

AN INTEGRATED MACHINE LEARNING-PROBABILISTIC APPROACH TO PREDICT BEACH VOLUME CHANGE

Aline Pieterse^{1*}, Hilario Castro Lara¹, Bereket Habtemariam¹, Rob Schepper¹, Gert Leyssen¹, and Annelies Bolle¹

In the DataBeach study a machine learning model was developed with the aim to improve the efficiency and sustainability of design of soft coastal defense projects. A morphological model based on machine learning was trained and tested to predict beach volume changes with significantly reduced computational time compared to traditional process-based models. The machine learning model was then applied, combined with a 'penalty function' for inclusion of morphological feedback, to predict beach volume changes for the study area of approximately 2 km alongshore and on a 10-year project timescale. In order to run many different scenarios for the 10-year prediction, a probabilistic methodology was developed to take into account the uncertainties in this time period. The machine learning-based model provides great benefits for probabilistic simulations, due to the lower computational time, compared to process-based numerical models such as XBeach, and a flexible way in which it can incorporate measurement data. The performance of the tested machine learning models was comparable to that of the short term volume predictions of XBeach. Comparison of 10-year predicted volume changes using the machine learning model and measured beach topography (LiDAR) showed good agreement between measured and predicted volume changes for the dry beach area (when accounting for nourishments), but overestimation in the beach volume change predictions for the intertidal beach. These differences are partially attributed to the poor performance on XBeach for long term, normal wave conditions, which are an important factor in the intertidal area.

Keywords: beach volume change; machine learning; probabilistic predictions

INTRODUCTION

Soft coastal protection such as beaches pose specific challenges with respect to design and maintenance. Beach volume change predictions for beach design and management are typically done using either (limited) historical data, or with process-based 2D or 3D numerical models (e.g. XBeach (Roelvink et al., 2009), Delft3D (Reniers et al., 2004), or TELEMAC (Hervouet, 2007)). These are computationally intensive and are therefore less suitable for probabilistic calculations. Simplified models, such as ShoreFOR (Davidson et al., 2013) and ShorelineS (Roelvink et al., 2020), generally only operate in one dimension (cross-shore or alongshore) and predict less accurately the expected maintenance volumes. More recently, machine learning (ML) techniques are providing new solutions to coastal engineering problems that were previously too computationally intensive or too complex to tackle (Goldstein et al., 2019; Zeinali et al., 2021). Machine learning and related methods include different types of neural networks, Gaussian Process Emulators (GPEs), random forest algorithms and Bayesian learning. Neural networks have been applied to coastal engineering problems in a few cases, e.g. wave overtopping has been predicted using neural networks (van Gent et al., 2007; Verhaeghe et al., 2008), and using GPEs (Pullen et al., 2018). Morphological changes, specifically beach volume changes, have however to our knowledge not previously been predicted using machine learning algorithms.

Probabilistic calculations on coastal morphological development have been performed several times in the past (Li et al., 2013; Ruggiero et al., 2006; Yan et al., 2018), but were typically done either with simplified models or with a low number of simulations due to the high computational cost. This topic has recently seen renewed attention due to the market development toward coastal defense projects that shift morphological risks to the contractor. However, such probabilistic calculations were never before attempted using machine learning algorithms.

In this study, machine learning algorithms are combined with a probabilistic approach to predict beach volume changes at a study site on the Belgian coast of approximately 2 km alongshore length, for a time period of 10 years.

STUDY AREA

The selected study area is located in Oostende-Bredene on the Belgian coast. Oostende is a coastal city located approximately in the middle of the 67 km long Belgian coast. It is one of its main economical hubs in which its harbor plays an important role. The town of Bredene is located east of the port of Oostende, and here the selected study area is located (red box in Figure 1). The study area is approximately 3 km wide and is at 2 km distance from the port of Oostende.

¹ International Marine and Dredging Consultants, Van Immerseelstraat 66, 2018 Antwerp, Belgium.
^{*} aline.pieterse@imdc.be

This residential area is protected from coastal hazards by a dune system, fronted by sandy coastline (Figure 1). Dunes are ca. 120-180 m wide and generally steep with an approximate slope of 1:2 (IMDC, 2019). The average dune elevation in the area is +15.9 m TAW (IMDC, 2021), while the intertidal area has a slope of approximately 1:50 (IMDC, 2019).

The Belgian coast is macrotidal, with a typical tidal range of 4 m. Low water levels vary between 0.14 and 0.85 m TAW from spring to neap tide, while high water levels range from 3.83 to 4.71 m TAW between spring and neap tide. Mean water level at the study site is 2.4 m TAW.

Sand is retained in the area by a groyne system. Groyne lengths within the study area are approximately 200 m. To the west of this area, groynes are longer (~250-270 m) while their length gradually decreases to the east of the project area, up to the point where they do not exist anymore. The inter-groyne spacing in the study area is approximately 250 m.



Figure 1. Location of the study area at Oostende-Bredene (Image source: Google Earth). The study area is indicated in red.

MACHINE LEARNING MODEL

A new morphological ML model was developed, aiming to maintain a high level of accuracy compared to a validated benchmark XBeach model, at much lower computational time. It was set up for the study site east of Oostende, Belgium. An extensive training dataset of the driving parameters (hydraulic input parameters, bathymetric parameters, groyne properties) has been generated for the selected site based on measured wave and water level conditions and surveyed bathymetric and topographic characteristics. The training dataset of response parameters (sedimentation-erosion volumes) was generated by a large number of short XBeach simulations. This dataset of driving and response parameters was used to train various regression ML algorithms (Random forest Regressor, XGBoost Regressor, and Extra trees Regressor).

In order to facilitate the set-up of the machine learning model, the study area was divided into alongshore stretches of 3 different elevations: -5 to 0 m TAW, 0 to 5 m TAW, and 5 to 10 m TAW, corresponding approximately to the subtidal foreshore, the intertidal area, and the supratidal high beach and dunes. In addition, the study area was separated into 7 cross-shore cells (a-g, from west to east), that are constrained by the presence of groynes on either side (Figure 2). The driving and response parameters described in the following sections are all defined for each of these 21 morphological cells.

Hydrodynamic conditions

The wave climate at Oostende-Bredene has been simplified into wave classes. These wave classes have representative significant wave height, peak period, and wave direction values, which are selected based on their frequency of occurrence. Frequency of occurrence values are calculated from timeseries measured at a directional wave buoy near the port of Oostende, which is located approximately 1 km offshore.

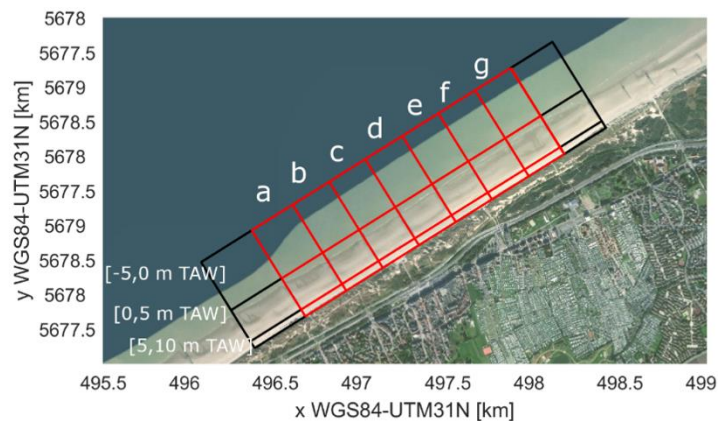


Figure 2. Morphological cells defined in the Oostende-Bredene area.

The wave conditions have been split into subcategories of normal and storm conditions. Normal conditions have significant wave heights between 0.25 m and 2.25 m, with intervals of 0.5 m. Associated peak periods range from 3 s to 11 s, at 1 s intervals, whereas associated wave directions range from 247.5 deg (WSW) to 45 deg (NE) (nautical coordinates) with intervals of 22.5 deg. This covers a range from -80.5 deg to 77 deg with respect to the shore normal, which is located at 328 deg (nautical coordinates). The normal wave conditions have been combined with water levels ranging from 0.3 m TAW to 4.3 m TAW, which is representative for an average tidal cycle (Vlaamse Hydrografie, 2013). An interval of 1 m was considered for the water levels. No storm surge component was thus assumed for such mild conditions. In total for the normal conditions 940 combinations of wave classes with water levels were used.

Storm conditions have significant wave heights between 2.75 m and 3.75 m at intervals of 0.5 m. Associated peak periods were taken from 8 s to 13 s, at 1 s intervals. Similarly to those from normal conditions, associated wave directions range from 247.5 deg (WSW) to 45 deg (NE) (nautical coordinates) with intervals of 22.5 deg. The storm wave conditions have been combined with higher water levels compared to the normal conditions. Water levels were taken from 3.3 m TAW to 6.3 m TAW, in order to account for storm surge. In total 120 different storm conditions were considered.

From the wave and water levels, 11 predictive hydraulic variables were derived. In addition to the already mentioned significant wave height, peak period, wave direction, and water level, other hydraulic variables were derived, such as run up, breaker height, and wave steepness.

Morphological parameters

The second set of driving parameters that were defined concerns the morphological characterization of the study area. Bathymetric and topographic characteristics are obtained from a dataset that combines LiDAR measurements at the Belgian coast, covering the intertidal, high beach and dune areas, with single-beam foreshore measurements and multibeam bathymetric offshore measurements. The dataset dates from 2015.

Using this bathymetric/topographic dataset, several variables were obtained per morphological cell. This included the initial volume present in the cell, the minimum and maximum elevations, and the area of the cell. The volumes in the cell were divided by the cell area in order to improve the comparison between cells, considering the large difference in area, especially between the alongshore stretches. In addition, the average cross-shore slope in the cell was calculated, as well as an alongshore maximum and minimum value. The variability of the beach within the cell was also defined for each cell. As a final part of the morphological parameters, a beach berm is present on the upper part of the beach in the study area. This berm is characterized by a steep seaward facing slope with a nearly horizontal top. If a berm is present in a morphological cell, the height of the berm is defined as an input parameter. For input into the machine learning model, the values of the parameters of the surrounding cells are considered as well, since in reality the processes within a cell also depend on the area around it (e.g. sand supply due to bypass around a groyne or deposition lower in the profile due to cross-shore sediment transport).

In addition to the morphological parameters of the beach and seabed, the geometrical characteristics of the groynes that are present in the area were also defined. Each morphological cell is bordered by a groyne on either side, and for each of these groynes a height, length and percentage coverage of the cell can be defined.

XBeach dataset

An XBeach model was set up in order to obtain the necessary response parameter for the machine learning model, i.e. sand volume changes at the study site within the different morphological cells. XBeach (Roelvink et al., 2009) is a nearshore two-dimensional depth-averaged numerical model for simulating hydrodynamics and morphodynamics at sandy coasts, initially developed to assess the impacts on beaches, dunes and barrier islands during storm and hurricane conditions (e.g. (McCall et al., 2010)). Therefore, its inception did not focus on the assessment of medium to long term morphodynamic patterns. XBeach was used in surfbeat mode, which has been extensively validated for predicting storm impacts on beach and dune morphology, but has also been validated for medium to longer term morphological changes. For example, Verheyen et al. (2014) and Gruwez et al. (2014) were able to calibrate an XBeach model to hindcast 1 year morphological changes in a groyne system in Ada, Ghana.

In total, 1060 short (3h) XBeach runs have been performed for the normal and storm conditions as described above. This gave resulting net sedimentation-erosion volumes for each morphological cell for each of these hydrodynamic conditions. In addition, a 1-year XBeach simulation was carried out to predict a timeseries of sand volume changes over a period of one year, to help in the machine learning model calibration for longer term.

An example of a sedimentation-erosion figure after a Simulations with normal wave conditions show limited morphological changes concentrated in a narrow beach strip, which falls in the 0-5 m TAW alongshore stretch. This is the location of a swash zone, which is limited in extension, as the water level for the simulation was is fixed for each 3h simulation. Bed level changes are generally less than 0.1 m. From the simulations with normal wave conditions it is also observed that oblique waves produce less disturbances on the seabed, as cross-shore processes are more limited in these cases. A sedimentation-erosion figure of a storm wave simulation shows berm erosion patterns with erosion occurring at the higher part of the beach, which leads to deposition on the lower part of the beach (Figure 3, left). As typically found for storm conditions, cross-shore transport dominates over long-shore transport. Similar to the normal wave conditions, storm waves that have a very oblique angle of incidence cause less bathymetric changes than more shore normal waves.

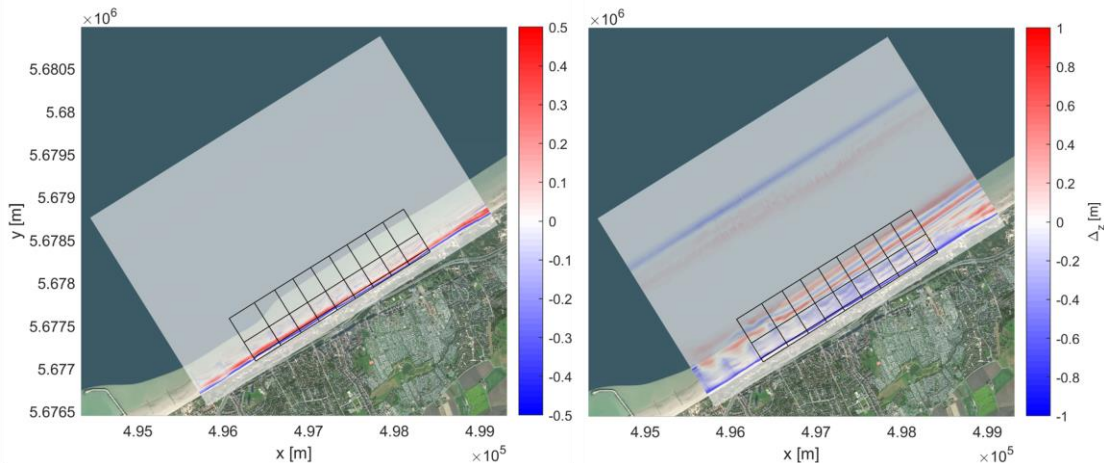


Figure 3. Left: Example of short XBeach simulation results for storm conditions. Right: sedimentation-erosion patterns after 1 year at Oostende-Bredene.

Sedimentation-erosion patterns for the 1-year XBeach run show significant erosion on the highest alongshore beach stretch (5-10 m TAW, Figure 3 (right)). Stronger erosion is observed in the western cells where a sand berm is present, explained by the large influence of avalanching due to the steeper slope. In the eastern cells, where the absence of a sand berm leads to milder slopes, avalanching plays a smaller role, and less erosion is observed in the upper stretch. At the middle elevations (0-5 m TAW), in the intertidal area, erosion is dominating as well, but spread over a larger area. In the subtidal area (-5 to 0 m TAW), reshaping of sand bars can be observed, with overall positive net volume changes, indicating that the sediment eroded from the upper parts of the beach are deposited here.

Machine learning model

The machine learning model was set up for the study area described above (Figure 2). In this study, three machine learning techniques were compared on their ability to predict the net sedimentation volumes: Random forest Regressor (Breiman, 2001), XGBoost Regressor (Chen and Guestrin, 2016),

and Extra trees Regressor (Geurts et al., 2006). Random forest is a machine learning algorithm that ranks the importance of each predictor included in a model by constructing a multitude of decision tree. Each node of a tree considers a different subset of randomly selected predictors, of which the best predictor is selected and split on. Extra tree regression is a non-parametric technique similar to random forest regressor. The main difference between the random forests and extra trees regressors is that extra trees uses a random value to split features, instead of computing the locally optimal split. This random split gives extra tree a lower variance than random forest, but also higher bias (Geurts et al., 2006). Finally, XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

In total, 67 initial input parameters were defined to train the machine learning model. Correlation plots were used to quantify the relationship between the variables and to identify a potential source of multicollinearity (e.g. Figure 4). The correlations show that the relation between net sedimentation and the bathymetric and groyne features is very small compared to the wave and water level parameters.

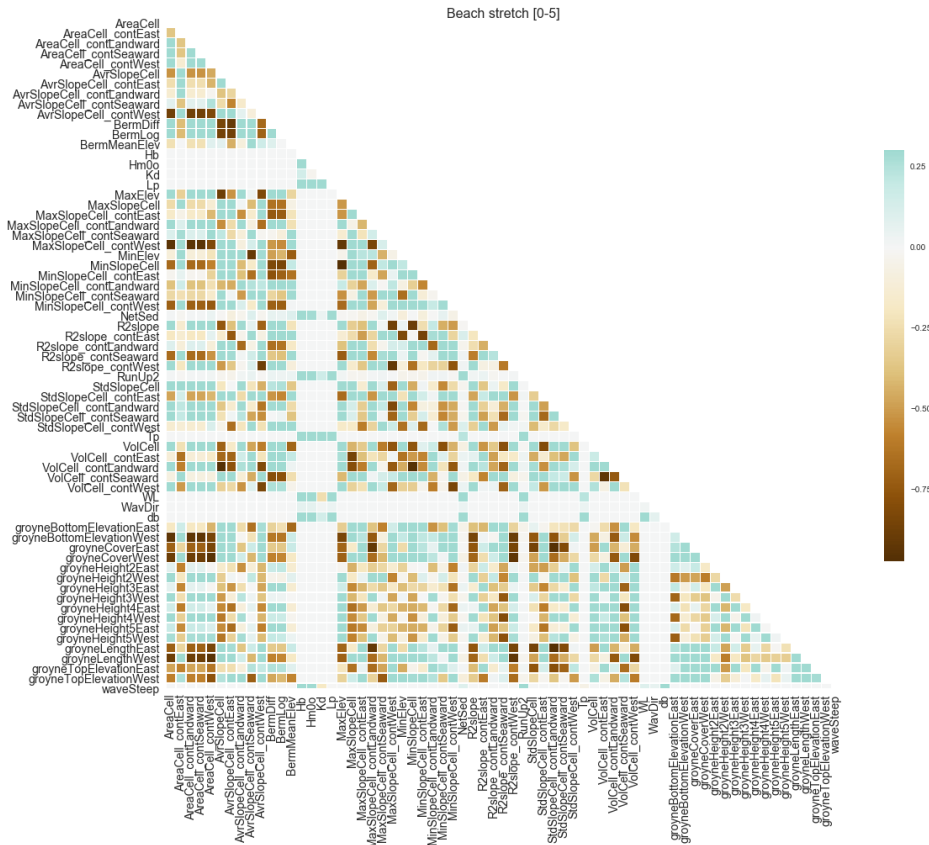


Figure 4. Correlation plot of feature for beach stretch 0 to 5 m TAW.

The training data was used to fit the three selected machine learning techniques for each alongshore beach stretch. The model fit was then evaluated using the mean squared error (Eq.1) and the adjusted R-squared (Eq.3).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

where y_i is the observed outcome and \hat{y}_i is the predicted outcome for observation i .

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$R_{adj}^2 = 1 - (1 - R^2) \left[\frac{n - 1}{n - (k + 1)} \right] \quad (3)$$

where n is the number of observations and k is the number of predictors in the model. The best possible score is 1.0, and the result can also be negative.

Model fits with all variables showed overfitting problems. Thus, in order to reduce model complexity and prevent overfitting, recursive feature elimination (RFE), which is a variable selection method, was used. In RFE features are ranked by their importance (i.e. percentage of variance explained, Figure 5)

and by recursively eliminating a small number of features per loop, RFE attempts to eliminate dependencies and collinearity that may exist in the model. Using RFE, 20 input variables were selected, for each of the three alongshore beach stretches. Overall, most hydraulic features and some bathymetry features, especially those related to slope and volume of the beach were selected as important parameters across all models.

The RFE selection reduced overfitting of the model, but some overfitting was still observed. The number of variables was thus further reduced by testing several combination of variables. Finally, the models were tuned for hyperparameters by using a cross-validation method. Figure 6 shows an example of the model prediction error and residuals for the Extra trees regression for the 0-5 m TAW alongshore stretch. The performance of the tested machine learning models was good compared to the short term volume predictions of XBeach: for the middle and upper alongshore beach stretches (0-5 m TAW and 5-10 m TAW) ca. 90% of the variability in the data was explained by the machine learning models. The performance of the machine learning models for the lower beach stretch (-5 to 0 m TAW) was comparatively poor, with ca. 60% of the variability explained.

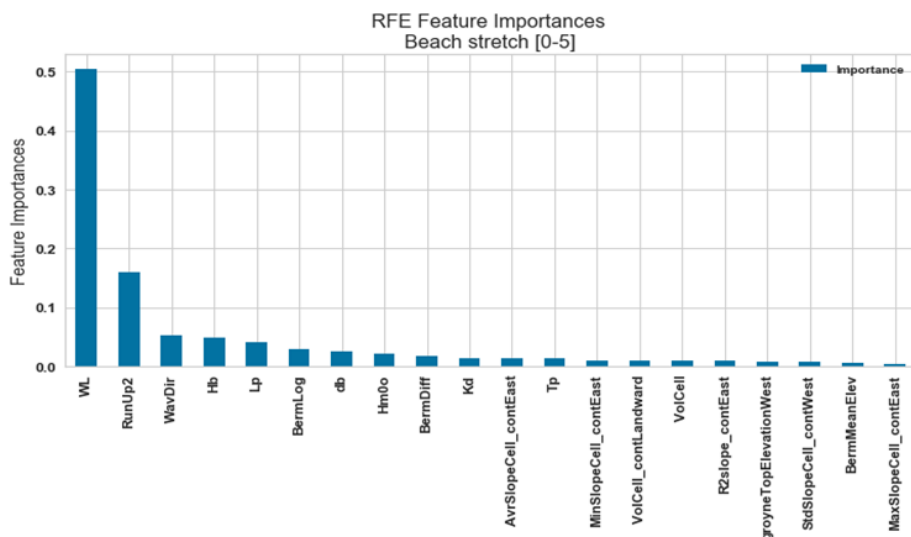


Figure 5. Feature importance for beach stretch 0-5 m TAW using recursive feature elimination technique.

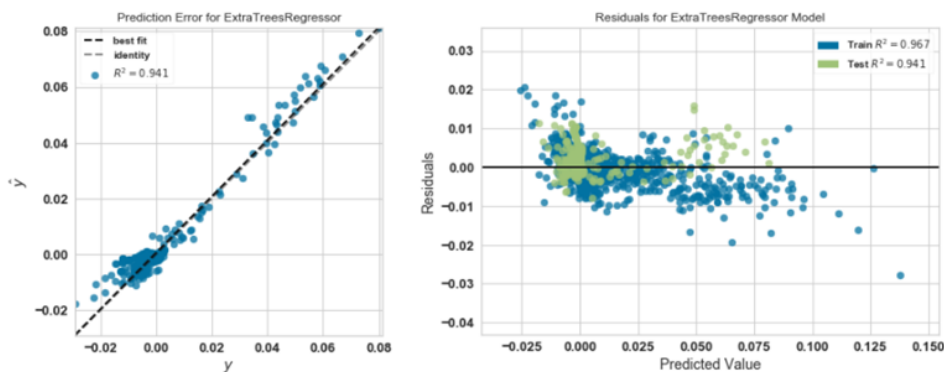


Figure 6. Model diagnostics for Extra trees regression, beach stretch 0-5 m TAW.

After initial training on the dataset with short time scales, the prediction period of the machine learning models was extended to 1 year by using the short predictions as time steps for the long term prediction. The available long term XBeach simulation was used for comparison. All machine learning models, as well as the XBeach model, predicted a cumulative net volume gain at the lower, subtidal beach stretch, and a net sediment loss at the middle (intertidal) and upper (supratidal) beach stretches. However, the machine learning predictions showed a general overprediction of volume change, compared to the 1-year XBeach simulation.

An underlying reason of the overestimation by the machine learning model is the fact that the machine learning models are trained using short XBeach simulations, where each simulation had the same initial state of the beach. Morphological changes due to previous forcings on the beach are thus not included in the machine learning model training. A ‘penalty function’ was therefore developed to account

for volume changes prior to each time step due to the governing hydrodynamic conditions. The principle of the penalty function is built upon the idea that when two storms occur at a beach in short succession, the erosion or volume change of the beach due to the second storm will be lower, as the beach is already more in equilibrium with the prevailing storm conditions. The penalty function is based on a weighted moving average filter of the dimensionless fall velocity:

$$\Omega = \frac{H_s}{w_s T_p} \tag{4}$$

where H_s and T_p are respectively the significant wave height and peak period, and w_s is the sediment fall velocity. The dimensionless fall velocity is then used in a dynamic equilibrium term:

$$\Omega_{eq} = \frac{\sum_{n=1}^{2\phi} \Omega_n 10^{-\frac{n}{\phi}}}{\sum_{n=1}^{2\phi} 10^{-\frac{n}{\phi}}} \tag{5}$$

in which n is the number of timesteps in the considered timeseries prior to the considered timestep, Ω_n is the dimensionless fall velocity at timestep n , and ϕ is the memory decay factor. This memory decay factor determines the number of antecedent timesteps that are considered (Davidson et al., 2013; Schepper et al., 2021). A large memory decay factor generates a slowly varying timeseries, considering and averaging many prior timesteps, while a small decay factor produces a timeseries with faster oscillations. Note that ϕ is a model free parameter that needs to be determined through calibration.

In a following step the disequilibrium of the dimensionless fall velocity is calculated, which is the difference between the instantaneous dimensionless fall velocity, and its equilibrium value. For a negative disequilibrium the current forcing conditions are larger than the considered antecedent forcing conditions and vice versa for a positive disequilibrium. For the scaling of the machine learning volume predictions with the penalty function, the wave height, peak period, wave incident angle, and water level are considered. Penalty function calibration was performed on a section of the one-year XBeach simulation. Including the penalty function significantly improved the machine learning model predictions. The best results were obtained with the Extra Trees Regressor (ETR, Figure 7), which was used in the remainder of the study.

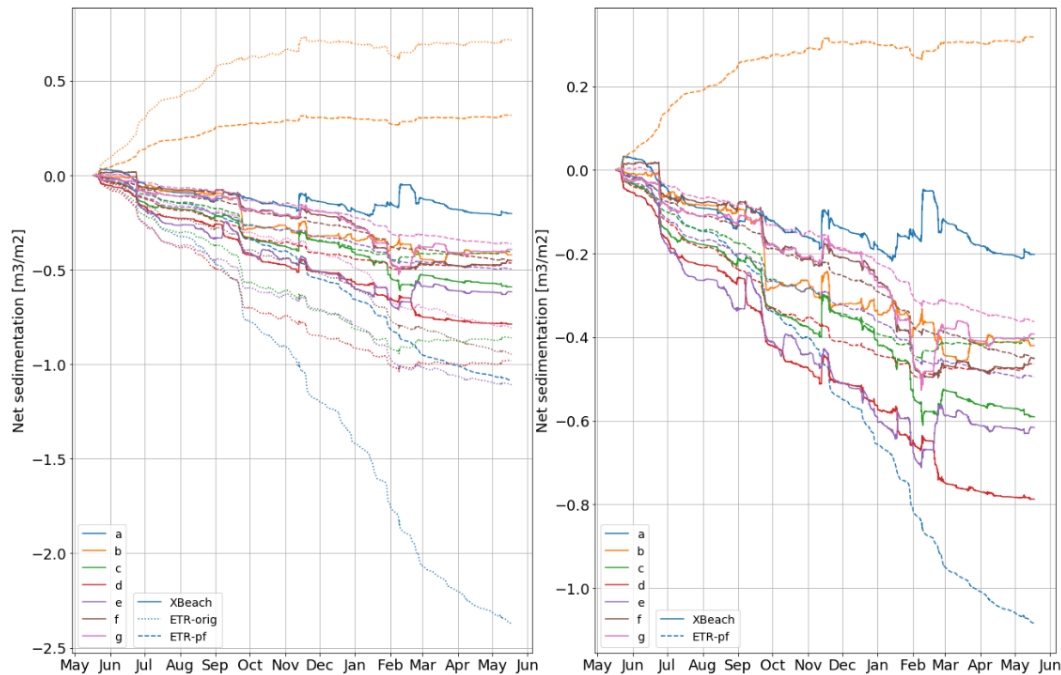


Figure 7. Comparison between 1-year XBeach results (solid), ML predictions without (dotted) and with (dashed) penalty function, per morphological cell (indicated by different colors) for beach stretch 0-5 m TAW.

PROBABILISTIC APPROACH

Based on the developed machine learning model and a statistical analysis of hydrodynamic conditions, a probabilistic design method was developed to assess beach volume changes on a soft coastal defense project scale (10 year project duration). The probabilistic method relies on a long period of input wave and water level timeseries at the study site.

Input data

In order to generate 10 year synthetic wave timeseries, a high quality and long nearshore timeseries is needed. In the nearshore close to the study site, a directional wave buoy is present that was also used for the frequency of occurrence of the input waves for XBeach and the machine learning model, from which wave timeseries can be obtained for a (relatively) short period. These data were combined with a long term (40 year) dataset obtained from ERA5 (Copernicus Climate Change Service (C3S), 2017), a comprehensive reanalysis by the European Centre for Medium-Range Weather Forecasts (ECMWF). It comprises the period from 1979 to near real time and is available at near-model resolution of 0.25° by 0.25° . The obtained long term offshore wave timeseries (1979-2021) is then translated to nearshore using a radial basis function. This is a real-valued function whose value depends only on the distance from a centre point. For the location at Oostende, multiple variants of the radial basis function are trained to reproduce the significant wave height and peak period at the nearshore wave buoy. Each variant is based on a different combination of offshore wave characteristics and environmental conditions. The variant that produces the best results is withheld for the offshore to nearshore wave transformation.

Long term water level timeseries were obtained from Flemish Banks Monitoring Network (Meetnet Vlaamse Banken, MVB) for the periods 1981-1995 and 2001-2020. For the time period 1981-1995 only high and low water levels were available and an average tidal profile was fitted through these values followed by a 2nd order polynomial interpolation to obtain hourly data. The water levels for the period 1995-2001 were obtained from a continental shelf model (iCSM), developed in TELEMAC-2D at IMDC (Chu et al., 2022, 2020). This model covers the North Sea, Celtic Sea, English Channel, Baltic Sea, and the Gulf of Bothnia. Water levels were extracted from the iCSM model near Oostende port.

The waves were divided into eight directional bins, in accordance to those used in the machine learning model training and XBeach simulations. An extreme value analysis was performed on the timeseries to determine the extreme wave climate per directional bin. For each directional bin, a set of the most extreme significant wave heights is selected with the Peak-over-Threshold (POT) method. In a next step, 10,000 timeseries of nearshore synthetic hydrodynamic boundary conditions are generated from the 40-year dataset. Timeseries were created both with and without extreme value replacement, based on a sampling procedure that maintained the statistical properties of the driving variables. Considering the long timeseries that is available, extreme waves are also assumed to be present in the dataset, without extreme value replacement.

Probabilistic predictions

Using the developed machine learning model, the sedimentation-erosion volumes per morphological cell and beach stretch were then predicted for the 10,000 timeseries of 10 years. A sensitivity analysis was performed to investigate the required number of 10-year periods to estimate the potential range in volume changes. The change in quantiles as a function of the number of simulations showed that performing 1000 simulations is sufficient to describe the cumulative density function of erosion volumes. This probabilistic approach is only possible with the machine learning-based model due to the lower computational time, compared to process-based numerical models. This resulted in the cumulative distribution function of volume changes, where the probability of a certain volume change per cell and per alongshore stretch can be derived (Figure 8).

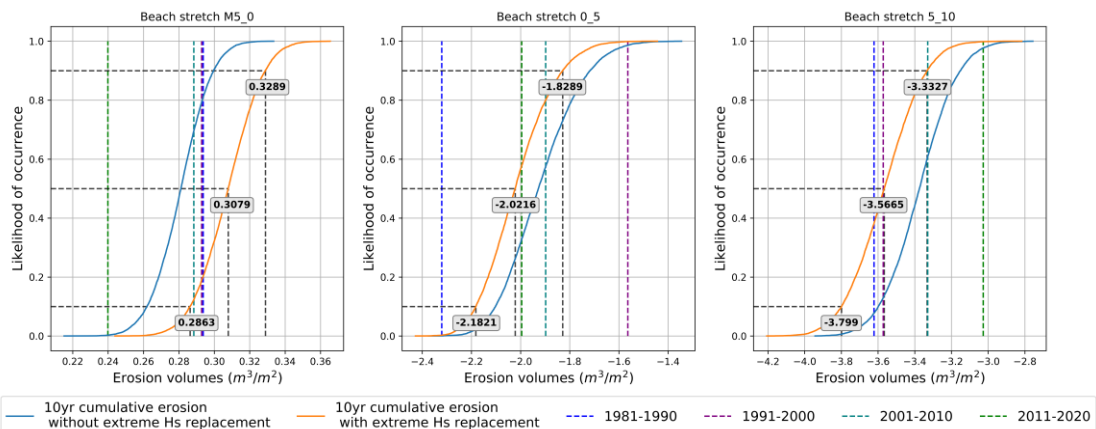


Figure 8. Total sediment loss prediction over 10 years per beach stretch. Black dotted line showing the 10th, 50th and 90th percentiles, and colored dotted lines show erosion volumes per historical 10 year time series.

VALIDATION

Oostende: Comparison with measurements

The 10-year predicted volumes from the machine learning algorithm are then compared with measured volume changes at the project site. To that end, a hindcast was performed for the period of 2011 to 2020. Measured waves were used to make a prediction of volume changes for this period. LiDAR data is used to calculate the volume changes that occurred at the project site. However, the beach east of Oostende has been nourished considerably during the past decade. These nourishments are not implemented in the machine learning model and are therefore subtracted from the measured volumes. Nourishment volumes per year and per coastal section were known, however, details on the distribution of the nourishments were not known. Two types of nourishments are distinguished in the area during this period by Division Coast: 1) beach nourishments, and 2) high beach nourishments. It is assumed that high beach nourishments fall completely in the upper beach stretch (5-10 m TAW), while beach nourishments are distributed equally over the middle (0-5 m TAW) and upper beach stretch. Measurements for the lowest, subtidal beach stretch were not available. In years that nourishments were carried out, these varied from ca. 25,000 m³ to 250,000 m³ within the project site. In addition, nourishments were done in the vicinity of the study site that could have had an influence on the development of the beach in the project area, but these are not considered in the volume comparison.

The comparison between the measured and predicted volume evolution for the upper beach stretch per morphological cell is shown in Figure 9. For this alongshore stretch, the predicted volume evolution (dashed line) corresponds quite well with the measured volume changes (thick solid lines), when accounting for the nourishments. The variability in the measured volumes for all cells is larger, but the overall trend is captured and the order of magnitude is similar compared to the predicted volumes. The comparison between predicted and measured volume change for the intertidal beach stretch (not shown) is not as good. This can probably be explained by the fact that the forcing is more variable for the intertidal zone, which is affected by all different water levels and wave conditions, and XBeach, which is used for training and as a benchmark model, is not able to reliably capture beach changes during calm conditions. In addition the intertidal area is influenced by the presence of the groynes. The upper beach stretch on the other hand is only active during the highest water levels and wave conditions. In addition it may be assumed that the training data for the upper beach stretch more correctly represents reality, since XBeach is developed to predict beach response during storm conditions. A comparison between the measured volume changes to the predicted volumes changes for the entire alongshore stretch further confirms the good agreement between measurements and predictions for the 5-10 m TAW stretch, and the overprediction for the 0-5 m TAW stretch (Figure 10).

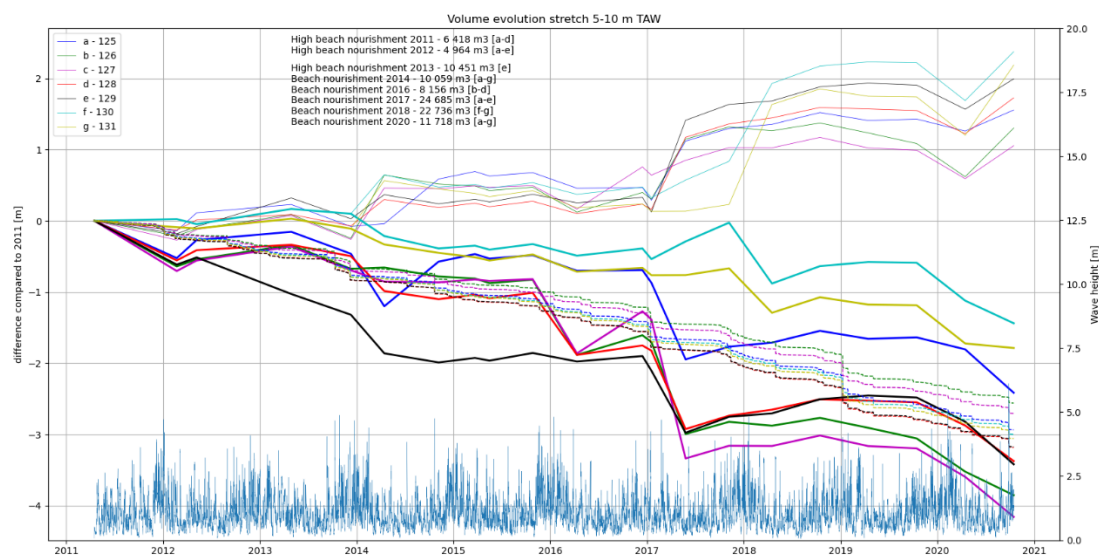


Figure 9. Measured (solid lines) / predicted (dashed lines) volume evolution per coastal cell (colours). The thin solid lines indicate the raw measured volume evolution. The thick solid lines indicate the volume evolution when accounting for nourishments.

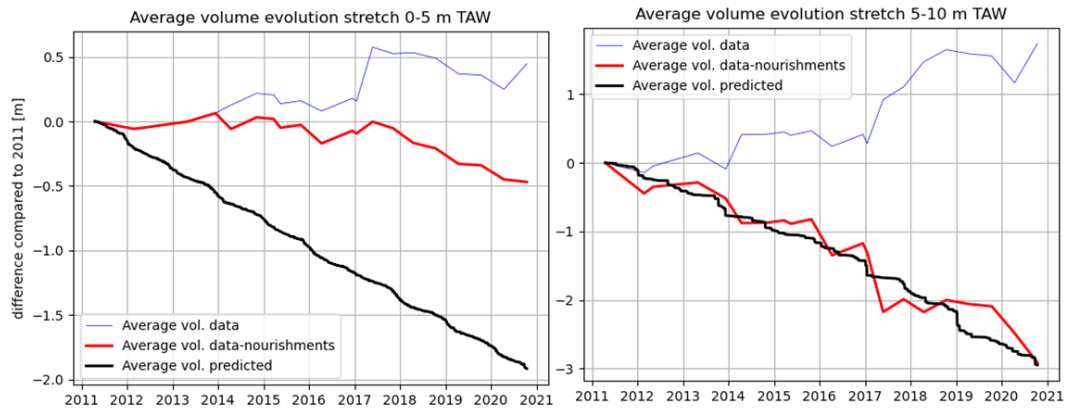


Figure 10. The average volume evolution at Oostende for beach stretch 0-5 m TAW (left) and 5-10 m TAW (right) with (blue) and without (red) nourishments, compared with the corresponding prediction (black).

Pilot case: Knokke-Zoute

In addition to the test zone at Oostende, the machine learning model and probabilistic method were also set up for a pilot zone, at Knokke-Zoute (Belgium), located approximately 5 km east of the port of Zeebrugge (Figure 11). All hydraulic and bathymetric input parameters were newly generated for this location to predict volume changes with the machine learning model. The statistical analysis of the wave conditions was also repeated for the probabilistic predictions. However, the calibration of the machine learning model was not repeated, and it was applied as-is to this new location. Compared to the study site in Oostende, the foreshore of the location in Knokke-Zoute is much steeper, while the intertidal zone is only slightly steeper and also has groynes that are used as boundaries between the morphological cells.



Figure 11. Location of the pilot-site at Knokke-Zoute (Image source: Google earth). The site is indicated in red.

Similar to the Oostende location, a 10-year volume change hindcast was then made for Knokke. The results of a 10-year prediction using wave data from 2009-2019 were then compared to available bi-yearly LiDAR data over a 10-year period of the intertidal and dry areas of the beach. At Knokke-Zoute, nourishments were also performed in the considered time period. These were done every year from 2009 to 2017, except 2016, and had volumes varying from 60,000 m³ to 180,000 m³ within the study area. Again a differentiation is made between beach nourishments and high beach nourishments, which were again assumed to be placed half/half in the intertidal zone and upper beach stretch, and fully in the upper beach stretch respectively.

Similar to the results at Oostende, a comparison of the volume evolution over the upper alongshore stretch at Knokke shows that the predicted and measured volumes (corrected for nourishments) correspond well (Figure 12, right). In the intertidal zone, this is again not the case, but now the prediction is an increase in volume in the middle beach stretch, while the data (corrected for nourishments) shows

a more stable situation (Figure 12, left). The machine learning model is thus not able to capture the changes in the intertidal zone.

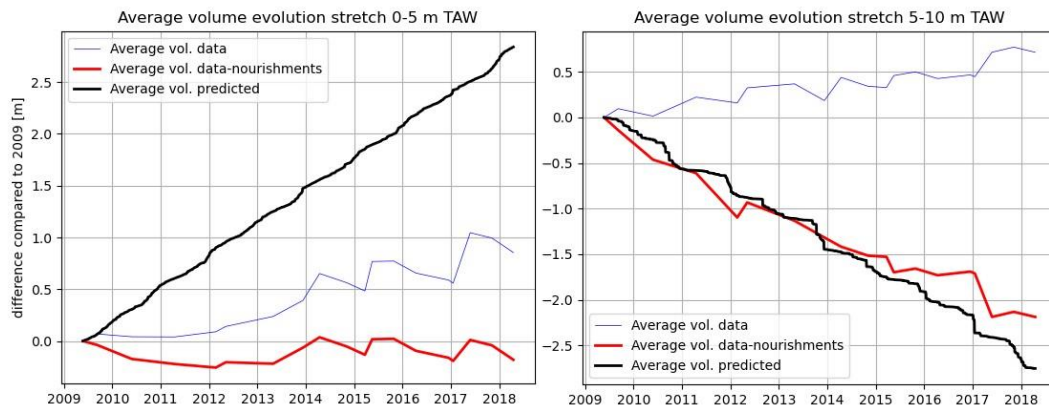


Figure 12. Predicted vs measured 10-year volume change in the 0-5 m TAW zone (left) and in the 5-10 m TAW zone (right) at Knokke-Zoute.

CONCLUSIONS

A pioneering machine learning-probabilistic method was developed to predict beach volume changes on a soft coastal defense project scale at Oostende-Bredene (Belgium). A morphological model based on machine learning was trained and tested to predict beach volume changes at a much lower computational expense compared to traditional process-based models. XBeach model simulations were used for model training and as a benchmark. The performance of the tested machine learning models was comparable to that of the short term volume predictions of XBeach: for the middle and upper alongshore beach stretches (0-5 m TAW and 5-10 m TAW) ca. 90% of the variability in the data was explained by the machine learning models. However, the performance for the lower beach stretch (-5 to 0 m TAW) was comparatively poor (ca. 60% variability explained). In the longer term, 1-year prediction, the machine learning model overestimated the volume changes on the beach in all alongshore stretches. A penalty function was developed to account for this overestimation. This penalty function is based on the disequilibrium between the current and antecedent forcing conditions, to connect the timesteps for the machine learning model. Wave height, peak period, water level, and wave incident angle were considered for the penalty function for a number of antecedent timesteps, depending on a memory decay factor that was calibrated. From the three tested machine learning models, the Extra trees Regressor showed the best correspondence with the 1-year XBeach volume change predictions.

The possible beach volume changes over a period of 10 years were then predicted using a probabilistic approach. This approach takes into account both normal and extreme climate and was tested at the main project site as well as a pilot case location. Two methods were considered, with and without replacement of historical storm conditions with synthetic storms. Both methods gave comparatively similar results, with mostly increased volume changes in the upper beach stretch for the synthetic storm method, due to the introduction of exceptional storm conditions. Comparison of 10-year predicted volume changes using the machine learning model and measured beach topography (LiDAR) shows good agreement between measured and predicted volume changes (dominated by erosion) for the upper beach stretch (when accounting for nourishments), but overestimation in the predicted volume changes for the intertidal beach. This was the case both for the study site at Oostende and the pilot case location at Knokke-Zoute.

Some challenges of the proposed methodology are that the machine learning model is data driven, which allows it to perform very fast calculations, but also limits its use on locations with different characteristics. The application on similar beaches at a different location on the Belgian coast was tested and gave comparable results. For completely different beaches a retraining of the model with local data is recommended. In addition, the volume change predictions in the more complex intertidal area (0-5 m TAW) for the long term are challenging. This becomes especially clear when comparing the volume change predictions to measurements. However, part of the discrepancy between the machine learning model results and the measurements are attributed to the poor performance of XBeach for long term, normal conditions (governing beach recovery), which was used for training of the machine learning model. In addition, the forcing mechanisms in this area are changing continuously over time with subsequent complex morphological interactions, which are difficult to predict with the machine learning

model. Hence, improvements in this area of the beach are required to be able to use the model for predictions. In addition, validation of the lower, subtidal beach stretch has to be carried out to be able to determine the performance of the model in this area.

ACKNOWLEDGMENTS

We thank VLAIO and the Blue Cluster for providing funding for the DataBeach project. Division Coast of the Agency for Maritime and Coastal Services of the Flemish government provided the measurement data used in this study. We also thank the DataBeach project partners: Fluves, DEME, and Vanbreda Risk & Benefits.

REFERENCES

- Breiman, L. 2001. Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen, T., and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd acm sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.48550/ARXIV.1603.02754>
- Chu, K., W.A. Breugem, and B. Decrop. 2022. Metocean Database of the North Sea, in: *Proceedings of the ASME 2022 41st International Conference on Ocean, Offshore and Arctic Engineering OMAE2022*.
- Chu, K., W.A. Breugem, T. Wolf, and B. Decrop. 2020. Improvement of a Continental Shelf Model of the North Sea. *Presented at the TELEMAC User Conference*, Antwerp, Belgium.
- Copernicus Climate Change Service (C3S). 2017. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. *Copernicus Climate Change Service Climate Data Store (CDS)*, date of access. <https://cds.climate.copernicus.eu/cdsapp#!/home>.
- Davidson, M.A., K.D. Splinter, and I.L. Turner. 2013. A simple equilibrium model for predicting shoreline change. *Coastal Engineering*, 73, 191–202. <https://doi.org/10.1016/j.coastaleng.2012.11.002>
- Geurts, P., D. Ernst, and L. Wehenkel. 2006. Extremely randomized trees. *Machine Learning*, 63, 3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- Goldstein, E.B., G. Coco, and N.G. Plant. 2019. A review of machine learning applications to coastal sediment transport and morphodynamics. *Earth-Science Reviews*, 194, 97–108. <https://doi.org/10.1016/j.earscirev.2019.04.022>
- Gruwez, V., B. Verheyen, P. Wauters, and A. Bolle. 2014. 2DH morphodynamic time-dependent hindcast modelling of a groyne system in Ghana. *Presented at the ICHE 2014 Hamburg - 11th International Conference on Hydrosience & Engineering*, Hamburg, Germany.
- Hervouet, J.-M. 2007. Hydrodynamics of free surface flows: modelling with the finite element method. John Wiley & Sons.
- IMDC. 2021. Ruimte-Impact maatregelen zeevering (No. I/RA/11505/20.108/API).
- IMDC. 2019. Complex Project Kustvisie. D06 Bouwtechnische ondersteuning. Huidige kustmorfologie en representatieve kustprofielen Vlaamse kust. (No. I/RA/11505/19.168/ROS/ROS).
- Li, F., P.H.A.J.M. van Gelder, D.P. Callaghan, R.B. Jongejan, C. den Heijer, and R. Ranasinghe. 2013. Probabilistic modeling of wave climate and predicting dune erosion. *Journal of Coastal Research*, 760–765. <https://doi.org/10.2112/SI65-129.1>
- McCall, R.T., J.S.M. Van Thiel de Vries, N.G. Plant, A.R. Van Dongeren, J.A. Roelvink, D.M. Thompson, and A.J.H.M. Reniers. 2010. Two-dimensional time dependent hurricane overwash and erosion modeling at Santa Rosa Island. *Coastal Engineering* 57, 668–683. <https://doi.org/10.1016/j.coastaleng.2010.02.006>
- Pullen, T., Y. Liu, P. Otinar Morillas, D. Wyncoll, S. Malde, and B. Gouldby. 2018. A generic and practical wave overtopping model that includes uncertainty. *Proceedings of the Institution of Civil Engineers - Maritime Engineering*, 171, 109–120. <https://doi.org/10.1680/jmaen.2017.31>
- Reniers, A.J.H.M., D. Roelvink, and E.B. Thornton. 2004. Morphodynamic modelling of an embayed beach under wave group forcing. *Journal of Geophysical Research*, 109, C01030. <http://doi.wiley.com/10.1029/2002JC001586>
- Roelvink, D., B. Huisman, A. Elghandour, M. Ghonim, and J. Reyns. 2020. Efficient Modeling of Complex Sandy Coastal Evolution at Monthly to Century Time Scales. *Frontiers in Marine Science*, 7, 535. <https://doi.org/10.3389/fmars.2020.00535>

- Roelvink, D., A. Reniers, A. van Dongeren, J. van Thiel de Vries, R. McCall, and J. Lescinski. 2009. Modelling storm impacts on beaches, dunes and barrier islands. *Coastal Engineering*, 56, 1133–1152. <https://doi.org/10.1016/j.coastaleng.2009.08.006>
- Ruggiero, P., J. List, D. Hanes, and J. Eshleman. 2006. Probabilistic shoreline change modeling, in: *Proc. 30th Int. Conf. Coastal Eng. Presented at the 30th Int. Conf. Coastal Eng.*, World Scientific Publishing Company, San Diego, California, USA, pp. 3417–3429. https://doi.org/10.1142/9789812709554_0288
- Schepper, R., R. Almar, E. Bergsma, S. De Vries, A. Reniers, M. Davidson, and K. Splinter. 2021. Modelling Cross-Shore Shoreline Change on Multiple Timescales and Their Interactions. *Journal of Marine Science and Engineering*, 9.
- van Gent, M.R.A., H.F.P. van den Boogaard, B. Pozueta, and J.R. Medina. 2007. Neural network modelling of wave overtopping at coastal structures. *Coastal Engineering*, 54, 586–593. <https://doi.org/10.1016/j.coastaleng.2006.12.001>
- Verhaeghe, H., J. De Rouck, and J. van der Meer. 2008. Combined classifier–quantifier model: A 2-phases neural model for prediction of wave overtopping at coastal structures. *Coastal Engineering*, 55, 357–374. <https://doi.org/10.1016/j.coastaleng.2007.12.002>
- Verheyen, B., V. Gruwez, N. Zimmermann, A. Bolle, and P. Wauters. 2014. Medium Term Time-Dependent Morphodynamic Modelling of Beach Profile Evolution in Ada, Ghana. *Presented at the ICHE 11th International Conference on Hydrosience & Engineering*, Hamburg, Germany.
- Vlaamse Hydrografie. 2013. Getijtafels 2014.
- Yan, D., S.-C. Kim, A.E. Frey, R.L. Permenter, and R. Styles. 2018. Probabilistic modeling of long-term shoreline changes in response to sea level rise and waves, in: *Scour and Erosion IX*. Presented at the Scour and Erosion IX.
- Zeinali, S., M. Dehghani, and N. Talebbeydokhti, 2021. Artificial neural network for the prediction of shoreline changes in Narrabeen, Australia. *Applied Ocean Research*, 107, 102362. <https://doi.org/10.1016/j.apor.2020.102362>