



A machine-learning hybrid-classification method for stratification of multidecadal beach dynamics

Víctor Rodríguez-Galiano, Emilia Guisado-Pintado, Antonio Prieto-Campos & Jose Ojeda-Zujar

To cite this article: Víctor Rodríguez-Galiano, Emilia Guisado-Pintado, Antonio Prieto-Campos & Jose Ojeda-Zujar (2022): A machine-learning hybrid-classification method for stratification of multidecadal beach dynamics, Geocarto International, DOI: [10.1080/10106049.2022.2110616](https://doi.org/10.1080/10106049.2022.2110616)

To link to this article: <https://doi.org/10.1080/10106049.2022.2110616>



Published online: 10 Aug 2022.



Submit your article to this journal [↗](#)



Article views: 70




View related articles [↗](#)



View Crossmark data [↗](#)



A machine-learning hybrid-classification method for stratification of multidecadal beach dynamics

Víctor Rodríguez-Galiano , Emilia Guisado-Pintado, Antonio Prieto-Campos and Jose Ojeda-Zujar

Departamento de Geografía Física y Análisis Geográfico Regional, Universidad de Sevilla, Sevilla, Spain

ABSTRACT

Coastal areas are one of the most threatened natural systems in the world. Environmental beach indicators, such as erosion and deposition rates of exposed beaches in Andalusia (640 km), were calculated using the upper limit of the active beach profile and detailed orthophotos (1:2500) for the periods 1956–1977, 1977–2001 and 2001–2011. A hybrid classification method, both supervised and unsupervised, based on machine-learning (ML) techniques was then applied to model beach response and dynamics for this 55-year period. The use of a K-means technique allowed stratification into four beach groups that have responded similarly in terms of coastline mobility and erosion/deposition patterns. Furthermore, the application of a classification and regression tree (CART) based on the K-means results helped to identify the threshold values for erosional and depositional rates and the period that characterises each cluster or stratum, enabling correct classification of 1415 out of 1509 beaches (93.77%).

ARTICLE HISTORY

Received 10 February 2022
Accepted 2 August 2022

KEYWORDS

Erosion rate; Andalusia; coast; artificial intelligence; regression tree

1. Introduction

Coastlines all around the world are exposed to anthropogenic and natural pressures with various consequences, such as erosion, flooding, socio-economic losses, and safety and health issues affecting coastal communities. Monitoring beaches provides extremely relevant information about the status of coastal systems and the sea-land interaction that enables us to adapt to, mitigate or prevent said consequences. This information is also used as input data for calculating statistical descriptors and for numerical models used to simulate coastal behaviour (wave, atmospheric, sediment-transport models, etc.). The analysis of beach profiles and the associated fluctuations in shorelines have been and continue to be one of the most commonly used methods in coastal sciences (Emery 1961, Jackson et al. 2016, Benavente et al. 2014, Splinter et al. 2018, Turner et al. 2016, Guisado-Pintado and Jackson 2020). In fact, there are several and diverse approaches to monitor beaches, ranging from the use of *in situ* studies techniques such as profile/shoreline extraction using GPS (Guisado-Pintado and Jackson 2018, Darwin et al. 2014, Klemas 2015, Guisado-Pintado and Jackson 2019, Guisado-Pintado et al. 2019), employing drones and

photogrammetry to extract volumetric and altimetry data (Guisado-Pintado et al. 2019, Casella et al. 2016, Zhou et al. 2017), the digitalisation of shorelines using orthophotos (Anders and Byrnes 1991, Crowell et al. 1991, Fernandez-Nunez et al. 2015), to the use of satellite imaging for studying long time series or extensive coastal regions (Ford 2013, Luijendijk et al. 2018, Vos et al. 2019, Dai et al. 2019, Pardo-Pascual et al. 2014, Sánchez-García et al. 2015). The development of new technologies for beach monitoring have changed both the type and quantity of information that can be obtained for the analysis of coastal systems.

An increased availability of coastal data on a finer scale with greater spatial coverage and an improved acquisition frequency, has highlighted the need for adequate statistical/computational techniques that allow the effective analysis of this data (e.g. Luijendijk et al. 2018, Montaña et al. 2020). As noted by Goldstein et al. (2019), this new wave of data in both coastal sciences and other fields has led to increased interest in empirical research known as ‘data-driven’ science. Here, machine-learning (ML) techniques play a central role in data analysis and the optimisation of results. To date, several ML methods have been used in coastal sciences, and, more specifically, in studies related to monitoring and predicting shoreline-related indicators. Some works have used Artificial Neural Networks (ANN), such as Hashemi et al. (2010) for predict beach profiles using wind and wave data; Rigos et al. (2016) investigated multiple coastline positions based on hydrodynamic inputs and the work from Tsekouras et al. (2015) which trained an ANN model to predict erosion as a function of bathymetry and storm variables. Others, like Carrero et al. (2014), tested the applicability of ANN in simulating past changes in coastal land use and to create future land-use scenarios based on changes in different climatic scenarios. Grimes et al. (2015) is an example of a work that used Genetic Algorithms (GA) in the analysis of coastline time series to examine the roles played by both internal dynamic agents (beach-face geometry) and external dynamic agents (wave forcing). However, Bayesian networks (BN) remain the most extensively used approach to date in predicting shoreline erosion rates and morphodynamic changes (Gutierrez et al. 2011, Gutierrez et al. 2015, Plant et al. 2016, Bulteau et al. 2015). For instance, Yates and Le Cozannet (2012) used a BN model to analyse the probability of how European coastlines would evolve in future (erosion, accretion, stabilisation) by combining dynamic variables (mean tidal range, rate of sea level rise and mean significant wave height) and physical variables (geology, geomorphology). Some other works of note in this area include: predicting coastal erosion related to storm events (Beuzen et al. (2018), Wilson et al. (2015), Hapke and Plant (2010), Beuzen et al. (2019), Beuzen et al. (2017)), or testing the relationship between erosion and sand nourishments (Giardino et al. 2019). Other studies that used Bayesian-based decision support systems include: Ferreira et al. (2019), which reported measures aimed at reducing risks associated with the occurrence of extreme events in southern Portugal; Pearson et al. (2017), which assessed wave-driven flooding hazards; Wright and Short (1984), which statistically determined beach state classification and Loureiro et al. (2013), which was based on sedimentological and hydrodynamic data from the Portuguese and Irish coastlines.

As stated above, in the past decade there has been an increase in the use of ML techniques to analyse and predict shoreline and beach response (erosion/deposition rates). Special mention requires de work from Burningham and French (2017) that used a combined shoreline trend and cluster-based segmentation analysis to understand mesoscale shoreline behaviour in eastern England. This study, to our knowledge, is the only one to date that has implemented a clustering approach (hierarchical clustering) to stratify shoreline behaviour at regional level. Despite this increase in the use of ML techniques, to the

best of our knowledge there are no examples of using regression trees (RT) to classify and stratify coastline behaviour. The main benefit of using a tree structure to perform classification/regression is that the tree structure can be viewed as a 'white box,' which is easier to interpret to understand the relationships between dependent and independent variables compared to other machine learning techniques (Villarin and Rodriguez-Galiano 2019, Coimbra et al. 2014).

This paper presents a new approach to understand mesoscale shoreline behaviour combining conventional shoreline change metrics with a hybrid classification method (both supervised and unsupervised) which combines the performance of K-means and a Classification Tree (CT) (Km-CART). The main objective of this work is to provide an integrated regional overview of medium- and long-term trends in erosion/deposition processes along different parts of the coast in Andalusia between 1956 and 2011. A specific objective is to propose a highly interpretable methodology to obtain an unbiased stratification of beaches according to their erosional/depositional dynamics, showing simple rules based on cut-off values of three different periods. The paper is structured as follows: section 1 an introduction and background of ML techniques and shoreline monitoring is presented, section 2 presented the study area and the methodological approach and section 3 results from shoreline dynamics and stratification with Km-CART are presented. Finally, section 4 is dedicated to the discussion the implications of using these techniques to classify multidecadal coastal dynamics in terms of erosion/accretion patterns and their performance is presented and discussed and in section 5 the main conclusions are outlined.

2. Materials and methods

2.1. Study area

The coast of Andalusia in southern Spain is a complex and highly dynamic system that stretches 917 km and has two basins facing the Atlantic Ocean to the west and the Mediterranean Sea to the east (Figure 1). It represents 30% of the Spanish coastline. The coast is highly diverse, with 66.7% comprising sandy beaches, 13% rocky cliffs and headlands, and 20% other coastal features such as marshes, estuaries and deltas.

The Atlantic coast is characterised by a relatively flat and gentle continental platform that extends 30–50 km to the 100-m isobath. Predominant coastal morphologies include estuaries (associated with the Guadiana, Guadalquivir, Piedras, Odiel and Barbate rivers), salt marshes and extensive sandy beaches (Guisado-Pintado et al. 2014). Furthermore, the presence of large rivers and a flat and wide continental shelf allows the development of aeolian landforms and mobile dunes, such as those found in Doñana National Park. Beaches are composed of fine sediments (medium-coarse sand). From a hydrodynamic perspective, the west coast can be defined as mesotidal and semidiurnal where tidal range varies from 1 m during neap tides to 4 m during spring tides. Spatially, the tides vary from 4.2 m at the mouth of the Guadiana River with minimal values reaching 0.8 m to the east near the Strait of Gibraltar. Modal waves have a bidirectional approach from SE to W-SW. A mean significant wave height (H_s) of 1.90 m with a period of 7 s is typical during the winter season, while in summer waves rarely exceed 1.07 m with short wave periods of 5–6 s (Guisado-Pintado et al. 2014). Winds are persistent and intense, switching from west to east. South-westerly storms are frequent, with 155 events registered in the decade between 2000–2010. These conditions result in beaches mostly of a dissipative



Figure 2. The proxy used to discriminate the contact between the backshore and the foredune (A), infrastructure (B) or cliff (C), depending on the coastal section. All examples are from the Andalusian coast.

has been subjected to a profound transformation that started in the 1970s. Human practices such as the abandonment of agriculture, strong regulation of river basins, and the development of coastal urbanisation and infrastructure (ports, promenades, breakwaters) have led to dissimilar effects on the western Atlantic coast and the eastern Mediterranean coast. Additionally, there is an unequal direct response to these changes along the Andalusian coast, which ultimately depends on the characteristics of the coastal landforms (exposure, presence of dunes, estuaries and deltas) and the intensity of human activity (urbanisation, % of coastal infrastructure, land use etc.)

2.2. Calculation of erosion rates in exposed Andalusian beaches

2.2.1. Coastline photointerpretation and digitalisation

The coastline photointerpretation for shoreline mapping requires the selection of an adequate feature that can serve as an indicator, this is what is normally called a shoreline proxy. Therefore, the proxy is a measure or indicator that properly reflects the real shoreline position and could be repetitively measured for assess shoreline evolution. The election of the shoreline proxy has to be driven by the geomorphological characteristics of the coastal sectors interrogated. Many proxies have been proposed in the literature for coastal erosion studies ranging from beach toe, wet/dry line, erosion scarp, stable vegetation line, dune toe, dune crest, cliff toe, cliff top, etc. (Pajak and Leatherman 2002, Fletcher et al. 2003, Boak and Turner 2005, among others). As stated by Del Río and Gracia (2013), it has to be considered that each shoreline proxy has its own advantages and disadvantages and that the final election of one or another should be considered all of them. Further, choosing the correct shoreline proxy to monitor shoreline evolution (erosion/deposition) depends on the goal of the study at hand. In this case, the proxy used is the contact between the backshore and dune (foredune, infrastructure or cliff), as it has been proved to be a robust proxy for medium- to long-term studies and, therefore, the most used proxy in the bibliography (Moore 2000, Del Río and Gracia 2013, Prieto et al. 2018) (Figure 2).

The comparison between the shoreline proxies among different dates, this is the shoreline position digitised from the available orthophotographies, allows assessing the changes

Table 1. Data sources.

Data source	Date	Pan/colour	Spatial resolution	Scale
American flight	1956–1957	Panchromatic	1 m	1:32.000
Inter-ministerial (IRYDA) flight	1977–1983	Panchromatic	0.5 m	1:18.000
Andalusian photogrammetric analogical flight	2001–2003	Panchromatic	0.5 m	1:40.000
QuickBird-Ikonos satellite orthoimage	2005	Colour + Near Infrared (NIR)	0.7 m	–
Andalusian photogrammetric flight in colour	2010–2011	Colour + NIR	0.5 m	1:40.000

in the shoreline position over time. Here, we refer to shoreline erosion when the proxy (shoreline digitised between the backshore and dune) experiments a landward movement between two dates (this is interpreted as shoreline retreat or erosion). The opposite situation, a seaward movement between two shoreline proxies extracted from different dates is interpreted as shoreline advance or deposition. From a geomorphological point of view, since the tool calculates horizontal changes in shoreline position, the result is a 2D analysis on the retreat or advance of the shoreline between years.

The entire study site (917 km) was digitised in the proprietary ArcGIS 10.3 software (ESRI, Redlands, CA, USA) at a scale of 1:2500 using aerial orthophotography from the periods 1956–1977, 1977–2001 and 2001–2011 (Table 1). Further, in order to improve the digitalisation, along cliff beach sectors without associated beaches, the 2005 *QuickBird-Ikonos* orthoimage was used to assign geometric continuity to the shorelines. This dataset given its spatial resolution (0.7 m) minimises relief displacement found in traditional orthophotos in the before mentioned areas.

The ETRS80 spatial reference system was used with a UTM zone 30 N projection, in accordance with Spanish Royal Degree 1071/2007. Once the coastlines had been digitised, they underwent a topographical quality-control process to confirm both their integrity and spatial continuity. The digitisation process was supplemented by relevant thematic data related to each segment of the coastline in line with a previously defined data model. This data model, which was based on an entity-relationship model, was used by the regional Administration to manage coastal areas and has been presented in several articles (Fernandez-Nunez et al. 2015, Prieto et al. 2018).

2.2.2. Error assessment

The digitalisation process, manual photo-interpretation of shorelines, can be affected by different variables that need to be considered when estimating the cumulative uncertainty in shoreline position (error). Following previous works (Del Río and Gracia 2013; Fletcher et al. 2003), we assume that the cumulative uncertainty in shoreline position digitised on an aerial photograph is the result of three simultaneous factors: the resolution, the scale and the photo-interpretation criterion (Table 2).

The resolution error ($\text{Error}_{\text{resolution}}$) is represented by the image resolution. The size of the minimum information unit, known as pixel, determines the resolution: the larger the pixels, the larger the uncertainty in interpreting shoreline position on the photograph. The scale error ($\text{Error}_{\text{scale}}$) is closely associated with the resolution of the screen set when digitising the shoreline (1680 × 1050). In this work, a 1:2500 scale was chosen to digitise the shorelines from the aerial photographs, as a result a mean 1.5 m continual error was obtained. The third component is called the photo-interpretation criterion ($\text{Error}_{\text{criteria}}$) and it's based on the mean distance obtained when different shorelines are digitalised by different photo-interpreters. Although the 90% of the digitalisation was carried out by a single photo-interpreter, two others helped in the process. The mean error obtained from this process was 1.25 m for photographs with sub-metric resolutions and 3 m for the period 1956-1957 (Table 2).

Table 2. Calculation of each error per orthophotography period (m).

Orthophotography period	Error _{resolution} (m)	Error _{scale} (m)	Error _{criteria} (m)	Error period (Ep) (m)
Period 1 1956–1957	1	1.5	3	3.5
Period 2 1977–1983	0.5	1.5	1.25	2
Period 3 2001–2003	0.5	1.5	1.25	2
Period 4 2010–2011	0.5	1.5	1.25	2

Once the three components of error are calculated, the total uncertainty in shoreline position on a certain photograph (Ep) is calculated as the quadratic sum of each component (e.g. Ruggiero and List 2009, Coyne et al. 1999, Del Río and Gracia 2013) according to Eq. 1.

$$Ep = \sqrt{Error_{resolution}^2 + Error_{Scale}^2 + Error_{Criteria}^2} \quad (1)$$

The errors calculated by means of Eq. 1 correspond to each individual period considered (that is the photographs of a given period), so since the tool compares two shoreline positions extracted from two separate photographs (e.g. P1 and P2), the error has to be cumulative (Anders and Byrnes 1991). Therefore, the total uncertainty of each rate calculated between two periods (two shoreline positions) needs to be calculated by considering the individual errors, as well as the time in years (T) between the two given photographs (Moore 2000, Fletcher et al. 2003, Del Río and Gracia 2013, Prieto et al. 2018) as stated in Eq. 2.

$$E_{rate} = \frac{\sqrt{Error_{P1}^2 + Error_{P2}^2}}{T} \quad (2)$$

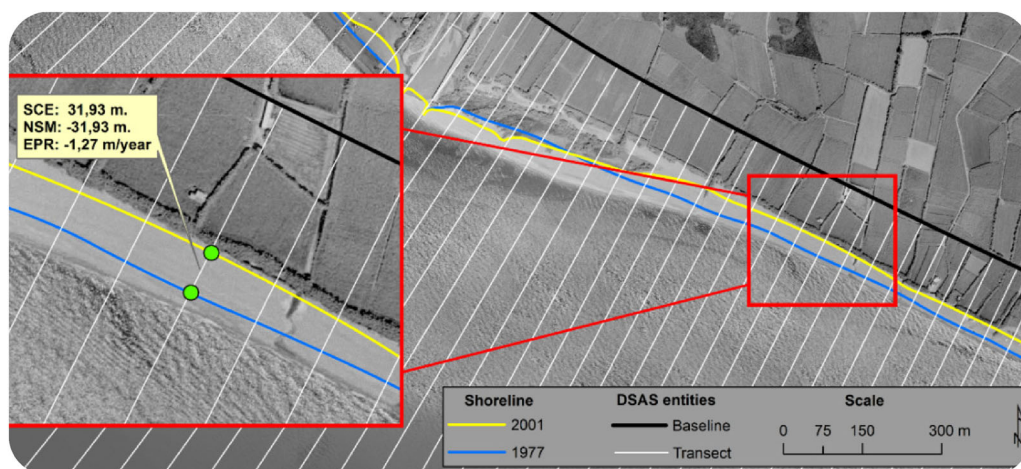
The calculated error of each rate (E_{rate}) is used as an input in the DSAS tool (Table 3) and incorporated in the calculations of the different indicators (see Thieler et al. 2009). The shoreline uncertainty (Ep) is then incorporated into the calculations for the standard error, correlation coefficient, and confidence intervals, which are provided for the simple and weighted linear regression methods.

2.2.3. Computation of coastline erosion rates

Rate calculation was performed using the Digital Shoreline Analysis System (DSAS) tool, which was developed by the United States Geological Survey (USGS) (Thieler et al. 2009). In order to determine erosion rates, a baseline was created and used to delineate orthogonal transects along the coast. The baseline was digitalised onshore, parallel to the shoreline, and was used to automatically generate 15,069 transects with equidistant spatial sampling at 50 m. Of these transects, 30% were manually modified to ensure that they remained orthogonal to the shoreline. Starting at the baseline, each transect intersects the various shorelines that were extracted using the orthophotos, thus allowing us to calculate the distance between them throughout the time period studied. The following beach response indicators were obtained (Figure 3): (i) Shoreline Change Envelope (SCE), distance (m) between the shoreline closest to and furthest from the baseline; (ii) Net Shoreline Movement (NSM), distance (m) between the oldest and newest shoreline; and (iii) End Point Rate (EPR), which is the NSM distance divided between the shoreline dates, thus obtaining the annual change rate (m/year). Finally, for the classification and regression tree, only the EPR indicator was used as it's indicative of a rate and therefore allows better comparison between dates.

Table 3. Error value per period considered (E_p) and for each rate (E_{rate}).

Time period	E_p (m)	E_{rate} (m/year)
1956–1977	4.03	0.20
1977–2001	2.83	0.12
2001–2011	2.83	0.33

**Figure 3.** Diagram of the DSAS approach showing the orthogonal transects (in white) used to calculate beach response indicators such as SCE, NSE, EPR (inset image). Shorelines from 2001 (yellow) and 1977 (blue) are used in this example.

2.3. Machine-learning techniques for stratifying beach response indicators

Machine learning is a field of research that falls under artificial intelligence and consists of the use of data to learn patterns and build models that enable us to infer the value of a target variable (i.e. erosion etc.). It therefore involves empirical models that are built based on information contained within the data, identifying relationships between the target variable and a series of variables that explain its behaviour. Unlike traditional statistical methods, machine-learning techniques allow us to identify complex, nonlinear relationships and forego previous assumptions about the statistical distribution of variables (i.e. normality). There are two types of issues that can be dealt with using machine learning, depending on whether the target variable is continuous or categorical. These allude to a regression process and a classification process, respectively. Classification methods, in turn, are subdivided into supervised and unsupervised, depending on whether or not the learning examples used during the training contain the class labels for said examples, and when these values are not present, the system seeks to group the examples into similar clusters (Sammuto and Webb 2017).

Beach erosion rates, despite being continuous variables by definition, are susceptible to stratification into groups with similar magnitudes (m/year) or specific behaviours over time. Thus, an erosion rate with a lower value over a specific period that then presents a greater value in a subsequent period may be indicative of a change in soil use, impact from coastal engineering and/or a change in sedimentary input within the system. On the other hand, a rate that does not change over periods may indicate stabilisation of the coastline due to anthropogenic measures (e.g. construction of seafront promenades or breakwaters) that prevent or inhibit coastline erosion. This stratification can be complex, especially when studying large swaths of coastline that exhibit a large number of patterns

controlled by various drivers, overwhelming experts' ability to establish a criterion. Rather than relying on expertise, unsupervised machine-learning classification methods can use standardised values for erosion rates as coordinates to locate each beach transect within a multidimensional space. The number of dimensions of said space is equal to the number of time periods for which the rates were measured. Therefore, two beaches with similar erosion rates across all time periods will appear very close to one another in this new feature space, which quantitatively expresses similarities in coastline behaviour in terms of their relative position (Hargrove and Hoffman 1999).

2.3.1. K-means algorithm

The K-means algorithm (Hartigan and Wong 1979) allowed us to establish separate groups of beaches that responded similarly over time in terms of coastline mobility and erosion/deposition patterns. The K-means algorithm is initially based on the number of groups specified by the user (k). An example is selected for each of these groups to serve as a reference, referred to as a seed. The seeds in our case are beach transects with erosion rates located on the extreme ends of the multidimensional feature space. Therefore, these initial seeds will correspond to beach transects with very high or very low variability. Once all beach transects have been examined sequentially and the seeds have been identified, each beach transect is assigned to a group represented by the closest seed using the Euclidean distance. Then, new coordinates are calculated for each cluster, calculating the mean for all erosion rates of the beach transects within each group. This allows us to regroup the transects based on their proximity to the reference beach transects. Thus, the references will progressively move towards the centre of each cluster. These processes are repeated iteratively until the number of beaches that change between assigned groups from one iteration to the next is very low, which means that the result is stable and that the transects within each group have a similar behaviour.

2.3.2. Classification and regression tree algorithm

Unsupervised machine-learning classification, such as K-means, enables us to group examples into homogeneous groups based on distances within the feature space. However, this grouping is based on centrality and not on identifying rules to establish separation boundaries between classes. Applying it to coastline studies allows us to establish groups, in this case coastal transects, with similar behaviour in response to the erosion/accretion variable within a time period. This method enables us to integrate large data volumes, which is becoming increasingly more common and necessary for effective coastline management, as well as to recognise behaviour patterns that can be used to create informed management policies and strategies on an intermediate (regional) scale.

Furthermore, a classification and regression tree (CART) model (Breiman et al. 1984) was built based on the K-means results to identify the erosional and depositional rate threshold values and the period that characterises each cluster or stratum. The categories obtained by the K-means algorithm and the erosion rates were combined into a set of input feature vectors to grow a CART model. A CART model comprises many different nodes. The primary node, known as the root, contains the data for all beach transects together with its group label. The interior nodes, collectively referred to as nonterminal nodes, are linked to decision stages to split the beach transects into more homogeneous nodes according to a rule. Finally, the terminal nodes (leaves) represent the final classes. Therefore, a CART represents a set of rules organised hierarchically into levels, which are successively applied down the tree from the root to a leaf.

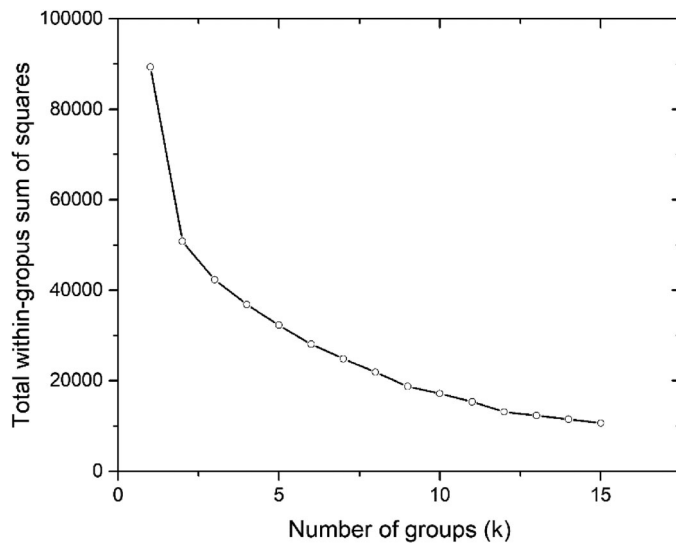


Figure 4. Sum of squared distances for different number of groups (k) in the K-means algorithm. The total within-groups sum of square (similarity between groups or classes) decreases dramatically when k is greater than 4. Models of increased complexity with more than 4 classes lead to more similar beach groups/classes.

2.3.3. Application of the Km-CART method

This study has combined the K-means and CART classification methods to create a hybrid method (supervised classification from automatic clusters) referred to from this point onwards as Km-CART, and models beach dynamics for three periods of interest (see Section 3.2.1 and 3.2.2). The Km-CART method learnt from the erosion and deposition rates of 1509 beach transects. All analyses were carried out in R software, using the ‘stats’ and ‘rpart’ packages. These beach transects were initially clustered using a K-means algorithm based on Euclidian distances in a 3D temporal feature space and corresponding to three EPR of the periods of interest (1956–1977, 1977–2001 and 2001–2011). Different models were built for different combinations of K-means outputs and CART model complexities. The K-means algorithm was built considering a number of clusters between 2 and 15. [Figure 4](#) shows the sum of the square differences when considering different numbers of k clusters. One can see that the error begins to converge slowly starting at 3 or 4 groups. The CART models were built using the outputs of the K-means algorithm (1509 beach transects) considering 3, 4, 5 and 6 classes, and validated using a 10-fold cross validation. With the goal of obtaining robust and generalisable models, we built all possible decision trees at a depth level of 2 and with a minimum number of observations per node between 10 and 50. A Km-CART model with 5 K-mean classes and a CART tree with 2 levels and 10 observations per node outperformed the rest (overall accuracy: 94.04%). This model was therefore selected for further analyses.

3. Results

3.1. Coastal evolution rates

Results are presented for the three time periods for which orthophotos were available (see [Table 1](#)): 1957–1977, 1977–2001 and 2001–2011. In the first time period (1957–1977), we observed that erosion was present in 42% of transects analysed (average of -1.43 ± 0.20 m/y) and deposition was only observed in 25% of transects ([Figure 5](#)) with

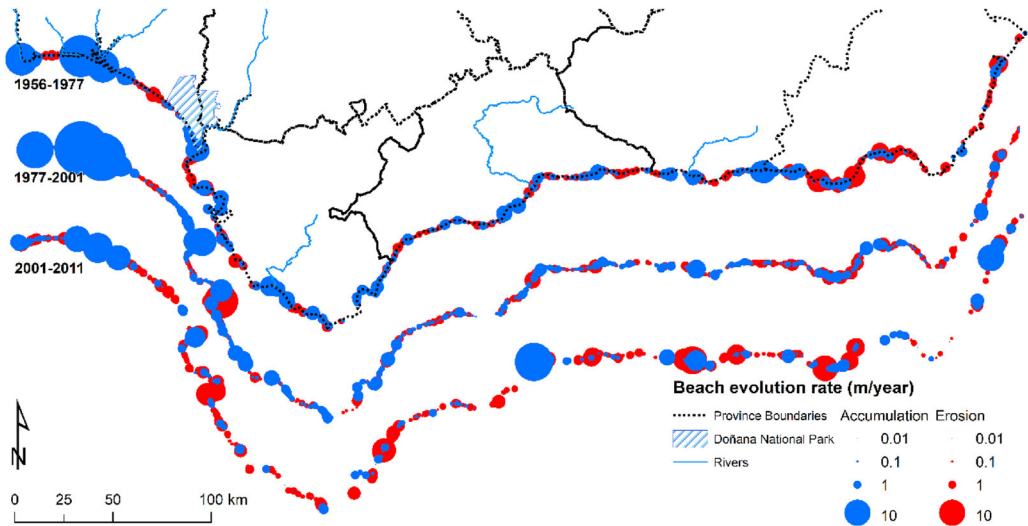


Figure 5. Beach response indicators (erosion and accretion rates, calculated as EPR) for the three periods studied (1956–1977, 1977–2001 and 2001–2011). The size of the circles (blue for deposition/accretion and red for erosion) represents the value of the rate (meters/year) and blank spaces represent stabilisation.

an average value of $+3.21 \pm 0.20$ m/y, reaching 30 ± 0.20 m/y in some sections of the Atlantic region. The second period (1977–2001) was characterised by an increased number of transects showing deposition (50%) and a small decrease in those showing erosion. However, the average accumulation rates were approximately $+3.45 \pm 0.12$ m/y, while erosion rates increased slightly compared to the previous period at approximately -1.55 ± 0.12 m/y, with a maximum rate of -19 ± 0.12 m/y on the Cádiz coastal sector. Lastly, the third period (2001–2011) saw a significant decline in the number of transects displaying both accretion (9%) and erosion (19%). This gave rise to a major increase in stable sectors of up to 72% (i.e. behaviour in these sectors did not change compared to the previous period). This change in dynamics may be due to the stabilisation of the proxy used as a result of the increased presence of coastal infrastructures. Another finding for this period was an increase in the mean erosion rate up to values of around -1.96 ± 0.23 m/y, located mainly on the Mediterranean coast.

In general, over the entire 55-year period studied, we observed an almost widespread pattern of erosive behaviour along the coast, accompanied by significant stabilisation of coastline sections over time (represented by blank spaces), this being primarily evident and significant in the last period studied (Figure 5). Average erosion and accumulation rates for the entire period studied were -1.17 m/y and 2.72 m/y, respectively. Sections with significant erosion prevailed in the first period, although at lower rates, while in the second period the number of transects showing erosion were almost equal to those showing accumulation, reaching particularly high erosion levels along the coasts of Cádiz and Huelva ($> -10 \pm 0.12$ m/y). Lastly, during the third period (2001–2011), most of the Mediterranean coast remained stable, with some sections displaying erosion along the coasts of Malaga and Granada, and some areas of accumulation along the coastline in Huelva and in certain Mediterranean deltas. For a full description of the results, see Prieto et al. (2018). This results are pretty much in line with those from Molina et al. (2019) that showed a negative net balance of $29,738.4 \text{ m}^2/\text{year}$ corresponding to the loss of 1784.30 km^2 of beach surface in the 1956–2016 period.

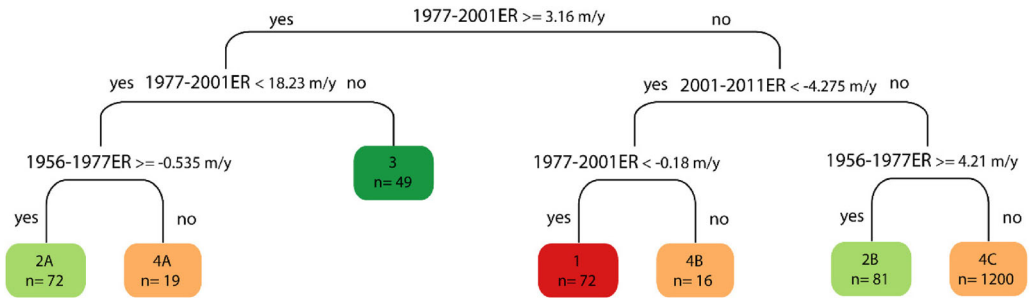


Figure 6. Km-CART model showing the different categories based on the erosion/accretion behaviour of all 1509 beach transects. Note that the second period (1977–2001) appears as the main driver of long-term beach behaviour. The model identified four classes which are colour coded.

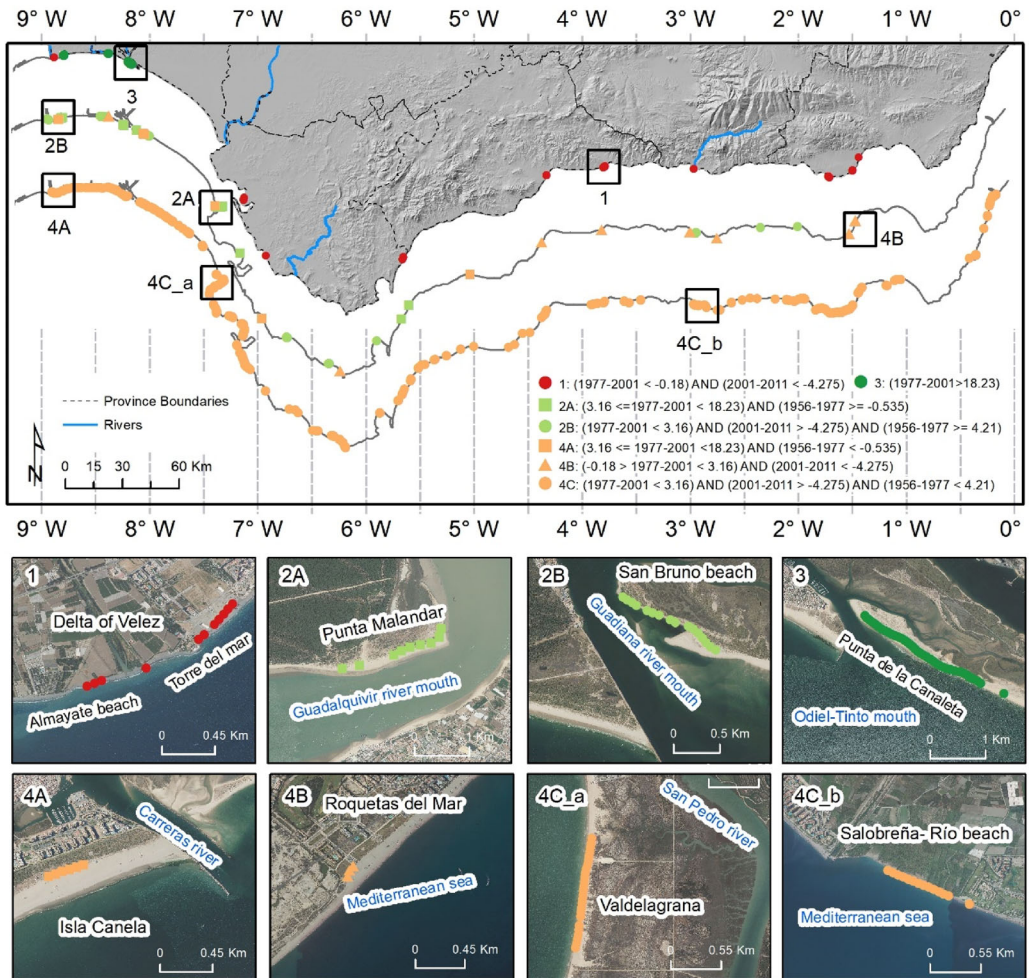


Figure 7. Spatial representation of beach response categories. Coloured shapes represent the different classes identified using the Km-CART model. Insets labelled as 1, 2A, 2B, 3, 4A, 4B, 4C_a and 4C_b are representative examples of each category.

Table 4. Cluster centroids extracted using a K-means algorithm for the three classes and periods studied.

Class	EPR 1956–1977 ± 0.20 (m/y)	EPR 1977–2001 ± 0.12 (m/y)	EPR 2001–2011 ± 0.23 (m/y)	Size	Error
1	−0.78	−5.04	−6.12	103	4383.838
2	5.55	5.68	2.85	129	13199.077
3	9.13	25.02	6.60	51	7003.972
4	−0.32	−0.01	−0.36	1226	12254.806

The mean EPR value (negative erosion and positive accretion) is shown for each class. Size refers to the number of beach transects in that class. The sum of squares within the cluster is computed as the summatory of the distance of the observation to their assigned cluster centres. A low value of sum of squares means homogeneity within the class.

3.2. Results from stratification of beach response indicators

The four categories resulting from the Km-CART model are presented in a decision tree diagram with two levels (Figure 6), which has also been geographically spatialised as a map (Figure 7). Both the diagram and the map are colour coded based on the clusters identified by the model. Also, the cluster centroids used to separate classes based on the K-means algorithm are shown in Table 4. Of the seven leaf nodes in the model, it can be observed that several belong to the same class. This occurred in Class 2 (patterns A and B) and Class 4 (patterns A, B and C), and is the result of the subclasses or patterns being defined by different rules, thus being conditioned by different drivers in some of the periods that were studied.

The classification tree was able to correctly classify 1415 of 1509 beach transects (93.77%) into the four classes identified by the K-means algorithm. Table 5 shows the confusion matrix obtained from the cross-validation that was performed during Km-CART model training. Commission and omission errors were low in general, except in classes 1 and 2. Class 1 was overestimated by 32.04% due to the inclusion of 34 beach transects that corresponded to Class 4. On the other hand, Class 2 was underestimated by 27.45% due to 39 beach transects that corresponded to Class 2 being included under Class 4. It was possible to observe how erroneous transitions between classes occurred primarily with Class 4 as it was the predominant class (Table 5).

The overall results showed that the period between 1977–2001 appears to be the main driver of long-term beach behaviour and discriminates between erosive and depositional shorelines for the whole dataset (1509 beach transects). Although all categories are clearly represented along the coast, 79.5% of beach transects fell under Class 4C, which is indeed found on both the Atlantic and Mediterranean coasts (Figures 6 and 7). The first level of the Km-CART model grouped beaches based on their behaviour in the second period (1977–2001), discriminating between those with an accretion rate greater than or equal to 3.16 ± 0.12 m/y (9.3% of beach transects) or less (90.7% of beach transects). The second level classified beach transects for the third period (2001–2011) with an erosion rate greater than or less than -4.275 ± 0.23 m/y (Figure 6). The third level discriminated between beach transects with an accretion rate greater than or equal to 4.21 ± 0.20 m/y for the first period (1956–1977) (classes 2B and 4C) and beach transects that experienced erosion ($EPR < -0.18 \pm 0.12$ m/y) for the second period (1977–2001), labelled as 1 and 4B. At the second level on the opposite side of the tree, beach transects in the second period (1977–2001) are discriminated based on whether they showed an accretion rate $> 18.23 \pm 0.12$ m/y (Class 3) and those with $EPR < 18.23 \pm 0.12$ m/y and that displayed an erosion rate in the first period $\geq -0.535 \pm 0.12$ m/y (Class 2A) or < -0.535 (Class 4A) (Figure 6).

Class 1, which represents 4.7% of beach transects ($n=72$), is comprised of systems that displayed erosion rates ranging between -0.22 and -17 ± 0.12 m/y in the second

Table 5. Confusion matrix for the Km-CART model.

	C1	C2	C3	C4	TOTAL	E. COMISIÓN
C1	70	1	0	32	103	32.04
C2	0	111	0	18	129	13.95
C3	0	2	49	0	51	3.92
C4	2	39	0	1185	1226	3.34
TOTAL	72	153	49	1235	1509	
E. OMISIÓN	2.78	27.45	0	4.05		

References appear in columns (test), while the predicted beach transect classes (map) appear in rows. Agreements between the reference and the map appear across the main diagonal, while the remaining cells are interpreted as errors. Commission errors report the percentage of beach transects that are misclassified on the map and omission errors report the percentage of beach transects that are missed on the map according to the test.

period (1977–2001) and $>-4.33 \pm 0.23$ m/y in the third period (2001–2011). The balance for all periods (1977–2011) was therefore negative with regard to erosion rates, with values ranging between $EPR = -0.36 \pm 0.12$ m/y and -8.69 ± 0.12 m/y. It primarily comprised Mediterranean deltas and beaches with a clear erosional trend over the past 55 years. This pattern, however, is intensified over the last decade for most of the beach transects in this category. The presence of coastal infrastructures (ports, seawalls and breakwaters) may have led to a disruption of the natural longshore drift resulting in highly erosive areas. An example shown in [Figure 7](#) (inset 1) is the Vélez River Delta located on the Mediterranean coast between the Almayate and Torre del Mar beaches in Malaga Province. Other examples include the Guadalfeo River (Motril) and beaches associated with the Punta Entinas-Sabinar system (Almeria Province on the east Mediterranean coast, see [Figure 1](#)). The cluster centroids based on the K-means algorithm (see [Table 4](#)) increased towards the third period (2001–2011), reaching $EPR = -6.12 \pm 0.23$ m/y.

For beach transects that had an accumulation rate in the second period (1977–2001) greater than 3.16 ± 0.12 m/y, the second level of the Km-CART model discriminated between beach transects with an accretion rate of less than 18.3 ± 0.12 m/y for the same period: Class 2 A (4.77% of beach transects) and Class 4 A (1.25% of beaches). Class 2 A is associated with the large depositional systems along the western coast of Andalusia ($EPR > 19 \pm 0.12$ m/y). The availability of sediments, both of a natural and anthropic origin, along with an intense longshore drift (W–SE) may have led to the existence and development of these coastal systems. A good example is Punta Malandar ([Figure 7, 2A](#)) at the mouth of the Guadiana River, where the longshore drift drives sediment transport, influencing the dynamicity of these shorelines located on the Atlantic margin. Class 4 A, located on the opposite side of the tree at this same level, is where the model classified beach transects that experienced erosion rates of less than -0.535 ± 0.20 m/y during the first period (1956–1977). This class is observed downstream of the littoral drift direction of the main Atlantic rivers (i.e. Tinto, Carreras, Odiel), where the presence of man-made dykes upstream may have led to sediment accumulation along these coastal sections after 1977. It also comprises beach transects with significant inter-annual fluctuations in shoreline position due to their proximity to large Atlantic estuaries in Huelva Province, such as the Guadiana and Guadalquivir rivers and, to a lesser extent, the Piedras River (see [Figures 1](#) and [7, 4A](#)).

The other pattern within this class was found in Class 2B (5.36%, $n = 81$), which corresponded to historically depositional and currently erosional beaches apparently due to artificial interruption (e.g. dams) of sediment input from estuaries. The majority corresponded to Atlantic estuarine systems that were traditionally responsible for sedimentary input, but which are currently experiencing a sediment deficit ($EPR > -4.275 \pm 0.23$ m/y). A typical example of this class is Isla Canela, located in the mouth of the Guadiana River

(Figure 7, 2B). Despite this class being found primarily along the Atlantic coast (associated with large estuaries), it was also found in some areas along Andalusia's Mediterranean coast, as these areas show fluctuations that may be result of the construction of leisure ports and fishing marinas, or the stabilisation of deltas for agricultural purposes. The cluster centroid for this class changed from 5.55 ± 0.20 m/y in the first period, to 5.68 ± 0.12 m/y in the second period and down to approximately 2.85 ± 0.23 m/y towards 2001–2011.

Class 3 (1977–2001 $EPR > 18.23 \pm 0.12$ m/y) was not well represented along the coast (3.24%). This class only occurred on the Atlantic side and was associated with significant sedimentary systems such estuaries, but with a very localised scope (i.e. the Carreras estuary, Rompido spit and the mouth of the Odiel-Tinto River, as shown in Figure 7, 3). This class also displayed deposition across all periods, with the second period (1977–2001) being particularly relevant, with EPR values of around 25 ± 0.12 m/y. This class presented a positive cluster centroid value for all periods, with values of 25.02 ± 0.12 m/y in the second period (1977–2001) where these systems underwent maximum accretion.

Class 4B, a subclass of Class 4, represented 1% ($n = 16$) of the beach transects in this study. These systems are historically depositional (between 1957–2001) as they fall within the area of influence of Mediterranean deltas, however in the third period (2001–2011) they showed intense erosion on the order of -4.35 ± 0.23 m/y and -10.7 ± 0.23 m/y ($-0.18 > 1977-2001 < 3.16$ and $2001-2011 < -4.275$). Examples of this class are the Guadalhorce River in Malaga (Figure 1), Torre del Mar (connected to the mouth of the Vélez River) and some sections (beach transects) in Roquetas de Mar in Almería (Figure 7, 4B). The erosion/deposition balance for these beaches is negative between 1956–2011.

Lastly, 4C (79.5%, $n = 1200$), the predominant subclass of Class 4, contains those systems that displayed deposition rates greater or equals to 4.21 ± 0.12 m/y for the first period and erosion rates less than -4.275 ± 0.23 m/y in the third period (2001–2011), and that also demonstrated different behaviour to the other periods (Figure 7). Most beach transects ($n = 550$, 45.8%) were traditionally stable or depositional given their orientation with respect to longshore drift. However, more recently (from 1977 onwards or from 2001–2011) these show severe erosion and a reduction in sediment input that may be the result of changes in soil use and intense urbanisation along the coast. The EPR for these beach transects reached values above -4.275 ± 0.23 m/y during the last period. Although these beach transects are broadly distributed throughout the Andalusian coast, urban beaches such as Valdelagrana (Cádiz) and Salobreña (Granada) stand out as some good examples (Figure 7, 4C_a,b). In fact, under the 4C subclass there were a large number of beach transects that demonstrated a slight variance in patterns: those that showed an erosive pattern since the first period (331); a subset that were traditionally depositional, but had displayed erosion rates in the last two periods (138); and those that were positive (depositional) in all three periods, but with very low rates that were borderline stable (120). On the other hand, this subset also included erosive beaches (with minimal range) that, over recent decades, had become depositional or stabilised (299). Class 4 displayed a negative centroid across all periods, but with very low values, especially in the second period (1977–2001) where the cluster centroid was almost zero (-0.01 ± 0.12 m/y), thus demonstrating the variance of the class and its subclasses, as well as the stability of some systems.

Given the great amount of beach transects classified under the class 4C, a new model was built for this class only with the aim of better understand their behaviour during the three periods. Results of this sub-model are presented in the Supplementary section.

4. Discussion

Coastal monitoring frequently involves the understanding of shoreline change through the calculation of conventional shoreline metrics over a specific time frame. The new advances in technologies have allowed to extend the analysis of changes in shoreline position to a broader scale (with higher temporal as well as spatial resolutions), which in turn could help to identifying trends and magnitudes of relative shoreline change over time (Burningham and French 2017). Notwithstanding, beyond the quantification of the rate of change (erosion/deposition vs advance or retreat of a given portion of coast) (Banks et al. 2017, Dai et al. 2019), understanding regional drivers, non-linear coastal changes or broader regional dynamics (Hapke et al. 2016) often require more sophisticated approaches.

In this study, the application of a new hybrid methodology that combines automatic K-means clustering and a supervised CART system to stratify beach behaviour (beach shoreline metrics) is applied to the coast of Andalusian. Results from shoreline dynamics indicators (analysed for three consecutive periods: 1957–1977, 1977–2001 and 2001–2011) obtained through the use of a proxy (contact between the backshore and dunes, fore-dunes, infrastructure or cliffs) enabled us to analyse the evolution of different sections of coastline over the past 55 years (Figure 7).

4.1. 55-years of shoreline change: the influence of anthropogenic activities

As demonstrated by many authors (e.g. Manno et al. 2016, Molina et al. 2019, Villar Lama and Ojeda 2007), the Andalusian coast has been influenced by accentuated regional land-use changes (agricultural-use land converted into urban areas and the proliferation of greenhouse agriculture on the Mediterranean deltas), massive coastal urbanisation beginning in the 1960s and with a peak during the 1970s (and the resulting construction of coastal infrastructure, such as ports, dikes, wharfs) as well as the artificial stabilisation of large sections of coastline (e.g. construction of seafront promenades and sand nourishment).

In this sense, the fact that the first period shows sections that are predominantly erosive (although at a medium/low rate of -1.43 ± 0.20 m/y) indicates the significant coastal transformation that occurred during the '60s and '70s when massive urban expansion was undertaken along the coastline, especially on the Mediterranean side, which was accompanied by the construction of water-regulating infrastructure (e.g. reservoirs) (Del Río and Malvarez 2017, Malvarez 2012, Garel and Ferreira 2011, Del Río et al. 2020). This fact, along with land-use changes associated with alterations in the agricultural production model and the artificialisation of the coastline (Villar Lama and Ojeda 2007), caused changes to sedimentary dynamics (and longshore drift flows due to the construction of dikes and wharfs), affecting the natural behaviour of many sections of coastline (Senciales and Malvarez 2003, Molina et al. 2019). Some examples on the Mediterranean coast includes the construction of Puerto Banús port (1970), groynes and breakwaters in the eastern coast of Malaga and Granada and the main dock of the port of Almerimar (Almeria) in 1978, among others (Molina et al. 2019). In fact, as stated by Manno et al. (2016) in the Mediterranean coast of Andalusia the length of armoured coastal sectors increased from 42.1 (1956) to 98.2 Km (1977). In 20 years, the numbers of ports doubled (from 10 in 1956 to 20 in 1977) and the progressive emplacement of groins constituted 1.3 km of armoured coastline in the 50s an increased to 20.6 km in 1977 (more than 47 groins built in two decades) (Manno et al. 2016).

In the second period studied (1977–2001), this sedimentary imbalance was evident in many sections that presented chronic erosion. However, the number of coastal transects that showed a positive balance (deposition) also increased, in most cases due to the presence of coastal engineering infrastructure that facilitate short-term beach stabilisation (Malvárez et al. 2015, Morales et al. 2004) or associated with the large Atlantic estuarine systems where accumulation values reached approximately 12 ± 0.12 m/y (Morales et al. 2004) (Figure 7). As some authors have demonstrated, the first stabilised areas in these coastal areas were observed with the appearance of seafront promenades and occupation of the Maritime Terrestrial Public Domain (DPMT) (Górgolas 2019, Malvarez 2012, Manno et al. 2016, Molina et al. 2019). In the Mediterranean during the 80 s-90s the armoured coastline increased to 182.3 km, which included new ports, enlargement of existing ones and emplacement of 97 new infrastructures (88 groins and 9 breakwaters) (Manno et al. 2016). During this period, it's also relevant to highlight the construction of several dams, such as the case of Vélez River where since 1988, several dams were emplaced in the basin of the river that significantly reduced fluvial sedimentary load, or the Benimar Dam (1988) in the Adra River (Almería coast), or the Guadalfeo mouth that lead to the emplacement of about 100 small groins along the eastern side of the delta (Prieto et al. 2012). This placement of coastal infrastructures and periodic nourishment works brought the opposite side of the deltas to show progradation (Molina et al. 2019, Guisado-Pintado and Malvárez 2015).

The last period is dominated by a predominance of this 'artificial' stabilisation of beaches that resulted in almost 50% of stable sections, especially along the Mediterranean coast where most beaches undergo artificial maintenance (i.e. coastal regeneration performed via continuous replenishment of beach sediments). In this period, some coastal transects still show predominantly erosive behaviour with an average value of -1.96 ± 0.23 m/y, while coastal sections showing accumulation are concentrated around the large Atlantic sedimentary systems. In terms of coastal infrastructures, in the Mediterranean coast, according to Manno et al. (2016) this period showed a slightly increased in the armouring coastlines which reaches 197.3 km with 947 groins/breakwaters and 39 seawalls and revetments observed in 2010.

The erosion/accumulation rates analysed using the End Point Rate indicator (EPR, m/y) are those that enabled the model to establish different clusters or classes that use rules to explain the behaviour of different beaches, grouping these according to a series of explanatory variables. The fact that the second period (1977–2001) was the determining period that divided the classes makes sense, as it was during that period when urban expansion was undertaken along the coastline. This infrastructures had a massive impact on systems compared to the previous period (construction of seafront promenades, dams, leisure ports, land-use changes) and caused beach stabilisation in some cases such as the sector in 'La Linea de la Concepción' at the eastern Mediterranean side of the Strait of Gibraltar (see Molina et al. (2019) for further details on the Mediterranean coast). In fact, the first classification (1977–2001 $EPR \geq 3.16 \pm 0.123$ m/y) displayed in Figure 6 already discriminates between systems with an erosive pattern (classes 1, 4B and 4C) and systems more prone to accumulation (2A, 2B, 3 and 4A).

In general, there was good agreement between the results of the Km-CART stratification into four classes and the general coastline dynamics observed along the Andalusian coast over the past 55 years, during which time periods of accretion and erosion were driven by both natural and anthropogenic processes. This is also in line with previous works developed in the Mediterranean coast of Andalusia (Molina et al. 2019), that showed that depositional beaches were essentially observed in areas associated with coastal

infrastructures (up-drift of ports and groins, breakwaters) whereas erosion is associated with to down-drift areas of ports and groins and along the mouths of largest rivers and deltas.

4.2. The complexity of shoreline dynamics: past behaviour and trends

The model shows that Class 4C was the largest, including 79.5% of beach transects studied in both the Atlantic and Mediterranean margin, with a predominantly erosive pattern over the past 55 years due to intense urbanisation activity and the artificialisation of the coastline (seafront promenades and breakwaters, see Manno et al. (2016) and Malvarez (2012) for detailed cases in the Mediterranean), which brings about the loss of natural elements such as dunes. This is related to what was observed for erosion rates, where we can see that the pattern of behaviour over past decades shifted towards chronic erosion, especially in Mediterranean environments. Examples of this category can be found at the westernmost coast of Malaga province where large number of coastal infrastructures and ports (e.g. La Duquesa, Estepona, and Puerto Banús) have been developed. Another example is Salobreña beach (Figure 7_4C_b) where since the construction of a dam (2005) constant erosion has been recorded despite the lately efforts carried out by beach nourishment initiatives (Félix et al. 2012). Similarly, the model classified 8% (2 A and 3) of sectors located at the mouths of large estuaries and adjacent areas as accretional systems over the past decade. A submodel only accounting for beach transects classified as 4C ($n = 1200$) was built to better understand the rules that describe these locations (see Supplementary section).

Classes 1 and 3 appeared as opposites to one another: Class 1 was associated with systems displaying erosion across all periods studied and principally associated with beaches in the Mediterranean. These beaches witnessed the expansion of tourism in the 1960s as well as significant land-use changes, as detailed by Senciales and Malvarez (2003) concerning the Vélez River Delta (Malaga Province), or the example of Adra river basin transformation during the last decades discussed by Molina et al. (2019). Class 3 represents large Atlantic sedimentary systems where there has been a pattern of accumulation over the past 50 years, with inter-annual and inter-decadal variations, and whose functioning appears to respond to the large accretional contribution made by rivers, such as the Odiel and Guadiana Rivers (Morales et al. 2019). These classes demonstrated very little variation within the class itself (Table 4), with all beach transects appearing under the same leaf node and thus sharing the same rules.

Other intermediate classes, subtypes of classes 2 and 4, are represented under 2B and 4 (A and B). Class 2B, which is very similar to Class 3 with cluster centroids showing lower deposition rates (Table 4), is an example of systems linked to river mouths that have also seen fluctuations resulting from river management, which is particularly evident starting in the 1970s. For their part, classes 4A and 4B are opposites. Class 4A demonstrated a positive change rate from 1956–2011 with an EPR between $+0.76$ and $+4.34 \pm 0.20$ m/y and corresponded to beaches in areas near Atlantic river mouths, while Class 4B was concentrated along the Mediterranean coast and showed a deficit for the entire period, which may be linked to the anthropogenic use of these beaches. Both cases are quite heterogeneous internally, as is evident from the error values in Table 4, meaning the beaches in these classes react to different internal rules. This could be explained by the fluctuations suffered by these beaches during the past decades and for the wave regime and long-shore drift. For instance, beach transects only separated few meters can depict opposite behaviour, this is those located up-drift could record accretion respect to down-drift areas were

erosive patterns are dominant (Jonathon et al. 2001). Further, coastal management practices in erosive areas commonly implies the progressive emplacement of new structures to prevent or reduce erosion (e.g. Manno et al. 2016, Malvarez 2012). The practice unlike fixing the problem it usually generates the so called ‘domino’ effect (Cooper et al. 2009) and normally implies periodic artificial nourishments to maintain this beach which in turns could be interpreted as progradation or accretion. In any case, and particularly in these classes, shoreline evolution is rarely uniform (Burningham and French 2017, Molina et al. 2019, Malvarez et al. 2015) and some beaches could experiment inversion of trends (from erosion to accretion) induced by the emplacement of coastal structures and/or artificial nourishments, e.g. Roquetas de Mar (Figure 7, 4B).

4.3. The Km-CART model: limitations and new steps

The classification tree was able to correctly classify 1415 out of 1509 beach transects (93.77%), which corresponded to the four classes identified by the K-means algorithm, while erroneous transitions between classes occurred primarily in Class 4, as shown in the confusion matrix (Table 5). This may be due not only to the fact that the latter class was the largest (1234 transects), but also that the study periods used for the beach response indicators had uneven durations (21, 24, and 10 years, respectively), which could be concealing inter- and intra-annual behaviours with more dynamic variability, for example those corresponding to classes 1 and 3. Furthermore, transects used for the discrimination of classes are displayed every 50 meters. Therefore, only groups which were composed of spatially continuous transects with similar patterns were considered to be representative of a class. Isolated transects within a beach section are not considered as representative of the behaviour of a beach. Future studies should consider contextual information to characterise beach classes.

The use of ML techniques include added benefits, such as: i) the ability to learn complex patterns, considering the nonlinear relationships between the study variable and auxiliary variables; ii) they are capable of providing generalisations, as they can be applied to new cases of the same problem; iii) they are capable of incorporating different types of variables into the analysis: continuous, ordinal, categorical; iv) statistical distribution is independent of the data (normality of variables is not assumed) (Coimbra et al. 2015, Burningham and French 2017, Goldstein et al. 2019). However, some algorithms, despite their potential to classify or estimate precise coastal information, can also be complex in their application or difficult to interpret. In this regard, we have used the Km-CART method to show a simple way to stratify beaches in terms of their erosive behaviour while also generating knowledge that can be understood by specialists and coastal managers. We have converted patterns using natural language, selecting the most important variables and their cut-off or threshold values. While the model we have developed here only applies to the Andalusian coastline, the proposed method can be applied to any coastal area and can be used to stratify erosion, as well as study the behaviour of other types of issues where no predefined classes exist, but where there are groups with different behaviours susceptible to analysis (response to storms, nearshore bar evolution, development and behaviour of foredunes).

Improvements could be applied to the stratification of beaches according to their erosional/depositional behaviour and the main drivers that affect their dynamics (such as wave climate). Furthermore, new predictive models could be built from beach stratifications and using different environmental variables, including land-use changes, storms magnitude and frequency, climatic indices (e.g. precipitation) etc. This will allow for

objective and more accurate identification of the main territorial and environmental drivers of coastline changes. Results showed good performance of machine-learning methods, and, more specifically, of Km-CART, in modelling medium- to long-term shoreline dynamics. Decision trees are algorithms that offer a great amount of transparency, as they provide information through the split rules used to separate examples into classes, identifying the most important variables for each class as well as the threshold values used to divide said variables. However, under certain circumstances this high interpretability can also be concealed due to increased noise sensitivity, which may negatively impact classification performance (Rodríguez-Galiano and Chica-Rivas 2014). There have been new developments in this regard that, despite being less interpretable, may provide a more precise stratification, such as Random Forest or Support Vector machines (Rodríguez-Galiano et al. 2012, Rodríguez-Galiano and Chica-Rivas 2014).

5. Conclusions

- The evolution of the Andalusian coast over time (the past 55 years) has been influenced by the dynamicity of large river systems (Atlantic estuaries), regional land-use changes (agricultural-use land converted into urban areas and greenhouse agricultural practices on Mediterranean deltas), massive coastal urbanisation beginning in the 1960s and 1970s (and the resulting construction of coastal infrastructure, such as ports, dikes, wharfs) as well as the artificial stabilisation of large sections of coastline (e.g. construction of seafront promenades).
- The complexity of shoreline dynamics (non-linear behaviour, intra-decadal changes) and the recent availability of datasets highlights the need of combining the quantification of shoreline metrics with any sort of analysis to extract trends of mesoscale beach behaviour.
- The proposed Km-CART methodology enables stratification and characterisation of complex, continuous multivariate transects of decadal beach behaviour into homogeneous groups for easier interpretation. Andalusian beaches were separated into four classes with different erosive behaviours according to the beach dynamic indicator, End Point Rate (EPR), during three periods: 1956–1977, 1977–2001, 2001–2011.
- Km-CART was able to correctly classify 1415 out of 1509 beach transects (93.77%), corresponding to the four classes based on the following rules: First, an initial classification was performed ($1977\text{--}2001 \text{ EPR} > = 3.16 \pm 0.12 \text{ m/y}$) to distinguish between systems with an erosive pattern (classes 1, 4B and 4C) and systems more prone to accumulation (2A, 2B, 3 and 4A).
- Then, each of the four classes had well-defined rules: Class 1 comprises beach transects that presented erosion in all periods; Class 2, with two subclasses: 2A for beach transects with accretion in all periods except for the first, and 2B for beach transects that are traditionally prograde but erosive in the last period; Class 3 comprises systems displaying significant accretion across all periods; Class 4, with three subclasses: 4A, beach transects with positive rates; 4B, historically prograde beaches that experienced intense erosion during the last period; and 4C, beach transects with a high amount of internal variability but with coastline recession in the last period (See supplementary section).
- The confusion matrix that was obtained based on cross-validation showed that commission and omission errors were low in general, with classes 1 and 2 being the exception. We observed how erroneous transitions between classes occurred primarily with Class 4. It is also possible that behaviours in shorter time series for this group of

beaches may be concealed by the number of beaches that fell into this category as well as the different durations of the periods studied.

This study is an analysis of the erosive state of the Mediterranean and Atlantic beaches along the Andalusian coastline, solely taking into account beach response indicators based on the computation of shoreline erosion rates. Future research areas may benefit from applying ML methodologies to identify territorial drivers that have contributed to changes in the erosive state of beaches, such as land-use changes, wave climate, coastal management approaches.

Apart from its ease of application and the interpretability of its results, the Km-CART method has many advantages for stratifying multidecadal erosional and depositional dynamics of beaches or for use in coastal studies, such as: (i) the ability to handle complex data from different statistical distributions, responding to nonlinear relationships between variables; and (ii) producing results that are relatively unaffected by redundant data or outliers.

The methodology presented in this study can be applied to other coastal areas and other types of issues where no predefined classes exist, but where groups need to be formed using different meaningful variables, identifying the most important variables in the process as well as the cut-off threshold values required to separate these into homogeneous groups.

This first stratification and the coherence observed within each of the four classes shows that regional monitoring programmes could achieve considerable cost savings by strategically targeting monitoring activity at representative sites within each class, as well as increasing the availability of high resolution orthophotos in order to better identify intra-annual shoreline changes.

Acknowledgements

The authors would like to express their gratitude for the financial support provided under projects RTI2018-096561-A-I00 and US-1262552, the former funded by the ‘Ministerio de Ciencia e Innovación,’ ‘Agencia Estatal de Investigación’ and the European Regional Development Fund (ERDF), and the latter funded by the ‘Junta de Andalucía’ and ERDF. This work is a contribution to the UNESCO IGCP Project 639 ‘Sea Level Change from Minutes to Millennia.’

Authors’ contributions

Rodríguez-Galiano and Guisado-Pintado devised the idea. Prieto-Campos photointerpreted and computed erosion rates. Rodríguez-Galiano developed the theory and performed modelling. Guisado-Pintado verified the analytical results. All authors discussed the results and contributed to the final manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Ministerio de Ciencia, Innovación y Universidades.

ORCID

Víctor Rodríguez-Galiano  <http://orcid.org/0000-0002-5422-8305>

References

- Anders FJ, Byrnes MR. 1991. Accuracy of shoreline change rates as determined from maps and aerial photographs. *Shore Beach*. 59:17–26.
- Banks S, Millard K, Behnamian A, White L, Ullmann T, Charbonneau F, Chen Z, Wang H, Pasher J, Duffe J. 2017. Contributions of actual and simulated satellite SAR data for substrate type differentiation and shoreline mapping in the Canadian arctic. *Remote Sens*. 9(12):1206.
- Benavente J, Plomaritis T, Del Río L, Puig M, Valenzuela C, Minuzzi B. 2014. Differential short- and medium-term behavior of two sections of an urban beach. *J Coast Res*. 70:621–626.
- Beuzen T, Harley MD, Splinter KD, Turner IL. 2019. Controls of variability in berm and dune storm erosion. *JGR Earth Surf*. 124(11):2647–2665.
- Beuzen T, Splinter KD, Marshall LA, Turner IL, Harley MD, Palmsten ML. 2018. Bayesian networks in coastal engineering: distinguishing descriptive and predictive applications. *Coast Eng*. 135:16–30.
- Beuzen T, Splinter KD, Turner IL, Harley MD, Marshall L. 2017. Predicting storm erosion on sandy coastlines using a Bayesian network. *Australasian Coasts and Ports 2017 Conference*, 102–108.
- Boak EH, Turner IL. 2005. Shoreline definition and detection: a review. *J Coast Res*. 214:688–703.
- Breiman L, Friedman JH, Olshen RA, Stone CJ. 1984. *Classification and regression trees*, Monterey, Belmont, CA: Chapman and Hall/CRC.
- Bulteau T, Baills A, Petitjean L, Garcin M, Palanisamy H, Le Cozannet G. 2015. Gaining insight into regional coastal changes on La Réunion island through a Bayesian data mining approach. *Geomorphology*. 228:134–146.
- Burningham H, French J. 2017. Understanding coastal change using shoreline trend analysis supported by cluster-based segmentation. *Geomorphology*. 282:131–149.
- Carrero R, Navas F, Malvárez G, Guisado-Pintado E. 2014. Artificial intelligence-based models to simulate land-use change around an estuary. *J Coast Res*. 70:414–419.
- Casella E, Rovere A, Pedroncini A, Stark CP, Casella M, Ferrari M, Firpo M. 2016. Drones as tools for monitoring beach topography changes in the Ligurian Sea (NW Mediterranean). *Geo-Mar Lett*. 36(2): 151–163.
- Coimbra R, Immenhauser A, Olóriz F, Rodríguez-Galiano V, Chica-Olmo M. 2015. New insights into geochemical behaviour in ancient marine carbonates (Upper Jurassic Ammonitico Rosso): novel proxies for interpreting sea-level dynamics and palaeoceanography. *Sedimentology*. 62(1):266–302.
- Coimbra R, Rodríguez-Galiano VF, Oloriz F, Chica-Olmo M. 2014. Regression trees for modelling geochemical data- an application to Late Jurassic carbonates (Ammonitico Rosso). *Comput Geosci*. 73: 198–207.
- Cooper A, Anfuso G, Del Río L. 2009. Bad beach management: European perspectives.
- Coyne MA, Fletcher CH, Richmond BM. 1999. Mapping coastal erosion hazard areas in Hawaii: observations and errors. *J Coast Res*. 28:171–184.
- Crowell M, Leatherman SP, Buckley MK. 1991. Historical shoreline change: error analysis and mapping accuracy. *J Coast Res*. 7:839–852.
- Dai C, Howat IM, Larour E, Husby E. 2019. Coastline extraction from repeat high resolution satellite imagery. *Remote Sens Environ*. 229:260–270.
- Darwin N, Ahmad A, Zainon O. 2014. The potential of unmanned aerial vehicle for large scale mapping of coastal area. *IOP Conf Ser: Earth Environ Sci*. 18:012031.
- Del Río J, Malvarez G. 2017. Coastal erosion and sediment accumulation in reservoir lakes: the case of La Concepción in the Headland Bay of Marbella, Málaga. In: *Geotemas*, ed. IX Jornadas de Geomorfología Litoral (Menorca, 2017), pp. 251–254.
- Del Río JL, Malvárez G, Navas F. 2020. Reservoir Lake effects on eroded littoral systems: the case of the Bay of Marbella, Southern Spain. *J Coast Res*. 95(sp1):443–447.
- Del Río L, Gracia F. 2013. Error determination in the photogrammetric assessment of shoreline changes. *Nat Hazards*. 65(3):2385–2397.
- Emery KO. 1961. A simple method of measuring beach profiles. *Limnol Oceanogr*. 6(1):90–93.
- Félix A, Baquerizo A, Santiago J, Losada M. 2012. Coastal zone management with stochastic multi-criteria analysis. *J Environ Manage*. 112:252–266.

- Fernandez-Nunez M, Díaz-Cuevas P, Ojeda J, Prieto A, Sánchez-Carnero N. 2015. Multipurpose line for mapping coastal information using a data model: the Andalusian coast (Spain). *J Coast Conserv.* 19(4): 461–474.
- Ferreira Ó, Plomaritis TA, Costas S. 2019. Effectiveness assessment of risk reduction measures at coastal areas using a decision support system: findings from Emma storm. *Sci Total Environ.* 657:124–135.
- Fletcher C, Rooney J, Barbee M, Lim S-C, Richmond B. 2003. Mapping shoreline change using digital orthophotogrammetry on Maui, Hawaii. *J Coast Res.* SI(38):106–124.
- Ford M. 2013. Shoreline changes interpreted from multi-temporal aerial photographs and high resolution satellite images: Wotje Atoll, Marshall Islands. *Remote Sens Environ.* 135:130–140.
- Garel E, Ferreira Ó. 2011. Effects of the Alqueva Dam on Sediment Fluxes at the Mouth of the Guadiana Estuary. *J Coast Res.* 64:1505–1509.
- Giardino A, Diamantidou E, Pearson S, Santinelli G, Heijer C. 2019. *A regional application of Bayesian modeling for coastal erosion and sand nourishment management.*
- Goldstein EB, Coco G, Plant NG. 2019. A review of machine learning applications to coastal sediment transport and morphodynamics. *Earth Sci Rev.* 194:97–108.
- Górgolas P. 2019. Del «urbanismo expansivo» al «urbanismo regenerativo»: directrices y recomendaciones para reconducir la herencia territorial de la década prodigiosa del urbanismo español (1997- 2007). Aplicación al Caso de Estudio Del Litoral Andaluz. *LI(199):*81–100.
- Grimes DJ, Cortale N, Baker K, Mcnamara DE. 2015. Nonlinear forecasting of intertidal shoreface evolution. *Chaos.* 25(10):103116.
- Guisado-Pintado E, Jackson DWT. 2018. Multi-scale variability of storm Ophelia 2017: the importance of synchronised environmental variables in coastal impact. *Sci Total Environ.* 630:287–301.
- Guisado-Pintado E, Jackson DWT. 2019. Coastal impact from high-energy events and the importance of concurrent forcing parameters: the cases of storm Ophelia (2017) and Storm Hector (2018) in NW Ireland. *Front Earth Sci.* 7:1–18.
- Guisado-Pintado E, Jackson DWT. 2020. Monitoring cross-shore intertidal beach dynamics using oblique time-lapse photography. *J Coast Res.* 95(sp1):1106–1110. 5.
- Guisado-Pintado E, Jackson DWT, Rogers D. 2019. 3D mapping efficacy of a drone and terrestrial laser scanner over a temperate beach-dune zone. *Geomorphology.* 328:157–172.
- Guisado-Pintado E, Malvárez G. 2015. El rol de las tormentas en la evolución morfodinámica del Delta del río Vélez: Costa del Sol, Málaga. *Geo-Temas.* 15:189–192.
- Guisado-Pintado E, Malvárez G, Navas F, Carrero R. 2014. Spatial distribution of storm wave energy dissipation for the assessment of beach morphodynamics. *Journal of Coastal Research.* 70:259–265.
- Guisado E, Malvárez GC, Navas F. 2013. *Morphodynamic environments of the Costa del Sol.* Madrid, Spain: BIOONE.
- Gutierrez B. T, Plant NG, Thieler ER. 2011. A Bayesian network to predict coastal vulnerability to sea level rise. *J Geophys Res.* 116(F2):1–15.
- Gutierrez BT, Plant NG, Thieler ER, Turecek A. 2015. Using a Bayesian network to predict barrier island geomorphologic characteristics. *J Geophys Res Earth Surf.* 120(12):2452–2475.
- Hapke CJ, Plant NG. 2010. Predicting coastal cliff erosion using a Bayesian probabilistic model. *Mar Geol.* 278(1-4):140–149.
- Hapke CJ, Plant NG, Henderson RE, Schwab WC, Nelson TR. 2016. Decoupling processes and scales of shoreline morphodynamics. *Mar Geol.* 381:42–53.
- Hargrove WW, Hoffman FM. 1999. Using multivariate clustering to characterize ecoregion borders. *Comput Sci Eng.* 1(4):18–25.
- Hartigan JA, Wong MA. 1979. Algorithm AS 136: a K-means clustering algorithm. *J R Stat Soc Ser C.* 28: 100–108.
- Hashemi MR, Ghadampour Z, Neill SP. 2010. Using an artificial neural network to model seasonal changes in beach profiles. *Ocean Eng.* 37(14-15):1345–1356.
- Jackson DWT, Cooper JAG, O'Connor M, Guisado-Pintado E, Loureiro C, Anfuso G. 2016. Field measurements of intertidal bar evolution on a high-energy beach system. *Earth Surf Process Landforms.* 41(8):1107–1114.
- Jonathon RM, Paul ER, David AH. 2001. Field measurements of sediment dynamics in front of a seawall. *J Coast Res.* 17:195–206.
- Klemas VV. 2015. Coastal and environmental remote sensing from unmanned aerial vehicles: an overview. *J Coast Res.* 315:1260–1267.
- Loureiro C, Ferreira Ó, Cooper JAG. 2013. Applicability of parametric beach morphodynamic state classification on embayed beaches. *Mar Geol.* 346:153–164.

- Luijendijk A, Hagenaars G, Ranasinghe R, Baart F, Donchyts G, Aarninkhof S. 2018. The state of the world's beaches. *Sci Rep.* 8(1):6641.
- Malvarez G. 2012. The history of shoreline stabilization on the Spanish Costa del Sol.
- Malvarez G, Navas F, Guisado-Pintado E, Jackson DWT. 2019. Morphodynamic interactions of continental shelf, beach and dunes: the Cabopino dune system in southern Mediterranean Spain. *Earth Surf Process Landforms.* 44(8):1647–1658.
- Malvarez G, Navas F, Guisado Pintado E. 2015. Procesos y dinámica costera en la ensenada de Marbella. In Vidal, J. R. & Lozano, M. C. N. (eds.) *Litoral de Andalucía: norma y naturaleza*. Huelva (Spain): Servicio de Publicaciones de la Universidad de Huelva.
- Manno G, Anfuso G, Messina E, Williams AT, Suffo M, Liguori V. 2016. Decadal evolution of coastline armouring along the Mediterranean Andalusia littoral (South of Spain). *Ocean Coast Manage.* 124: 84–99.
- Molina R, Anfuso G, Manno G, Gracia Prieto FJ. 2019. The mediterranean coast of Andalusia (Spain): medium-term evolution and impacts of coastal structures. *Sustainability.* 11(13):3539.
- Montaño J, Coco G, Antolínez JAA, Beuzen T, Bryan KR, Cagigal L, Castelle B, Davidson MA, Goldstein EB, Ibaceta R, et al. 2020. Blind testing of shoreline evolution models. *Sci Rep.* 10(1):2137.
- Moore LJ. 2000. Shoreline mapping techniques. *J Coast Res.* 16:111–124.
- Morales J, Rodríguez-Ramírez A, Sedrati M. 2019. Beaches of Huelva: dynamic processes, sediments and management.
- Morales JA, Borrego J, Ballesta M. 2004. Influence of harbour constructions on morphosedimentary changes in the Tinto-Odiel estuary mouth (south-west Spain). *Environ Geol.* 46(2):151–164.
- Pajak MJ, Leatherman S. 2002. The high water line as shoreline indicator. *J Coast Res.* 18(2):329–337.
- Pardo-Pascual JE, Almonacid-Caballer J, Ruiz LA, Palomar-Vázquez J, Rodrigo-Aleman R. 2014. Evaluation of storm impact on sandy beaches of the Gulf of Valencia using Landsat imagery series. *Geomorphology.* 214:388–401.
- Pearson SG, Storlazzi CD, Van Dongeren AR, Tissier MFS, Reniers AJHM. 2017. A Bayesian-based system to assess wave-driven flooding hazards on coral reef-lined coasts. *J Geophys Res Oceans.* 122(12): 10099–10117.
- Plant NG, Robert Thieler E, Passeri DL. 2016. Coupling centennial-scale shoreline change to sea-level rise and coastal morphology in the Gulf of Mexico using a Bayesian network. *Earth's Future.* 4(5):143–158.
- Prieto A, Cuevas P, Fernandez Nunez M, Jose O. 2018. Methodology for improving the analysis, interpretation, and geo-visualisation of erosion rates in coastal beaches—Andalusia, Southern Spain. *Geosciences.* 8:1–17.
- Prieto A, Jose O, Rodríguez-Polo S, Gracia F, Del Río L. 2012. *Procesos erosivos (tasas de erosión) en los deltas mediterráneos andaluces: herramientas de análisis espacial (DSAS) y evolución temporal (servicios OGC)*.
- Rigos A, Tsekouras GE, Vousdoukas MI, Chatzipavlis A, Velegrakis AF. 2016. A Chebyshev polynomial radial basis function neural network for automated shoreline extraction from coastal imagery. *ICA.* 23(2):141–160.
- Rodríguez-Galiano VF, Chica-Rivas M. 2014. Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models. *Int J Digital Earth.* 7(6):492–509.
- Rodríguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J Photogramm Remote Sens.* 67:93–104.
- Ruggiero P, List JH. 2009. Improving accuracy and statistical reliability of shoreline position and change rate estimates. *Journal of Coastal Research.* 255:1069–1081.
- Sammut C, Webb GI. 2017. (eds.) *Encyclopedia of machine learning and data mining*. New York: Springer.
- Sánchez-García E, Pardo-Pascual JE, Balaguer-Beser A, Almonacid-Caballer J. 2015. Analysis of the shoreline position extracted from landsat Tm and Etm+ imagery. *ISPRS - Int Arch Photogramm Remote Sens Spatial Inf Sci.* XL-7/W3:991–998.
- Senciales J, Malvarez G. 2003. La desembocadura del río Vélez (provincia de Málaga, España). Evolución reciente de un delta de comportamiento mediterráneo. *Cuaternario y Geomorfología: Revista de la Sociedad Española de Geomorfología y Asociación Española Para el Estudio Del Cuaternario.* 17(1–2): 47–61.
- Splinter KD, Kearney ET, Turner IL. 2018. Drivers of alongshore variable dune erosion during a storm event: Observations and modelling. *Coast Eng.* 131:31–41.

- Thieler ER, Himmelstoss EA, Zichichi JL, Ergul A. 2009. The digital shoreline analysis system (DSAS) Version 4.0 - An ArcGIS extension for calculating shoreline change. *Open-File Report* - ed. Reston.
- Tsekouras GE, Rigos A, Chatzipavlis A, Velegrakis A. 2015. A Neural-fuzzy network based on hermite polynomials to predict the coastal erosion. Cham: Springer International Publishing, pp. 195–205.
- Turner I, Harley M, Short A, Simmons J, Bracs M, Phillips M, Splinter K. 2016. A multi-decade dataset of monthly beach profile surveys and inshore wave forcing at Narrabeen. *Sci Data*. 3:160024.
- Villar Lama A, Ojeda J. 2007. Evolución del suelo urbano/alterado en el litoral de Andalucía (Evolution of urbanization on the Coast of Andalusia). *Geofocus: Revista Internacional de Ciencia y Tecnología de la Información Geográfica*. 7.
- Villarin MC, Rodriguez-Galiano VF. 2019. Machine learning for modeling water demand. *J Water Resour Plann Manage*. 145(5):04019017.
- Vos K, Splinter KD, Harley MD, Simmons JA, Turner IL. 2019. CoastSat: a google earth engine-enabled python toolkit to extract shorelines from publicly available satellite imagery. *Environmental Modelling & Software*. 122:104528.
- Wilson KE, Adams PN, Hapke CJ, Lentz EE, Brenner OT. 2015. Application of Bayesian networks to hindcast barrier island morphodynamics. *Coast Eng*. 102:30–43.
- Wright LD, Short AD. 1984. Morphodynamic variability of surf zones and beaches: a synthesis. *Mar Geol*. 56(1-4):93–118.
- Yates M, Le Cozannet G. 2012. *Brief communication "Evaluating European Coastal Evolution using Bayesian Networks"*.
- Zhou X, Wang G, Bao Y, Xiong L, Guzman V, Kearns TJ. 2017. Delineating beach and dune morphology from massive terrestrial laser-scanning data using. *J Surv Eng*. 143(4):04017008.