SATELLITE-DERIVED BATHYMETRY USING CONVOLUTIONAL NEURAL NETWORKS AND MULTISPECTRAL SENTINEL-2 IMAGES

Y. A. Lumban-Gaol[1,2,*], K. A. Ohori[1], R. Y. Peters[1]

[1] Faculty of Architecture and the Built Environment, Delft University of Technology, The Netherlands -
yustisiardhitasari@gmail.com, (K.Ohori, R.Y.Peters)@tudelft.nl
[2] Geospatial Information Agency (BIG), Jl. Raya Jakarta-Bogor Cibinong, Indonesia

Commission III, WG III/4

KEY WORDS: Bathymetry, Shallow water, SDB, CNN, Sentinel-2, Multispectral images.

ABSTRACT:

Satellite-Derived Bathymetry (SDB) has been used in many applications related to coastal management. SDB can efficiently fill data gaps obtained from traditional measurements with echo sounding. However, it still requires numerous training data, which is not available in many areas. Furthermore, the accuracy problem still arises considering the linear model could not address the non-relationship between reflectance and depth due to bottom variations and noise. Convolutional Neural Networks (CNN) offers the ability to capture the connection between neighbouring pixels and the non-linear relationship. These CNN characteristics make it compelling to be used for shallow water depth extraction. We investigate the accuracy of different architectures using different window sizes and band combinations. We use Sentinel-2 Level 2A images to provide reflectance values, and Lidar and Multi Beam Echo Sounder (MBES) datasets are used as depth references to train and test the model. A set of Sentinel-2 and in-situ depth sub-image pairs are extracted to perform CNN training. The model is compared to the linear transform and applied to two other study areas. Resulting accuracy ranges from 1.3 m to 1.94 m, and the coefficient of determination reaches 0.94. The SDB model generated using a window size of 9x9 indicates compatibility with the reference depths, especially at areas deeper than 15 m. The addition of both short wave infrared bands to the four visible bands in training improves the overall accuracy of SDB. The implementation of the pre-trained model to other study areas provides similar results depending on the water conditions.

1. INTRODUCTION

Bathymetry data in shallow water areas is essential for coastal management purposes. A number of bathymetry survey techniques exist, such as single or multibeam echo sounder and Lidar bathymetry. However, these methods still leave data gaps due to various reasons. Single beam echo sounding (SBES) data are limited to the sounding lines, where gaps occur in between the sounding lines. Multibeam echo sounding (MBES) measures denser depth values than SBES, but it still cannot reach areas with very shallow depth, such as along the coastline or coral reefs areas. Lidar bathymetry is capable of measuring these areas and producing higher resolution data than other methods, but the survey plan must consider numerous factors to minimize possible gaps in the final result Quadros (2016).

Optical remote sensing images are a promising alternative data source to extract water depth information. Satellite-Derived Bathymetry (SDB) is a way to model water depth in shallow water areas using multispectral imagery. The advent of SDB and its developments have been known to fill data gaps that occur with echo sounding. Furthermore, SDB is a low-cost technique since remote sensing images are used instead of field surveys. It also has few environmental impacts and risks to personnel or equipment since the model can be derived without directly accessing the shallow water area.

SDB works based on the principle of light attenuation in the water column. The light passes through the water and hits the water bed before it arrives at the sensor. The bottom reflectance can be retrieved from satellite images after removing the

atmospheric scattering, surface reflection, and in-water scattering. The bottom reflectance is then used to extract water depth values.

There are two approaches of SDB: analytical and empirical. The analytical method is based on modelling the behaviour of light penetration in water. It requires several optical properties of shallow water region, such as the attenuation coefficient, backscatter coefficient, coefficient of suspended and dissolved materials, and bottom reflectance Spitzer and Dirks (1986); Gao (2009). Since it needs many parameters and assumptions, many studies used empirical or mixed-method to perform SDB.

The empirical method uses the relationship between reflected radiation and in-situ depth empirically without considering light transmission in water. Two well-known methods exist, which are linear transform Lyzenga (1985); Lyzenga et al. (2006) and ratio transform Stumpf et al. (2003). Much research on SDB uses these methods Hamylton et al. (2015); Kabiri (2017); Traganos et al. (2018). Some studies modified Lyzenga85's method to improve the accuracy Lyzenga et al. (2006); Kanno et al. (2012); Kanno and Tanaka (2012); Kanno et al. (2013), or used a different regression technique, such as Geographically Weighted Regression Vinayaraj et al. (2016). These methods rely on a linear relationship between water reflectance and depth, while it may not be entirely linear due to bottom types variation and noise. In the last five years, studies on SDB have started to use machine learning approaches to address the non-linear relationship, such as Random Forest Manessa et al. (2016); Sagawa et al. (2019); Tonion et al. (2020) and Support Vector Machines Misra et al. (2018); Manessa et al. (2018); Tonion et al. (2020).

* Corresponding author

The methods above use per pixel reflectance-depth pair, while the spatial correlation may also influence SDB computations assuming that adjacent pixels have a relationship towards depth. That being said, it is necessary to consider the spatial and non-linear relationship in building SDB computations. Existing methods do not address both factors at the same time. Thus, this study used another machine learning method, Convolutional Neural Networks (CNN), to estimate shallow water depth. CNN has been widely used for classification purposes. In this study, CNN was used to solve the regression problem. CNN's ability to capture the connection between neighbouring pixels and the non-linear relationship in the training process makes it interesting to be used for SDB. We try various configurations of the CNN architecture to investigate the accuracy of CNN for SDB. We verify the CNN model by comparing the accuracy with the linear transform method. Furthermore, we implemented the CNN model in different study areas.
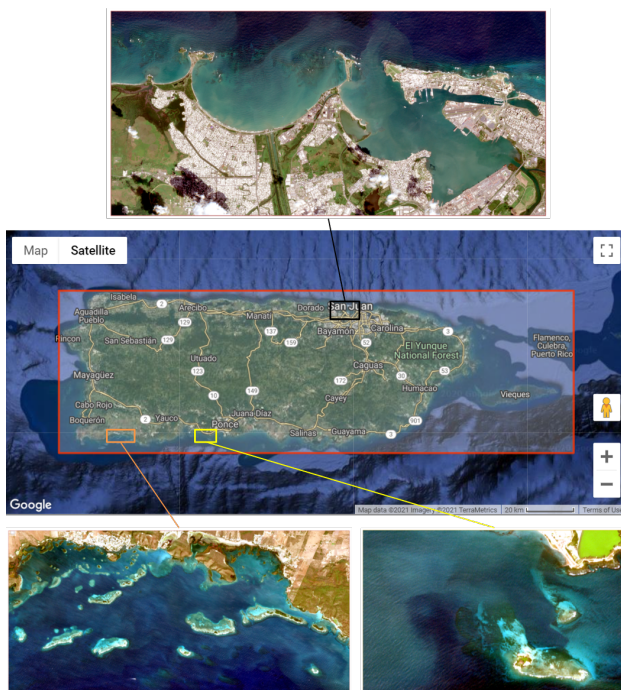
2. MATERIALS AND METHOD

Figure 1. The area of study on several locations around Puerto Rico main island, with the visualization from Sentinel-2 images. Bottom-right, bottom-left, and top are Ponce, Southwest, and San Juan respectively.

2.1 Study area

The study was conducted in three locations around Puerto Rico main island as shown in Figure 1. Two sites are located in the south part and one in the north, more precisely Ponce, the Southwest part of the island, and San Juan with an area of 12 km^2, 79 km^2, and 100 km^2 respectively. These sites were chosen considering the availability of in-situ bathymetry, satellite images, as well as water conditions. Ponce and Southwest Puerto Rico are considered to have low to moderate turbidity; meanwhile San Juan is categorized as having high turbidity based on its natural colour composite from the image. This study uses Ponce as the main site to compare CNN model architectures. Then, SDB model accuracies are compared between all sites.
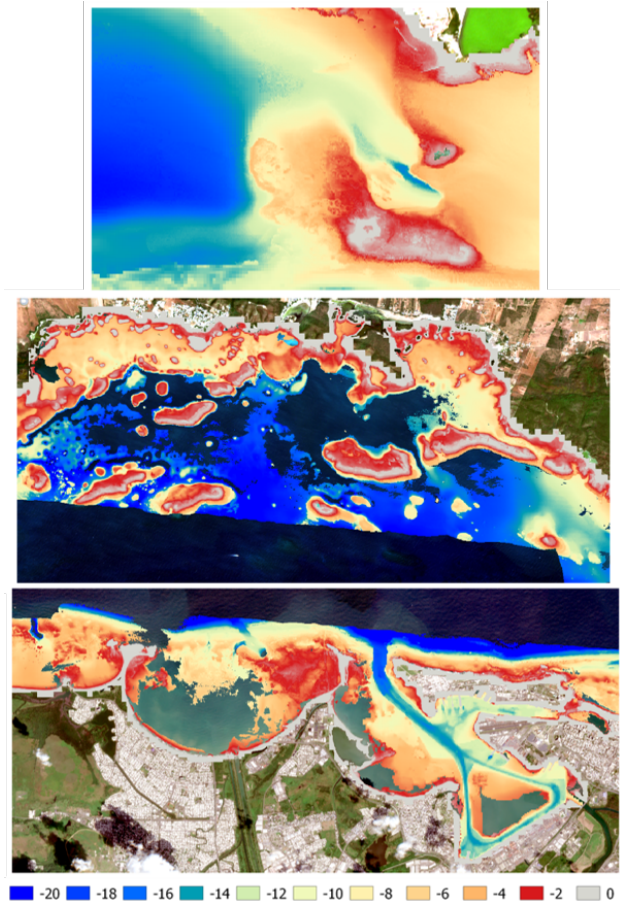

Figure 2. In-situ depth acquired from Lidar or MBES. From top to bottom: Ponce, Southwest, San Juan.

2.2 Data

The study used Sentinel-2 Level-2A images and bathymetry data measured using MBES and Lidar. These data sets are open and accessible through the Copernicus[1] and the National Oceanic and Atmospheric Administration (NOAA)[2] open-access websites. In addition, NOAA also provides tides data that are used for depth correction[3].

Sentinel-2 Level-2A images are corrected for radiometric, geometric, and atmospheric distortion, and provide surface reflectance pixels. It consists of 13 spectral bands with spatial resolutions between 10, 20, and 60 m. This study uses various combinations of nine spectral bands as listed in Table 1. Since clouds affect SDB calculations, images used in this paper were filtered on the cloudy percentage level, where an image with the least clouds was selected for each site. The filtering process results in an image acquired on 3 January 2020 for use in the Ponce and Southwest area and 2 February 2020 for San Juan. The visualization of natural colour composite, red-green-blue (RGB), on each location is shown in Figure 1.

In-situ depth data sets were taken in 2018. Ponce and San Juan have both MBES and Lidar data available, ranging from 0 to roughly 20 m in Ponce and 30 m in San Juan. The spatial resolution of the two data is different, specifically 1 m for Lidar

[1] https://scihub.copernicus.eu/
[2] https://coast.noaa.gov/digitalcoast/data/home.html
[3] https://tidesandcurrents.noaa.gov/

Table 1. Sentinel-2 multispectral bands used in this study.

Spectral band	Wavelength (nm)	Resolution (m)
Band 2 - Blue	490	10
Band 3 - Green	560	10
Band 4 - Red	665	10
Band 5 - Vegetation Red Edge (VRE1)	705	20
Band 6 - Vegetation Red Edge (VRE2)	740	20
Band 7 - Vegetation Red Edge (VRE3)	783	20
Band 8 - NIR	842	10
Band 11 - Short Wave Infrared (SWIR1)	1,610	20
Band 12 - Short Ware Infrared (SWIR2)	2,190	20

and 32 m for MBES. Both data complement each other so that our sites have a complete set of depths in shallow water areas, especially in the Ponce area, as shown in Figure 2. Meanwhile, only Lidar is available in the Southwest, having a similar depth range to Ponce. All measurements are compliant with the IHO Standards for Hydrographic Surveys (S-44) 5th Edition. The measurements of Lidar pass uncertainty standards for special and 1a Order.

2.3 Method

We try different CNN hyper-parameters and architectures in order to find the optimum CNN architecture for extracting shallow water depth. The linear transform method was used in comparison to assess the accuracy of the SDB model generated from CNN. Additionally, the CNN architecture was implemented in a different number of channels and different locations. Here we describe the workflow of this study.
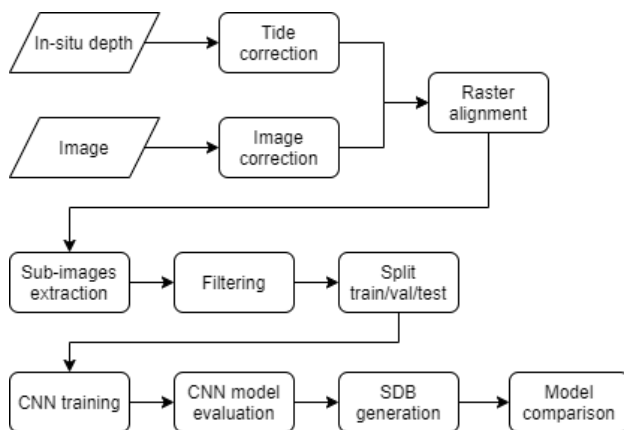

Figure 3. General workflow of this study.

As shown in Figure 3, a number of sequential data pre-processing steps need to be implemented. The spatial resolution for all data sets was resampled to 10 m, following the satellite image product. To synchronize disparate times between in-situ and satellite images, a tidal correction was applied to the MBES and Lidar bathymetry data depending on the satellite image sensing time.

The satellite images were corrected using average deep water pixels to extract the bottom reflectance Lyzenga et al. (2006). This method assumes that there is no bottom reflectance present in deep water considering the attenuation of light through water.

Deep water is usually represented as darker pixels. This paper selected this area manually based on image visualization. Then, the corrected value is given by Green et al. (2000):

$$X(\lambda)_i = log(L(\lambda)_i - mean(L_\infty(\lambda)_i)) \qquad (1)$$

where $X(\lambda)_i$ is transformed radiance of band λ at ith point, $L(\lambda)_i$ and $L_\infty(\lambda)_i$ are the pixel values in shallow and deep water respectively.

Sub-images for training and testing   Sub-images for SDB model generation
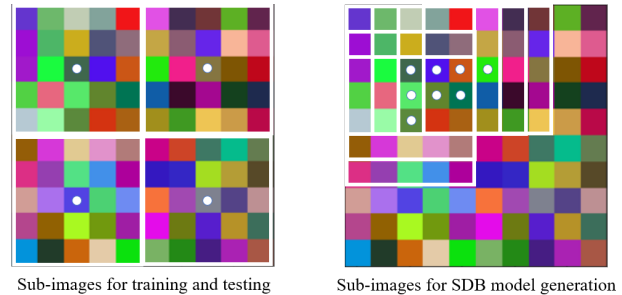
Figure 4. An illustration of sub-image extraction for training data (left) and SDB model generation (right).

Subsequently, we apply raster alignment to depth and multispectral images to avoid shifting between pixels. Then, sub-images with a particular window size were extracted from the data. Figure 4 provides an example of the sub-image extraction for a window size of five and stride of five, note that this paper use stride of three. As we can see, the amount of training data was reduced by striding the image pixels. Three different window sizes, i.e. 5×5, 7×7, and 9×9, are implemented during CNN architecture building process. Furthermore, Lidar bathymetry still captures the elevation on the land, so we filter the elevation to include only areas up to 2 m above sea level. At the end of the pre-processing stage, the sub-images were separated into 80% for training, including validation, and 20% for testing.

CNN architecture. The main difference in CNN for classification and regression is in the output activation layer function. We use a linear activation function in the last layer instead of softmax or other activations that are usually used in the classification task. We use ReLU as the activation function for other layers to address the non-linear trend between reflectance and water depth. We also implement batch normalization in our architecture to reduce overfitting. As for the hyper-parameters, we tried a different number of configurations. They include learning rate, batch size, and dropout rate.

We evaluate the training results by monitoring the validation accuracy on each epoch, in this case Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Based on our experiment, the validation accuracy in the training process was optimal with a learning rate of 0.0001, batch size of 512, and a dropout rate of 0.3. Thus, these parameters were used for further study in this paper.

To build up a complete architecture, we try different numbers of convolutional layers, kernel sizes, and pooling layers. Table 2 provides a list of the configurations that were implemented. Our experiments suggest that the training cannot converge with only a single convolutional layer and not yielded considerable validation accuracy. Thus, experiments with a single convolutional layer were excluded from this paper. We limit the number of convolutional layers and kernel size to three since our maximum window size is 9x9. The use of the pooling layer intends

to reduce the processing time without adding parameters during the training process. We also examined whether it can improve the quality of the result.

Table 2. Different CNN architectures used in this study.

Architecture	Conv2D	Kernel size	Polling layer
CNN1	2x	2x2	No
CNN2	2x	2x2	Yes
CNN3	3x	3x3	No
CNN4	3x	3x3	Yes

Up to this stage, we used the Ponce site with only three channels: red, green, and blue. Furthermore, we tried different combinations of bands, as shown in Table 3, to see whether they can improve our model or not.

Table 3. Band combinations.

Channels	Number of channels	Bands
RGB	3	Red, Green, Blue
RGBN	4	Red, Green, Blue, NIR
RGBNSS	6	Red, Green, Blue, NIR, SWIR1, SWIR2
All bands	9	Red, Green, Blue, NIR, SWIR1, SWIR2, VRE1, VRE2, VRE3

Model verification. The assessment of the SDB model from CNN was carried out by comparing the predicted and in-situ depth from testing data. We used a standard accuracy assessment by estimating the Root Mean Square Error (RMSE) to measure the error and the coefficient of determination (R^2) to measure the variance of the model and ground truth values. These metrics were compared to the linear transform method to verify the SDB-CNN model. Furthermore, we implemented the CNN with the same architecture to generate the SDB model of other areas of interest.

3. RESULTS AND DISCUSSIONS

Using 10,820 sub-images in the training process, Table 4 depicts the accuracy of different CNN architectures for 2,705 testing sub-images. The RMSE ranges from 1.48 m to 1.94 m. The use of the pooling layer in CNN2 and CNN4 speeds up the training execution time but did not improve the accuracy. A deeper network shows small improvements on the result, but the validation accuracy is smoother for three than two network layers.

Table 4. Accuracy assessment on different CNN architectures.

Models	5x5		7x7		9x9	
	RMSE	R2	RMSE	R2	RMSE	R2
CNN1	1.59	0.91	1.58	0.92	1.55	0.93
CNN2	1.94	0.89	1.63	0.91	1.55	0.93
CNN3			1.53	0.92	1.48	0.94
CNN4					1.64	0.90

Based on the metric accuracy, a larger window size increases the result's accuracy by approximately 10 cm. If we compare the predicted depth in Figure 5 and the in-situ depth in Figure 2, we can see that larger window sizes improve the result, especially in deeper depth areas. Figure 6 shows the predicted depth over the reference depth for overall depth within the same test data sets. It illustrates that the CNN model tends to fit the data. In comparison, with more training data, the linear transform method results in an overall accuracy of 4.05 m with R^2 of 0.41. For comparison, a linear transform using the same amount

of in-situ data as CNN yielded a lower accuracy, roughly 5.30 m with R^2 of 0.13. As we can see in Figure 6, the linear transform method cannot fit the values as well, mainly in the very shallow, between 0 to 2 m, or in depths deeper than 10 m. It is likely that the linear transform method cannot capture the non-linearity due to bottom reflectance variation in this area since it predicts very shallow depths to be deeper and deep areas to be shallower.

However, we can also see that some significant errors are still present in Figure 5. For example, there is a significant variation in depth in the east part that is not visible in the reference data. The difference between prediction and reference in this area is up to 10 m. However, there are no peculiar patterns of reflectance and depth distribution in the area. The errors might reflect the difficulty of CNN in predicting a particular depth range. Table 5 points out that the accuracy tends to increase as the depth increases since the bottom reflectance component becomes less accurate with increasing depth.

Table 5. SDB based CNN model accuracy over different depth ranges.

Depth ranges (m)	RMSE (m)
0-5	0.61
5-10	1.53
10-15	2.07
15-20	1.22

Furthermore, Table 6 depicts the impact of different band combinations on the accuracy of depth calculations. The accuracy was increased again with larger window sizes, especially for a window size of 9×9 with six bands. As we can see in Figure 7, a different number of channels affects the SDB model. If we focus on the large error present in the eastern part, we can see that the model can obtain a better prediction as we increase the number of bands to train. Additionally, Figure 8 illustrates the accuracy for depts ranging from 0m to 20m. The results demonstrate that the combination of RGBNSS outperforms other configurations, except in the depth range between 1 m and 2 m, and depths deeper than 17 m. It can be due to the lack of sufficient training data or because too many bands may cause a bias in a certain depth range of the CNN model. In general, the combination of RGB, RGBN, and RGBNSS have similar accuracies up to 7 m. Deeper than that, SWIR bands contribute to keeping the accuracy lower than others. Meanwhile, the use of all bands did not indicate any particular improvement compared to other combinations.

Table 6. Accuracy assessment of SDB using CNN with different number of channels.

Channels	5x5		7x7		9x9	
	RMSE	R2	RMSE	R2	RMSE	R2
RGB	1.59	0.91	1.53	0.92	1.48	0.94
RGBN	1.64	0.91	1.40	0.93	1.37	0.94
RGBNSS	1.63	0.91	1.53	0.91	1.31	0.94
All bands	1.68	0.90	1.55	0.91	1.45	0.94

Two cross profiles were generated in order to assess SDB models over in-situ measurements. As shown in Figure 9, profile section a-b indicates that the SDB produced a good fit to the measurements in depths shallower than 14 m. In the deeper ranges, SDB tends to be more distracted. Meanwhile, profile section b-c describes the trend over the area shallower than 14m with land present in between. The cross section illustrates a good fit in the SDB model as it can detect the coastline area quite precisely.
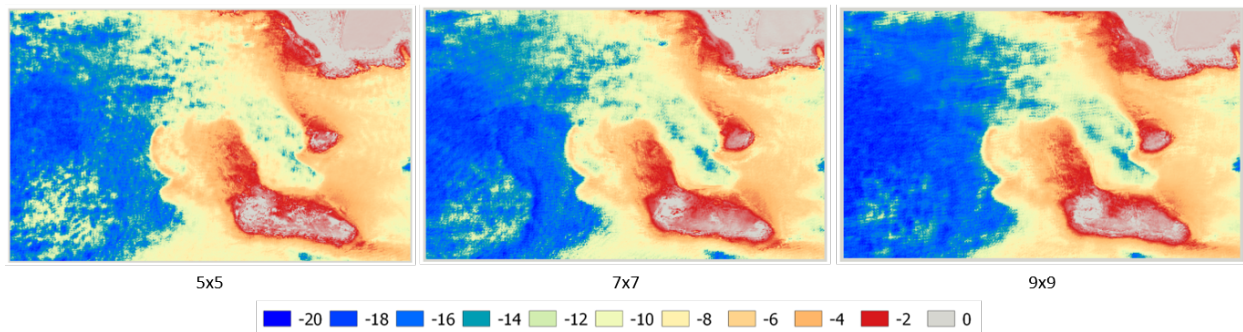
Figure 5. Visualization of SDB result generated with different window sizes.
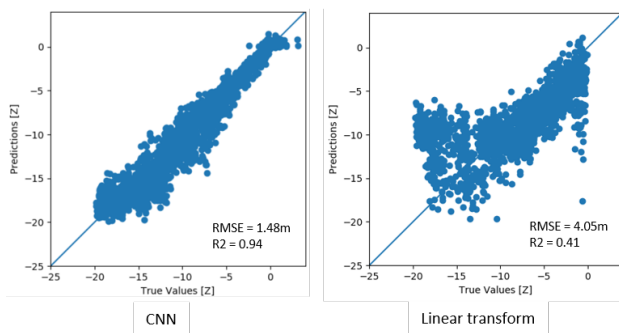

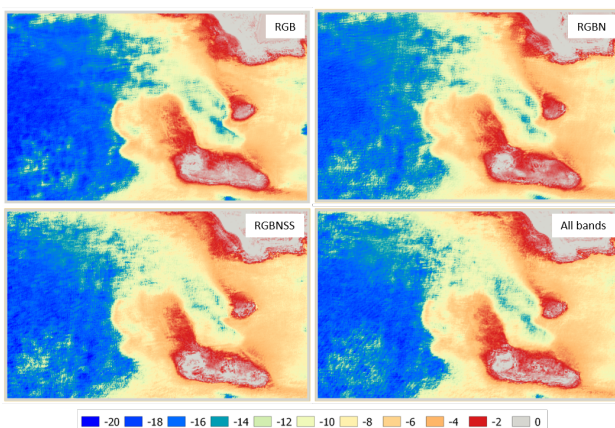Figure 6. Depth prediction vs ground truth.

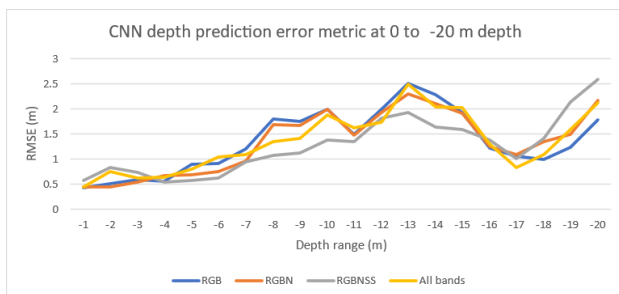Figure 7. SDB model generated using different number of channels.

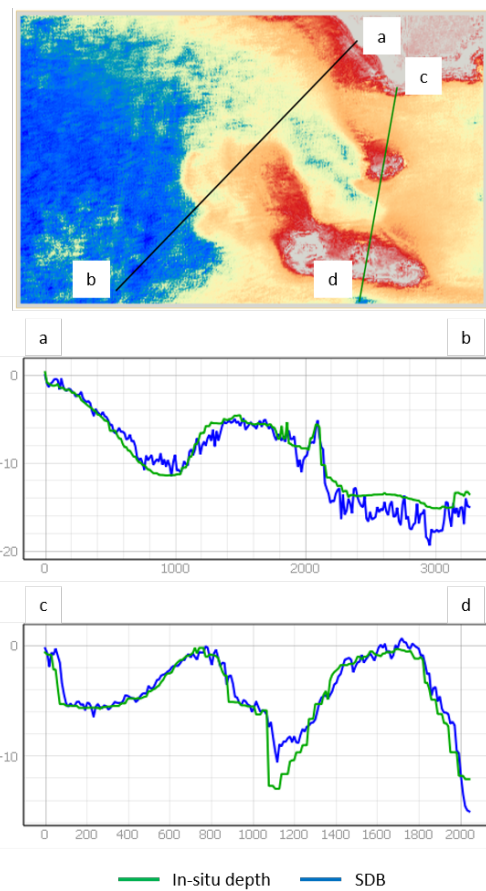Figure 8. CNN accuracy metric on different number of channels per depth range 1 m.

Figure 9. Cross profile showing comparison between SDB model from CNN and the in-situ measurements.
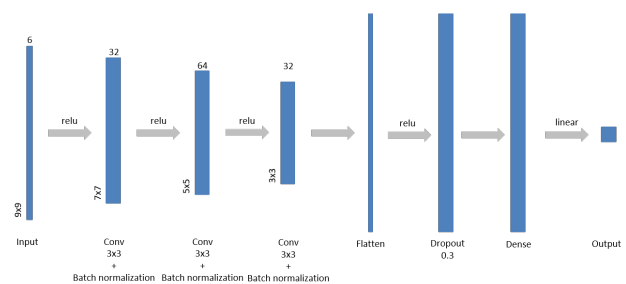

Figure 10. Final CNN architecture for SDB.

Based on the experiments on several CNN architectures, we implemented the CNN architecture shown in Figure 10 in other sites. We used CNN4 with a 9x9 window size and RGBNSS
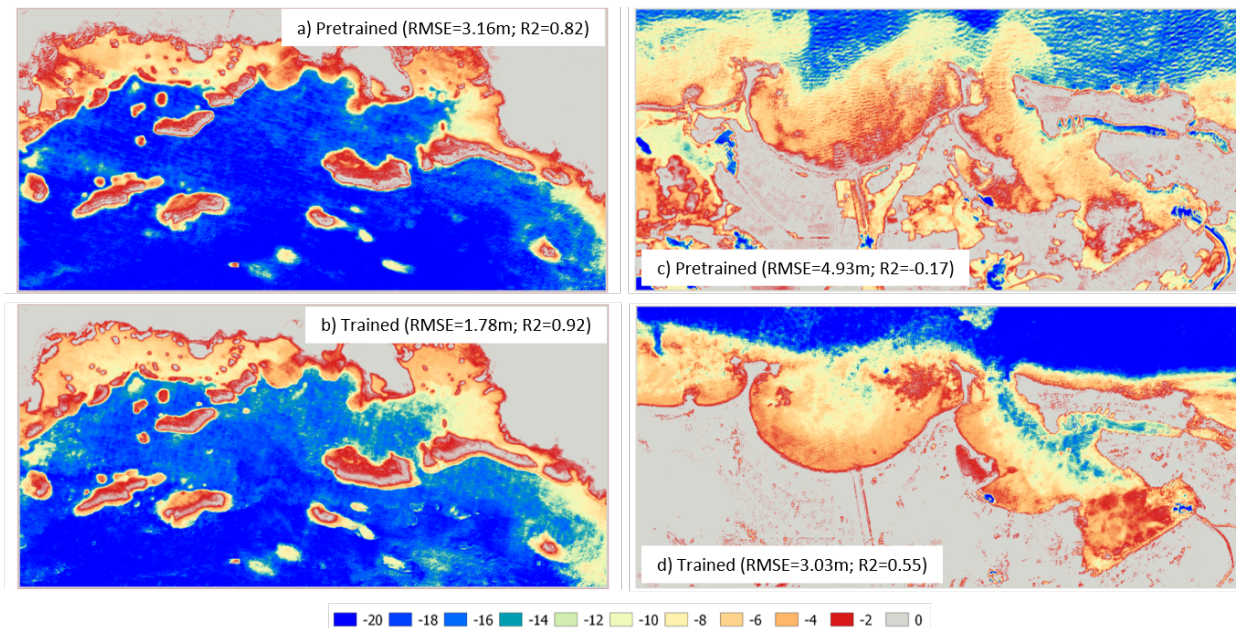
Figure 11. SDB results in Southewest (a, b) and San Juan (c, d).

bands combination as our final architecture. For other areas, we implemented the pre-trained CNN model obtained using the Ponce site first before we train the model using depth data from the area itself. As illustrated in Figure 11a, by using the pre-trained model, the Southwest SDB model has an accuracy of 3.16 m with R^2 of 0.82. The model indicates a similar trend with Ponce since the Southwest area is located near Ponce, in the south part of the Puerto Rico main island, so the water condition is similar as well as the reflectance since they came from the same image.

On the contrary, the pre-trained model is not suitable for the San Juan area, as shown in Figure 11c. The coefficient of determination yields a negative value. It means that the SDB model has an opposite trend than the in-situ measurements. San Juan has a large port with different and more turbid water conditions than Ponce and the Southwest. This condition causes different variations in the reflectance, so the relationship with depth is less clear. Using training data from the San Juan area, the SDB model improved but is still missing some details. For example, there is a clear channel of the entry to the port shown in Figure 1 bottom while it is missing in Figure 11d. The predicted result did not indicate a precise depth near the coastline since, in turbid water, the reflectance is noisier due to sediments in the water column.

4. CONCLUSIONS AND FURTHER WORK

In this paper, we tested different configurations of CNN architectures to generate SDB models. The results were discussed and compared to the linear transform method and in different locations having different water conditions. The accuracy of different architectures ranges from 1.31 m to 1.94 m, with R^2 values between 0.89 and 0.94. Table 7 depicts the overall CNN configuration that obtained the lowest RMSE with the highest R^2.

Based on the analysis, SDB generated from CNN outperforms the linear transform method. A larger window size improves the SDB model, especially in areas deeper than 15 m. The use

Table 7. Summary of CNN architecture.

Parameters	Setting
Window size	9x9
Number of channels	6 (RGBNSS)
Convolution layers	3
Kernel size	3x3
Activation	ReLu
Padding	Valid
Batch normalization	Yes
Pooling layer	No
Dropout rate	0.3
Dense layer	1
Output activation	Linear
Learning rate	0.0001
Batch size	512

of both SWIR bands in the training process raises the overall depth accuracy, especially between 2 m and 16 m. However, CNN results are worse where more noise is present, such as in turbid water.

Some of our observations regarding the implementation of the pre-trained model in other sites suggest that it works in neighbour areas with similar water conditions, but is not suitable to be used in different conditions. The model probably needs to be tested in areas farther apart. Further work still needs to be done in order to perform training with more data, preferably combining the data from several sites having different water conditions; or using multi-temporal images to enrich the variation of the reflectance and depth data pairs as data input for training.

ACKNOWLEDGEMENTS

References

Gao, J., 2009. Bathymetric mapping by means of remote sensing: Methods, accuracy and limitations. *Progress in Physical Geography*, 33(1), 103–116.

Green, E. P., Mumby, P. J., Edwards, A. J., Clark, C. D., 2000. *Remote Sensing Handbook for Tropical Coastal Management.*

Hamylton, S. M., Hedley, J. D., Beaman, R. J., 2015. Derivation of high-resolution bathymetry from multispectral satellite imagery: A comparison of empirical and optimisation methods through geographical error analysis. *Remote Sensing*, 7(12), 16257–16273.

Kabiri, K., 2017. Discovering optimum method to extract depth information for nearshore coastal waters from Sentinel-2A imagery-case study: Nayband Bay, Iran. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W4), 105–110.

Kanno, A., Koibuchi, Y., Isobe, M., 2012. Shallow Water Bathymetry from Multispectral Satellite Images: Extensions of Lyzenga's Method for Improving Accuracy. *Coastal Engineering Journal*, 53(4), 431–450.

Kanno, A., Tanaka, Y., 2012. Modified lyzenga's method for estimating generalized coefficients of satellite-based predictor of shallow water depth. *IEEE Geoscience and Remote Sensing Letters*, 9(4), 715–719.

Kanno, A., Tanaka, Y., Kurosawa, A., Sekine, M., 2013. Generalized Lyzenga's Predictor of Shallow Water Depth for Multispectral Satellite Imagery. *Marine Geodesy*, 36(4), 365–376.

Lyzenga, D. R., 1985. Shallow-water bathymetry using combined lidar and passive multispectral scanner data. *International Journal of Remote Sensing*, 6(1), 115–125.

Lyzenga, D. R., Malinas, N. P., Tanis, F. J., 2006. Multispectral bathymetry using a simple physically based algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8), 2251–2259.

Manessa, M. D. M., Haidar, M., Hartuti, M., Kresnawati, D. K., 2018. Determination of the Best Methodology for Bathymetry Mapping Using Spot 6 Imagery: a Study of 12 Empirical Algorithms. *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, 14(2), 127.

Manessa, M. D. M., Kanno, A., Sekine, M., Haidar, M., Yamamoto, K., Imai, T., Higuchi, T., 2016. Satellite-Derived Bathymetry Using Random Forest Algorithm and Worldview-2 Imagery. *Geoplanning: Journal of Geomatics and Planning*, 3(2), 117.

Misra, A., Vojinovic, Z., Ramakrishnan, B., Luijendijk, A., Ranasinghe, R., 2018. Shallow water bathymetry mapping using Support Vector Machine (SVM) technique and multispectral imagery. *International Journal of Remote Sensing*, 39(13), 4431–4450. https://doi.org/10.1080/01431161.2017.1421796.

Quadros, N., 2016. Technology in Focus: Bathymetric Lidar.

Sagawa, T., Yamashita, Y., Okumura, T., Yamanokuchi, T., 2019. Shallow water bathymetry derived by machine learning and multitemporal satellite images. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 8222–8225.

Spitzer, D., Dirks, R., 1986. Remote sensing for resources development and environmental management: Classification of bottom composition and bathymetry of shallow waters by passive remote sensing. *Proceedings of the Seventh International Symposium, Enschede, Netherlands*, 2, 775–777.

Stumpf, R. P., Holderied, K., Sinclair, M., 2003. Determination of water depth with high-resolution satellite imagery over variable bottom types. *Limnology and Oceanography*, 48(1part2), 547–556.

Tonion, F., Pirotti, F., Faina, G., Paltrinieri, D., 2020. A machine learning approach to multispectral satellite derived bathymetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 5(3), 565–570.

Traganos, D., Poursanidis, D., Aggarwal, B., Chrysoulakis, N., Reinartz, P., 2018. Estimating satellite-derived bathymetry (SDB) with the Google Earth Engine and sentinel-2. *Remote Sensing*, 10(6), 1–18.

Vinayaraj, P., Raghavan, V., Masumoto, S., 2016. Satellite-Derived Bathymetry using Adaptive Geographically Weighted Regression Model. *Marine Geodesy*.