Predicting beach profiles with machine learning from offshore wave reflection spectra

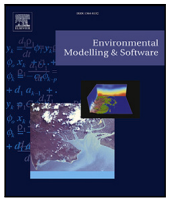
Elsa Disdier, Rafael Almar, Rachid Benshila, Mahmoud Al Najar, Romain Chassagne, Debajoy Mukherjee, Dennis G Wilson

Environmental Modelling and Software 183 (2025) 106221

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Predicting beach profiles with machine learning from offshore wave reflection spectra

Elsa Disdier [a], Rafael Almar [a,*], Rachid Benshila [a], Mahmoud Al Najar [a,b,c], Romain Chassagne [d], Debajoy Mukherjee [e,f], Dennis G. Wilson [b]

[a] Laboratory of Spatial Geophysics and Oceanography Studies (CNES/CNRS/IRD/UPS), University of Toulouse, 14 Avenue Edouard Belin, Toulouse, 31400, France
[b] ISAE-SUPAERO, University of Toulouse, 10 Avenue Edouard Belin, Toulouse, 31400, France
[c] Earth Observation Lab, The French Space Agency (CNES), 18 Avenue Edouard Belin, Toulouse, 31400, France
[d] BRGM (French Geological Survey), 3 Av. Claude Guillemin, Orleans, 45100, France
[e] Indian Institute of Technology Kharagpur (IIT Kharagpur), Kharagpur, 721302, India
[f] Texas A and M University, 18 Avenue Edouard Belin, Toulouse, 31400, United States of America

ARTICLE INFO

Keywords:
Machine learning
ANN
Bathymetry
Sand bars
Beach slope
Wave reflection

ABSTRACT

Tracking and forecasting changes in coastal morphology is vital for development, risk reduction, and overall coastal management. One challenge of current coastal research and engineering is to find a method able to accurately assess the bathymetry profile along the coast and key parameters such as slope and sandbars. Traditional bathymetry measurements are obtained through echo-sounding, which is time-consuming, hazardous and costly. Using a variety of simulated cases, we test the potential of machine learning and in particular Neural Networks to reconstruct the coastal bathymetry profile from offshore sensed waves, based on shore-based wave reflection. Features such as foreshore slope, curvature, sandbars amplitude and positions can be captured.

1. Introduction

Coastal areas are currently facing environmental and resource problems aggravated by anthropogenic pressure and over-exploitation. The environmental context of extreme events (eg, floods and coastal erosion) combined with anthropogenic pressure is a limiting factor for coastal development and flood-risk exposure (Oppenheimer et al., 2019). The state and evolution of the nearshore bathymetry over time is an important element in determining coastal risk exposure, that must be considered for coastal development and planning, but the acquisition of in-situ measured data is time-consuming and expensive, and not always possible due to the energetic nature of breaking waves in the coastal zone. Hence, performing bathymetric measurements is a challenge. Reliable, fast, and inexpensive estimates of coastal ocean depth are increasingly necessary for coastal areas (Cesbron et al., 2021; Turner et al., 2021). Specifically, knowledge of morphological characteristics (e.g. beach slope) holds significant importance in various applications, including the computation of wave runup and overtopping, as well as the assessment of longshore sediment transport (Kamphuis et al., 1986). Furthermore, the presence and amplitude of sandbars have proven to be of utmost significance, not only for amphibious operations but also as a natural protective mechanism in response to storms (Ruessink et al., 2009; Elgar et al., 2001).

Recently, remote sensing tools have emerged as sources of inexpensive data that can be used to estimate and forecast coastal dynamics and features globally (Benveniste et al., 2019; Melet et al., 2020), including traditionally low-data regions (Almar et al., 2023). These tools range from shore-based (Almar et al., 2009; Holman et al., 2013) or drone-mounted video cameras (Bergsma et al., 2019b), to space-borne satellite constellations (Almar et al., 2019b, 2021; Bergsma et al., 2019a). The present work focuses on the ability to estimate the coastal bathymetry profile based on off-the-coast wave measurements. After the pioneering work by Iribarren and Nogales (1949), further studies (Madsen and Plant, 2001; Almar et al., 2019a) have provided insights on the potential to reconstruct near-shore bathymetry profiles from distant measurements of reflected waves. However, research efforts have yet to provide a deterministic system of equations which captures and models the physical relation between coastal bathymetry and wave reflections offshore.

Deep Learning (DL) is a field of machine learning algorithms which has seen many improvements over the last decade, from simple Artificial Neural Networks (ANN) to complex deep models. DL has demonstrated impressive capabilities in a variety of different fields (Goodfellow et al., 2016), especially in Remote Sensing (Ma et al., 2019). Recent works have applied deep learning to the problem of bathymetry

* Corresponding author.
E-mail address: rafael.almar@ird.fr (R. Almar).

https://doi.org/10.1016/j.envsoft.2024.106221
Received 25 March 2024; Received in revised form 3 September 2024; Accepted 18 September 2024
Available online 23 September 2024

estimation using water color (Sagawa et al., 2019) and wave kinematics (Al Najar et al., 2022), as well as estimating wave-driven morphologies (Goldstein et al., 2019) and shorelines (Calkoen et al., 2021). In this work, we evaluate the feasibility of using simple deep learning models to predict coastal bathymetry from offshore wave reflection spectra. We create a synthetic dataset with a wide range of bathymetries and wave conditions and train a three-layer ANN model to predict bathymetry features. We perform feature importance analysis on the ANN model.

In Section 2, we begin by describing the process of generating the dataset and examining how each bathymetry feature influences the wave spectra. Section 3 then outlines two distinct approaches to predicting bathymetry. The first approach involves a direct prediction of bathymetry, while the second approach focuses on predicting bathymetry features and subsequently deriving the bathymetry values from these predictions. Additionally, this section provides a comprehensive overview of the learning methods utilized for training. The outcomes of the various approaches employing Neural Networks are presented in Section 4. In Section 5, we apply the models to real bathymetry profiles. Section 6 engages in a detailed discussion that encompasses the experimentation with real data, highlighting the encountered challenges in selecting input variables and accurately predicting sandbars.

2. Data

Neural Networks require vast amounts of data for their training, on the order of thousands of labeled examples. The availability of measured coastal bathymetry profiles and their corresponding wave data is somewhat limited. Nonetheless, a viable approach involves the generation of synthetic beach profiles and the simulation of waves through a model. This section presents the process of generating such data and offers an analysis of the resulting spectra depending on bathymetry feature values.

2.1. Dataset generation

We generate bathymetry profiles consisting of 1001 data points, corresponding to a span of one kilometer. These profiles are computed using feature values which are selected randomly on the following ranges:

- slope $s \in [0.015, 0.050]$
- curve $c \in [0.5, 1.5]$
- number of sandbars $S \in \{0, 1\}$
- sandbar height (m) $S_h \in [0, 25]$
- sandbar width (m) $S_w \in [0, 800]$
- sandbar position (m) $S_x \in [0, 700]$

Then, we generate the flat bathymetry $\mathbf{B}$ from a randomly generated positional vector $\mathbf{X}$ by first adjusting $\mathbf{X}$ according to slope : $\mathbf{a} = \frac{\mathbf{X}}{|\mathbf{X}|}|\mathbf{X}|^c$. This vector a is then used to calculate the flat bathymetry B without sandbars:

$$\mathbf{B} = -s\left(\mathbf{a} - (\max(\mathbf{a}) - \max(\mathbf{X}))\frac{\mathbf{a}}{\max(\mathbf{a})}\right) + \epsilon \tag{1}$$

where $\epsilon = 0.0001$. The final profile is obtained by adding each sandbar iteratively:

$$\mathbf{B} = \mathbf{B} + S_h e^{\frac{-(\mathbf{X} - S_x dx + L)^2}{S_w}} \tag{2}$$

where S_x, S_h, and S_w are respectively the position, height and width of the sandbar, dx denotes the interval distance of $\mathbf{X}$ (spatial resolution, here 1 m), and L denotes the beach length. For all bathymetries studied in this work, L is set to 1000 m.

The representativeness of the synthetic training data set is key. A fully-non-linear Boussinesq wave-resolving model, FUNWAVE-TVD (Shi et al., 2022), is used in order to simulate wave incidence and reflection. To this end, FUNWAVE-TVD is initialized using a Jonswap distribution including a default $\gamma = 3.3$ width, and random wave conditions on the significant wave height (Hs), the peak wave period (Tp), and is run using the previously-generated bathymetry profiles as inputs. Hs and Tp are randomly assigned between 0.5 and 2 m, and 5 and 18 s respectively, with no direct relationship between Hs and Tp. The spectra of the resulting incident and reflected waves are extracted at 512 frequencies.

Incoming and outgoing wave spectra were derived from simulation output extracted from a Radon transform (Almar et al., 2014b) using 512 frequencies within a given range. The real spectra, derived from directional wave observations, are set to match the same frequencies.

To effectively predict sandbars, we determined that the models must be trained on data that have a sandbar with a significant enough impact on waves. Thus, we focused our attention on the subset of 1-bar data points where the presence of sandbars satisfies the breaking condition:

$$\gamma = \frac{H_s}{S_d} > 0.4 \tag{3}$$

where H_s denotes the significant height of the incident wave and S_d is the bar depth (Miche, 1944; Battjes, 1974). We further discuss sandbar prediction in Section 6.2.

The final synthetic dataset contains 8991 beach profiles with associated bathymetry feature values (slope, curve, number of sandbars, height, position, and width) and wave spectra (incident and reflected). 80% of the dataset is allocated for training and 20% is reserved for testing. Fig. 1 illustrates an example of the data.

2.2. Dataset analysis

Given the absence of a deterministic model that establishes a direct link between wave spectra and bathymetry, our focus lies in investigating whether the parameters exert any influence on bathymetric characteristics. To accomplish this, we compute quantiles at the 1/3 and 2/3 levels for key parameters such as slope, curve, sandbar amplitude, sandbar width, and sandbar position. This computation allows us to visualize and analyze the average reflected/incident spectra ratio of data residing below the 1/3 quantile and above the 2/3 quantile for each parameter.

Looking at Fig. 2, it is clear that parameter values have an impact on the ratio of spectra. These substantial disparities in wave spectra attributed to varying parameter values suggest that it is relevant to predict bathymetry from wave spectra. The difference between low-curve and steep curve is the most obvious. Notably, a steep curve tends to attenuate high frequencies. The effect of the slope value is mostly noticeable at frequencies around 0.3 Hz. In fact, steep-slope data present a higher peak at these frequencies than low-slope data, which have their highest peak at low frequencies, close to 0 Hz. We also see that steep slope data present a small bump at high frequencies, around 0.9 Hz. The effect of the position is the least noticeable. We can suppose that it will lead to worse prediction scores.

We can notice that data with low and steep slopes have similar spectra shapes to data with low and high sandbars respectively. Then we can suppose that a steep slope and a high sandbar have a similar impact on the reflected spectra, which seems logical. We note the similarity between the low and steep slope spectra and the low and high sandbar spectra, which suggests that a steep slope or a high sandbar would lead to similar impacts on the reflected spectra. This raises the question of whether deep learning algorithms can effectively differentiate between a steep slope and a high bar.

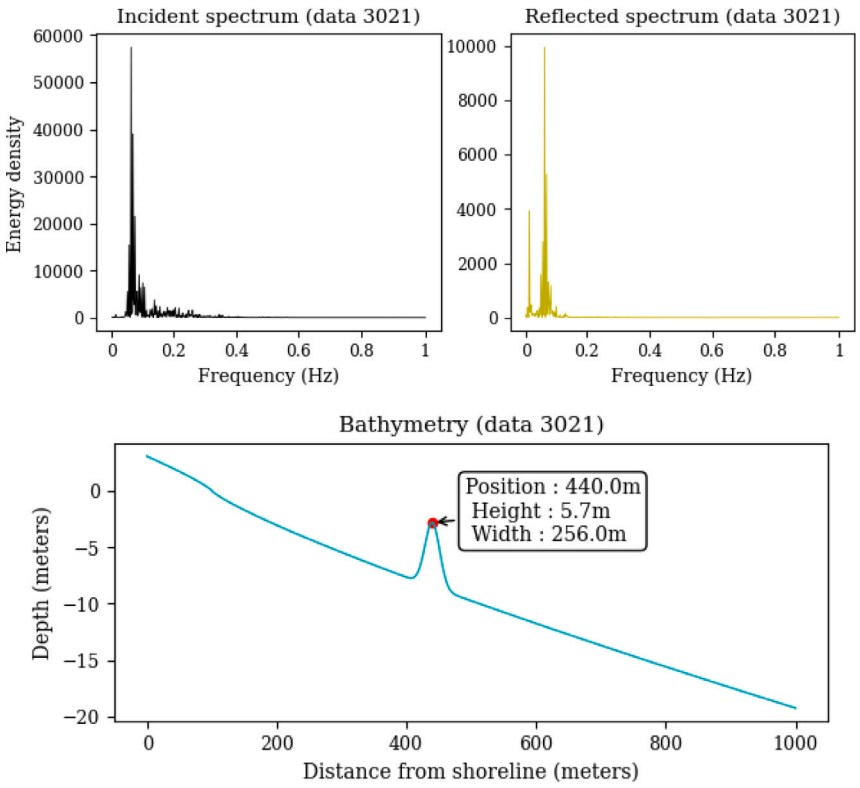


Fig. 1. Representative example of the spectral data (top) and corresponding bathymetry data (bottom).

3. Methods

The aim of our deep learning models is to predict bathymetry or some of the bathymetry features from wave spectra. Various combinations of inputs and outputs were explored and evaluated. In this Section, the different approaches to predict the bathymetry profile from wave spectra are presented, and a review of the machine learning techniques used is given.

3.1. Problem statement

We make use of the incident and reflected wave spectra as inputs to our models, representing an input of size 1024. The ratio and the difference between these two spectra were also tested as input because these operations divide the input size by two, and because it was thought that the essential information lay in the comparison between the incident and reflected spectra. To further reduce the input size, we also tested resampling the spectra to fewer frequencies (32,64,128 or 256). We normalize all inputs and outputs using Z-score normalization.

We study two different formulations of the problem of bathymetry estimation: we consider the prediction of either the bathymetry profile directly, or the bathymetry features, i.e. slope, curve, number of sandbars, sandbar height, sandbar width, and sandbar position. The first approach is a regression task that returns the 1001 values of the bathymetry profile (Section 4.1). The second approach allows to recover bathymetry by using the predicted values in the bathymetry generation formulas detailed in Section 2.1. However, the second approach requires separate cases with and without sandbars. We considered two solutions to this problem. The first solution is to consider the sandbar height, width, and position parameters as 0 for all data without sandbars (Section 4.2). This could introduce bias into the data, so we also explore the use of a two-model prediction scheme. We consider the prediction of the number of sandbars as a classification task (Section 4.3), then train two separate models for feature prediction. One model performs a regression task on no-bar data to predict slope and curve (Section 4.4). The other model performs a regression task on one-bar data to predict slope, curve, sandbar height, sandbar width, and sandbar position (Section 4.5).

3.2. Learning methods

A number of ML algorithms have been applied in previous studies for water depth estimation from observation data. In particular, ANN's (Sandidge and Holyer, 1998; Liu et al., 2015, 2018; Lumban-Gaol et al., 2021; Collins et al., 2021), and Random Forests (Manessa et al., 2016; Sagawa et al., 2019; Tonion et al., 2020; Mudiyanselage et al., 2022) have been commonly applied in similar studies and have generally shown promising performances. While we explore using both algorithms, our analysis focuses on ANN's as these have consistently outperformed their RF alternatives in our tests. We use the well-known Scikit-learn library (Pedregosa et al., 2011) for all models and training.

Neural networks are particularly well-suited for tasks involving a large number of inputs and multiple outputs due to their intrinsic properties and ability to learn complex patterns from data. For our models, we tested different three layer architectures, represented as (first layer size, second layer size, third layer size). Considering the input sizes, which are 1024 when taking both incident and reflected spectra and 512 when taking their ratio or difference, we tested the following layer sizes: (1024, 512, 256), (512, 256, 128), (256, 128, 64), and (128, 64, 32). For all architectures, we used the ReLU activation function and the Adam optimization method (Kingma and Ba, 2014). We used a grid search to find optimization hyperparameters, such as the L2 regularization parameter α and the learning rate.

In the following, we evaluate our models according to two separate metrics for classification and regression tasks. For classification, model accuracy is computed as the proportion of correctly predicted samples among the complete dataset, according to:

$$Accuracy = \frac{(n_{TP} + n_{TN})}{(n_{TP} + n_{TN} + n_{FP} + n_{FN})} \quad (4)$$

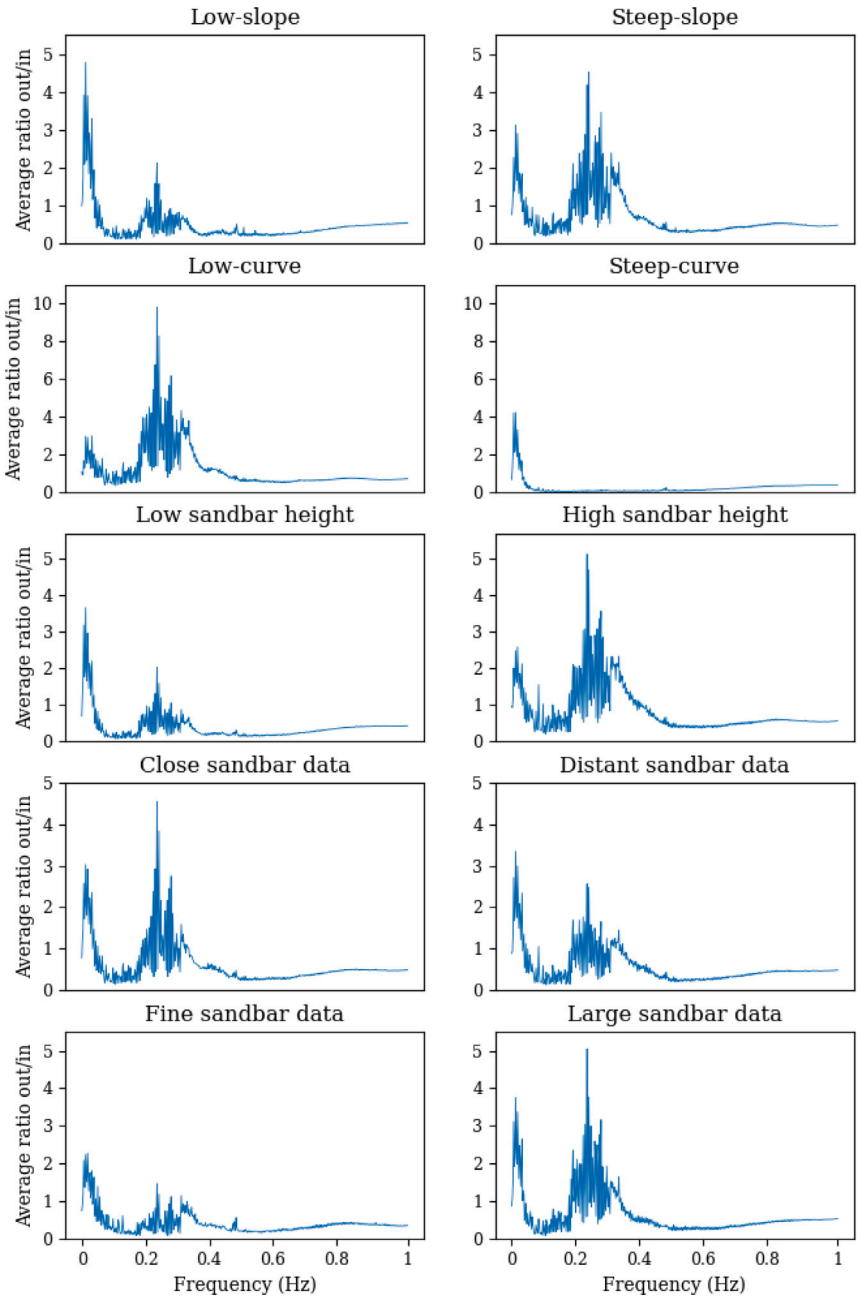


Fig. 2. Average of incident and reflected spectra ratios for data with low and large feature values.

where n_{TP} and n_{TN} correspond to the number of True Positives and True Negatives, and n_{FP} and n_{FN} to the numbers of False Positives and False Negatives, respectively. While the coefficient of determination (also known as R2 score) is used to evaluate the models over regression tasks, and is computed according to:

$$R2 = 1 - \frac{\sum_i^N (y_i - m_i)^2}{\sum_i^N (y_i - \overline{y})^2} \tag{5}$$

where N is the number of samples, y_i corresponds to the target variable, $\overline{y}$ is the mean of the target variable over the full dataset, and m_i corresponds to model predictions.

4. Results

This section presents our results on bathymetry and bathymetric feature prediction using ANN's. The different problem formulations presented in Section 3 are examined and we highlight the problem formulations that yielded the most accurate predictions. A summary of results is also provided to compare the different approaches.

4.1. Coastal bathymetry profile estimation

We first consider the task of directly predicting the full bathymetry profile from the spectral data. These bathymetry profile prediction results were obtained with an ANN of 3 hidden layers of sizes (128,64, 32), $\alpha = 0.0001$, and a learning rate of 0.001. Default settings were used for the other hyperparameters. Incident and reflected wave spectra were taken as network input. This model obtained an R^2 score of 0.90 on the test data.

In the predicted bathymetry profiles in Fig. 3, we note that the neural network predicts the curve and slope fairly well, but not the sandbars. We therefore explore the second approach which consists in

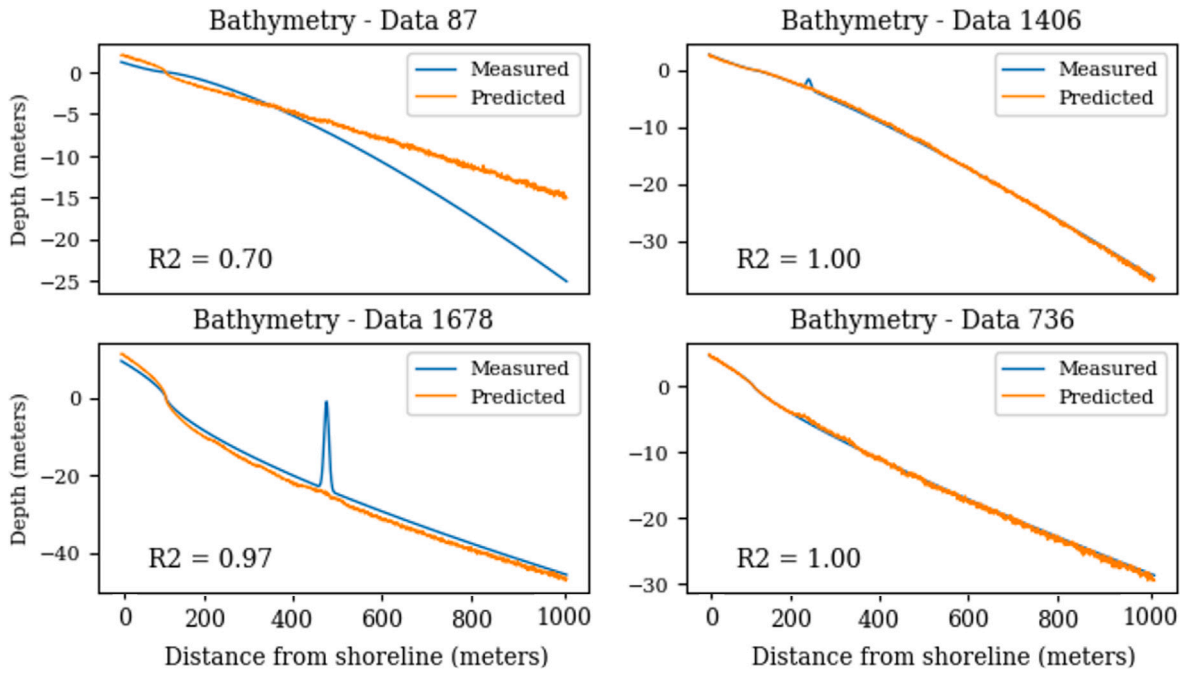


Fig. 3. Representative examples of Neural Network bathymetry predictions among the test set.

predicting bathymetry features, meaning slope, curve, number of sandbars, their height, width, and position, with the idea of reconstructing bathymetry from these feature values.

4.2. Bathymetry features estimation on all data

We first examine the prediction of the bathymetry features described in Section 2.1 for all generated profiles, those with and without a sandbar. In order to represent the sandbar features S_x, S_h, and S_w, we set the value of each feature to 0 for all data where no sandbar is present.

The ANN architecture used for this prediction of these 5 features was 3 hidden layers of sizes (1024,512,256), $\alpha = 0.0001$, and a learning rate of 0.001. We obtain on the test set an R^2 correlation of 0.80 for slope, 0.87 for curve, 0.83 for sandbar position, 0.89 for sandbar height, and 0.69 for sandbar width prediction.

While the correlation scores are high, we note in Fig. 4 that the predicted sandbar features are not close to zero for profiles with no sandbar. Consequently, the algorithm often predicts a sandbar when it should not. We see and example in Fig. 5 for profile 840. For this case, we also notice that the predicted curve and slope are less accurate. We therefore consider splitting the problem into two separate predictions: the number of sandbars, and the features for the two separate sets of data, those with and those without a sandbar.

4.3. Sandbar number classification

We first consider the ability of the model to predict whether or not there is a sandbar based on spectral data, without predicting the features of the sandbar or bathymetry. These sandbar number classification results were obtained with an ANN architecture of 3 hidden layers of sizes (128,64,32), $\alpha = 0.0001$, and a learning rate of 0.001. Incident and reflected wave spectra were taken as network input.

We obtained on the test set an accuracy score of 0.94. A confusion matrix over the test set is presented in Table 1. We note that, while the model has a slight bias towards predicting a sandbar when none is present, it is able to reliably predict the presence of a sandbar. We consider that the bias towards false positives may be in the use of a threshold on sandbar depth S_d to consider if a sandbar is present or not, as described in Section 2.1 (Eq. (3)).

Table 1
Confusion matrix of the number of bar classification for 0 and 1 bar data using Neural Network.

Number of bars	Measured 0	Measured 1
Predicted 0	717	51
Predicted 1	56	975

4.4. Bathymetry features estimation on 0-bar data

Given that we can reliably predict if a sandbar is present or not, we next consider separate models trained on data with and without sandbars. We first consider the case of bathymetry feature estimation for profiles where no sandbar is present. For this case, we therefore only need to predict two bathymetry features: the curve c and slope s. The ANN architecture used for the prediction of these 2 features was 3 hidden layers of sizes (1024, 512, 256), $\alpha = 0.0001$, and a learning rate of 0.001. The reflected/incident spectra ratio was taken as network input. We obtained an R^2 score of 0.89 for slope and 0.95 for curve prediction on the test set (see Figs. 6 and 7).

4.5. Bathymetry features estimation on 1-bar data

We next consider the case of predicting bathymetry features for profiles with a sandbar. For these profiles, we need to predict the sandbar features (height, width, position), in addition to the slope and curve. The ANN architecture used for the prediction of these 5 features was 3 hidden layers of sizes (1024,512,256), $\alpha = 0.0001$, and a learning rate of 0.001. We obtained an R^2 score of 0.80 for slope, 0.82 for curve, 0.74 for sandbar position, 0.89 for sandbar height, and 0.71 for sandbar width prediction (see Figs. 8 and 9).

4.6. Summary of results

The results of the main experiments conducted are summarized in Table 2. As input, we examined the use of incident and reflected wave spectra, denoted in the table as “spectra”, or the ratio between reflected and incident spectra, denoted as “ratio”. Table 2 specifies whether the experiment was conducted on profiles with no sandbar, profiles with a

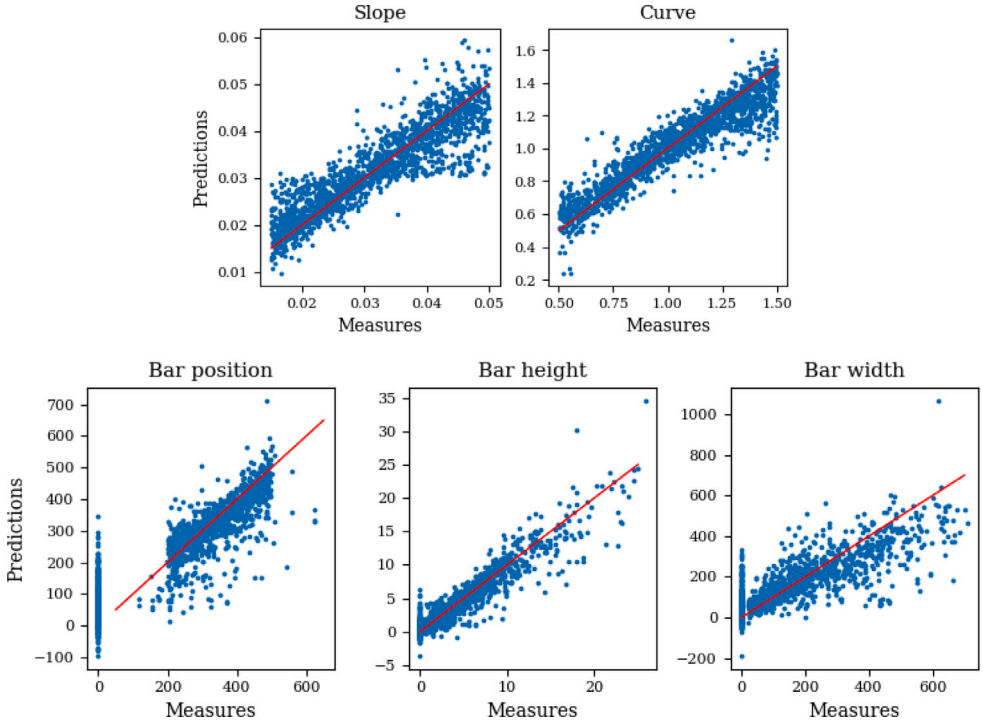


Fig. 4. Predicted versus measured features when using an ANN on all data with 0 values for no sandbar profiles.

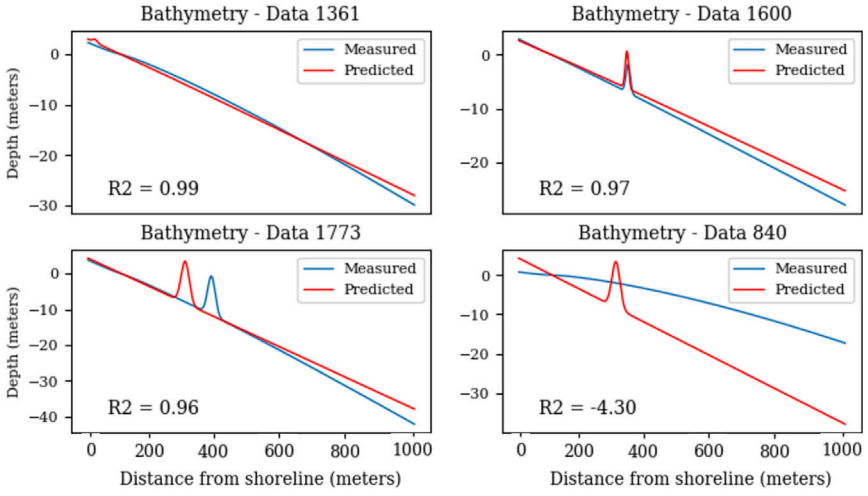


Fig. 5. Bathymetry reconstruction from ANN feature predictions on 6 example profiles.

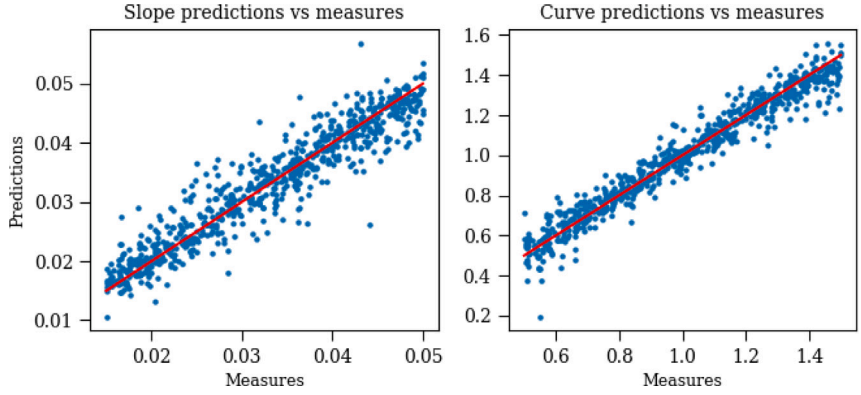


Fig. 6. Predicted versus measured slopes and curves for no-bar data using an ANN.

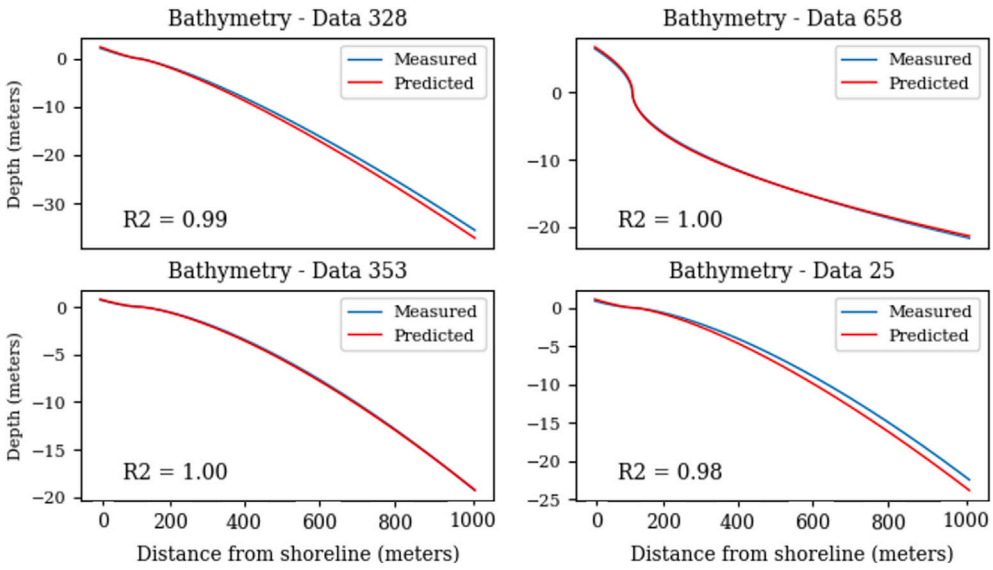


Fig. 7. Bathymetry reconstruction from ANN feature prediction on 6 test profiles with no sandbar.

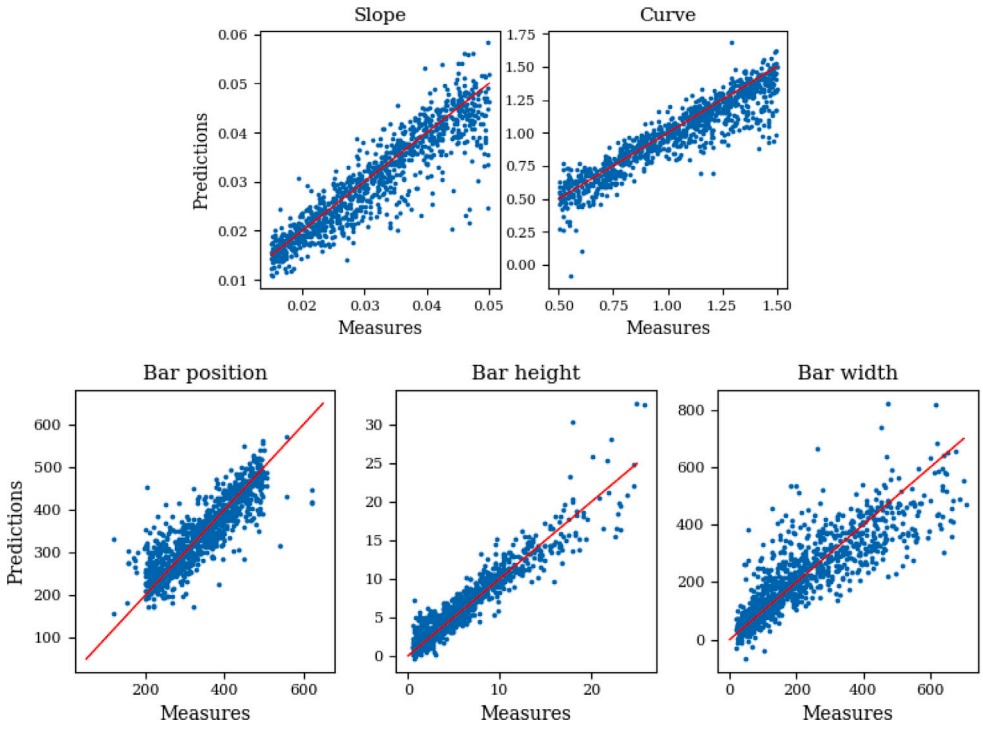


Fig. 8. Predicted versus measured features for one-bar data using an ANN.

sandbar, or all data combined. We note that utilizing spectra as input tends to result in higher prediction scores in most instances compared to the use of the spectral ratio as input, except in the case of feature prediction for data with no sandbar.

5. Analysis

We next test the proposed models on 4 in-situ bathymetry measurements collected using echo-sounder in diverse beach types worldwide. Specifically, we study Truc Vert (2008 ECORS experiment, case 1, Almar et al., 2010), Saint Louis (Senegal, case 2, Ndour et al., 2020), Narrabeen (Australia, case 3, Turner et al., 2016) and Grand Popo (Benin, case 4, Abessolo et al., 2020; Almar et al., 2014a) with associated representative wave conditions obtained during field experiments. We employ the FUNWAVE model, similar to the approach used with the simulated bathymetries (Section 2.1), to generate waves and extract their incident and reflected spectra. The typical values of Hs (significant height of offshore waves) and Fp (frequency peak of incident waves) for each coast were used for the simulation. Consequently, these are semi-real data, as the bathymetries are measured in-situ, but the spectra are simulated.

For each real case, we use the bathymetry and the spectra of incident and reflected waves, in the same shape as the simulated data, so that we can apply the models trained on simulated data. We test the predictions using ANN models from two problem formulations. In Fig. 10, we show results from the ANN model which directly predicts the full bathymetry profile over all data (Section 4.1). In Fig. 11,

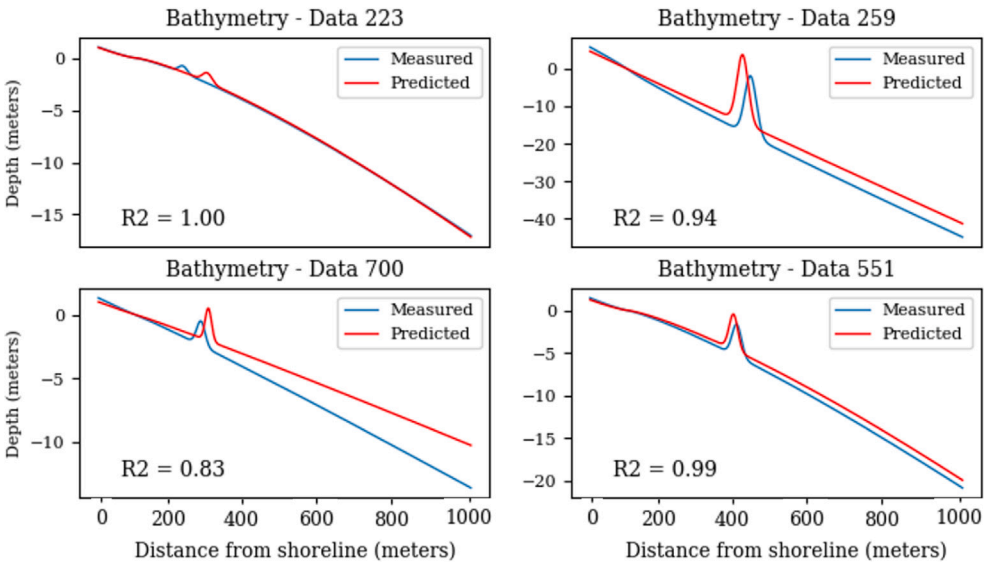


Fig. 9. Bathymetry reconstruction from ANN feature predictions on 4 test profiles with a sandbar.

Table 2
Regression (R^2) and classification (accuracy †) scores for all models. **Bold** indicates highest performance.

Data	Problem formulation		Neural Network	
	Output	Input	Spectra	Ratio
No sandbar	Features	Slope	0.77	**0.89**
		Curve	0.86	**0.95**
Sandbar	Features	Slope	**0.80**	0.74
		Curve	**0.82**	0.80
		Position	**0.74**	0.56
		Height	**0.89**	0.80
		Width	**0.71**	0.55
All data	Bathymetry		**0.90**	0.87
	Number of bars		**0.94**†	0.89†
	Features	Slope	0.80	**0.82**
		Curve	**0.87**	**0.87**
		Position	**0.83**	0.77
		Height	**0.89**	0.81
		Width	**0.69**	0.56

we show results from the ANN model trained on single-bar data that predicts slope, curve and sandbar features (Section 4.5).

We first note that the real bathymetry profiles, shown in Fig. 10, can differ from the simulated profiles by being more complex. We therefore expect the model to have more difficulty in reconstructing the bathymetry. We note in Fig. 10 that the model which directly predicts bathymetry generates far more noise than in the simulated cases. In addition, the predicted bathymetry does not match the measured bathymetry as closely as in the simulated cases. However, we observe that the average trend of the bathymetry is captured well, leading to a high overall R^2, such as in the third profile shown in Fig. 10.

In Fig. 11, we note that the prediction of bathymetry features also leads to a strong prediction of the average profile trend and is less noisy than the direct profile prediction shown in Fig. 10. However, it is still less consistent with measured bathymetry than it is in simulated cases. In addition, the model tends to predict sandbars where there are none, as for profile 4, or in a form that does not correspond to reality, as for profile 2. In future work, we plan on modifying the sandbar generation process, detailed in Eq. (2), in order to allow for a wider and more realistic variety of sandbar features.

6. Discussion

6.1. Input selection

With the aim of reducing the input size and obtaining better results, multiple inputs were tested as explained in Section 3.1. We first explored the use of the ratio or the difference of the spectra in order to divide the input size by two, using as input the ratio of the incident and reflected wave. We also consider downsampling the spectral data to further reduce the input size.

The use of the spectral ratio showed prediction performances close to those obtained with the two full spectra. This is in contrast to the use of spectral difference as input, which did not result in any accurate models. Using the current experimental setup and model architecture, the use of the raw incident and reflected wave spectra leads to the best prediction scores, except for the feature prediction task using 0-bar data, as presented in Section 4.4.

Moreover, resampling the spectra on fewer frequencies showed to reduce performance. Indeed, the more we reduced the number of frequencies used, the lower the prediction scores. However, reducing the size of the input space allowed us to have smaller inputs and therefore shorter computation times. While much of the work presented here makes use of the full spectra as inputs, future work should further explore the use of resampled spectra in order to reduce the computational cost of model training and application.

6.2. Difficulties in the sandbar prediction

This work made use of a dataset of synthetic bathymetry profiles as well as simulated wave conditions in order to study the ability of ML to estimate bathymetric features based on wave incidence and reflection spectra. The development of this dataset went through multiple stages which are not presented in this work, and was informed by the empirical results obtained using each dataset setup. The final dataset used to train and test the models presented in this work was composed of bathymetry profiles with either a single sandbar or none, where the sandbars are generated such that they would impact the simulated wave signatures by causing wave breaking. While the development of this dataset is not detailed in this work, we briefly discuss these steps in this section in order to highlight different directions for future work, which we believe would enhance the real-world applicability of the methods presented here even further.

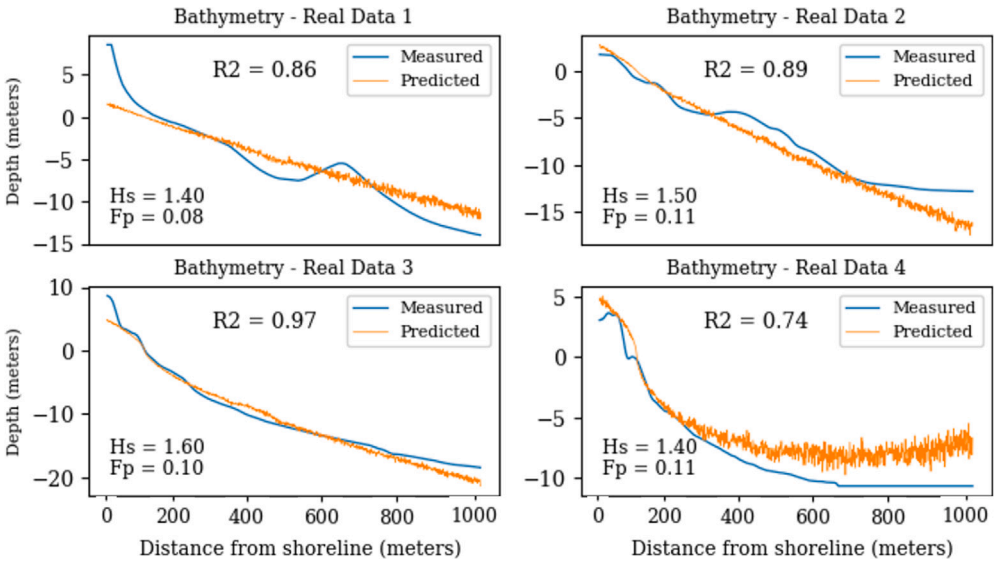


Fig. 10. Bathymetry predictions on real data when directly predicting the full profile with an ANN.

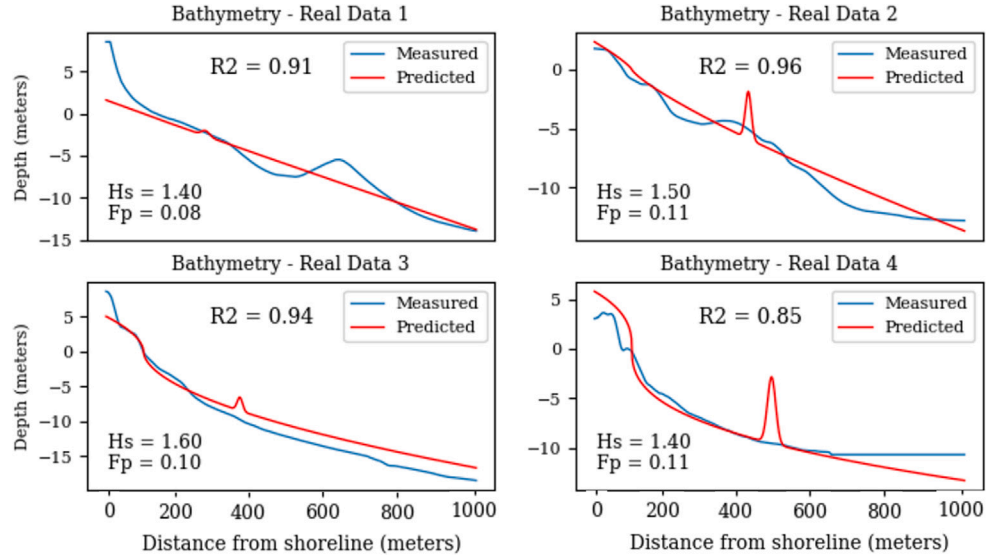


Fig. 11. Bathymetry predictions on real data when predicting profile features with an ANN.

Much of the work on the design of this dataset was focused on the impact of sandbars on final model performance. We tested including multiple sandbars (up to 3 in a single bathymetry profile) and found that the model was able to accurately predict the bathymetry profile slope and curve (R^2 scores near 0.90), while the sandbar feature predictions were inaccurate and led to negative R^2 scores. Limiting the number of possible sandbars to a maximum of one improved the model's ability to estimate sandbar height, width, and position ($R^2 \simeq 0.40$). We further examined the impact of constraining the sandbar generation to single sandbars which cause clear impact on the wave signature through wave breaking (criterion given by (3)). Using this dataset to train the models, we find large improvements in model performance across all tasks. Future development of the methodology presented here should further explore expanding the synthetic dataset to include additional and more realistic morphological configurations.

While we studied the use of real bathymetry profiles, throughout this work we use synthetic spectral data. We believe that the method and overall results will apply to observational spectral data, revealing a link between the spectral data and bathymetry. However, such a model will require training on observational data and corresponding bathymetry; such a dataset does not currently exist, to our knowledge. With this work, we hope to demonstrate that the collection of such spectral data for bathymetry study is of use, and a future direction is to train a machine learning model on this data.

7. Conclusion

This work presents an empirical study on the use of Machine Learning to estimate the near-shore bathymetry and morphological features (i.e., sandbars) based on simulated coastal wave incidence and reflection spectra. Furthermore, the problem was studied from two different perspectives. ANNs were first used to directly estimate the bathymetry profile using spectral inputs and were found to achieve significant estimation accuracy ($R^2 \simeq 0.90$). However, the model was unable to reconstruct low-level features such as the existence and characteristics of sandbars in the target data. Our second approach aimed at indirectly reconstructing the bathymetry profile by training the model to estimate the low-level features individually, rather than the complete bathymetry profile. Using this configuration, the model achieves significant R^2 scores of 0.9 when predicting the bathymetry

profile slope and curve, and R^2 scores of 0.89, 0.74, and 0.71 for sandbar height, position, and width, respectively.

Finally, this work presents the first steps towards the development of a novel technique for bathymetry inversion based on wave incidence and reflection spectra, which are measured off-shore. Such a technique can be used to create systematic measurements of bathymetry profiles at low cost and in areas where the application of current satellite-derived bathymetry techniques is limited by data-related issues (e.g. near-shore turbidity, clouds). Here, we made use of synthetic bathymetry profiles and the FUNWAVE simulator in order to create a synthetic dataset for model training and validation. To maximize the real-world applicability of the method presented here, future works should further develop the synthetic bathymetry generation procedure to produce more representative and realistic data, which would facilitate the transfer of the model to real data. Overall, the methodology presented here shows great promise for enhancing our ability to effectively monitor coastal evolution.

CRediT authorship contribution statement

Elsa Disdier: Writing – original draft, Formal analysis, Data curation. **Rafael Almar:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization. **Rachid Benshila:** Validation, Supervision, Methodology, Investigation, Data curation. **Mahmoud Al Najar:** Methodology, Investigation. **Romain Chassagne:** Supervision, Investigation, Formal analysis, Conceptualization. **Debajoy Mukherjee:** Methodology, Investigation, Formal analysis, Data curation. **Dennis G. Wilson:** Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All codes and data presented in this study are freely available at doi: https://zenodo.org/records/13832043.

Acknowledgment

This research was supported by the ANR-22-ASTR-0013-01 GLOBCOASTS project.

References

Abessolo, GO, Almar, Rafaël, Jouanno, Julien, Bonou, F, Castelle, B, Larson, M, 2020. Beach adaptation to intraseasonal sea level changes. Environ. Res. Commun. 2 (5), 051003.

Al Najar, Mahmoud, Benshila, Rachid, Bennioui, Youssra El, Thoumyre, Grégoire, Almar, Rafael, Bergsma, Erwin WJ, Delvit, Jean-Marc, Wilson, Dennis G, 2022. Coastal bathymetry estimation from sentinel-2 satellite imagery: Comparing deep learning and physics-based approaches. Remote Sens. 14 (5), 1196.

Almar, Rafael, Bergsma, Erwin W.J., Maisongrande, Philippe, de Almeida, Luis Pedro Melo, 2019b. Wave-derived coastal bathymetry from satellite video imagery: A showcase with pleiades persistent mode. Remote Sens. Environ. (ISSN: 0034-4257) 231, 111263. http://dx.doi.org/10.1016/j.rse.2019.111263, URL https://www.sciencedirect.com/science/article/pii/S0034425719302822.

Almar, Rafael, Bergsma, Erwin WJ, Thoumyre, Gregoire, Baba, Mohamed Wassim, Cesbron, Guillaume, Daly, Christopher, Garlan, Thierry, Lifermann, Anne, 2021. Global satellite-based coastal bathymetry from waves. Remote Sens. 13 (22), 4628.

Almar, Rafael, Blenkinsopp, Chris, Almeida, Luis Pedro, Bergsma, Erwin W.J., Catalan, Patricio A., Cienfuegos, Rodrigo, Viet, Nguyen Trung, 2019a. Intertidal beach profile estimation from reflected wave measurements. Coast. Eng. (ISSN: 0378-3839) 151, 58–63. http://dx.doi.org/10.1016/j.coastaleng.2019.05.001, URL https://www.sciencedirect.com/science/article/pii/S0378383919300870.

Almar, Rafael, Bonneton, Philippe, Senechal, Nadia, Roelvink, Dano, 2009. Wave celerity from video imaging: a new method. In: Coastal Engineering 2008. World Scientific Publishing Company, ISBN: 978-981-4277-36-5, pp. 661–673. http://dx.doi.org/10.1142/9789814277426_0056, URL https://www.worldscientific.com/doi/abs/10.1142/9789814277426_0056.

Almar, R, Castelle, B, Ruessink, BG, Sénéchal, N, Bonneton, P, Marieu, V, 2010. Two- and three-dimensional double-sandbar system behaviour under intense wave forcing and a meso–macro tidal range. Cont. Shelf Res. 30 (7), 781–792.

Almar, Rafael, Hounkonnou, Norbert, Anthony, Edward J, Castelle, Bruno, Senechal, Nadia, Laibi, Raoul, Mensah-Senoo, Trinity, Degbe, Georges, Quenum, Mayol, Dorel, Matthieu, et al., 2014a. The grand popo beach 2013 experiment, Benin, West Africa: from short timescale processes to their integrated impact over long-term coastal evolution. J. Coast. Res. (70), 651–656.

Almar, Rafael, Michallet, Herve, Cienfuegos, Rodrigo, Bonneton, Philippe, Tissier, Marion, Ruessink, Gerben, 2014b. On the use of the Radon Transform in studying nearshore wave dynamics. Coast. Eng. 92, 24–30.

Almar, Rafael, Stieglitz, Thomas, Addo, Kwasi Appeaning, Ba, Kader, Ondoa, Gregoire Abessolo, Bergsma, Erwin WJ, Bonou, Frédéric, Dada, Olusegun, Angnuureng, Donatus, Arino, Olivier, 2023. Coastal zone changes in West Africa: challenges and opportunities for satellite earth observations. Surv. Geophys. 44 (1), 249–275.

Battjes, Jg A., 1974. Surf similarity. In: Coastal Engineering 1974. pp. 466–480.

Benveniste, Jérôme, Cazenave, Anny, Vignudelli, Stefano, Fenoglio-Marc, Luciana, Shah, Rashmi, Almar, Rafael, Andersen, Ole, Birol, Florence, Bonnefond, Pascal, Bouffard, Jérôme, et al., 2019. Requirements for a coastal hazards observing system. Front. Mar. Sci. 6, 348.

Bergsma, Erwin W.J., Almar, Rafael, Maisongrande, Philippe, 2019a. Radon-augmented sentinel-2 satellite imagery to derive wave-patterns and regional bathymetry. Remote Sens. (ISSN: 2072-4292) 11 (16), 1918. http://dx.doi.org/10.3390/rs11161918, URL https://www.mdpi.com/2072-4292/11/16/1918, Number: 16 Publisher: Multidisciplinary Digital Publishing Institute.

Bergsma, Erwin W.J., Almar, Rafael, Melo de Almeida, Luis Pedro, Sall, Moussa, 2019b. On the operational use of UAVs for video-derived bathymetry. Coast. Eng. (ISSN: 0378-3839) 152, 103527. http://dx.doi.org/10.1016/j.coastaleng.2019.103527, URL https://www.sciencedirect.com/science/article/pii/S0378383918303582.

Calkoen, Floris, Luijendijk, Arjen, Rivero, Cristian Rodriguez, Kras, Etienne, Baart, Fedor, 2021. Traditional vs. machine-learning methods for forecasting sandy shoreline evolution using historic satellite-derived shorelines. Remote Sens. 13 (5), 934.

Cesbron, Guillaume, Melet, Angélique, Almar, Rafael, Lifermann, Anne, Tullot, Damien, Crosnier, Laurence, 2021. Pan-European satellite-Derived Coastal bathymetry—Review, user needs and future services. Front. Mar. Sci. 8, 740830.

Collins, Adam M, Geheran, Matthew P, Hesser, Tyler J, Bak, Andrew Spicer, Brodie, Katherine L, Farthing, Matthew W, 2021. Development of a fully convolutional neural network to derive surf-zone bathymetry from close-range imagery of waves in Duck, NC. Remote Sens. 13 (23), 4907.

Elgar, Steve, Gallagher, Edith L., Guza, R.T., 2001. Nearshore sandbar migration. J. Geophys. Res.: Oceans 106 (C6), 11623–11627.

Goldstein, Evan B., Coco, Giovanni, Plant, Nathaniel G., 2019. A review of machine learning applications to coastal sediment transport and morphodynamics. Earth-Sci. Rev. (ISSN: 0012-8252) 194, 97–108. http://dx.doi.org/10.1016/j.earscirev.2019.04.022, URL https://www.sciencedirect.com/science/article/pii/S001282521830391X.

Goodfellow, Ian, Bengio, Yoshua, Courville, Aaron, 2016. Deep Learning. In: Adapt. Comput. Mach. Learn., MIT Press.

Holman, Rob, Plant, Nathaniel, Holland, Todd, 2013. cBathy: A robust algorithm for estimating nearshore bathymetry. J. Geophys. Res.: Oceans (ISSN: 2169-9291) 118 (5), 2595–2609. http://dx.doi.org/10.1002/jgrc.20199, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jgrc.20199, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgrc.20199.

Iribarren, C.R., Nogales, C., 1949. Protection des Ports, Paper Presented at Xviith International Navigation Congress, Permanent Int. Assoc. of Navig. Congr., Lisbon, Portugal.

Kamphuis, J.W., Davies, M.H., Nairn, R.B., Sayao, O.J., 1986. Calculation of littoral sand transport rate. Coast. Eng. 10 (1), 1–21.

Kingma, Diederik P., Ba, Jimmy, 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Liu, Shan, Gao, Yong, Zheng, Wenfeng, Li, Xiaolu, 2015. Performance of two neural network models in bathymetry. Remote Sens. Lett. 6 (4), 321–330.

Liu, Shan, Wang, Lei, Liu, Hongxing, Su, Haibin, Li, Xiaolu, Zheng, Wenfeng, 2018. Deriving bathymetry from optical images with a localized neural network algorithm. IEEE Trans. Geosci. Remote Sens. 56 (9), 5334–5342.

Lumban-Gaol, Y.A., Ohori, K.A., Peters, R.Y., 2021. Satellite-derived bathymetry using convolutional neural networks and multispectral sentinel-2 images. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 43, 201–207.

Ma, Lei, Liu, Yu, Zhang, Xueliang, Ye, Yuanxin, Yin, Gaofei, Johnson, Brian Alan, 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177.

Madsen, A.J., Plant, N.G., 2001. Intertidal beach slope predictions compared to field data. Mar. Geol. 173 (1–4), 121–139.

Manessa, Masita Dwi Mandini, Kanno, Ariyo, Sekine, Masahiko, Haidar, Muhammad, Yamamoto, Koichi, Imai, Tsuyoshi, Higuchi, Takaya, 2016. Satellite-derived bathymetry using random forest algorithm and worldview-2 Imagery. Geoplan.: J. Geomat. Plan. 3 (2), 117.

Melet, Angelique, Teatini, P, Le Cozannet, Gonéri, Jamet, C, Conversi, A, Benveniste, J, Almar, Rafael, 2020. Earth observations for monitoring marine coastal hazards and their drivers. Surv. Geophys. 41, 1489–1534.

Miche, M., 1944. Mouvements ondulatoires de la mer en profondeur constante ou décroissante. Ann. Ponts Chaussées 1944, pp (1) 26-78, (2) 270-292, (3) 369-406.

Mudiyanselage, SSJD, Abd-Elrahman, A, Wilkinson, B, Lecours, V, 2022. Satellite-derived bathymetry using machine learning and optimal Sentinel-2 imagery in South-West Florida coastal waters. GISci. Remote Sens. 59 (1), 1143–1158.

Ndour, A, Ba, K, Almar, A, Almeida, P, Sall, M, Diedhiou, PM, Floc'h, F, Daly, Christopher, Grandjean, P, Boivin, J-P, et al., 2020. On the natural and anthropogenic drivers of the Senegalese (West Africa) low coast evolution: Saint Louis beach 2016 COASTVAR experiment and 3D modeling of short term coastal protection measures. J. Coast. Res. 95 (SI), 583–587.

Oppenheimer, Michael, Glavovic, Bruce, Hinkel, Jochen, Van de Wal, Roderik, Magnan, Alexandre K, Abd-Elgawad, Amro, Cai, Rongshuo, Cifuentes-Jara, Miguel, Deconto, Robert M, Ghosh, Tuhin, et al., 2019. Sea level rise and implications for low lying islands, coasts and communities.

Pedregosa, Fabian, Varoquaux, Gaël, Gramfort, Alexandre, Michel, Vincent, Thirion, Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg, Vincent, et al., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Ruessink, B.G., Pape, L., Turner, I.L., 2009. Daily to interannual cross-shore sandbar migration: Observations from a multiple sandbar system. Cont. Shelf Res. 29 (14), 1663–1677.

Sagawa, Tatsuyuki, Yamashita, Yuta, Okumura, Toshio, Yamanokuchi, Tsutomu, 2019. Satellite derived bathymetry using machine learning and multi-temporal satellite images. Remote Sens. 11 (10), 1155.

Sandidge, Juanita C., Holyer, Ronald J., 1998. Coastal bathymetry from hyperspectral observations of water radiance. Remote Sens. Environ. 65 (3), 341–352.

Shi, Fengyan, Kirby, James T, Harris, Jeffrey C, Grilli, Stephan T, 2022. FUNWAVE-TVD. Young.

Tonion, F., Pirotti, F., Faina, G., Paltrinieri, D., 2020. A machine learning approach to multispectral satellite derived bathymetry. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. 3, 565–570.

Turner, Ian L, Harley, Mitchell D, Almar, Rafael, Bergsma, Erwin WJ, 2021. Satellite optical imagery in Coastal Engineering. Coast. Eng. 167, 103919.

Turner, Ian L, Harley, Mitchell D, Short, Andrew D, Simmons, Joshua A, Bracs, Melissa A, Phillips, Matthew S, Splinter, Kristen D, 2016. A multi-decade dataset of monthly beach profile surveys and inshore wave forcing at Narrabeen, Australia. Sci. Data 3 (1), 1–13.