saveliev A A and Smith G M 2009 *Mixed effects models and extensions in ecology with R*: Springer Science & Business Media)
[83] Wei T, Simko V, Levy M, Xie Y, Jin Y and Zemla J 2017 Package 'corrplot' *Statistician* **56** e24
[84] Phillips S J, Anderson R P and Schapire R E 2006 Maximum entropy modeling of species geographic distributions *Ecological Modelling* **190** 231-59
[85] Wang B, Xu Y and Ran J 2017 Predicting suitable habitat of the Chinese monal (Lophophorus lhuysii) using ecological niche modeling in the Qionglai Mountains, China *PeerJ* **5** e3477
[86] Phillips S J and Dudík M 2008 Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation *Ecography* **31** 161-75
[87] Blondel P and Gómez Sichi O 2009 Textural analyses of multibeam sonar imagery from Stanton Banks, Northern Ireland continental shelf *Applied Acoustics* **70** 1288-97

[88] Fielding A H and Bell J F 1997 A review of methods for the assessment of prediction errors in conservation presence/absence models *Environmental conservation* **24** 38-49
[89] Swets J A 1988 Measuring the accuracy of diagnostic systems *Science* **240** 1285-93
[90] Hosmer Jr D W, Lemeshow S and Sturdivant R X 2013 *Applied logistic regression* vol 398: John Wiley & Sons)
[91] Reiss H, Cunze S, König K, Neumann H and Kröncke I 2011 Species distribution modelling of marine benthos: a North Sea case study *Marine Ecology Progress Series* **442** 71-86
[92] Naimi B and Araújo M B 2016 sdm: a reproducible and extensible R platform for species distribution modelling *Ecography* **39** 368-75
[93] Li Y, Shao W and Jiang J 2021 Predicting the potential global distribution of Sapindus mukorossi under climate change based on MaxEnt modelling *Environmental Science and Pollution Research*
[94] Rahimian Boogar A, Salehi H, Pourghasemi H and Blaschke T 2019 Predicting Habitat Suitability and Conserving Juniperus spp. Habitat Using SVM and Maximum Entropy Machine Learning Techniques *Water* **11**