remote sensing

MDPI

Article

Surveying Nearshore Bathymetry Using Multispectral and Hyperspectral Satellite Imagery and Machine Learning

David Hartmann [1], Mathieu Gravey [2], Timothy David Price [3], Wiebe Nijland [3] and Steven Michael de Jong [3,*]

[1] Royal Netherlands Aerospace Centre (NLR), Anthony Fokkerweg 2, 1059 CM Amsterdam, The Netherlands; david.hartmann@nlr.nl
[2] Institute for Interdisciplinary Mountain Research, Austrian Academy of Sciences, 6020 Innsbruck, Austria; research@mgravey.com
[3] Faculty of Geosciences, Utrecht University, Princetonlaan 8A, 3584 CW Utrecht, The Netherlands; t.d.price@uu.nl (T.D.P.); w.nijland@uu.nl (W.N.)
* Correspondence: s.m.dejong@uu.nl; Tel.: +31-30-2532749

Abstract: Nearshore bathymetric data are essential for assessing coastal hazards, studying benthic habitats and for coastal engineering. Traditional bathymetry mapping techniques of ship-sounding and airborne LiDAR are laborious, expensive and not always efficient. Multispectral and hyperspectral remote sensing, in combination with machine learning techniques, are gaining interest. Here, the nearshore bathymetry of southwest Puerto Rico is estimated with multispectral Sentinel-2 and hyperspectral PRISMA imagery using conventional spectral band ratio models and more advanced XGBoost models and convolutional neural networks. The U-Net, trained on 49 Sentinel-2 images, and the 2D-3D CNN, trained on PRISMA imagery, had a Mean Absolute Error (MAE) of approximately 1 m for depths up to 20 m and were superior to band ratio models by ~40%. Problems with underprediction remain for turbid waters. Sentinel-2 showed higher performance than PRISMA up to 20 m (~18% lower MAE), attributed to training with a larger number of images and employing an ensemble prediction, while PRISMA outperformed Sentinel-2 for depths between 25 m and 30 m (~19% lower MAE). Sentinel-2 imagery is recommended over PRISMA imagery for estimating shallow bathymetry given its similar performance, much higher image availability and easier handling. Future studies are recommended to train neural networks with images from various regions to increase generalization and method portability. Models are preferably trained by area-segregated splits to ensure independence between the training and testing set. Using a random train test split for bathymetry is not recommended due to spatial autocorrelation of sea depth, resulting in data leakage. This study demonstrates the high potential of machine learning models for assessing the bathymetry of optically shallow waters using optical satellite imagery.

Keywords: satellite-derived bathymetry; hyperspectral PRISMA; multispectral Sentinel-2; machine learning; convolution neural networks

Academic Editor: Andrzej Stateczny

Received: 15 November 2024
Revised: 13 December 2024
Accepted: 2 January 2025
Published: 15 January 2025

Citation: Hartmann, D.; Gravey, M.; Price, T.D.; Nijland, W.; de Jong, S.M. Surveying Nearshore Bathymetry Using Multispectral and Hyperspectral Satellite Imagery and Machine Learning. *Remote Sens.* **2025**, *17*, 291. https://doi.org/10.3390/rs17020291

1. Introduction

Accurate surveying of nearshore bathymetry is relevant for a range of applications, such as monitoring benthic habitats, assessing the vulnerability of the coast to hazards like storms and tsunamis and for engineering and infrastructural works [1–3]. Conventional methods of mapping the nearshore bathymetry include ship-sounding and airborne LiDAR. However, data acquisition using these techniques is laborious, not very efficient and, for some inaccessible areas, unsuitable [4]. The costs of ship-based bathymetry acquisition

depend on project size and the type of water body but generally cost range between USD 2300 and 3500 per square kilometer [5]. As a result, data of over half of the world's nearshore bathymetry are unavailable [6]. Also, for a number of applications, multitemporal measurements of bathymetry are required to gain insight into bathymetric changes over time. The Seabed 2030 project was launched to accelerate bathymetric mapping and to produce a complete global public bathymetric dataset by 2030 [7]. However, relying on conventional bathymetry mapping techniques makes this goal unreachable for 2030.

Optical and microwave satellite imagery provide a promising alternative to conventional methods because they allow for the assessment of bathymetry at detailed spatial resolutions (10 to 30 m) and temporal scales up to days or weeks. Satellite-derived bathymetry using optical sensors has been extensively studied since the 1970s [8]. The interaction between radiance and water is complex, making bathymetry surveys with satellite imagery far from easy [9]. Optical satellite sensors capture the upwelling radiant flux that is reflected from water bodies on the Earth's surface and the atmosphere in the solar part of the spectrum (400 to 2500 nm) and that originates from the water column, atmospheric scattering, and the specular reflection of sunlight on the water surface [10]. Figure 1 presents a conceptual overview of the processes and fluxes involved. When mapping bathymetry using optical satellite imagery, it is essential to consider the various factors that influence the observed signal.

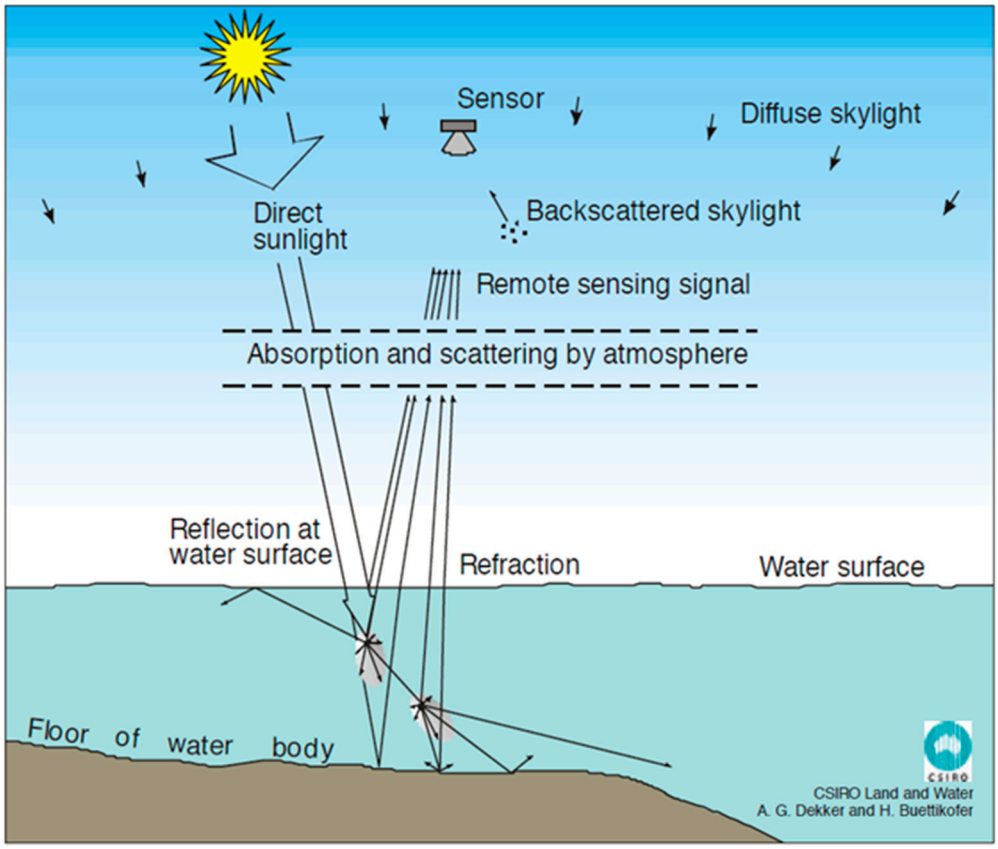


Figure 1. A schematic diagram of the various processes that contribute to the measured signal in optically shallow water. The arrows schematically indicate the direction the direction of radiance. Reprinted with permission from ref [9]. Copyright 2024, Springer Nature.

Methods to derive bathymetry with satellite imagery can be divided into two categories: color-based or empirical methods and analytical methods. Analytical methods typically derive bathymetry by simulating the radiative transfer of light in the water column (e.g., [11,12]). They consider various optical properties of water, such as the attenuation

coefficient and backscattering, which are required as input parameters [13]. The flow radiative transfer model is a commonly used analytical model and it requires the input of the spectral signatures of suspended and dissolved materials and the bottom reflectance [14]. Another type of analytical model is based on the principles of wave kinematics, where the depth is predicted as a function of wave celerity and wavelength [6]. Analytical methods can give accurate results, but accurately accounting for all these processes in a water body is complex and often impracticable. They often require several in situ parameters related to the optical properties and the seabed, making them challenging to implement in practice [15].

Therefore, empirical approaches such as spectral band ratioing are frequently used in practice, indirectly accounting for all these factors, and they employ reference datasets for training the models and for calibration. Traditional empirical models are based on the principle of varying attenuation rates of different wavelengths in water [8]. Lyzenga developed the groundwork for empirical bathymetry. Stumpf et al. [4] improved upon this by developing a band ratio model, which reduced the influence of bottom albedo on the depth estimations, and by reducing the number of tunable constants. A drawback of empirical methods is that they need to be trained or calibrated with in situ data, which are often sparse and imperfect for remote sensing data [16]. Another approach is the use of complex machine learning methods such as random forest models [17,18] and deep learning models, such as convolutional neural networks (CNNs) [19,20]. These methods are promising and have not yet been widely applied to (hyperspectral) imagery for bathymetry modeling.

Despite significant advancements in bathymetry modeling, no comprehensive studies have compared the performance of various empirical methods, ranging from traditional spectral techniques to complex machine learning models, using both multispectral and hyperspectral satellite imagery over an extensive area. In particular, the application of convolutional neural networks (CNNs) to hyperspectral imagery and the use of CNNs trained on large multitemporal multispectral datasets like Sentinel-2 for bathymetry mapping remain largely unexplored. Therefore, this study aims to evaluate the accuracy of nearshore bathymetry mapping in optically shallow and clear waters using both multispectral (Sentinel-2) and hyperspectral (PRISMA) imagery, employing a range of methods including traditional spectral approaches and advanced machine learning techniques (XGBoost v2.1.3, CNN) for a study area in southwestern Puerto Rico.

2. Study Area

A study area with mostly clear waters, covering approximately 100 km^2 and located southwest of Puerto Rico (Figure 2), was selected for this study for a number of reasons. The 40 to 50 km-wide continental shelf here is characterized by a wide range of depths, generally less than 30 m, but with maxima up to 50 m. The shelf has a variety of bottom types, such as coral reefs, sand, volcanic rock and benthic habitats including seagrasses and different species of red, green or brown algae [21]. Together, this provides a large and diverse area for training and testing the various models for estimating depth. The tidal range in the study area is generally small, with a daily maximum of 40 cm and a yearly maximum of 55 cm [22] excluding the need for selecting imagery at a specific tide, which largely increases image suitability. Wave heights are typically around 1 m, with maxima reaching up to 4 m during storms and hurricanes [23].

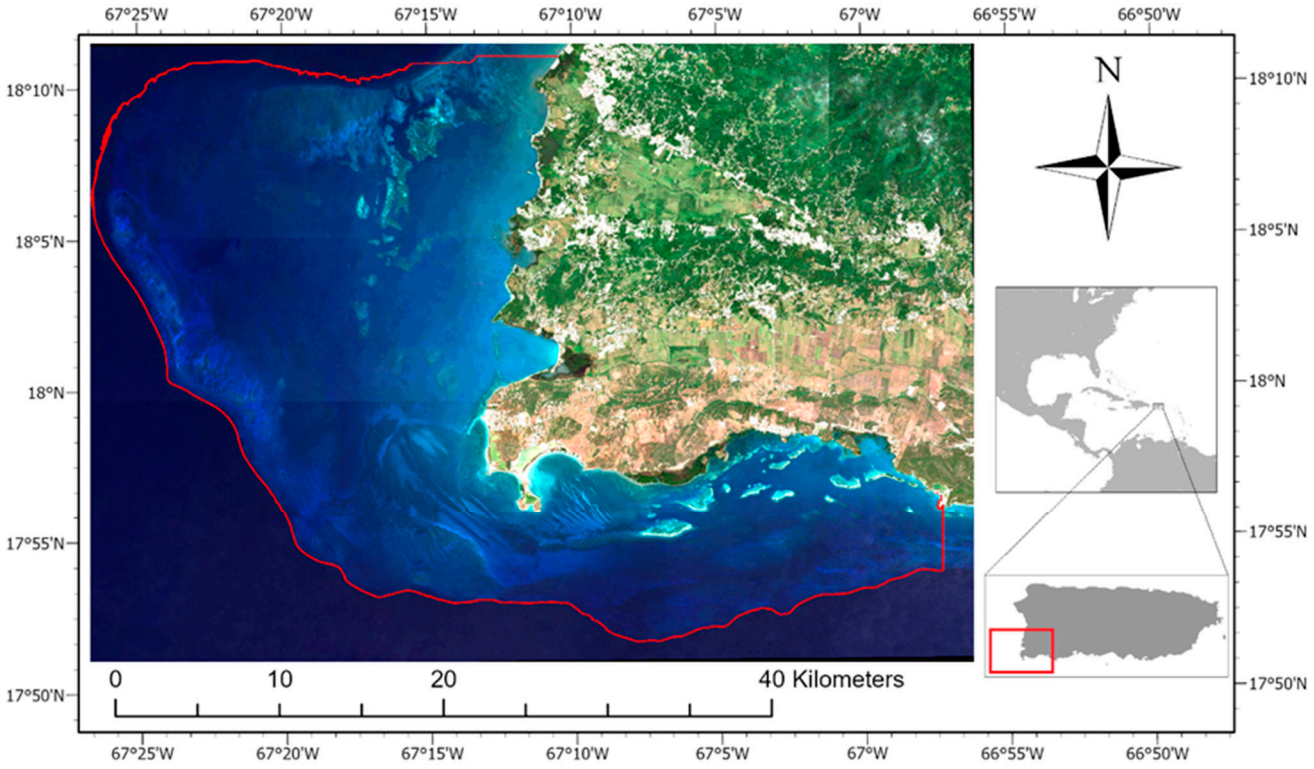


Figure 2. The inset on the right shows the location of the study area in Central America. The left figure shows a Sentinel-2 RGB composite of the study area. The red contour line delineates the continental shelf area.

3. Materials and Methods

3.1. Data

To evaluate the suitability of optical remote sensing for bathymetric surveying, we used ESA Sentinel-2 multispectral images and hyperspectral images acquired by the spaceborne PRISMA sensor and operated by the Italian Space Agency (ASI). Below, we briefly discuss the datasets used, shown in Figures 2 and 3. The main statistics of the datasets are presented in Table 1.

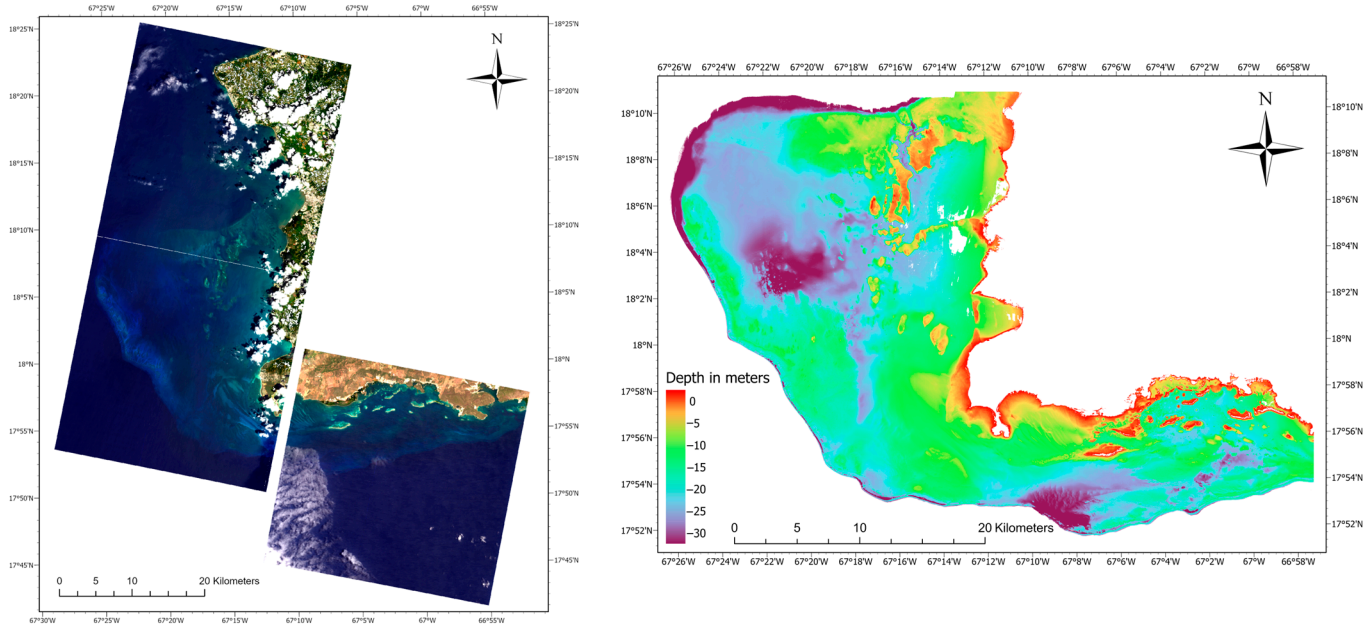
Table 1. Characteristics of the datasets used in this study.

Data	Processing Level	Resolution (m)	Spectral Bands	Revisit Time	Number of Images Used	Acquisition Dates	Note
Sentinel-2	L2A	10 m RGBN, 20 m SWIR	12	5 days	129	2018–2023	Forty-nine low-cloud images selected for multitemporal models.
PRISMA	L2D	5 m PAN, 30 m VNIR SWIR	240	29 days	3	2020	Requires tasking. No continuous data stream.
NOAA Airborne LiDAR	MLLW	3 m	1 (green)	-	1	2006	MAE of 1.2 m in depth measurements.

3.1.1. ESA Sentinel-2

The Sentinel-2 mission consists of two identical satellites, Sentinel-2A and Sentinel-2B [24]. The two satellites operate simultaneously, phased at 180 degrees to each other, and have a sun-synchronous orbit with a swath of 290 km. The satellites are equipped with a multispectral push-broom instrument (MSI). The revisit time in Puerto Rico is five days on average. The MSI captures data in 13 spectral bands that range from visible to short-wave infrared wavelengths. A total of 129 Level-2A surface reflectance images of the study area

were processed from Google Earth Engine ranging between 2018 and 2023 with a cloud cover of less than 5%. Band 10, SWIR Cirrus, is used for atmospheric correction and thus is not included in the Level-2A data product. Few summer images were available due to high cloud cover.



(**a**) (**b**)

Figure 3. Datasets used in this study comprising time-series of three available ASI-PRISMA hyperspectral images of the study area [25] (**a**); Bathymetry map measured with airborne LiDAR instrument by NOAA [26], used as reference data in this study (**b**). Sentinel-2 data are shown in Figure 2.

3.1.2. ASI-PRISMA

PRISMA, a medium-resolution hyperspectral satellite, was launched in March 2019 [25] and operates on a user-driven request system. Users can request images of an area of 30 by 30 km. It has a sun-synchronous orbit and a revisit time of 29 days, nearly six times longer than Sentinel-2's constellation revisit time. The sensor has 240 bands, with 66 in the VNIR spectrum (450–1010 nm) and 173 bands in the SWIR spectrum (925–2497 nm). VNIR bands have a spectral width between 8.9 and 13.2 nm (SNR > 200). SWIR bands have a spectral width between 8.9 and 15.28 nm (SNR > 100). Three PRISMA images from a close point in time (March and April) in 2020 were selected.

3.1.3. NOAA LiDAR Bathymetry Reference Data

The reference bathymetric LiDAR dataset was obtained from the National Oceanic and Atmospheric Administration (NOAA) dataset [26]. For data collection, NOAA used the LADS Mk II LiDAR Airborne System, between 7 April and 15 May 2006, to measure the bathymetry of the Puerto Rican shelf up to approximately a 50 m depth. The measured depths were converted to the Mean Lower-Low Water (MLLW) using tidal data [26]. The MLLW represents the average depth of all low tides during a 19-year recording period [27]. The bathymetry reference data have a spatial resolution of 3 m. This airborne LiDAR reference dataset from NOAA has a reported estimated Mean Absolute Error (MAE) of 1.2 m in the study region. Given the fact that no other complete reference dataset is available and given this high spatial resolution, this LiDAR dataset was used as the reference set for all computed bathymetry products.

3.2. Preprocessing

The preprocessing workflow included several steps to ensure data quality for bathymetry estimation. For Sentinel-2 and PRISMA, atmospherically corrected data products were used. Sentinel-2 images were processed with the Sen2Cor algorithm (Level-2A) and for PRISMA we used the Level-2D product.

For Sentinel-2 imagery, a median composite of the 129 images was created in Google Earth Engine (GEE v.7.3.1) to reduce noise in the data, such as sediment flows, clouds, waves and pollutants such as oil from ships. Sunglint was mitigated by the median compositing approach, which effectively filtered out glint and other noise across multiple images. For both Sentinel-2 and PRISMA, cloud and water masks were created using a near-infrared (NIR) threshold of 0.03 (band 55, 849 nm for PRISMA) and a red band threshold (band 34, 636 nm for PRISMA) of 0.01 to mask cloud shadows. The median Sentinel-2 composite resulted in 8,760,015 pixels. For deep learning methods where using multiple Sentinel-2 images was beneficial, 49 images with minimal noise and cloud cover were selected. No explicit sunglint correction was applied to these images, based on the hypothesis that the deep learning models would implicitly learn to mitigate sunglint-related noise during training.

PRISMA imagery required geometric calibration due to geolocation errors of up to 100 m. Coregistration to the Sentinel-2 dataset was performed using 10,000 tie points and the green band (559 nm) in ENVI's (v.5.6) image registration workflow. Of the PRISMA imagery, only the VNIR bands were extracted (450–1010 nm) due to the negligible radiance of SWIR in water and to reduce the high dimensionality. Sunglint correction was not applied to the PRISMA data due to the absence of visible sunglint. Exploratory analysis of the sunglint correction method proposed by Kutser et al. [28] did not show performance improvement. The resulting PRISMA mosaic contained 908,583 pixels. For the deep learning models, both Sentinel-2 and PRISMA data were z-score normalized.

The NOAA LiDAR bathymetry data were resampled to 30 m when used as reference data for PRISMA imagery and to 10 m for the Sentinel-2 imagery. Nearest-neighbor was applied as the resampling technique, as it preserves the sharp edges and boundaries in the data. A drawback of this technique is the potential loss of information. Due to the limited tidal range in the study area of 0.4 m [22], no tidal corrections were applied. Additionally, the median Sentinel-2 composite reduces the influence of tidal variability, approximating a more typical water level condition [17].

3.3. Band Ratioing Methods

Two conventional spectral band ratio methods were used here to assess nearshore bathymetry, the method developed by Stumpf et al. [4] and that by Ma et al. [29].

The Stumpf method [4] is based on the Beer–Lambert law for estimating bathymetry in shallow clearwater environments, stating that light is attenuated exponentially with depth in the water column. To reduce the influence of bottom albedo, this Stumpf method takes advantage of the fact that different wavelengths of light are attenuated at varying rates in water. By taking the ratio of the reflectance of the green and blue bands, with different water attenuation coefficients, the influence of bottom albedo and variable bottom types is reduced. The equation used is as follows:

$$z(i) = m_1 \frac{\ln(nR_w(Green))}{\ln(nR_w(Green))} + m_0 \quad (1)$$

where z is the estimated depth at pixel i, m_1 is a tunable constant to calibrate the depth ratio to absolute depth, R_w is the water-leaving reflectance, m_0 is an offset for a depth of 0 m,

and n is a fixed constant for all areas. The value of n is chosen to assure that the logarithm will be positive under any condition.

The Ma method [29] practices an algorithm for estimating clear and shallow coastal water bathymetry using hyperspectral imagery. This approach builds upon the Stumpf method by assessing the influences of changing bottom type and depth on the spectral curve. When the water depth is fixed but the bottom type changes, the shape of the spectral curve changes. When the bottom type is equal and the depth varies, the reflectance values will change but the spectral shape remains largely unaltered [30]. To distinguish the different influences of changes in water depth and bottom type on spectral curves, ref. [29] developed a method that utilizes a reference spectrum. As the reference spectrum, they used the spectral signature of a water area where the depth is close to 0 m. They then compared the spectral curve of each pixel (from 480 nm to 610 nm) to the reference spectrum by calculating the Pearson correlation coefficient (CC) and similarity coefficient (SC), as further explained in Appendix B. This method uses the following equation [29]:

$$z(i) = m_1 \frac{\ln(nSC(i))}{\ln(nCC(i))} + m_0 \quad (2)$$

where $m1$ and $m0$ are tunable constants to calibrate the ratio to depth. Here, five reference points per 2 m depth interval up to a depth of 30 m were used to tune these parameters. A disadvantage of this approach is the selection of an accurate 0 m depth reference spectrum, which introduces subjectivity and is challenging for dynamic coastal environments.

3.4. XGBoost

XGBoost (eXtreme Gradient Boosting) is a powerful machine learning algorithm widely used in various remote sensing applications [31–33]. XGBoost uses the predictions of multiple underfitted decision trees to create a single prediction. The XGBoost algorithm has several advantages over other machine learning techniques. It is known to be robust to outliers and noise, it is capable of handling many features, and its predictive power is immune to multicollinearity. The latter is relevant for hyperspectral imagery containing many highly correlated bands. Drawbacks of the algorithm include the need for more training data than the band ratio models and the requirement of hyperparameter tuning to achieve optimal performance [34]. Here, an XGBoost model was used to model bathymetry with the PRISMA mosaic and median Sentinel-2 composite.

3.5. Deep Learning Methods

Deep learning is a subset of machine learning based on artificial neural networks (ANNs). ANNs are composed of node layers with an input layer, one or multiple hidden layers, and an output layer. ANNs are trained with a backpropagation algorithm that adjusts the weights of connections between nodes to minimize a loss function in each iteration [35]. A convolutional neural network (CNN) is a type of ANN containing convolutional layers and is commonly used in image analysis tasks. In a CNN, a filter of size n by n is shifted across the image to extract features rather than analyzing the entire image at once, as in a fully connected neural network. The filters extract feature maps that represent different patterns in the image. In the context of bathymetry, such patterns can relate to variations in depth, turbidity and bottom types. These feature maps are fed to the final layer, which applies an activation function to produce the final predictions. In this study, a relatively shallow CNN with 5 layers and 393,185 parameters was trained to predict the bathymetry.

U-Net is a CNN architecture developed by [36]. Its original purpose was for medical image segmentation, but it has been applied successfully for regression and classification

purposes in other fields, including remote sensing and bathymetry studies [19,37,38]. The U-Net architecture consists of two main parts: the encoder path and the decoder path. The goal of the encoder is to capture increasingly complex and abstract patterns, such as water clarity and bottom type. Here, an Xception-type U-Net model with 2,053,185 parameters was applied, as introduced by [39]. The last layer consists of a 1 by 1 convolutional layer with a linear activation function to acquire the bathymetry predictions in patches of 64 × 64 pixels. The architecture of the model is shown in Figure A3.

To analyze the PRISMA mosaic, a 2D-3D CNN model was created. Conventional CNNs contain 2D convolutional layers like the shallow CNN and Xception U-Net models presented above. These layers perform convolutions only in the spatial dimensions, namely height and width. Having high-dimensional data like hyperspectral imagery, 2D CNNs extract good discriminating feature maps less efficiently [40]. Three-dimensional convolutional layers can learn spatial–spectral features by capturing the relationships in the spatial and spectral dimensions at once. The latter is beneficial for hyperspectral data due to the high volume of bands [40]. In this study, a 2D-3D CNN inspired by the HydridSN model [40] was made to estimate bathymetry from PRISMA imagery. The 2D convolutional layers after the 3D convolutional layers focus on extracting spatial information and use less parameters. The architecture of the model is shown in Figure A4.

3.6. Training Procedures and Validation

The study area was divided into four areas with similar pixel counts as shown in Figure A1 for Sentinel-2 and Figure A2 for PRISMA. Three areas served as the training set and the remaining one served as the test set, iteratively. Such spatial separation of training and testing sets ensures independent sets, which is a requirement for accurate model assessment [41]. While many similar studies use random train–test splits [17,38,42,43], this leads to data leakage, and thus to spatial overfitting and an overestimated model performance due to spatial autocorrelation present in bathymetry [43,44].

The band ratio models were calibrated using 10 random reference depth points of the LiDAR dataset, per 2 m depth interval (0–40 m), resulting in 200 calibration points. Adding more points did not improve performance, as the model's performance had plateaued. For the XGBoost models, the parameters were tuned with GridSearchCV, which performs an exhaustive search over a specified parameter grid. For Sentinel-2, the band ratio model and XGBoost model were trained using the median image.

For the deep learning models, several convolutional neural network (CNN) architectures were evaluated through a systematic experimental approach. The tested models included multiple U-Net variants, shallow 2D CNNs, 2D-3D CNNs, ResNet18, ResNet50 and VGG16. Initial evaluations were conducted on one of the test areas to identify the architectures that demonstrated the best accuracy and stability. The Xception-type U-Net for Sentinel-2 imagery and the 2D-3D CNN for PRISMA imagery were selected based on their superior performance in these tests. Further fine-tuning of these selected architectures was performed by systematically adjusting the patch size, batch size, learning rate, network layers, input dimensions and kernel sizes, varying one parameter at a time while keeping the others fixed. This iterative process allowed for a detailed exploration of model configurations and optimization of each component. Compared to the automated hyperparameter tuning used for the XGBoost models, this fine-tuning process was significantly more time-consuming and resource-intensive.

The deep learning models were trained using the Adam optimizer with a learning rate of 0.001. In total, 25% of the training data were reserved for validation, and early stopping was applied after 15 epochs of no improvement (up to 100 epochs). The Root Mean Squared Error (RMSE) was used as the loss function. For Sentinel-2 imagery, Xception-type U-Net

was trained on patches of 64 × 64 pixels with a 75% overlap to reduce edge artifacts, resulting in approximately 73,000 patches derived from 49 low-cloud images, and also on the median Sentinel-2 image. A batch size of 32 was optimal from empirical findings. For PRISMA imagery, the 2D-3D CNN model processed patches of 16 × 16 pixels spanning 62 VNIR bands (450–1010 nm), yielding approximately 11,000 patches. A batch size of 8 showed the highest performance. All training was conducted on a T4 GPU environment with 52 GB RAM available in Google Colab, using TFRecord files for efficient data streaming of the image patches.

To account for variability in CNN performance due to random initialization, each deep learning model was trained five times. Both the median and the best-performing runs were reported to provide a robust and fair comparison [45]. Model performance was evaluated using MAE and RMSE. A high-level flowchart of the methodology employed in this study is shown in Figure 4.

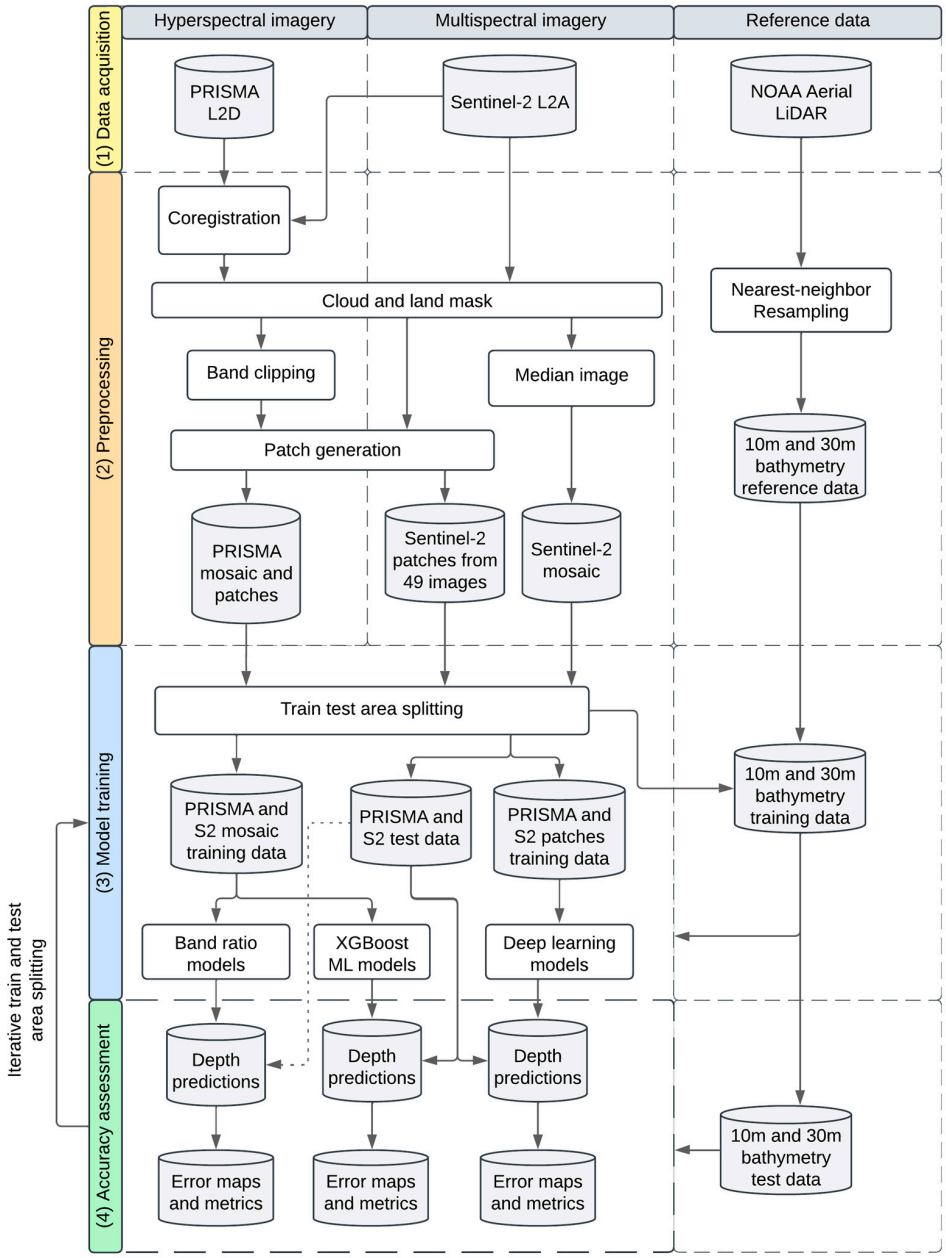


Figure 4. High-level flowchart of the methodology employed in this study, with four main steps: data acquisition, preprocessing, model training and accuracy assessment.

4. Results

In this study, we predicted the nearshore bathymetry in the clear water of southwest Puerto Rico using Sentinel-2 and hyperspectral PRISMA satellite imagery using spectral band ratio models, XGBoost models and CNNs. Two novel depth estimation methods were presented and evaluated: the U-Net model trained on a substantial number of Sentinel-2 images and the 2D-3D CNN trained on PRISMA imagery. The predicted water depth maps are presented in Figures A5 and A6 for the various algorithms and images trained by Sentinel-2 (Figure A5) or PRISMA (Figure A6). In this section, we present the quantitative results of the accuracy assessment for each of the models.

In general, bathymetry estimation using optical remote sensing tends to decrease in accuracy with increasing water depth. Table 2 shows the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) with increasing depth for all the models based on Sentinel-2 and PRISMA imagery. Mapping accuracy ranges from 0.78 m for shallow water (5–10 m) to 13.50 m for deep water (35–40 m).

Table 2. The Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in m per depth interval of the Sentinel-2 (S-2) and PRISMA (PR) models. The last column shows the average MAE and RMSE per model. The best-performing model per depth range is highlighted in bold.

Model	Metric	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40	Avg
S-2 Stumpf et al. [4]	MAE	1.21	2.09	2.93	3.72	3.18	3.65	6.42	11.73	3.24
	RMSE	1.64	2.59	3.34	4.10	3.86	4.67	7.55	12.64	4.16
PR Stumpf et al. [4]	MAE	2.18	2.11	2.53	3.04	2.87	3.98	6.69	12.37	3.00
	RMSE	3.75	2.71	3.14	3.85	3.83	4.85	7.82	13.38	4.11
PR Ma et al. [29]	MAE	1.39	1.58	2.28	2.82	2.47	3.33	6.77	12.86	2.65
	RMSE	1.96	1.99	2.74	3.42	3.37	4.23	7.57	13.50	3.68
S-2 XGBoost	MAE	1.03	1.84	2.37	2.57	2.91	5.10	6.29	8.22	2.77
	RMSE	1.39	2.36	3.00	3.25	3.80	5.85	7.36	9.51	3.77
PR XGBoost	MAE	0.98	1.38	1.45	2.03	1.88	3.31	5.81	9.60	2.00
	RMSE	1.52	1.96	1.95	2.81	2.67	4.11	6.98	11.53	3.04
S-2 shallow CNN	MAE	1.12	1.53	1.63	2.12	2.44	4.29	5.55	**7.43**	2.28
	RMSE	1.44	1.97	2.21	2.63	3.18	5.05	6.34	**8.11**	3.15
S-2 U-Net	MAE	**0.85**	**0.78**	**0.84**	**1.08**	**1.62**	2.87	5.62	10.15	**1.42**
	RMSE	**1.15**	**1.02**	**1.13**	**1.49**	**1.97**	3.21	**5.98**	10.44	**2.14**
PR 2D-3D CNN	MAE	1.06	0.80	1.03	1.38	**1.62**	**2.33**	**5.46**	8.44	1.52
	RMSE	1.54	1.22	1.46	2.04	2.33	**3.00**	6.09	9.54	2.43

4.1. ESA Sentinel-2-Trained Models

Figures 5 and 6 show the results for the Sentinel-2-trained models of Stumpf [4], XGBoost and U-Net. The predicted depth maps for these three methods are shown in Figures A5 and A6. Figure 6 shows XY plots of predicted depth versus reference depth (LiDAR) and error depth for these three models. The scatterplots of predicted versus reference depths demonstrate a positive linear correlation for all three models with the highest density of points around the X = Y line. Reference points spread everywhere around the X = Y line indicate that errors occur at lower and larger depths for all models. As the depth increases, the spread increases, suggesting larger errors with increasing depth. Beyond depths of 30 m the performance of all three models largely deteriorates, especially for the Stumpf model (Figure 5a,b); this is less so for the XGBoost model (Figure 5c,d) and the relative best performance is by the U-Net model (Figure 5e,f). The lower model performance with increasing depth is most probably caused by the high spectral similarity of reflectance at greater depths.

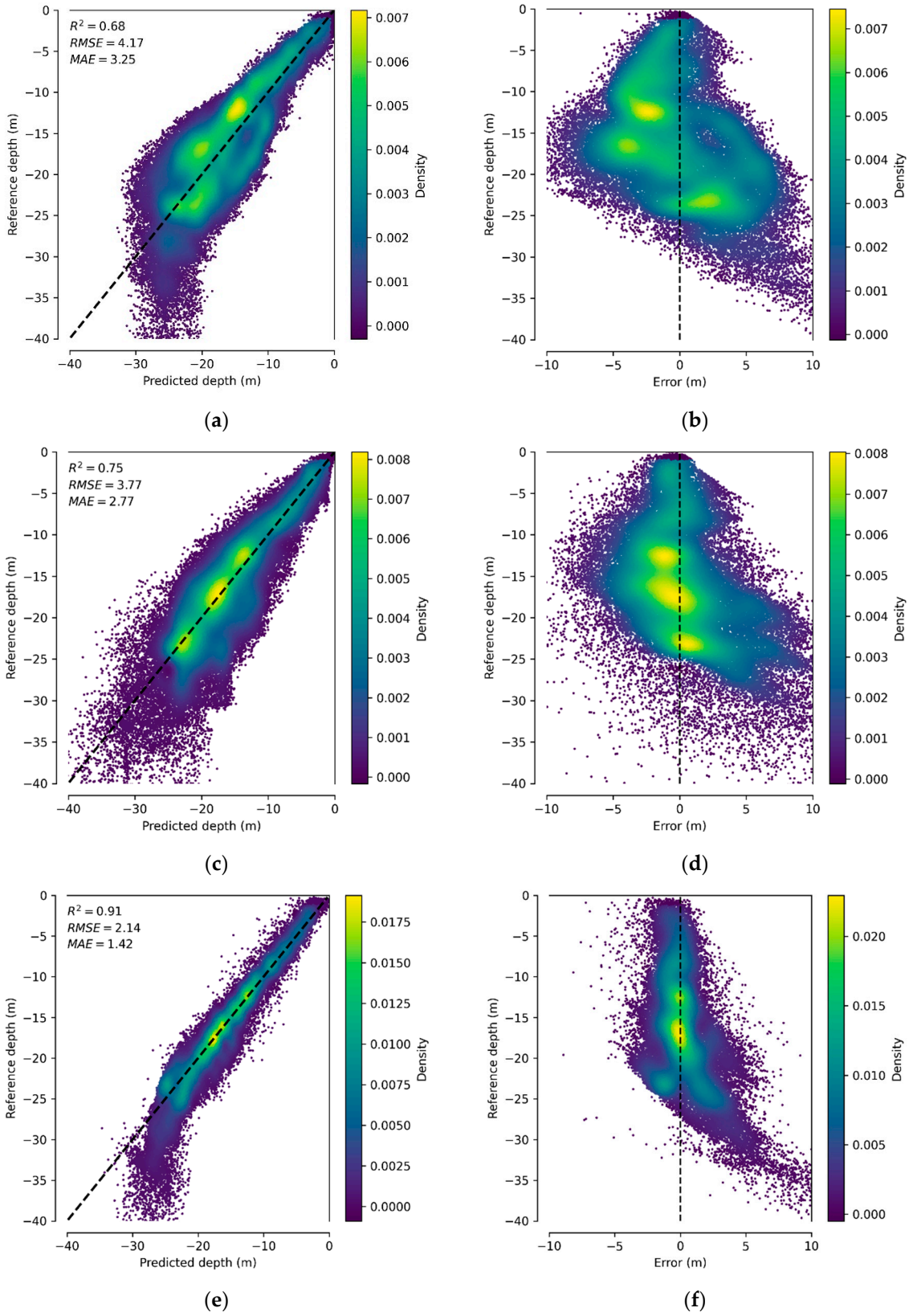


Figure 5. XY-plots of the predicted depth (trained with Sentinel-2) versus the reference depth (LiDAR) and error depth for the Stumpf (**a**,**b**), XGBoost (**c**,**d**) and U-Net (**e**,**f**) models. (**a**) Stumpf: Reference vs. predicted depth. (**b**) Stumpf: Reference depth vs. error. (**c**) XGBoost: Reference vs. predicted depth. (**d**) XGBoost: Reference depth vs. error. (**e**) U-Net: Reference vs. predicted depth. (**f**) U-Net: Reference depth vs. error. The dashed lines in the figures represent zero errors (1:1 line).

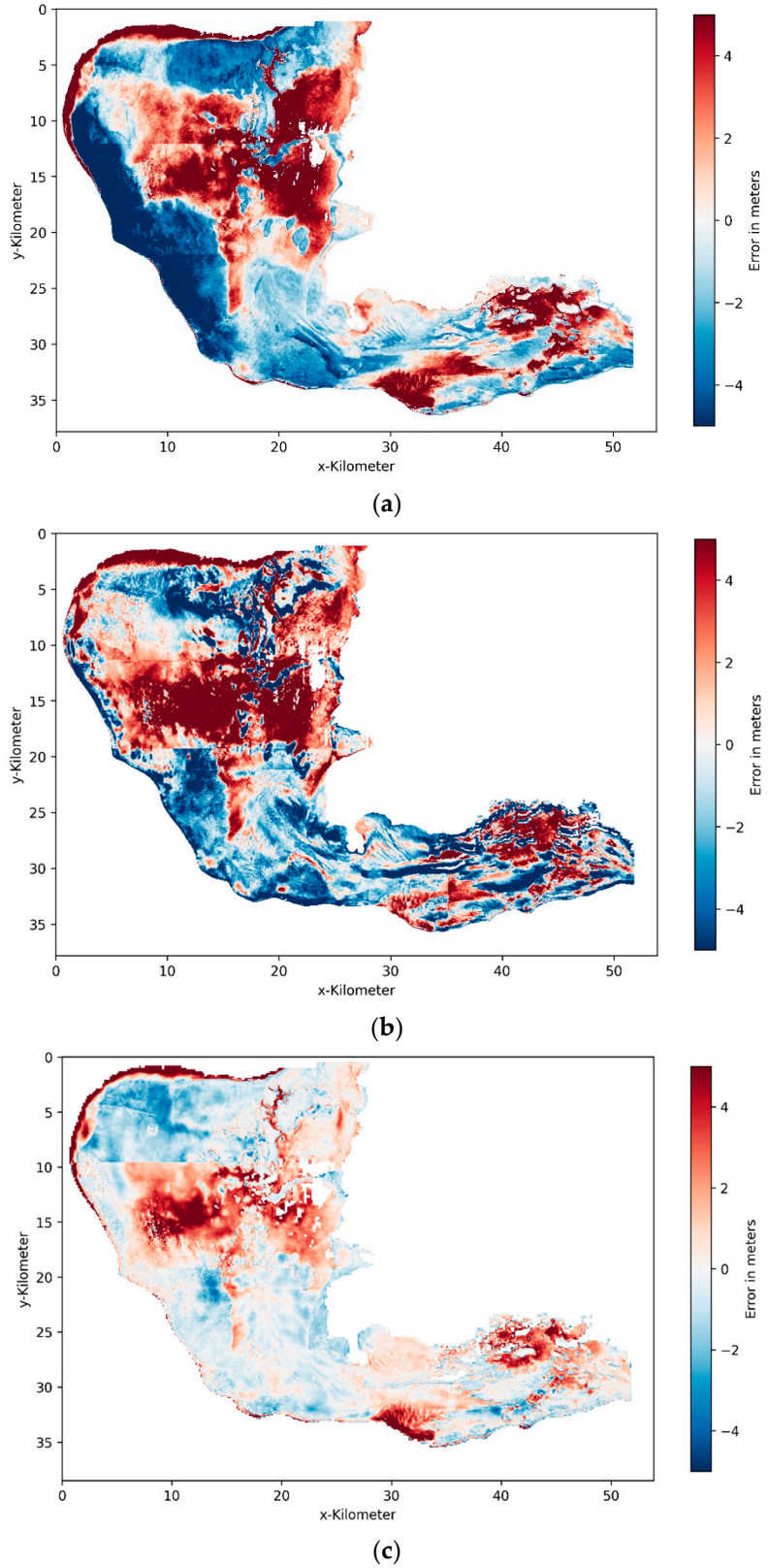


Figure 6. Error maps of predictions from Stumpf (**a**), XGBoost (**b**) and U-Net (**c**) models trained with Sentinel-2 imagery versus the LiDAR reference depth. The visible horizontal lines are artificial and due to spatial train–test splitting from Figure A1. (**a**) Stumpf predictions error map. (**b**) XGBoost predictions error map. (**c**) U-Net predictions error map.

The error maps in Figure 6 show that the three models perform quite differently across the different areas. The error map for the Stumpf model (Figure 6a) does well in the

southwestern part of Puerto Rico at the peninsula of Cabo Rojo. This area is a sheltered bay and relatively calm and stable over time, with low levels of sedimentation and other environmental factors that affect water clarity. For depths up to 20 m, the MAE of the model in this area is 1.15 m and the RMSE is 1.44 m. The error map for the XGBoost model predictions (Figure 6b) shows similar error locations to the Stumpf model. The model underpredicts south of the Guanajibo River due to turbid water, as well as in the southeast area with numerous above-water structures. Shallower regions are generally overpredicted. For all three error maps, sharp transitions in errors are visible, which represent the boundaries between the area splits.

For the Stumpf model, the overall MAE and RMSE of the model are 3.2 m and 4.1 m, respectively, while the MAE for shallow depths, between 0 and 5 m, is 1.3 m and increases gradually towards larger depths. Its performance does not increase when calibrating with all datapoints. The XGBoost model shows an overall MAE of 2.7 m (Table 2), which is 14% better than the Stumpf model. The RMSE is 3.7 m with an R2 of 0.75. For depths up to 20 m, the MAE is 2.21 m and the RMSE is 2.87 m, which is also an improvement of, respectively, 16% and 20% compared to the Stumpf approach. Using a random train–test split for an XGBoost model resulted in a 36.7% lower MAE compared to spatially segregated splits, showing the importance of independent train and test sets. Resampling the Sentinel-2 imagery to 30 m resolution while maintaining the reference dataset at 10 m showed similar performance.

The shallow CNN trained on the median image yielded an RMSE of 3.15 m and an MAE of 2.28 m (Table 2). The shallow CNN trained with 49 images has an RMSE of 3.15 m and an MAE of 2.39 m. A Sentinel-2 2D-3D CNN model with the same architecture as the PRISMA variant achieved similar performance. When the shallow CNN was trained only on pixels shallower than 20 m, the RMSE dropped to 1.82 m and the MAE to 1.39 m. This is a reduction of 22% and 21% compared to the XGBoost model for the same depth range.

The U-Net model was trained five times with the 49 Sentinel-2 images and performed best. For each of the five model runs, MAE and RMSE were calculated for each of the 49 predicted images. Taking the median prediction of the 49 predictions of the best-performing model run resulted in an MAE of 1.42 m and RMSE of 2.14 m. The MAE of the model run with the median performance was 1.74 m and the RMSE was 2.50 m. The predictive ability remained high up to a 25 m depth, with a decrease in performance for greater depths, as visible in Figure 5e. When training the U-Net for depths shallower than 20 m, the MAE decreased to 1.1 m and the RMSE to 1.5 m for this depth interval. When the U-Net was trained on the median Sentinel-2 image, the MAE was 2.90 m and the RMSE was 4.06 m, which is worse than when trained on 49 images. This suggests that employing an ensemble method and calculating the median prediction yields higher performance compared to using a single prediction model of the median image.

The error map of the best-performing U-Net model is shown in Figure 6c. The spatial patterns in the errors are similar to the previous two models, as also here, the model underpredicts the depth for turbid water and areas deeper than 20 m. However, the degree of overprediction and underprediction has improved over the other models. For example, the predictions south of the outflow of the Guanajibo River have improved. While the other models underpredicted the depth here by up to 5 m due to the sediment flows, the U-Net model predicted the bathymetry with a 1.5 m accuracy in this region. The U-Net is designed to use contextual information from the input image through its encoding and decoding paths, which may have helped to recognize the sediment flows. The map of the standard deviations of the 49 predictions is shown in Figure A7. It is visible that in areas where the model predictions are more variable or uncertain, such as periodically turbid regions, the prediction errors are higher. Each unit increase in standard deviation

corresponds to a 1.05 m increase in prediction error ($p < 0.05$). A histogram of the errors of the Sentinel-2 models' predictions is visible in Figure 7.

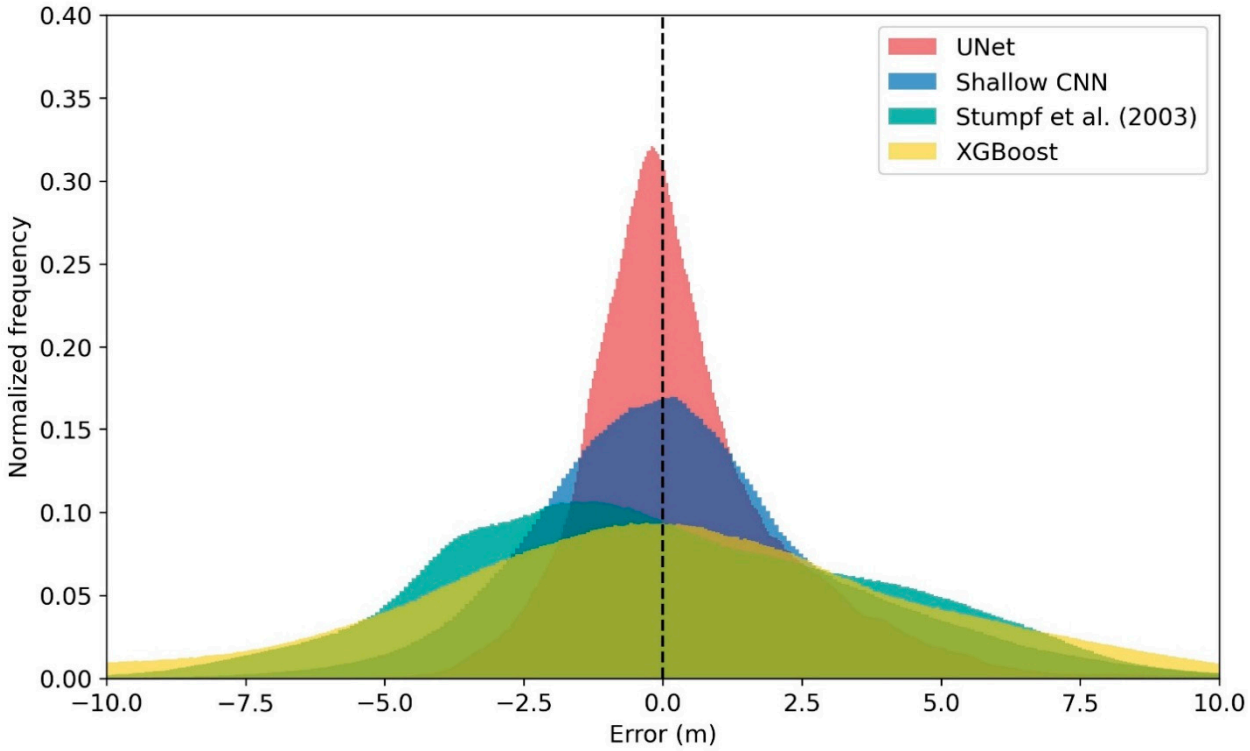


Figure 7. Histograms of predicted depth errors in meters for different models (U-Net, Shallow CNN, Stumpf et al. [4] and XGBoost) with Sentinel-2 imagery. The dashed vertical line represents zero error, with overprediction (predicted depth larger than the LiDAR reference) shown to the left and underprediction to the right.

4.2. ASI-PRISMA-Trained Models

Figures 8 and 9 show the results for the PRISMA-trained models of spectral band ratioing of [29], XGBoost and 2D-3D CNN. Figure 8 shows XY plots of predicted depth versus reference depth (LiDAR) and Figure 10 shows the error depth maps for these three models. The predicted depth maps for these three methods are shown in Figure A6. The XY plots show again for all three methods a linear trend with an increasing error with depth. The performance of the PRISMA models is highest for the 2D-3D CNN, with an MAE of 1.52 m for the best-performing model run, which is 7.8% worse than the best Sentinel-2 U-Net model run. The MAE is 1.58 m for the median-performing 2D-3D CNN run. This is followed by PRISMA XGBoost (2.00 m). From these three methods the Ma method has the largest MAE (2.65 m), but it outperforms the Sentinel-2 Stumpf model by 18%. For water depths over 20 m, the Ma method also outperforms the Stumpf, the Sentinel-2 XGBoost and the Sentinel-2 shallow CNN models described in Section 4.1. This better performance for larger water depth, 25 to 30 m, might be caused by the high spectral resolution of PRISMA's spectral bands in green and blue and the low noise levels. Calibrating the Ma band ratio model with the entire reference dataset did not lead to improved accuracy. In comparison, the XGBoost model trained with 10,000 pixels resulted in an MAE of 2.38, which was worse than the Ma method with the same number of pixels. However, training XGBoost with 100,000 pixels achieved similar results to using the complete dataset from all three splits (2.01 m). When the 2D-3D CNN was trained on 100,000 pixels instead of circa 2,800,000, its MAE was 2.49 m, an increase of 36%.

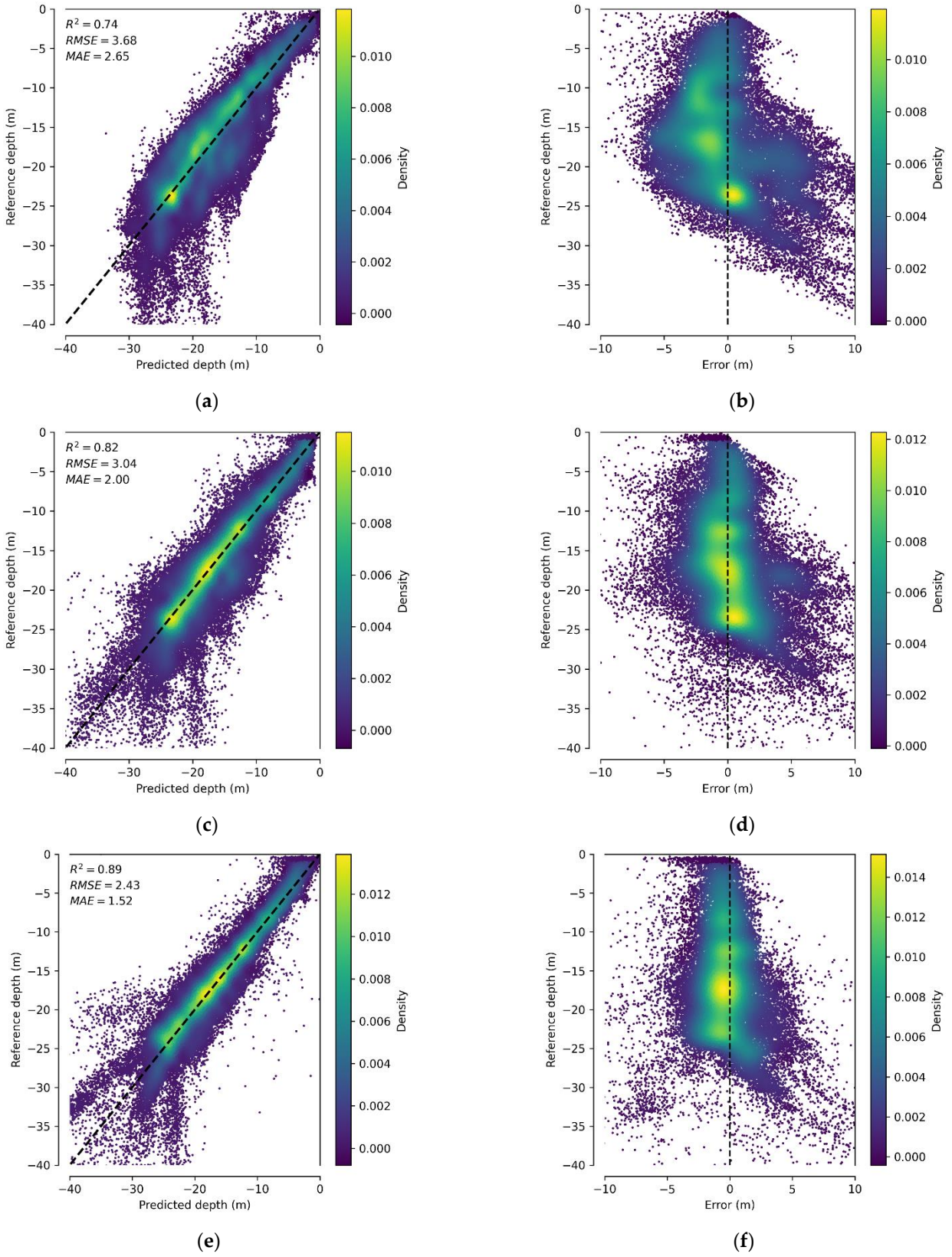


Figure 8. XY plots of the predicted depth (trained with PRISMA) versus the reference depth (LiDAR) and error depth for the models Ma (**a**,**b**), XGBoost (**c**,**d**) and 2D-3D CNN (**e**,**f**). (**a**) Ma: Reference vs. predicted depth. (**b**) Ma: Reference depth vs. error. (**c**) XGBoost: Reference vs. predicted depth. (**d**) XGBoost: Reference depth vs. error. (**e**) A 2D-3D CNN: Reference vs. predicted depth. (**f**) A 2D-3D CNN: Reference depth vs. error. The dashed lines in the figures represent zero errors (1:1 line).

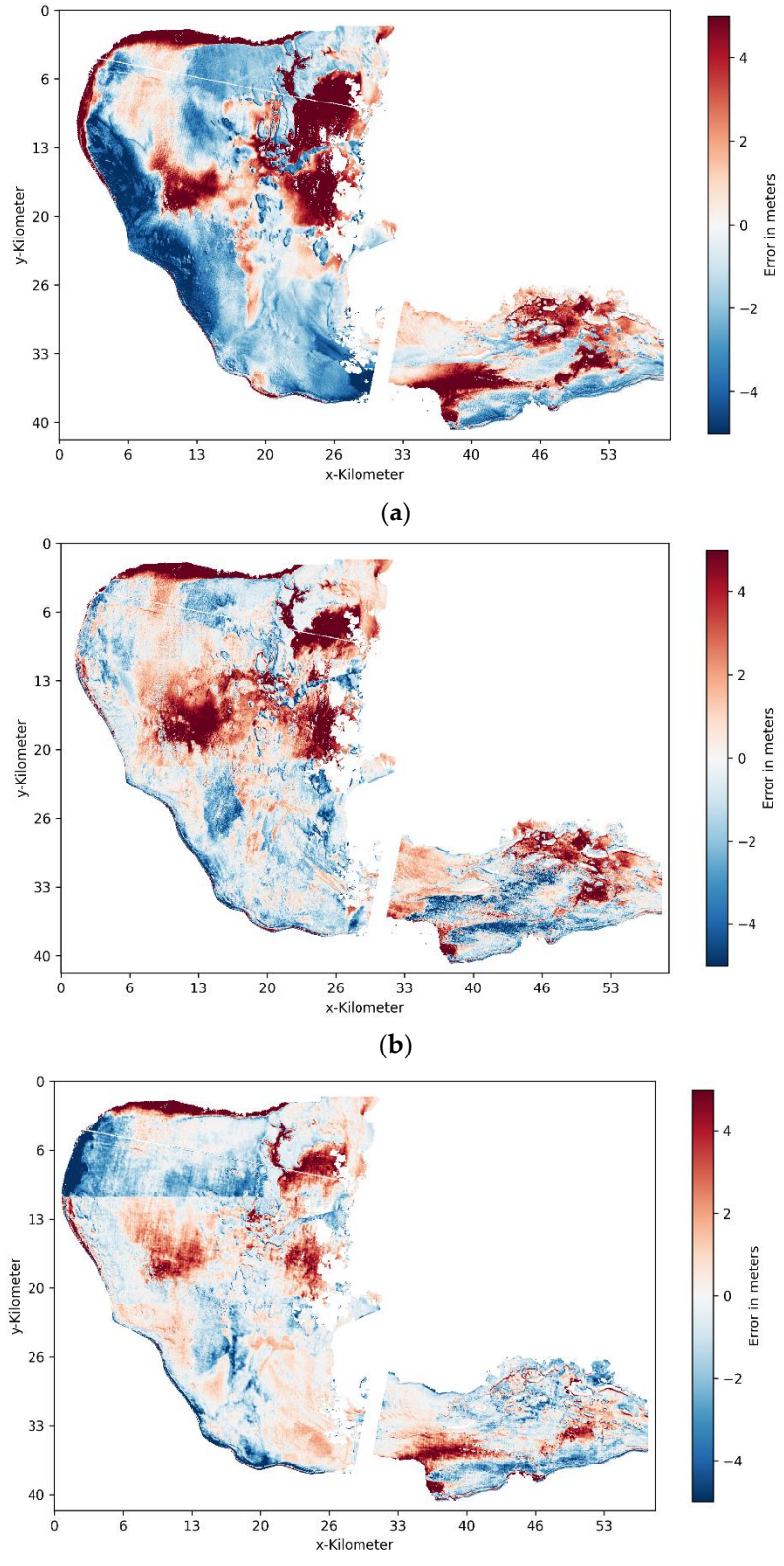


Figure 9. Error maps of predictions from Ma (**a**), XGBoost (**b**) and 2D-3D CNN (**c**) models trained with PRISMA imagery versus the LiDAR reference depth. The visible horizontal lines are artificial and due to spatial train–test splitting from Figure A2. (**a**) Ma predictions error map. (**b**) XGBoost predictions error map. (**c**) A 2D-3D CNN predictions error map.

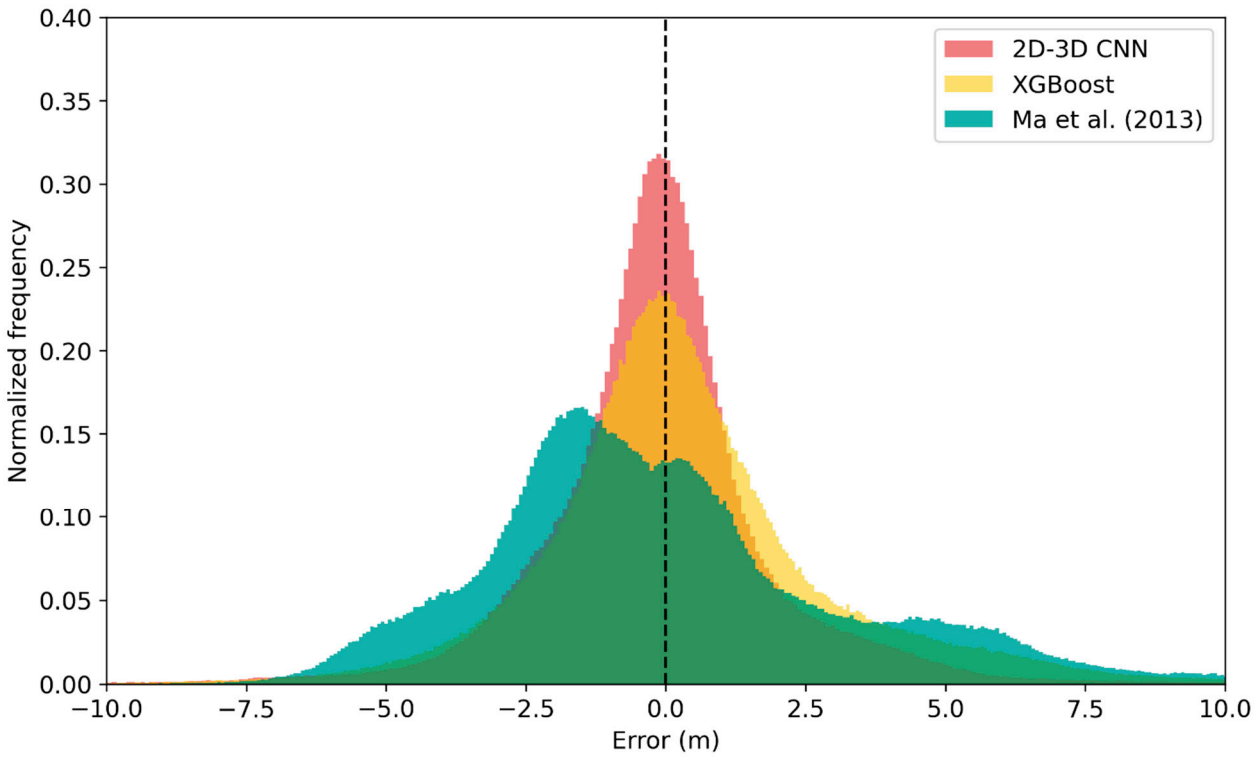


Figure 10. Histograms of predicted depth errors in meters for different models (2D-3D CNN, XGBoost and Ma et al. [29]) with PRISMA imagery. The dashed vertical line represents zero error, with overprediction (predicted depth larger than the LiDAR reference) shown to the left and underprediction to the right.

The error maps In Figure 9 show that each of the three models has fairly similar patterns with underpredictions in the turbid water near the outlet of the river and overpredictions in shallow, near-coastal waters. In Figure 9b, XGBoost shows overpredictions for water depth in the deep area in the southern part of the study area, while the other models show underpredictions. An explanation for this has not yet surfaced and might be difficult to find due to the black box character of the models. The error map for the 2D-3D CNN model in Figure 9c shows overprediction in the northern area and like the other models, underpredictions near the river outlet. A histogram of the errors of the PRISMA models' predictions is presented in Figure 10.

5. Discussion

We review here the performances of the used bathymetry models, compare them and discuss each model's advantages, limitations and potential improvements. Our results show that more complex machine learning methods perform better than conventional spectral band ratio models. These findings are in line with other published studies [15,17,20,46]. Below, we discuss the pros and cons of our methods, as summarized in Table 3.

The Stumpf et al. band ratio model [4], the most conventional method for estimating nearshore bathymetry, was shown to have the lowest performance. Beyond 20 m depth, the log ratios between the blue and green bands (Equation (1)) generally deviate from the assumed linear relationship. As a result, when these deep points are included for calibration, the shallow depths become overpredicted and deeper areas (>20 m) are underpredicted. Large underpredictions are also observed in turbid areas. These uncertainties were also observed in our other models and are reported in the literature, e.g., [18,43,47]. Regions with good predictions are normally calm and stable water areas, e.g., south of the peninsula of Puerto Rico. This highlights the importance of selecting a larger and more diverse study

area to assess the performance of a model and emphasizes the advantage of working with multitemporal remote sensing datasets. The Stumpf et al. model [4] assumes that the ratio of green and blue at a given depth is constant, despite changes in bottom albedo. However, this is not always true. If different bottom types have different green–blue reflectance ratios, the assumption of constant change in reflectance for different bottom types does not hold. Ref. [48] supports this by stating that the Stumpf method does not work whenever the change in reflectance is dominated more by the environmental state than the water depth. The high spatial variability in bottom types in the study area was one reason for the high errors with this model. A further limitation of the band ratio models is the assumption of homogeneous water and environmental conditions such as water clarity and atmospheric conditions. Under variable conditions, the linear relationship between the log ratio of the green and blue bands and the depth is not maintained [49]. Furthermore, the models rely on single pixel analysis and do not use the spatial information, even though bathymetry has an inherent spatial component that can be used for predictions [43].

Table 3. Summary of the advantages, disadvantages and accuracy of empirical models.

Model Type	Advantage	Disadvantage	Accuracy (MAE)
Band ratios	• Able to calibrate with few reference points (<100) • Transparency: easy to interpret causes of errors • Simple to implement	• Single pixel analysis; does not use neighborhood depth information from surrounding pixels • Only captures log-linear relationships • Severe underprediction of turbid waters (>5 m) • Requires calibration for each distinct water body	<20 m: ~1.8 m 20–30 m: ~2.6 m
XGBoost	• High performance for PRISMA imagery • Can deal well with many bands • Can establish nonlinear relationships • Optimal model through an exhaustive search	• Single pixel analysis; does not use neighborhood depth information from surrounding pixels • Highly sensitive to training leakage in presence of spatial autocorrelation • Strong underpredictions of turbid waters • Needs more training data (>10,000 points)	<20 m: ~1.7 m 20–30 m: ~2.2 m
CNNs	• Best performance • Extracts spatial information • Generalization potential • Hierarchical learning can ignore environmental noise	• Requires large amount of training data (>500,000 points) • Time-consuming trial and error optimization • Computationally expensive • Still underpredicts turbid waters	<20 m: ~0.9 m 20–30 m: ~1.8 m

The XGBoost models outperformed both band ratio models. PRISMA XGBoost yielded a 25% lower MAE than the Ma et al. band ratio model [29]. The XGBoost model is able to capture nonlinear relationships between spectral bands' reflectance and water depth, while band ratio models capture log-linear relationships. XGBoost makes use of small spectral differences in the PRISMA imagery that may not be apparent in basic band ratio models and this will result in improved depth predictions, especially for larger depths. The XGBoost algorithm has so far not extensively been used for bathymetry prediction. A random forest approach, a similar tree-based model, outperformed a linear band ratio model for Oahu, Hawaii [20]. An XGBoost limitation is the need for a large training dataset because it must learn the underlying relationships between reflectance and bathymetry from scratch. In contrast, the band ratio models are based on a predefined relationship between the spectral bands and the water depth and require only calibration with a small number of in situ points. Without a sufficiently large training dataset, complex machine learning models are not usable. The XGBoost model, like the band ratio models, does not account for spatial information.

The CNN models performed better than the XGBoost model and a fully connected ANN. This indicates the importance of spatial information for modeling bathymetry, consistent with [42]. The Xception-type U-Net achieved superior performance over the shallow

CNN with a lower MAE of 25%, when trained with 49 Sentinel-2 images. The U-Net is superior in learning the more complex relationships between the reflectance and water depth. The architecture of the U-Net, with its symmetric encoder and decoder structure, enables learning patterns at multiple levels of abstraction. This increases its capacity to learn spatial contexts and may be related to the improved predictions of sediment flows south of the Rio Guanajibo River. While the other models underpredicted the depth here strongly due to the higher reflectance of turbid water, the U-Net model predicted the bathymetry within 1.5 m accuracy in this region. However, the same error patterns persist as in the simpler models, as depths larger than 20 m and turbid waters are still mostly underestimated. When the Xception-type U-Net was trained on only the median image, it had a 25% higher MAE than the shallow CNN trained on the same data. Complex models like the U-Net have a higher risk of overfitting when trained with smaller datasets, as they have more parameters to learn from the data. In this case, the shallow CNN, with fewer parameters, was better suited to the limited training data of the median image. When an Xception-type U-Net is trained with 49 Sentinel-2 images, circa 300 million pixels, underfitting and computational efficiency become the predominant concerns [35]. The 2D-3D CNN trained with PRISMA imagery had a 24% lower MAE than the XGBoost model trained on the same data. It also performed better than a 2D CNN with a similar number of parameters. One reason for the improved performance of the 2D-3D CNN could be its ability to capture spatial and spectral features of the imagery simultaneously. Although the 2D CNN also captures spectral information, it treats each band as a separate 2D image. It is not designed to explicitly capture the interactions between different bands, unlike 3D convolutional layers [40].

This study focused on Puerto Rico due to its large area, varied bathymetry and bottom types, and the availability of both hyperspectral and reference data. For future research, extending the study area to include multiple regions with diverse water conditions and bottom types is recommended. Training deep learning models on such varied datasets or calibrating them locally for specific regions will help evaluate their generalizability and consistency in model performance across different regions.

5.1. Multispectral Versus Hyperspectral

Multispectral imagery has widely been studied for nearshore bathymetry studies [4,17,48], whereas hyperspectral imagery is far less used [29,50]. The only study that used PRISMA imagery for bathymetry estimation is from [51]. No previous studies have compared multispectral and hyperspectral satellite imagery for bathymetry estimation for the same study area as far as we know. Our results indicate that complex Sentinel-2 models outperform PRISMA models for depths up to 20 m, with an MAE of 0.9 m versus 1.1 m. This higher performance is likely due to the better availability of Sentinel-2 data, which allows for a more complex model and the implementation of ensemble prediction across 49 images. The PRISMA models utilized one mosaic from the three available images. The best-performing single mosaic Sentinel-2 model had an MAE of 2.28 m and thus was less accurate than the PRISMA 2D-3D CNN. Based on these results, we anticipate that hyperspectral models outperform Sentinel-2 models when more images are available. For depths between 25 and 30 m, PRISMA performs better than Sentinel-2. The PRISMA 2D-3D CNN had an MAE of 2.33 m for this depth interval, while the Sentinel-2 U-Net model had an MAE of 2.86 m. The PRISMA sensor's higher spectral resolution and higher SNR may allow for more accurate differentiation between the small differences in reflectance of depths between 25 m and 30 m. For depths up to 25 m, the U-Net with Sentinel-2 and the 2D-3D CNN with PRISMA imagery performed similarly. Experiments where Sentinel-2 imagery was resampled to 30 m showed similar performance to 10 m resolution with

XGBoost, suggesting that spatial resolution was not a determining factor. The pros and cons of using multispectral (Sentinel-2) versus hyperspectral (PRISMA) models, based on our study, are summarized in Table 4.

Table 4. Advantages, disadvantages and accuracy of multispectral Sentinel-2 and hyperspectral PRISMA models.

Model Type	Advantage	Disadvantage	Accuracy (MAE)
Sentinel-2	• High data availability allows for monitoring, ensemble predictions and diverse input conditions • Higher spatial resolution • Less data processing • Performances up to 25 m depth better than PRISMA models	• Worse performance for depths > 25 m • Worse performance for single mosaics	<20 m: ~0.9 m 20–30 m: ~1.9 m
PRISMA	• Higher SNR and spectral resolution • Higher performance for depths of 25–30 m • Better performance with single mosaics	• Laborious preprocessing • Limited data availability • Large image size	<20 m: ~1.1 m 20–30 m: ~1.8 m

5.2. Reference Data

A serious limitation of machine learning models is that they require large training datasets to learn and to prevent overfitting [35]. Extensive and usable LiDAR or ship-sounding datasets are limited. The IceSat-2 satellite LiDAR may provide an alternative reference data source for some regions, although intensive preprocessing is required to correct for noise [52]. IceSat-2 coverage is scarce and is provided in strips of approximately 90 m. Therefore, the IceSat-2 satellite is more commonly used as a dataset to calibrate band ratio models, which require few calibration points [52]. Ref. [53] modeled the bathymetry of the Virgin Islands with a shallow 2D CNN using IceSat-2 data as reference data and achieved an MAE of 1.80 m for depths up to 30 m, indicating the potential of IceSat-2 for training. Frugal learning techniques and training on top of foundation models offer promising directions to address the challenge of limited reference data. These approaches can efficiently leverage sparse labeled datasets or pre-trained models to potentially improve bathymetric predictions [54].

In this study, we used the NOAA airborne LiDAR datasets as the reference. This dataset has a relatively high error with an MAE of 1.2 m in our study area. Two more recent NOAA airborne LiDAR surveys dating from 2018 and 2019 have an MAE of less than 0.20 m [3], but due to high turbidity levels during data acquisition, large parts of our study area were not covered. Therefore, we used the 2006 dataset here as it provided complete coverage, was the best available and was considered sufficiently accurate for comparing the SDB methods and products. Furthermore, it was assumed that bathymetry is stable over time. However, the large temporal offset between the satellite images and the ground reference LiDAR represents a possible error source. As time increases between the LiDAR measurements and the acquisition time of the satellite imagery, currents and strong weather events can change the bathymetry. Ideally, reference data from shortly before the acquisition time of the satellite imagery are used [49]. Nonetheless, satellite imagery and LiDAR with similar temporal offsets have been used successfully in SDB studies before, e.g., [42,55].

5.3. Turbidity

An important constraint of successful bathymetry surveys using optical remote sensing is turbidity. Empirical-based methods for modeling bathymetry are only applicable for relatively clear water regions. Using median image products provide a solution for conditions when water is only periodically turbid due to filtering out most of the turbidity.

When training the U-Net with many images with low turbidity, the model can capture relevant patterns in patches containing less noise. However, for regions with constant turbidity, such multitemporal techniques are not viable [56]. Analytical models are more suitable for turbid waters than empirical models [13]. For example, the SAMBUCA algorithm [12] for hyperspectral imagery achieved an RMSE of 0.86 m for the turbid waters of Moreton Bay, Australia [13]. The downside is that such analytical models require the estimation of numerous parameters that often rely on field sampling. For empirical bathymetry methods, including the Normalized Difference Turbidity Index (NDTI) [57] could reduce errors associated with turbidity. The NDTI is defined by the ratio of the difference and sum of the red and green spectral bands. The index reflects the concept that more turbid water has a greater spectral response of red light relative to green light compared to clearer water. Including the NDTI as a predictor and explicit feature representation can therefore help to differentiate turbid from clear water, consequently improving the model's predictions in turbid conditions. Future studies may investigate the extent to which the inclusion of the NDTI improves model performance and assess its impact on more complex machine learning models.

6. Conclusions

This manuscript presents a study modeling the nearshore bathymetry of southwest Puerto Rico using multispectral Sentinel-2 and hyperspectral PRISMA satellite imagery, applying band ratio models, XGBoost models and convolutional neural networks. Two novel depth estimation methods were proposed: the U-Net model trained on a substantial number of Sentinel-2 images and the 2D-3D CNN trained on PRISMA imagery.

These complex models outperformed the band ratio and XGBoost models, achieving an MAE of approximately 1 m. These results illustrate that complex models outperform simpler models and highlight the value of spatial information in bathymetry modeling. Although the CNNs reduced overprediction in more turbid regions, such waters remain difficult for bathymetry modeling. Other regions which exhibited larger errors were areas with depths larger than 20 m, due to the strong light attenuation at these depths. For depths up to 20 m, Sentinel-2 demonstrated sub-1 m prediction accuracy, surpassing PRISMA. However, for greater depths between 25 and 30 m, PRISMA outperformed Sentinel-2 due to its high spectral resolution, which enables the distinction of small spectral differences at larger depths. The spatial resolution appeared not to be a determining factor for estimating bathymetry, as when Sentinel-2 was resampled to 30 m from its regular 10 m resolution, the performance remained similar. Sentinel-2 is the here-recommended data source for modeling optically shallow bathymetry because of its comparable performance, much larger data availability and ease of handling compared to PRISMA. The advantage of the large temporal resolution of Sentinel-2 imagery is that it allows for more complex models and ensemble predictions, outperforming single predictions. These findings implicate the potential of optical satellite imagery as an alternative to the conventional ship-sounding and airborne LiDAR techniques for estimating bathymetry at large spatial and temporal scales for coastal clear waters. As a result, optical satellite imagery could play a crucial role in creating a global coastal bathymetry dataset. For further research, it is recommended to use area-segregated training and test splits instead of random training and test splits to account for spatial autocorrelation and prevent overly optimistic model evaluations. Furthermore, it is recommended to include imagery from more regions when training complex neural networks to increase their generalizability potential, or to locally calibrate the trained models on different regions. Finally, incorporating self-supervised learning as a solution for the lack of in situ bathymetry data would be a promising direction for future research.

Author Contributions: Conceptualization, D.H., M.G., T.D.P., W.N. and S.M.d.J.; methodology, D.H. and M.G.; software, D.H.; validation, D.H.; formal analysis, D.H.; investigation, D.H., M.G., T.D.P. and S.M.d.J.; resources, D.H.; data curation, D.H.; writing—original draft preparation, D.H. and S.M.d.J.; writing—review and editing, S.M.d.J., M.G., T.D.P., W.N. and D.H.; visualization, D.H.; supervision, M.G., T.D.P. and S.M.d.J.; project administration, D.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The tests were conducted on publicly open datasets, including Sentinel-2 data from ESA, PRISMA data from ASI and Airborne LiDAR data from NOAA. The source code of the SDB models is available at https://github.com/dehartmann/SDB_comparison (accessed on 30 December 2023). All scripts used to generate plots and results are included in the repository.

Acknowledgments: The authors wish to acknowledge the valuable help of Logambal Madhuanand of Utrecht University and of Erwin Bergsma of the CNES in Toulouse.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

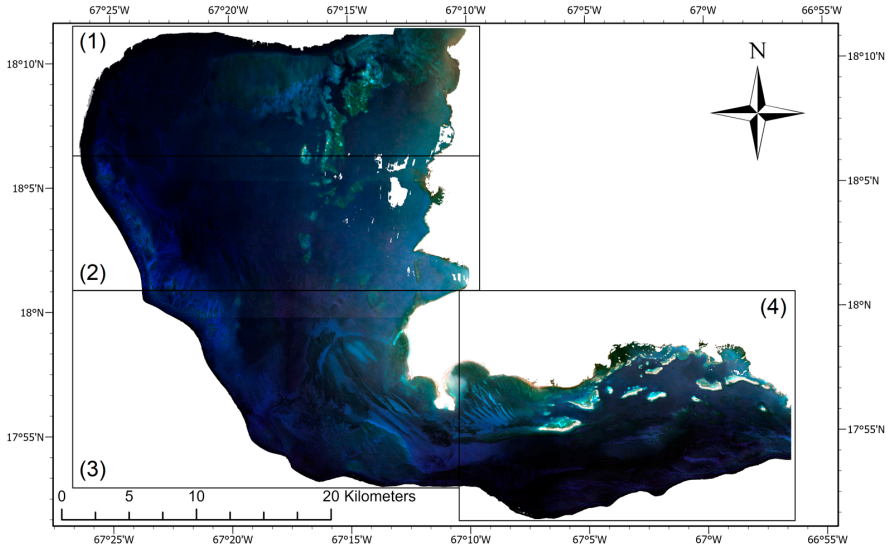


Figure A1. Map of cloud and water masked Sentinel-2 RGB mosaic with the four splits.

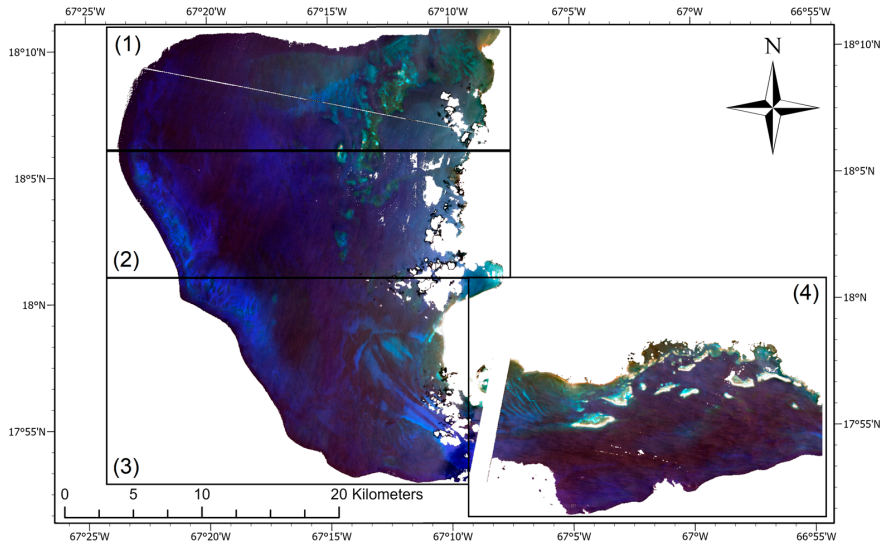


Figure A2. Map of cloud and water masked PRISMA RGB mosaic with the four splits.

Appendix B Method of Ma et al. [29] to Compare Reference Spectra with Image Pixel Spectra

Ref. [29] developed a method that utilizes a reference spectrum to distinguish the different influences of changes in water depth and bottom type on spectral curves. As the reference spectrum, the spectral signature of a water area where the depth is close to 0 m depth is selected. Next, they compared the spectral curve of each pixel, from 480 nm to 610 nm, to this reference spectrum by calculating the Pearson correlation coefficient (CC) and similarity coefficient (SC). The CC measures the similarity in the shape and magnitude between the water pixels' spectral curve and the 0 m depth reference spectrum. A higher CC value indicates greater similarity in magnitude and shape between the pixel's spectral curve and the reference spectrum, suggesting a shallower depth. The SC is the cosine of the spectral angle and measures similarity based on the spectral angles rather than the magnitude of the spectral curves. As a result, the SC is better at distinguishing bottom types, as it is not affected by changes in magnitude due to depth variations. A higher SC indicates greater similarity in spectral angles between the pixel and reference spectrum, suggesting a similar bottom type to the reference spectrum [29]. The CC and SC are calculated as follows:

$$CC(i) = \frac{\sum_{j=1}^{N} \left(R_{i,j} - \overline{R}_i\right)\left(R_{0,j} - \overline{R}_0\right)}{\sqrt{\sum_{j=1}^{N}\left(R_{i,j} - \overline{R}_i\right)^2 \sum_{j=1}^{N}\left(R_{0,j} - \overline{R}_0\right)^2}} + 1 \tag{A1}$$

$$SC(i) = \frac{\sum_{j=1}^{N} \left(R_{i,j} * R_{0,j}\right)}{\sqrt{\sum_{j=1}^{N} R_{i,j}^2}\sqrt{\sum_{j=1}^{N} R_{0,j}^2}} + 1 \tag{A2}$$

where N is the number of all bands between 480 and 610 nm, i.e., for ASI PRISMA hyperspectral imagery, the spectral bands are 16 to 30. $R_{i,j}$ is the reflectance of band j of pixel i; $R_{0,j}$ is the reference reflectance of band j. In this study, the 0 m depth reference spectrum of Van der Werff (personal communication) was employed. R_i represents the average reflectance of all bands of the pixel i and R_0 for the reference spectrum. In these two equations, a constant is added to ensure positive values.

Appendix C Deep Learning Architectures Used in This Study

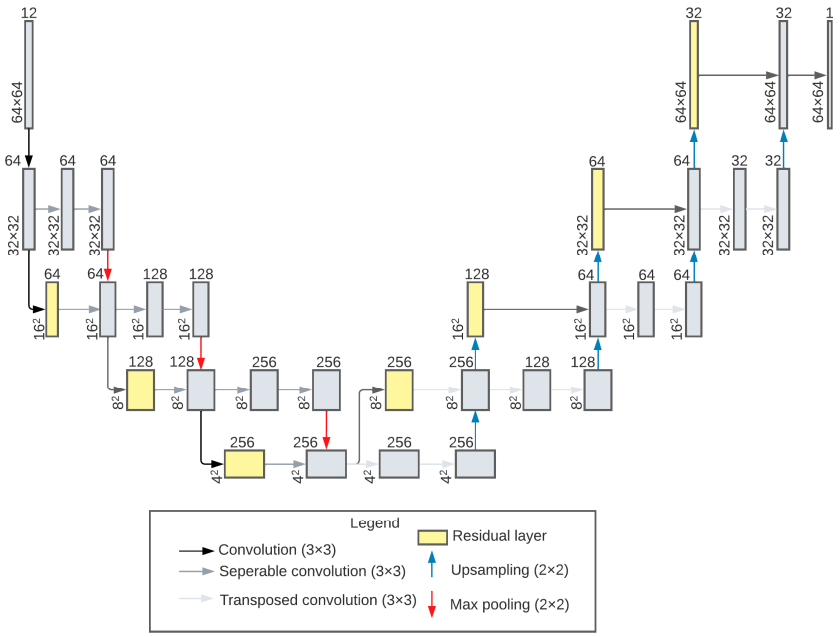


Figure A3. The architecture of the Xception U-Net used in this study.

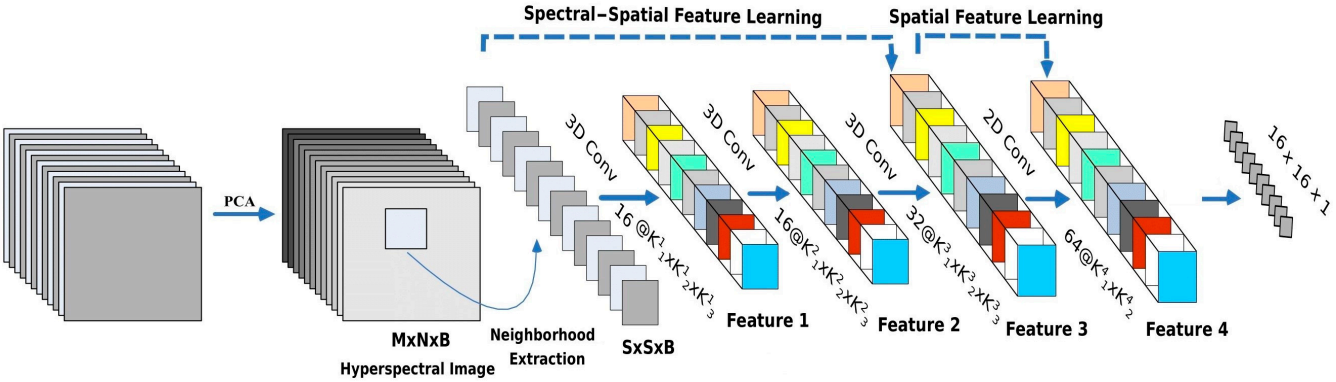
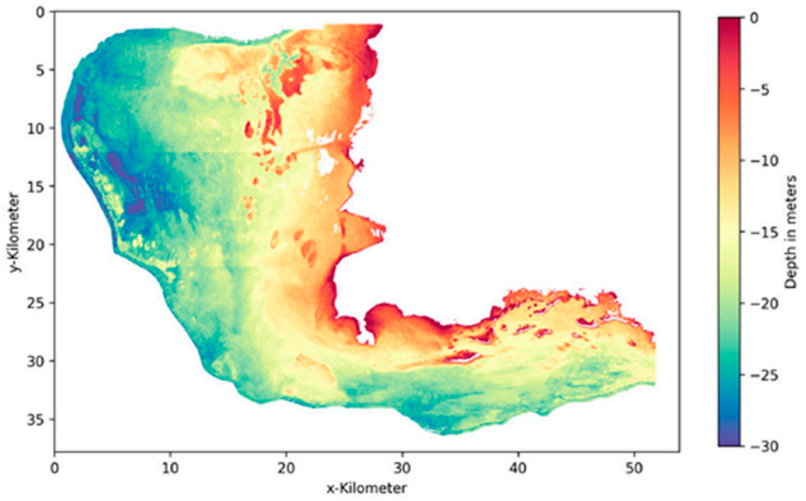
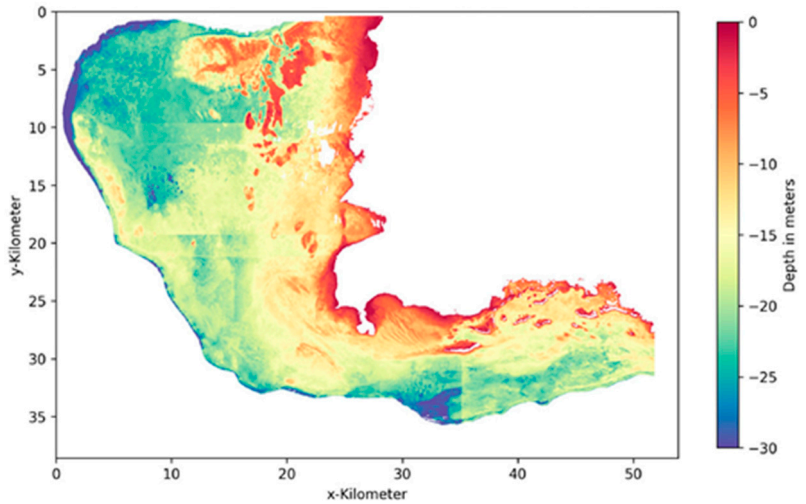


Figure A4. The architecture of the 2D-3D CNN used in this study. Figure adapted from [40].

Appendix D Prediction Maps of the Different Models

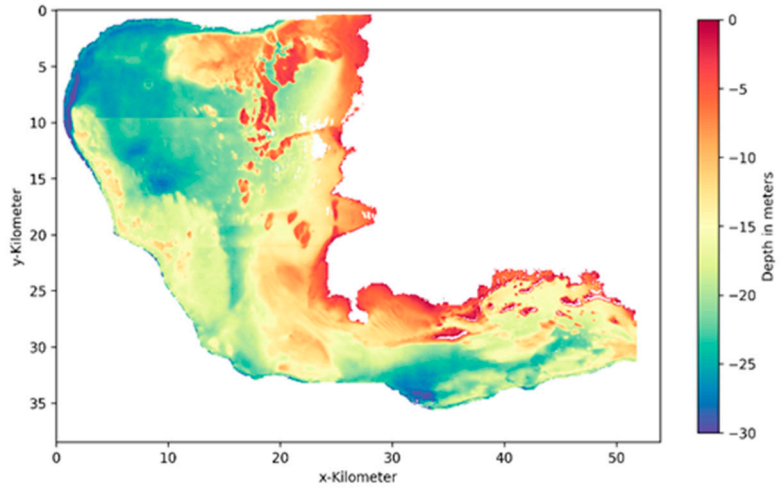


(**a**)



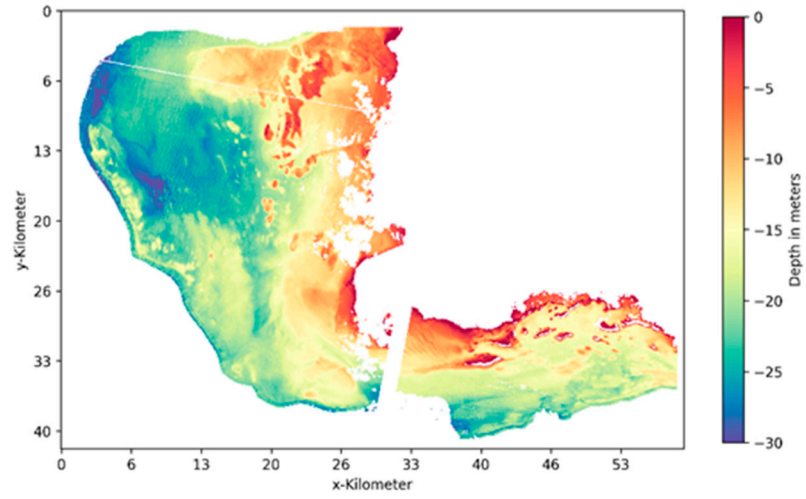
(**b**)

Figure A5. *Cont.*

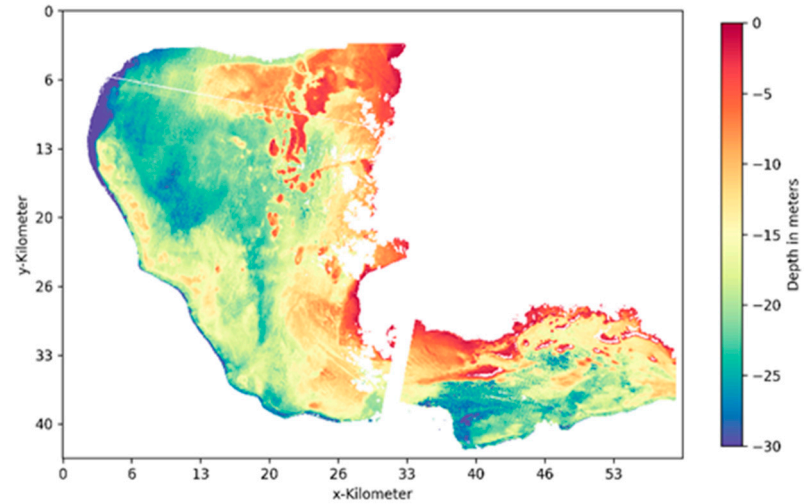


(**c**)

Figure A5. Maps of depth predictions from Stumpf (**a**), XGBoost (**b**) and U-Net (**c**) models trained with Sentinel-2 imagery. (**a**) Stumpf depth prediction map. (**b**) XGBoost depth prediction map. (**c**) U-Net depth prediction map.

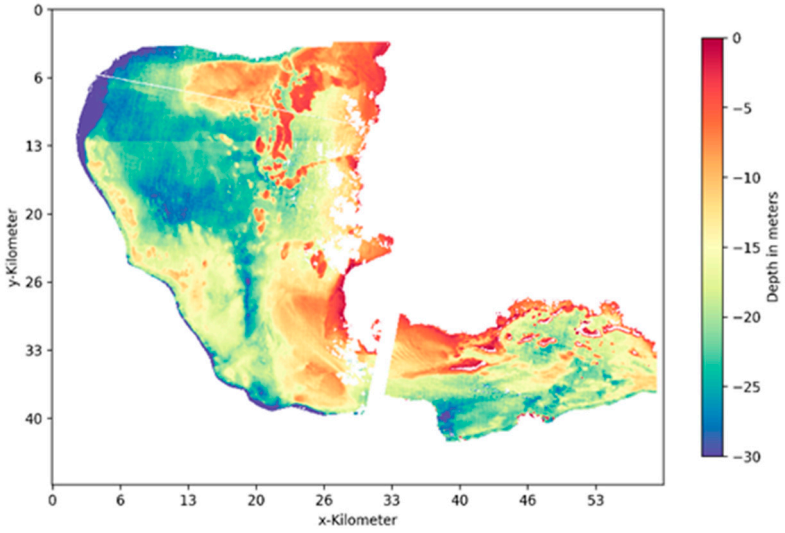


(**a**)



(**b**)

Figure A6. *Cont.*



(**c**)

Figure A6. Maps of depth predictions from Ma (**a**), XGBoost (**b**) and 2D-3D CNN (**c**) models trained with PRISMA imagery. (**a**) Ma depth prediction map. (**b**) XGBoost depth prediction map. (**c**) 2D-3D CNN depth prediction map.

Appendix E

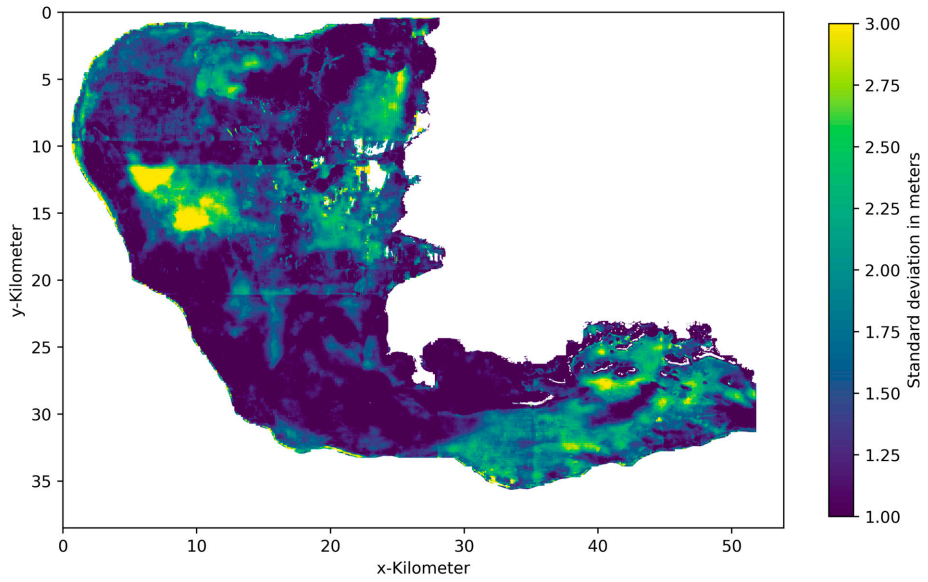


Figure A7. Standard deviation map of U-Net predictions on 49 Sentinel-2 images.

References

1. Che Hasan, R.; Ierodiaconou, D.; Laurenson, L.; Schimel, A. Integrating multibeam backscatter angular response, mosaic and bathymetry data for benthic habitat mapping. *PLoS ONE* **2014**, *9*, e97339. [CrossRef]
2. Wedding, L.M.; Friedlander, A.M.; McGranaghan, M.; Yost, R.S.; Monaco, M.E. Using bathymetric lidar to define nearshore benthic habitat complexity: Implications for management of reef fish assemblages in Hawaii. *Remote Sens. Environ.* **2008**, *112*, 4159–4165. [CrossRef]
3. NOAA. How Is Bathymetric Data Used? 2023. Available online: https://oceanservice.noaa.gov/facts/bathyuses.html (accessed on 10 April 2023).
4. Stumpf, R.P.; Holderied, K.; Sinclair, M. Determination of water depth with high-resolution satellite imagery over variable bottom types. *Limnol. Oceanogr.* **2003**, *48 Pt 2*, 547–556. [CrossRef]
5. Texas Water Development Board. Priority Areas Along the Texas Coast for Bathymetry Data Acquisition. 2023. Available online: https://recovery.texas.gov/documents/planning-studies/tiff-bathymetry-recommendtion-2.pdf (accessed on 26 June 2023).

6. Almar, R.; Bergsma, E.W.; Thoumyre, G.; Baba, M.W.; Cesbron, G.; Daly, C.; Lifermann, A. Global satellite-based coastal bathymetry from waves. *Remote Sens.* **2021**, *13*, 4628. [CrossRef]
7. GEBCO. Seabed 2030 Project. 2023. Available online: https://www.gebco.net/about_us/seabed2030_project/ (accessed on 28 January 2024).
8. Lyzenga, D.R. Passive remote sensing techniques for mapping water depth and bottom features. *Appl. Opt.* **1978**, *17*, 379–383. [CrossRef] [PubMed]
9. Dekker, A.G.; Brando, V.E.; Anstee, J.M.; Pinnel, N.; Kutser, T.; Hoogenboom, J.; Peters, S.W.M.; Pasterkamp, R.; Vos, R.J.; Olbert, C.; et al. *Imaging Spectrometry of Water. Imaging Spectrometry: Basic Principles and Prospective Applications*; Springer: Dordrecht, The Netherlands, 2001; pp. 307–359.
10. Miller, R.L.; Del Castillo, C.E.; McKee, B.A. *Remote Sensing of Coastal Aquatic Environments*; Springer: Dordrecht, The Netherlands, 2005; Volume 511.
11. 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. [CrossRef]
12. Lee, Z.; Carder, K.L.; Mobley, C.D.; Steward, R.G.; Patch, J.S. Hyperspectral remote sensing for shallow waters: 2. deriving bottom depths and water properties by optimization. *Appl. Opt.* **1999**, *38*, 3831–3843. [CrossRef] [PubMed]
13. Dekker, A.G.; Phinn, S.R.; Anstee, J.; Bissett, P.; Brando, V.E.; Casey, B.; Fearns, P.; Hedley, J.; Klonowski, W.; Lee, Z.P.; et al. Intercomparison of shallow water bathymetry, hydro-optics, and benthos mapping techniques in Australian and Caribbean coastal environments. *Limnol. Oceanogr. Methods* **2011**, *9*, 396–425. [CrossRef]
14. Jawak, S.D.; Vadlamani, S.S.; Luis, A.J. A synoptic review on deriving bathymetry information using remote sensing technologies: Models, methods and comparisons. *Adv. Remote Sens.* **2015**, *4*, 147–162. [CrossRef]
15. Ashphaq, M.; Srivastava, P.K.; Mitra, D. Review of near-shore satellite derived bathymetry: Classification and account of five decades of coastal bathymetry research. *J. Ocean Eng. Sci.* **2021**, *6*, 340–359. [CrossRef]
16. McIntyre, M.L.; Naar, D.F.; Carder, K.L.; Donahue, B.T.; Mallinson, D.J. Coastal bathymetry from hyperspectral remote sensing data: Comparisons with high resolution multibeam bathymetry. *Mar. Geophys. Res.* **2006**, *27*, 129–136. [CrossRef]
17. Sagawa, T.; Yamashita, Y.; Okumura, T.; Yamanokuchi, T. Satellite derived bathymetry using machine learning and multi-temporal satellite images. *Remote Sens.* **2019**, *11*, 1155. [CrossRef]
18. Misra, A.; Ramakrishnan, B. Assessment of coastal geomorphological changes using multi-temporal satellite-derived bathymetry. *Cont. Shelf Res.* **2020**, *207*, 104213. [CrossRef]
19. Najar, M.A.; Benshila, R.; Bennioui, Y.E.; Thoumyre, G.; Almar, R.; Bergsma, E.W.; Wilson, D.G. Coastal bathymetry estimation from sentinel-2 satellite imagery: Comparing deep learning and physics-based approaches. *Remote Sens.* **2022**, *14*, 1196. [CrossRef]
20. Lumban-Gaol, Y.; Ohori, K.A.; Peters, R. Extracting coastal water depths from multi-temporal Sentinel-2 images using convolutional neural networks. *Mar. Geod.* **2022**, *45*, 615–644. [CrossRef]
21. Bauer, L.J.; Edwards, K.; Kendall, M.S.; Roberson, K.K.; Tormey, S.; Battista, T.A. Shallow-Water Benthic Habitats of Southwest Puerto Rico. 2012. Available online: https://repository.library.noaa.gov/view/noaa/875 (accessed on 10 March 2023).
22. Glynn, P. Ecology of a Caribbean coral reef. the porites reef-flat biotope: Part i. meteorology and hydrography. *Mar. Biol.* **1973**, *20*, 297–318. [CrossRef]
23. Coastal Data Information Program. 181—Rincon, Puerto Rico, Station Summary. 2023. Available online: http://cdip.ucsd.edu/themes/cdip?zoom=auto&tz=UTC&llfmt=dm&numcolorbands=10&palette=cdipclassic&high=6.096&r=999&un=1&pb=1&d2=p9&u2=s:181:st:1:v:summary (accessed on 30 June 2023).
24. Copernicus. Copernicus Open Data Space. 2023. Available online: https://dataspace.copernicus.eu/ (accessed on 17 December 2023).
25. ASI. PRISMA User Manual Issue 1.3. 2023. Available online: http://prisma.asi.it/missionselect/docs.php (accessed on 17 December 2023).
26. Battista, T. *Water Depth and Reflectance Collected from LADS MK II Airborne System in Caribbean Sea and Puerto Rico from 2006-04-07 to 2006-05-15*; Dataset (NCEI Accession 0153360); NOAA National Centers for Environmental Information: Asheville, NC, USA, 2006.
27. NOAA. NOAA Tides & Currents. 2016. Available online: https://tidesandcurrents.noaa.gov/ (accessed on 13 March 2023).
28. Kutser, T.; Vahtmäe, E.; Praks, J. A sun glint correction method for hyperspectral imagery containing areas with non-negligible water leaving NIR signal. *Remote Sens. Environ.* **2009**, *113*, 2267–2274. [CrossRef]
29. Ma, S.; Tao, Z.; Yang, X.; Yu, Y.; Zhou, X.; Li, Z. Bathymetry retrieval from hyperspectral remote sensing data in optical-shallow water. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 1205–1212. [CrossRef]
30. Holden, H.; LeDrew, E. Measuring and modeling water column effects on hyperspectral reflectance in a coral reef environment. *Remote Sens. Environ.* **2002**, *81*, 300–308. [CrossRef]

31. Zamani Joharestani, M.; Cao, C.; Ni, X.; Bashir, B.; Talebiesfandarani, S. Pm2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. *Atmosphere* **2019**, *10*, 373. [CrossRef]
32. Georganos, S.; Grippa, T.; Vanhuysse, S.; Lennert, M.; Shimoni, M.; Wolff, E. Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 607–611. [CrossRef]
33. Susa, T. Satellite derived bathymetry with sentinel-2 imagery: Comparing traditional techniques with advanced methods and machine learning ensemble models. *Mar. Geod.* **2022**, *45*, 435–461. [CrossRef]
34. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
35. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; Available online: http://www.deeplearningbook.org (accessed on 12 December 2024).
36. Ronneberger 2016, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–Miccai 2015: 18th International Conference 2015, Munich, Germany, October 5–9, 2015, Proceedings, Part iii 18*; Springer: Cham, Switzerland, 2015; pp. 234–241.
37. Mandlburger, G.; Kölle, M.; Nübel, H.; Soergel, U. Bathynet: A deep neural network for water depth mapping from multispectral aerial images. *PFG–J. Photogramm. Remote Sens. Geoinf. Sci.* **2021**, *89*, 71–89. [CrossRef]
38. Nicolas, K.M.; Drumetz, L.; Lef'evre, S.; Tiede, D.; Bajjouk, T.; Burnel, J.-C. Deep learning–based bathymetry mapping from multispectral satellite data around europa island. In *European Spatial Data for Coastal and Marine Remote Sensing: Proceedings of International Conference Eucomare 2022-Saint-Malo 2022, France*; Springer: Cham, Switzerland, 2022; pp. 97–111.
39. Chollet. Image Segmentation with a u-Net-like Architecture. Available online: https://keras.io/examples/vision/oxford_pets_image_segmentation/ (accessed on 20 March 2023).
40. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. Hybrids: Exploring 3-d–2-d CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 277–281. [CrossRef]
41. Meyer, H.; Reudenbach, C.; Wöllauer, S.; Nauss, T. Importance of spatial predictor variable selection in machine learning applications–moving from data reproduction to spatial prediction. *Ecol. Model.* **2019**, *411*, 108815. [CrossRef]
42. Cahalane, C.; Magee, A.; Monteys, X.; Casal, G.; Hanafin, J.; Harris, P. A comparison of Landsat-8, Rapideye and Pleiades products for improving empirical predictions of satellite-derived bathymetry. *Remote Sens. Environ.* **2019**, *233*, 111414. [CrossRef] [PubMed]
43. Monteys, X.; Harris, P.; Caloca, S.; Cahalane, C. Spatial prediction of coastal bathymetry based on multispectral satellite imagery and multibeam data. *Remote Sens.* **2015**, *7*, 13782–13806. [CrossRef]
44. Kapoor, S.; Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* **2023**, *4*, 100804. [CrossRef]
45. Bouthillier, X.; Delaunay, P.; Bronzi, M.; Trofimov, A.; Nichyporuk, B.; Szeto, J.; Mohammadi Sepahvand, N.; Raff, E.; Madan, K.; Voleti, V.; et al. Accounting for variance in machine learning benchmarks. *Proc. Mach. Learn. Syst.* **2021**, *3*, 747–769.
46. Duan, Z.; Chu, S.; Cheng, L.; Ji, C.; Li, M.; Shen, W. Satellite-derived bathymetry using Landsat-8 and Sentinel-2a images: Assessment of atmospheric correction algorithms and depth derivation models in shallow waters. *Opt. Express* **2022**, *30*, 3238–3261. [CrossRef]
47. Pacheco, A.; Horta, J.; Loureiro, C.; Ferreira, O. Retrieval of nearshore bathymetry from Landsat 8 images: A tool for coastal monitoring in shallow waters. *Remote Sens. Environ.* **2015**, *159*, 102–116. [CrossRef]
48. Lyons, M.; Phinn, S.; Roelfsema, C. Integrating Quickbird multi-spectral satellite and field data: Mapping bathymetry, seagrass cover, seagrass species and change in moreton bay, australia in 2004 and 2007. *Remote Sens.* **2011**, *3*, 42–64. [CrossRef]
49. Freire, R.R. Evaluating Satellite Derived Bathymetry in Regard to Total Propagated Uncertainty, Multitemporal Change Detection, and Multiple Non-Linear Estimation. Ph.D. Thesis, University of New Hampshire, Durham, NH, USA, 2017.
50. Lee 2017, Z.P.; Casey, B.; Parsons, R.; Goode, W.; Weidemann, A.; Arnone, R. Bathymetry of shallow coastal regions derived from space-borne hyperspectral sensor. In Proceedings of the Oceans 2005 MTS/IEEE, Washington, DC, USA, 17–23 September 2005; pp. 2160–2170.
51. Alevizos, E.; Le Bas, T.; Alexakis, D.D. Assessment of PRISMA level-2 hyperspectral imagery for large scale satellite-derived bathymetry retrieval. *Mar. Geod.* **2022**, *45*, 251–273. [CrossRef]
52. Ma, Y.; Xu, N.; Liu, Z.; Yang, B.; Yang, F.; Wang, X.H.; Li, S. Satellite-derived bathymetry using the icesat-2 lidar and sentinel-2 imagery datasets. *Remote Sens. Environ.* **2020**, *250*, 112047. [CrossRef]
53. Zhong, J.; Sun, J.; Lai, Z.; Song, Y. Nearshore bathymetry from icesat-2 lidar and sentinel-2 imagery datasets using deep learning approach. *Remote Sens.* **2022**, *14*, 4229. [CrossRef]
54. Awais, M.; Naseer, M.; Khan, S.; Anwer, R.M.; Cholakkal, H.; Shah, M.; Yang, M.H.; Khan, F.S. Foundational models defining a new era in vision: A survey and outlook. *arXiv* **2023**, arXiv:2307.13721.
55. Vinayaraj 2023, P.; Raghavan, V.; Masumoto, S. Satellite-derived bathymetry using adaptive geographically weighted regression model. *Mar. Geod.* **2016**, *39*, 458–478. [CrossRef]

56. Caballero, I.; Stumpf, R.P. Confronting turbidity, the major challenge for satellite-derived coastal bathymetry. *Sci. Total Environ.* **2023**, *870*, 161898. [CrossRef]
57. Lacaux, J.; Tourre, Y.; Vignolles, C.; Ndione, J.; Lafaye, M. Classification of ponds from high- spatial resolution remote sensing: Application to rift valley fever epidemics in senegal. *Remote Sens. Environ.* **2007**, *106*, 66–74. [CrossRef]