Benha Journal of Applied Sciences (BJAS) **print: ISSN 2356–9751**
Vol. (9) Issue (5) (2024), (49-56) **online: ISSN 2356–976x**
http://bjas.bu.edu.eg

Exploring Best Practices in Machine Learning Approaches for near-shore bathymetry modeling: Insights from the Egyptian Mediterranean Coast

Nasef M. Aly[1,2], Ramadan Kh. Abel-maguid[1] and Aly M. Elnaggar[1]

[1]Transportation Department, Faculty of Engineering, Alexandria University, Egypt.
[2]M.Sc. Candidate,
E-Mail: NasefM.Aly@alexu.edu.eg

Abstract:

Many coastal areas, especially in developing countries or those with limited marine activity, lack detailed depth measurements. Past data in these areas may be incomplete or outdated, making it difficult to create accurate seafloor maps. This is important for the preliminary design of coastal structures. This study aims to explore the best way to use satellite images and open-source software to create Satellite-Derived Bathymetry (SDB) models. Our approach uses three machine learning algorithms (KNN, RF, MLR) to analyze satellite images of different areas. The images come from open-source databases. We use the closest truth data to the targeted area to train the algorithms to predict the unseen data. Our research shows that using satellite data to measure water depth can accurately determine depths of up to 27 meters. Furthermore, our assessment reveals mean absolute errors averaging 0.72 meters and root mean square errors averaging 1.0 meter, with accuracies around 94.6% for both samples. Random Forest (RF) performed better than KNN and MLR. In Sample El-Dabaa, RF performed well with Landsat-08 Single Scene image. The area has rocky cliffs with seagrass, steep slopes, and strong wave movement. In Sample El-Arish, RF's best results came with single scene image from Landsat-08. This area has sandy soil, gentle slopes, and gentle wave movement. Generally, usage of Single Scene image or Median Value image with ML algorithm depends on the seafloor dynamics.

Keywords: Satellite Derived Bathymetry, Nearshore, Machine learning, Mediterranean Sea, Open-source

1. Introduction:

Several coastal zones, particularly those in developing nations or with minimal marine traffic, are deficient in precise depth surveys. The deficiency in previous data within these regions might be insufficient or antiquated, as evidenced by online Admiralty charts, particularly in certain sections of the map close to the shoreline in shallow water. Posing challenges in generating precise maps of the seabed. In the initial design of buildings such as breakwaters, piers, and harbours, it is very important to have accurate depth measurements. These depths provide information on the sediment distribution, underwater topography and probable threats that may occur, which allows the designing of resilient structures to coastal dynamics, wave action, and sediment transport processes by the construction engineer. Moreover, knowledge concerning depths spatial distribution is required for evaluating stability and durability of structures locating along the coast against such influences as scouring, sedimentation, and sea level rise. Consequently, through integrating the correct data on how deep different areas are, constructors provide that structures on shores serve their purposes within a long time in safe conditions leading to endurance of coasts and continuity of growth that does not harm the environment. Traditional approaches towards the estimation of shallow water depth have been limited by several compromises and constraints. There exist restrained geological spreads for conventional approaches like ship-based surveys and aerial Laser Imaging, Detection, and Ranging (LiDAR) [10]. The limitation of conventional methods is that they are both slow and labour intensive hence; they cannot be applied globally [8,11]. Consequently, new methods have evolved that use satellite data to generate bathymetric charts [12,13]. Land-viewing sensors might not always have excellent spatial resolution when it comes to oceanic pixels located near shorelines whereas land-viewing sensors may not have great rates in terms of time leading to fewer number of satellite images which are cloud-free although this may particularly occur more often in areas such as rainy coastal tropics where cloudiness is frequent and high [7].

Moreover, bathymetric algorithms based on satellite telemetry, often draw on field data calibration [8,9,11,16,17] or a large number of calculations aimed at physically modeling general properties of the water column [10,12,13,14,15,18,19], and are frequently restricted to a single scene portrayed in either multispectral or hyperspectral images taken by space orbiters. Estimation of shallow water bathymetry confronts multiple challenges which are meticulously

taken care of through implementation of Machine Learning Algorithms and Open-Source Datasets. There is one of them - Google Earth Engine (GEE), conceived as a powerful cloud-based computational platform that allows easy access to high-coverage

global-scale analysis-ready satellite reflectance datasets. GEE has facilitated many global-scale products such as land cover, forest change, water surface extent and urban land use [20,21]. Specifically, Landsat-08 surface reflectance dataset in GEE is heavily employed in terrestrial environment studies[6]. Despite this fact, it is not entirely clear whether the Landsat-08 dataset in GEE is suitable for direct performance of shallow water bathymetry. Moreover, Machine learning techniques are very good at discovering complex patterns within high dimensional datasets, by utilizing advanced mathematical approaches to detect the correlations which could be difficult for man to notice. By employing sophisticated models like Random Forest algorithm as well as ensemble techniques, they are

able to grasp non-linear relationships between different attributes improving on their ability to make precise predictions. In this research paper we have come up with a fresh technique of shallow water depth mapping using Google Earth Engine, Machine Learning Algorithms within the coastlines' marine habitats based on field data close to the point of observation. This study tries to create a useful and reliable way of mapping bathymetry that takes into account seabed features.

2. Material and Methods:

2.1 Study Sites and Data:

We tested our bathymetric mapping method across the Egyptian Mediterranean Shoreline and choose two samples that represent the most geomorphology (Sandy bed, Rocky cliffs bed) of the shoreline regarding a Paper was published [4]. in addition to the slope of the seabed and the waves movement in these areas. One of the Samples is EL-Dabaa Area. The Seabed features in this area are about Rocky cliffs, Steep Slope, High Waves movements and partially distributed seagrass into the sample called *Posidonia oceanica*[5]. On the other hand, there is another sample in El-Arish. The seabed features in EL-Arish are about sandy soil, gentle slope and gentle wave movements.

Our Ground Truth Data that we use to validate the Method is bathymetry data collected and referred to the Chart Datum called Lowest- Astronomic Tide (LAT). These seabed depth maps are often used in coastal design and ship-navigation. We have around 247,256 depth points in both samples.

2.2 Satellite Imagery:

Open-source satellite imagery, such as Landsat-8, is commonly utilized in remote sensing analyses, but requires correction before use. These corrections include geometric, atmospheric and radiometric corrections, with an additional option for cloud cover adjustment regarding specific applications[3]. Fortunately, Google Earth Engine (GEE) platform offers geometric, radiometric corrected Landsat-8 images ("LANDSAT/LC08/C01/T1-SR") with customizable cloud cover selection[1]. Utilizing GEE, users can download single scene images with cloud cover less than 10% closely aligned with observed in-situ data, as well as median value images over a three-month period encompassing the in-situ observation time.

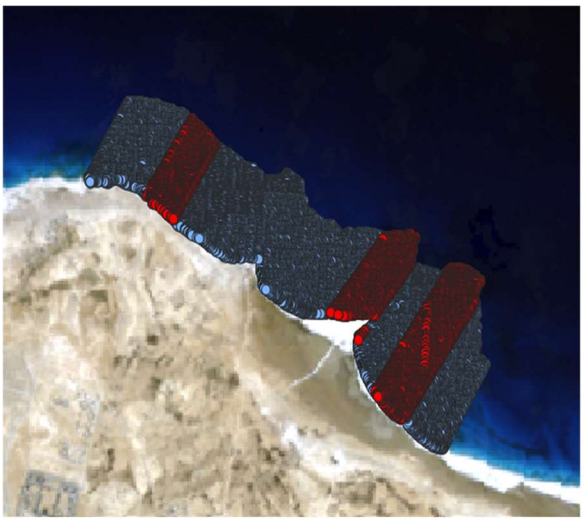


Fig (1) El-Dabaa Area with In-situ data (red dots represent the ML training points)

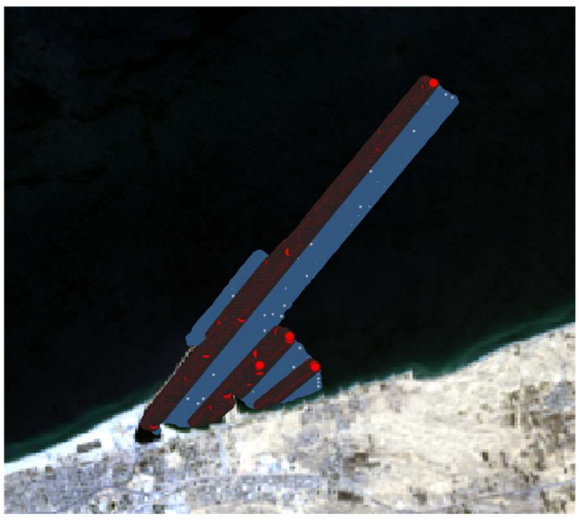


Fig (2) El-Arish Area with In-situ data (red dots represent the ML training points)

Table (1) overview of In-situ Data Features

In-situ Data	El-Arish	El-Dabaa
Observation time	23-08-2021	26-04-2019
Depth Points	168,385	77,922
Depth Range	15 m	27 m
Water Chart Datum	LAT	LAT
Grid	5x5 m	10x10 m
Datum	WGS-84	WGS-84
Projection	UTM Zone 36N	UTM Zone 35N
Sea Albedo	Sand	Rocky cliffs with seagrass

2.3 Machine Learning Algorithms and Bathymetry Estimation Method:

Machine learning (ML) is increasingly recognized as a valuable research tool for practitioners in GIS and remote sensing, providing enhanced flexibility in processing vast datasets. ML techniques have garnered widespread adoption within the remote sensing community, particularly in the processing of longitudinal high-resolution satellite data and/or datasets accompanied by high-resolution in-situ observations. The few of the studies on satellite-derived bathymetry using ML algorithms have been shown in article review with details of algorithm and accuracy achieved [3]. We use three of the most often machine learning algorithms to detect the depth. They are Random Forest (RF), K-Nearest Neighbourhood (KNN), and Multiple Linear Regression (MLR). Random Forest (RF) is a versatile and widely used machine learning algorithm known for its ability to handle complex datasets and provide robust predictions by aggregating multiple decision trees. Multiple Linear Regression (MLR) is a classic statistical technique used to model the relationship between a dependent variable and multiple independent variables, providing insights into the linear associations between them. K-Nearest Neighbors (KNN) is a simple yet powerful machine learning algorithm that classifies data points based on the majority vote of their nearest neighbors, making it particularly effective for classification tasks in both supervised and unsupervised learning scenarios.

Our proposed methodology for bathymetry estimation comprises two stages. Initially, we identify the most suitable machine learning (ML) models that demonstrates reliable performance. Subsequently, we partition the dataset into two subsets: 30% for training the ML model and generating a predicted bathymetry map for the designated area, while the remaining 70% of the dataset serves as unseen data for the model evaluation. The training ML and testing process is made with help of open-source Gui available online on Git-hub platform that use mixture of open-source python libraries [23].

Table (2) shows the characteristics of satellites images and depth points used in training and test

Samples		El-Arish		El-Dabaa	
Satellite Images		Single Scene	Median Value	Single Scene	Median Value
Landsat-08 Satellite Imagery	**Acquisition Date**	01-07-2021	21-06-2021 to 21-09-2021	28-04-2019	21-03-2019 to 21-06-2019
	Cloud Cover	Less 10%	Cloud-free	Less 10%	Cloud-free
	Spatial Resolution	30 m	30 m	30 m	30 m
	Projection	UTM Zone 36N	UTM Zone 36N	UTM Zone 35N	UTM Zone 35N
	Datum	WGS-84	WGS-84	WGS-84	WGS-84
ML Validation (Stage Two)	**Training depth Points**	21,690	21,690	50,966	50,966
	Unseen Depth Points	56,232	56,232	117,419	117,419

Table (3): shows the ML equations and parameters that have been used

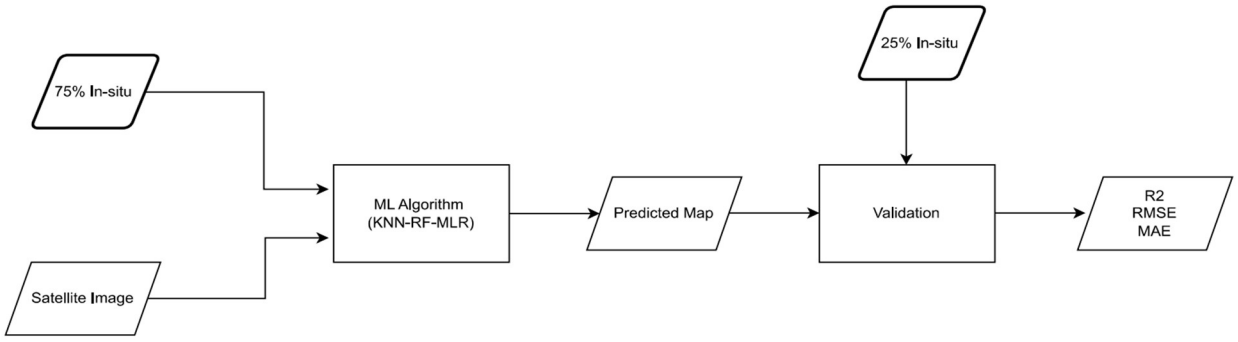
ML Algorithms	Equation	Parameters
Random Forest (RF)	$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i$	Number of Trees (300), Criterion (MSE), Bootstrap (True), Random state (0)
K-Nearest Neigh-bourhood (KNN)	$\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i$	Number of neighbours (5), Weights (Distance), Leaf Size (30)
Multiple Linear Regression (MLR)	$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$	Fit Intercept (True), Normalize (False), Copy X (True)

Table (4) Overview of the statistics accuracy metrics

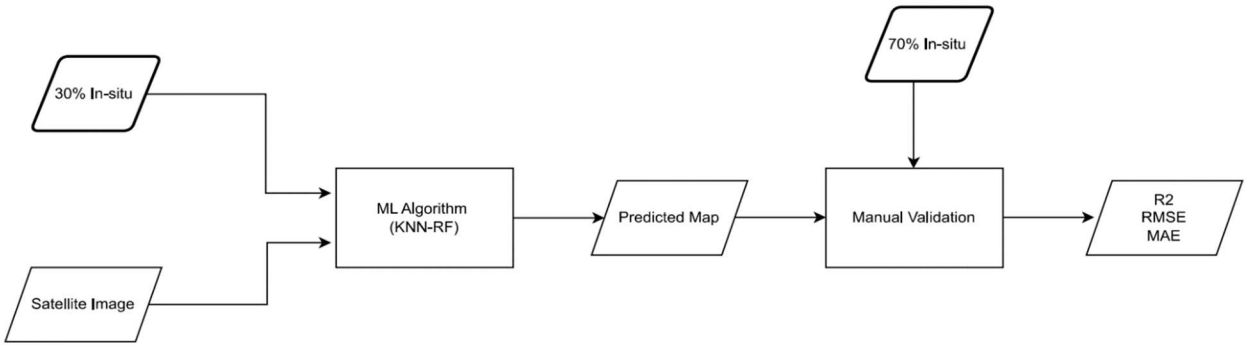
Accuracy Metric	Mathematical Formula
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$
R2	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

Where:

- n is the number of observations,
- yi is the actual value of the dependent variable for observation i,
- $y\hat{}i$ is the predicted value of the dependent variable for observation i,
- $\bar{y}$ is the mean of the observed values of the dependent variable.



Fig(3): Flow Chart displays the Stage One Process



Fig(4): Flow Chart displays the Stage Two Process

Now, we have generated a predicted bathymetry map for the unseen data, it is imperative to assess the accuracy of these maps to evaluate their reliability. This involves extracting the predicted depth points from the generated map and comparing them with the corresponding ground truth data in the 70% unseen dataset through manual validation. we used statistical equations that are often used [3]. We use Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and R-Squared (R2). RMSE measures the average deviation between predicted and actual values, giving more weight to larger errors due to squaring. It's important because it provides a measure of the model's accuracy in predicting numerical outcomes, with lower RMSE indicating better performance.

In addition, MAE calculates the average absolute difference between predicted and actual values, providing a direct measure of prediction error. It's important as it offers a straightforward assessment of model performance, unaffected by the direction of errors, making it easy to interpret. R2 quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables in a regression model. It's important because it indicates how well the independent variables explain the variability of the dependent variable, with higher values suggesting a better fit of the model to the data.

3. Results:

3.1 (Stage One) Validation of the proposed ML models:

After training the selected ML models on all In-situ data and all satellite images for both samples, we found that for the El-Arish area, KNN and Random Forest (RF) models exhibit high R2 values close to 0.996, in addition to lower RMSE and MAE indicating strong predictive performance, while the Multiple Linear Regression (MLR) model shows lower R2 values and high RMSE and MAE Similarly, for the El-Dabaa area, KNN and RF models demonstrate excellent performance with R2 values around 0.996, whereas MLR exhibits slightly lower R2 values and too higher values of RMSE and MAE. So, we decide to use KNN and RF models for further evaluation on Unseen Data.

3.2 (Stage Two) Validation of the best-fit models on Prediction Unseen Data:

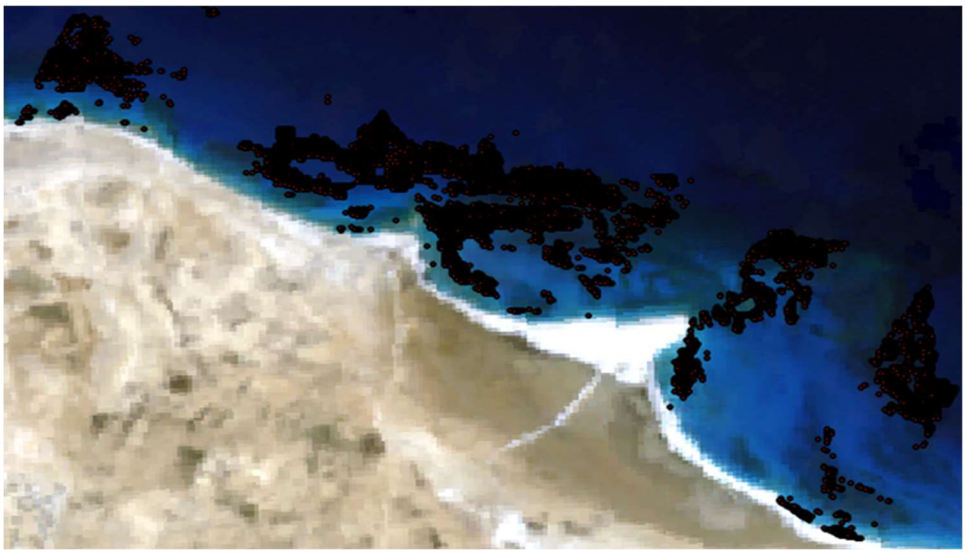
In our methodology, 30% of the in-situ data was partitioned for model training, whereas 70% was set aside for manual validation of model predictions on unseen data. As shown in Table (5), the Random Forest (RF) model consistently performed well across different samples and conditions. Notably, for the El-Arish Area and El-Dabaa, utilizing the Single Scene image yielded superior results compared to the Median Value image, while for the El-Dabaa Area, the Median Value image exhibited excellent performance solely with the both models rather than the median value image of El-Arish. Therefore, for broader application in SDB (Shallow Depth Bathymetry) tasks, the RF Algorithm is recommended as the best-fit ML model. Additionally, regarding the selection of satellite imagery, it is advisable to experiment with both Single Scene and Median Value images in your specific application and opt for the most effective option. Exploring the discrepancies observed in the El-Arish Area and EL-Dabaa Area, characterized by significant sedimentation and seagrass movements, revealed that the Single Scene image outperformed the Median Value image. Conversely, in the El-Dabaa Area, spatial analysis of absolute value of predicted depth errors that are greater than RMSE, particularly significant around seagrass areas, suggests a potential need for classifying sea albedo into seagrass and rocky cliffs. This could involve training ML models separately on each class or employing more advanced ML techniques such as Artificial Neural Networks [3].

Table (5) Stage One: displays the validation parameters used during the training of machine learning models to select the Best-Fit Models for testing on Unseen Data

Samples	ML Models	Landsat-08					
		Single Scene			Median Value		
		R2	RMSE	MAE	R2	RMSE	MAE
El-Arish	**KNN**	0.995	0.188	0.116	0.996	0.185	0.115
	RF	0.996	0.175	0.109	0.996	0.173	0.108
	MLR	0.817	1.188	0.937	0.870	1.000	0.746
El-Dabaa	**KNN**	0.996	0.427	0.300	0.996	0.434	0.302
	RF	0.996	0.425	0.298	0.996	0.421	0.297
	MLR	0.759	3.231	2.677	0.844	2.600	2.154

Table (6) Stage Two: presents the validation metrics for the best-performing machine learning models on unseen test data.

Samples	Best-fit ML models	Landsat-08					
		Single Scene			Median Value		
		R2	RMSE	MAE	R2	RMSE	MAE
El-Arish	KNN	0.919	0.771	0.595	0.870	1.080	0.790
	RF	0.924	0.779	0.580	0.901	0.931	0.700
El-Dabaa	KNN	0.964	1.337	0.924	0.947	1.595	1.081
	RF	0.967	1.229	0.850	0.946	1.427	0.861



Fig(5) Showing El-Dabaa Area with points that have error value greater than RMSE

4. Discussion:

Our investigation uncovered notable disparities in model effectiveness. In the El-Arish Area, employing Single Scene imagery outperformed the use of Median Value imagery, possibly due to substantial sedimentation in the region. Similarly, in El-Dabaa Area, which exhibited characteristics similar to El-Arish Area owing to seagrass dynamics, the Single Scene image provided a more accurate portrayal of the evolving seabed landscape.

The El-Dabaa image demonstrated outstanding performance, Residual Analysis of predicted depth errors highlighted significant discrepancies, particularly around seagrass habitats, suggesting the potential for refining model training by classifying sea albedo into distinct categories such as seagrass and rocky cliffs. This finer classification could enhance model accuracy and better capture underwater terrain complexities.

To address these challenges, future research endeavors could explore advanced machine learning techniques like Artificial Neural Networks (ANN) to enhance bathymetric model predictive capabilities. Furthermore, ongoing experimentation with various satellite imagery types, including both Single Scene and Median Value images, will be pivotal for optimizing model performance across diverse underwater landscapes. Overall, our findings underscore the importance of considering site-specific characteristics and selecting appropriate modeling approaches to effectively address the complexities of bathymetric estimation in shallow depth areas.

5. Conclusion:

Overall, this research has shed light on the SDB and the usage of ML Algorithms with in-situ Calibration. We found that:

1- Random Forest Algorithms can rely on in most SDB conditions.
2- Usage of Single Scene image or Median Value image with ML algorithm depends on the seafloor, so you may test both for the area of interest then choose the best-fit image.
3- Variance in sea albedo requires deep learning algorithms like (ANN) or separate the image to different category then predict each category separately .

6. References:

[1] Li, J., Knapp, D. E., Lyons, M., Roelfsema, C., Phinn, S., Schill, S. R., & Asner, G. P. (2021). Automated global shallowwater bathymetry mapping using google earth engine. Remote Sensing, 13(8). https://doi.org/10.3390/rs13081469

[2] Alevizos, E. (2020). A combined machine learning and residual analysis approach for improved retrieval of shallow bathymetry from hyperspectral imagery and sparse ground truth data. Remote Sensing, 12(21), 1–16. https://doi.org/10.3390/rs12213489

[3] Ashphaq, M., Srivastava, P. K., & Mitra, D. (2021). Review of near-shore satellite derived bathymetry: Classification and account of five decades of coastal bathymetry research. In Journal of Ocean Engineering and Science (Vol. 6, Issue 4, pp. 340–359). Shanghai Jiaotong University. https://doi.org/10.1016/j.joes.2021.02.006

[4] Hereher, M. E. (2015). Coastal vulnerability assessment for Egypt's Mediterranean coast. Geomatics, Natural Hazards and Risk, 6(4), 342–355. https://doi.org/10.1080/19475705.2013.845115

[5] Telesca, L., Belluscio, A., Criscoli, A., Ardizzone, G., Apostolaki, E. T., Fraschetti, S., Gristina, M., Knittweis, L., Martin, C. S., Pergent, G., Alagna, A., Badalamenti, F., Garofalo, G., Gerakaris, V., Louise Pace, M., Pergent-Martini, C., & Salomidi, M. (2015). Seagrass meadows (Posidonia oceanica) distribution and trajectories of change. Scientific Reports, 5. https://doi.org/10.1038/srep12505

[6] Waleed, M., & Sajjad, M. (2023). On the geospatial cloud-based platforms for disaster risk management: A global scientometric review of google earth engine applications. International Journal of Disaster Risk Reduction, 97, 104056. https://doi.org/10.1016/J.IJDRR.2023.104056

[7] Purkis, S.J. Remote Sensing Tropical Coral Reefs: The View from Above. Annu. Rev. Mar. Sci. 2018, 10, 149–168.

[8] Stumpf, R.P.; Holderied, K.; Sinclair, M. Determination of Water Depth with High-resolution Satellite Imagery over Variable Bottom Types. Limnol. Oceanogr. 2003, 48, 547–556.

[9] Lyzenga, D.R.; Malinas, N.P.; Tanis, F.J. Multispectral Bathymetry Using a Simple Physically Based Algorithm. IEEE Trans. Geosci. Remote Sens. 2006, 44, 2251–2259.

[10] Zhao, J.; Barnes, B.; Melo, N.; English, D.; Lapointe, B.; Muller-Karger, F.; Schaeffer, B.; Hu, C. Assessment of Satellite-Derived Diffuse Attenuation Coefficients and Euphotic Depths in South Florida Coastal Waters. Remote Sens. Environ. 2013, 131, 38–50.

[11] Traganos, D.; Poursanidis, D.; Aggarwal, B.; Chrysoulakis, N.; Reinartz, P. Estimating Satellite-Derived Bathymetry (SDB) with the Google Earth Engine and Sentinel-2. Remote Sens. 2018, 10, 859.

[12] Brando, V.E.; Anstee, J.M.; Wettle, M.; Dekker, A.G.; Phinn, S.R.; Roelfsema, C. A Physics Based Retrieval and Quality Assessment of Bathymetry from Suboptimal Hyperspectral Data. Remote Sens. Environ. 2009, 113, 755–770.

[13] Wei, J.; Wang, M.; Lee, Z.; Briceño, H.O.; Yu, X.; Jiang, L.; Garcia, R.; Wang, J.; Luis, K. Shallow Water Bathymetry with Multi-Spectral Satellite Ocean Color Sensors: Leveraging Temporal Variation in Image Data. Remote Sens. Environ. 2020, 250, 112035.

[14] Dekker, A.G.; Phinn, S.R.; Anstee, J.; Bissett, P.; Brando, V.E.; Casey, B.; Fearns, P.; Hedley, J.; Klonowski, W.; Lee, Z.P. Intercomparison of Shallow Water Bathymetry, Hydro-optics, and Benthos Mapping Techniques in Australian and Caribbean Coastal Environments. Limnol. Oceanogr. Methods 2011, 9, 396–425.

[15] Klonowski, W.M.; Fearns, P.R.; Lynch, M.J. Retrieving Key Benthic Cover Types and Bathymetry from Hyperspectral Imagery. J.Appl. Remote Sens. 2007, 1, 011505.

[16] Albright, A.; Glennie, C. Nearshore Bathymetry From Fusion of Sentinel-2 and ICESat-2 Observations. IEEE Geosci. Remote Sens.Lett. 2020.

[17] Parrish, C.E.; Magruder, L.A.; Neuenschwander, A.L.; Forfinski-Sarkozi, N.; Alonzo, M.; Jasinski, M. Validation of ICESat-2 ATLAS

Bathymetry and Analysis of ATLAS's Bathymetric Mapping Performance. Remote Sens. 2019, 11, 1634.

[18] Hamylton, S.; Hedley, J.; Beaman, R. Derivation of High-Resolution Bathymetry from Multispectral Satellite Imagery: A Comparison of Empirical and Optimisation Methods through Geographical Error Analysis. Remote Sens. 2015, 7, 16257–16273.

[19] Lee, Z.; Weidemann, A.; Arnone, R. Combined Effect of Reduced Band Number and Increased Bandwidth on Shallow Water Remote Sensing: The Case of Worldview 2. IEEE Trans. Geosci. Remote Sens. 2013, 51, 2577–2586.

[20] Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 2013, 342, 850–853.

[21] Shelestov, A.; Lavreniuk, M.; Kussul, N.; Novikov, A.; Skakun, S. Exploring Google Earth Engine Platform for Big Data Processing: Classification of Multi-Temporal Satellite Imagery for Crop Mapping. Front. Earth Sci. 2017, 5, 17.

[23] Rifqi Muhammad Harrys, Open-Source Bathymetry Gui. https://doi.org/10.5281/zenodo.10511745