ISBN: 978-92-990075-6-3

6th EnvImeko – IMEKO TC19 – Second Edition
Workshop on Environmental Instrumentation and Measurements
October 28, 2016, Reggio Calabria, Italy

EVOLUTIONARY POLYNOMIAL REGRESSION MODEL FOR THE PREDICTION OF COASTAL DYNAMICS

D. E. Bruno[1], *E. Barca*[1], *R. M. Goncalves*[2], *A. Lay-Ekuakille*[3], *S. Maggi*[1], *and G. Passarella*[1]

[1]*CNR-IRSA, Water Research Institute, Viale F. De Blasio 5, 70132 Bari, Italy, delia.bruno@ba.irsa.cnr.it*

[2]*Department of Cartography Engineering, Federal University of Pernambuco, 50670-901 Recife, Brazil*

[3]*Department of Innovation Engineering, Università del Salento, Via Monteroni, 73100 Lecce, Italy*

Abstract: The effective protection of the coastal ecosystem requires a detailed knowledge of the morphological evolution of the coastal environment. Several probabilistic models have been developed in the last decades to implement a reliable statistical forecasting of coastline dynamics. In this work, the non-linear Evolutionary Polynomial Regression (EPR) model has been used for the first time to evaluate the short-term dynamics of the shoreline from a set of measured shoreline positions in previous years. A comparison of the mean known shoreline positions with those predicted by the model, together with their confidence and prediction intervals, can be used to assess the reliability of the estimation by the EPR model.

Keywords: Evolutionary polynomial regression, Multilinear regression, Coastal dynamics, Marine regression/transgression.

1. INTRODUCTION

Coastal dynamics represents an important environmental threat. Regression and transgression impact the whole coastal ecosystem and affect the natural equilibrium between sea and freshwater. Any technical action aimed to protect the coastal region requires a deep knowledge of the features, magnitude, evolution and impact of coastal dynamics [1]. From a geological point of view, the coastal morphology is the result of events, such as uplift and eustatic fluctuations, that occurred during the Quaternary period. At regional scale, the coastal morphology can be considered the effect of the combined action of inner and outer factors, such as lithology, tectonics and past climate change. At basin scale, the coastal evolution is explained as the result of Holocenic events, usually linked to changes of river course, prevailing winds, marine currents, tides and, more recently, to anthropogenic environmental changes [2]. Studies of coastal evolution have several different research aims, such as the development of simulation and forecasting scenarios to mitigate the risks of erosion in coastal areas of high environmental and touristic value. This last task is particularly challenging, given the stochastic nature of the considered phenomena and the frequent lack of information about the various factors involved in coast erosion, which often act simultaneously. Knowledge of the past evolution of a coastline sector can strongly support the accurate prediction of its future configuration.

Since the 90's, probabilistic methods, such as the extrapolation of historical trends, have been applied to predict future shoreline positions [3]. More recently, stochastic models have been developed to handle the uncertainty associated with long-term forecasting. In particular, the scientific literature reports several different approaches based on fuzzy-logic numerical models [4], Bayesian networks [5] or Neural Network Approximation (NNA) models [6]. Bheeroo *et al.* [7] estimate the risk of erosion of a shoreline sector using the Digital Shoreline Analysis System (DSAS), a computer software for calculating shoreline change implemented and freely provided by the United States Geological Survey (USGS).

In this work, the multiple non-linear Evolutionary Polynomial Regression (EPR) model has been applied to the evaluation of the coastal dynamics along the Ionian coast of the Apulia region in Italy. EPR is an evolutionary data-driven technique, capable of exploring an almost infinite number of possible model structures and to identify a set of reliable polynomial functions. Although this approach is purely stochastic and not physically-based, after a proper training phase it can provide a number of reliable regression models, all capable to forecast the evolution of a given shoreline sector.

The EPR approach has already been applied in several different environmental areas, such as for quantitative and qualitative groundwater assessment or hydrological time-series analysis but, to the best of our knowledge, it has never been used to model coastline dynamics. The results of evolutionary polynomial regression model have been validated by comparing the observed position of the last known year with the position estimated by EPR.

On the basis of the main error statistics, the EPR model shows an acceptable residual error. A further improvement of the forecasting capabilities of the model could be obtained by expanding the available set of starting data.

2. MATERIALS AND METHODS

The EPR model has been applied to predict the dynamical evolution of a given shoreline sector from the known po-

ISBN: 978-92-990075-6-3

Variable	Symbol	Unit
Transect ID	Tr	—
Initial transect position	P_i	m
Final transect position	P_f	m
Initial year	Y_i	years
Final year	Y_f	years
Elapsed years	Δt	years

Table I: Independent variables used in the coastline EPR model.

sitions of the given shoreline at multiple time frames. All calculations have been done from within the EPR MOGA-XL Microsoft Excel^{TM} add-in [8], a very comfortable tool for setting the model parameters and the input and output data set from within a well-known and user-friendly spreadsheet environment and for launching the main EPR application that performs the actual data modeling task.

The analysis has been performed by following the steps listed below:

1. discretization of the coastline sector by transects;
2. determination of the transects positions at N different years by means of digital orthophotos;
3. definition of the of independent variables;
4. set up of the input dataset and of supporting information for the first $N-1$ years;
5. calculation of the resulting EPR model;
6. statistical assessment of the model.

In the first step, the sector of shoreline under consideration must be discretized in transects. A good balance between reliability of results and computational time can be achieved by using about 100 transects, where the length of each transect is less than 50 m [9].

The second step is essential for a good and reliable analysis and prediction of the shoreline changes. The aim of this step is to accurately measure for each available year the position of the transect with respect to a fixed reference point. The scientific and technical literature proposes various approaches to this task, based on the support and accuracy of the geographical information [10]. The proposed method is purely statistical and does not consider the physical phenomena at the basis of the shoreline changes. The method takes any kind of useful information from the input dataset of the measured variables, which, consequently, must be numerically significant, precise and carefully organized. At this stage, most of the historical available representations of the considered coastline (i.e., maps, aerial photographs, *in-situ* GPS surveys and image data derived from remote sensing platforms) need to be accurately collected and investigated.

Concerning the third step, Table I reports the list of the independent variables used in this work. The computational tools used for the determination of the EPR model are able to discard the not robust or intrinsically dependent variables.

The fourth and fifth steps are strictly related. Once the matrix of input values has been set, the calculation of the EPR model can be started. After a training stage, EPR usually converges to the best model for the available data, chosen among

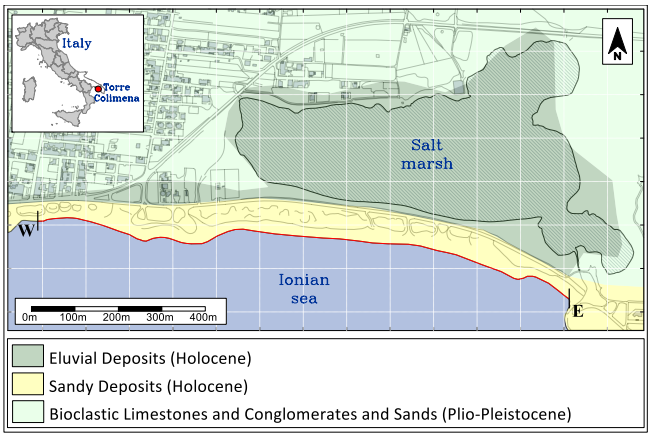

Figure 1: Geological sketch of the study area. The coastline sector considered in this work extends from W to E.

a wide list of non-linear functions. In this particular case, EPR has been applied to the estimation of the shoreline position at the N-th known year.

The model validation carried out at the last step of the proposed methodology, allows to assess the capability of EPR to predict future configurations of the considered coastline. The validation has been carried out using some classic error statistic indexes.

2.1. Evolutionary Polynomial Regression

The EPR is a data-driven technique based on evolutionary computation which deals with pseudo-polynomial structures representing a true physical system [11]. At first, EPR uses an evolutionary procedure based on a genetic algorithm (GA) to search for suitable model structures, then passes through a linear regression step, which computes the models constants by means of a least squares optimization. A typical compact formulation of the EPR model is

$$y = \sum_{j=1}^{m} F\left(X, f(X), a_j\right) + a_0, \tag{1}$$

where y is the dependent variable, F is the function produced by the optimization process, f is an user-defined function, X is the matrix of dependent variables, a_j is an adjustable parameter for the j-th term, a_0 is an optional bias and m is the total number of terms of the expression.

The model described above has been implemented in a software package, ERP-MOGA (release 1.0) [11], that works in a mixed environment to take advantage of the graphical facilities of Microsoft Excel and of the advanced computational capabilities of MathWorks MATLAB^{TM}. In order to run EPR-MOGA, besides the obvious training dataset, two distinct sets of parameters that control the evolutionary procedure and the linear regression steps must be assigned: the general model structure, the number of model terms, the range of the polynomial exponents, the regression type, the coefficients of the estimation method and the optimization strategy. A reasonable setting of such parameters affects positively the run time of the procedure.

ISBN: 978-92-990075-6-3

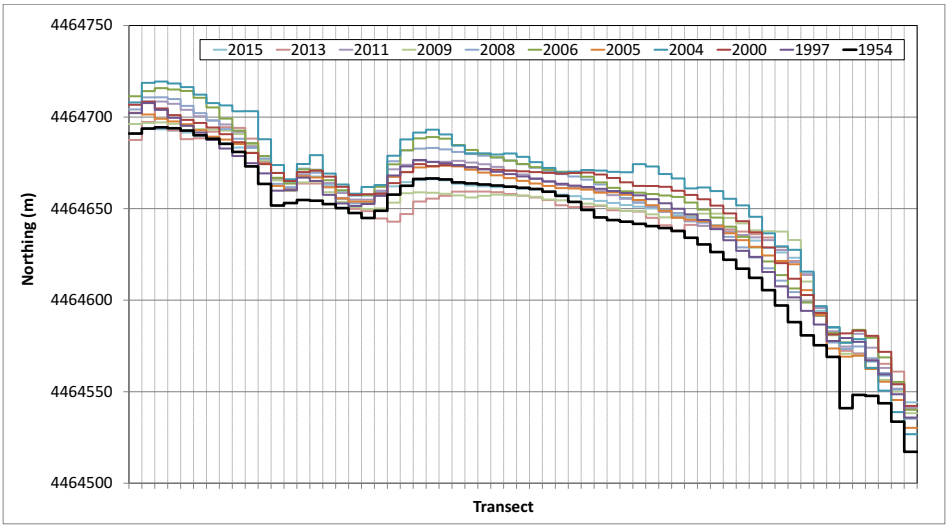

Figure 2: Schematization of the study area: transects and shorelines from 1954 to 2015.

3. CASE STUDY

3.1. Study area

As shown in Figure 1, the study area is located nearby the town of "Torre Colimena", on the Ionian coast of the Salento peninsula in Apulia, southern Italy. The coastline extends in the West-East direction for about 1.200 m and it is almost entirely sandy. Nevertheless, at the western end, it is characterized by an evident low rocky cliff, while at the other extreme a sandy-rocky mixed sector is present. The backshore area is characterized by the presence of lush Mediterranean vegetation, which makes this area very important for the preservation of priority habitats. These habitats are strongly representative of the biogeography of the area, but unfortunately they are located in high-risk alteration zones, in particular since this environment is characterized by a strong tourist vocation.

From a geological point of view, the Salento peninsula topographically constitutes the lower part of the Apulian foreland and is characterized by the alternation of high and low structural blocks, bounded by high-angle faults, mostly in NW-SE direction [12]. The Salento peninsula emerged discontinuously during the late Pleistocene, leading to the formation of the Ionian coastal plains and it is still emerging by up to 0.25 mm/year nearby the town of Taranto, several kilometers north of the study area [13]. Concerning the stratigraphic arrangement, the "Calcareniti del Salento" (Plio-Pleistocene), characterized by bio-clastic limestones, sands and conglomerates, overlap the limestones and dolomitic limestones of the "Calcare di Altamura" (Creta–Senon). At a local scale, between W and E (Figure 1), Holocene deposits of blue-gray sands, partially cemented coastal dunes and eluvial deposits emerge on bioclastic limestone.

The petrographic and mechanical characteristics of these rocks make the coast particularly vulnerable to erosion phenomena. At the transition zone from groundwater to marine water, very aggressive iper-karst processes have been observed and documented, as well as a large number of sinkholes along the whole coast of the peninsula. In particular, a large karst depressions has formed just behind the dunes of the study area and is now a salt marsh of about 0.2 km^2, directly connected to the sea (Figure 1). Due to the intrinsic breakability of calcarenite, a frequent event is the fall off of wide sectors of the cliffs, mainly provoked by the combined mechanic action of the sea and by dissolution phenomena.

The considered coastal sector is characterized by a unimodal regime with a clear predominance of currents from SSE. A noteworthy phenomenon during the summer is the appearance of waves with modest height generated by local winds from the fourth quadrant.

3.2. Shoreline data

The shoreline changes are related to the dynamics of the sea level, therefore the time scale chosen for the analysis strongly depends on the main goals of the investigation. In this study, the shorelines have been detected from satellite imagery and orthorectified aerial photographs taken at eleven different years, namely in 1954, 1997, 2000, 2004, 2005, 2006, 2008, 2009, 2011, 2013, 2015. All the orthorectified images have been processed in a GIS software application, setting the same reference system and cartographic projection, WGS84 and UTM (Zone 33N). Given the West-East orientation of the considered coastal sector, sixty-one transects following a north-south direction have been overlapped to each image, splitting the coastline sector W–E into 20 m long segments (Figure 2). Each transect is characterized by a numerical ID, an E value (UTM East coordinate) and eleven values of N (UTM North coordinate), one for each of the considered years. The proposed EPR method has been applied to determine a regression model capable to describe the shoreline dynamics discretized as described above.

3.3. Marine regression and transgression

Figure 2 shows an almost persistent transgression of the coastline with respect to 1954, leading to an overall shore loss during the whole considered period. However, during the con-

ISBN: 978-92-990075-6-3

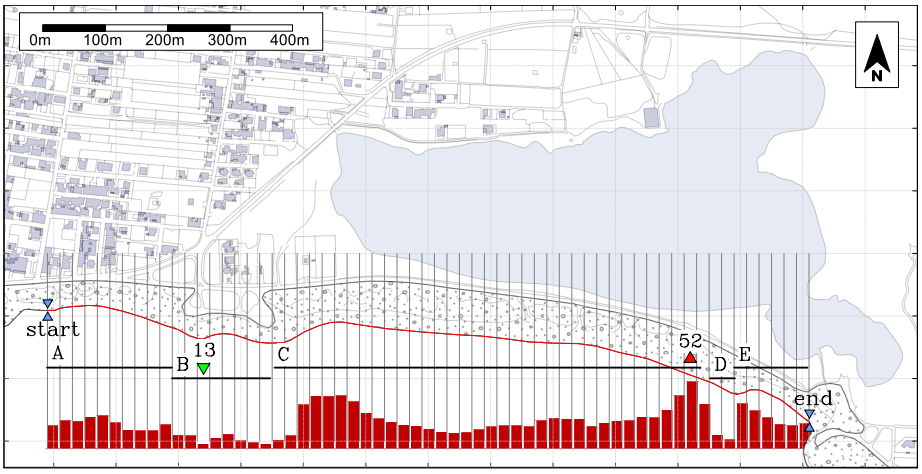

Figure 3: Dynamics of the coastal sector between 1954 and 2015. The red bars represent the variability rate for each of the 61 transects (black vertical lines). Sectors B and D show marine regression, sectors A, C and E transgression. The green and red arrows mark the transects characterized by the largest marine regression and transgression, respectively.

sidered time frame the annual rate of transgression decreases from -0.24 km^2/year to -0.02 km^2/year and, starting from about 2000, the coast dynamics inverts its trend and a gradual, almost constant in time, recover of the shore becomes evident. The choice of 1954 as the starting year for this analysis has been done because most of the coastal anthropic changes, e.g., in terms of construction of road, rail and tourism infrastructures, began only in the following decade. Figure 3 represents the percent transect variability during the whole considered period. Inspecting the histogram, it is evident that the whole coastal sector can be divided into five main sections, characterized by a different susceptibility to coastal dynamics (sections A to E in Figure 3), where the coast section between transects #10 and #20, together with transects #54 and #55, seems to be more stable and transect #52 is the most dynamic.

4. RESULTS AND DISCUSSION

The EPR-MOGA software provided about 20 different models, characterized by similar coefficients of determination R^2 and by different lengths, in terms of number of monomials that compose the expressions determined by the package. In fact, the choice of a model should consider not only the value of R^2, that measures the reliability of the calculated regression, but also the complexity of the expression determined by the package, because of practical computational problems. The model used in this work has a coefficient of determination $R^2 = 0.958$ and is composed by seven simple monomials,

$$\begin{aligned} P_f = {} & 0.0018351 \times \Delta t^2 - 0.010854 \times Y_f^{1.5} \\ & + 1609.8708 \times P_i^{0.5} - 0.60481 \times Tr^2 \\ & + 9.7052 \times Tr^{1.5} - 53.1258 \times Tr \\ & + 106.9025 \times Tr^{0.5} + 1063980.158. \end{aligned} \tag{2}$$

The standard error of the coefficients of the calculated model is always below 10% and in most cases is between 3 and 5%.

A preliminary evaluation of the problem has led to set some options and constraints for the optimization. In particular, (1) static regression has been preferred to dynamic regression, because of the irregular time interval between the observed coastlines, (2) the maximum number of terms of the model has been set to 11 and (3) the exponents of the monomials have been constrained between -3 and 3, with a step of 0.5.

The EPR model has been used to forecast the position of the 61 transects in 2015, starting from the known positions in the 10 previous years. The average of the 10 estimated values,

$$\widehat{P_f} = \sum_{P_i} \frac{P_f}{N}, \tag{3}$$

where N is the number of observed coastlines, has been considered as the most reliable position. Figure 4 compares the observed positions of the coastal sector in 2015 to the estimated positions $\widehat{P_f}$, with a 95% prediction interval (PI). The Figure clearly shows the accuracy of the estimation: for most transects the observed coastline is very close to the estimated one and, whenever some departure is visible, it is always well within the prediction interval of the estimation.

In order to rigorously validate the model, a statistical error analysis of the estimation (or residual) error,

$$r_k = y(x_k) - y^*(x_k), \tag{4}$$

has been performed, where $k = 1, \ldots, 61$ is the transect number and $y(x_k)$ and $y^*(x_k)$ are the measured and estimated coast positions in 2015, for each transect k.

Table II reports the computed values of the Mean Bias Error (MBE), the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) indices [14], which are related to the mean magnitude of the error and should be as small as possible. The gaussianity of the distribution of r_k was also positively tested [15]. Table II shows that the model provides a Gaussian error distribution characterized by a mean error (MBE) smaller than or equal to the values reported in literature for similar problems [10]. The minimum absolute error

ISBN: 978-92-990075-6-3

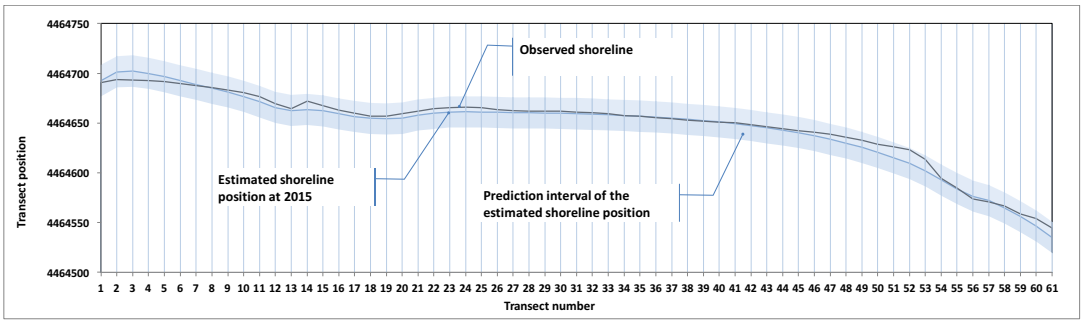

Figure 4: Comparison of (black) the observed coastline in 2015 with (blue) the coastline estimated with EPR.

Error statistics	EPR	Unit
Guassianity	yes	—
MBE	−2.0449	m
Standard deviation	4.4181	m
Minimum absolute error*	0.0165	m
Maximum absolute error*	14.0141	m
RMSE	4.8354	m
MAE	3.6322	m
ρ	0.9943	—
ρ_C	0.9915	—

*transect

Table II: Main statistic indexes of the error analysis. The minimum and maximum absolute error are relative to a transect.

of the EPR model is 0.0165 m. The EPR model also gives small values for the RMSE and MAE indices. Lastly, the coefficients of correlation and concordance, ρ and ρ_C, [16], between the observed and estimated values in 2015 are both close to 1, confirming that EPR provides very reliable results. In conclusion, the above analysis of the errors indicates that the proposed model performs more than satisfactorily when using the previously observed coastlines as input. Given the shoreline position in the ten previous years, it is possible to use EPR to reliably forecast the coastal evolution in the near future years.

An open question regards the extent of degradation of the model forecast when the time window is pushed forward into the future with respect to the observation period.

To evaluate the reliability of future predictions, the average transect position during the observation period and the related confidence interval (CI), has been calculated and has been used as a reference band. Future estimated transect positions whose prediction interval (PI) doesn't overlap with the CI of the corresponding average position in the past are not considered reliable.

Figure 5 compares the estimated shoreline in 2015 with its average position during the observation period. The predic-

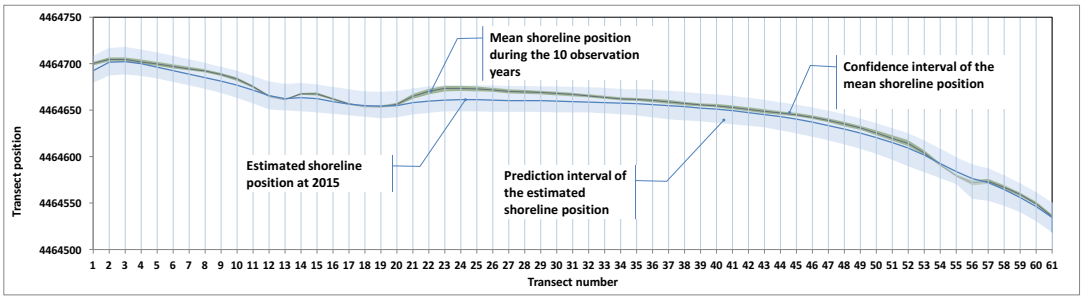

Figure 5: Estimated coastline in 2015 and mean position of the coastline during the observation period (1954-2013).

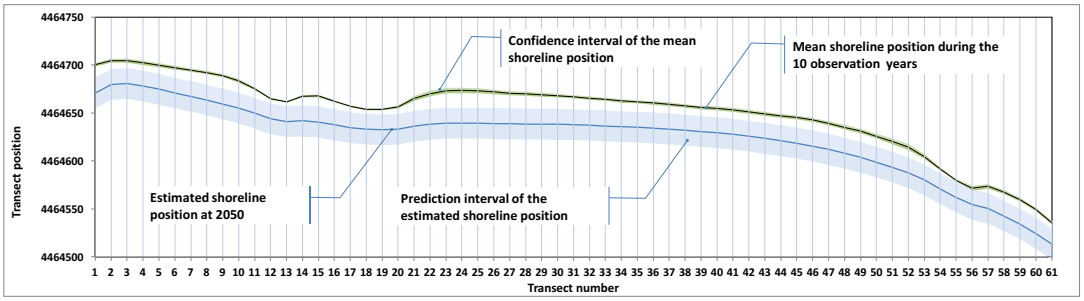

Figure 6: Estimated coastline in 2050 and mean position of the coastline during the observation period (1954-2013).

ISBN: 978-92-990075-6-3

tion band of the former always overlaps the confidence band of the latter, assessing the reliability of this estimation. On the other hand, considering the estimated coastline in 2050 (Figure 6), it is clear that the two calculated bands never overlap. Thus, at the present state of the analysis, this forecast cannot be considered reliable enough to be useful.

5. CONCLUSIONS

A multiple non-linear regression model calculated by EPR has been used to estimate the dynamics of a coastal sector of about 1.2 km, located on the Ionian coast of the Apulia region. The coast sector has been discretized in 61 transects, each 20 m long. Starting from 10 known coast positions from 1954 to 2013, the proposed model has been used to estimate the position of the transects in 2015. A statistical error analysis demonstrated the reliability of the results. The proposed approach is purely statistical and consequently it does not take into account any physical, environmental, anthropic factors capable of inducing an unforeseeable impact on the coastal dynamics. The EPR model has been also used to forecast the coast position in 2050. However, this forecast cannot be considered reliable enough, being too far ahead in time and space from the cluster of observed coastlines.

REFERENCES

[1] D. L. Passeri, S. C. Hagen, S. C. Medeiros, M. V. Bilskie, K. Alizad and D.Wang, "The dynamic effects of sea level rise on low-gradient coastal landscapes: A review", Earth's Future, vol. 3, no. 6, pp. 159–181, 2015.

[2] C. J. Hapke, M. G. Kratzmann and E. A. Himmelstoss, "Geomorphic and human influence on large-scale coastal change", Geomorphology, vol. 199, pp. 160–170, 2013.

[3] M. Crowell, B. C. Douglas and S. P. Leatherman, "On forecasting future US shoreline positions: a test of algorithms", Journal of Coastal Research, pp. 1245–1255, 1997.

[4] A. Altunkaynak, "Predicting water level fluctuations in Lake Michigan-Huron using wavelet-expert system methods", Water resources management, vol. 28, no. 8, pp. 2293–2314, 2014.

[5] B. T. Gutierrez, N. G.Plant and E. R. Thieler, "A Bayesian network to predict coastal vulnerability to sea level rise", Journal of Geophysical Research: Earth Surface, vol. 116, no. F2, 2011.

[6] D. I. Gopinath and G. S. Dwarakish, "Wave prediction using neural networks at New Mangalore Port along west coast of India", Aquatic Procedia, vol. 4, pp. 143–150, 2015.

[7] R. A. Bheeroo, N. Chandrasekar, S. Kaliraj and N. S. Magesh, "Shoreline change rate and erosion risk assessment along the Trou Aux Biches–Mont Choisy beach on the northwest coast of Mauritius using GIS-DSAS technique", Environmental Earth Sciences, vol. 75, no. 5, pp. 1–12, 2016.

[8] O. Giustolisi and D. A. Savic, "A symbolic data-driven technique based on evolutionary polynomial regression", Journal of Hydroinformatics, vol. 8, no. 3, pp. 207–222, 2006.

[9] O. Ferreira, T. Garcia, A. Matias, R. Taborda and J. A. Dias, "An integrated method for the determination of set-back lines for coastal erosion hazards on sandy shores", Continental Shelf Research, vol. 26, no. 9, pp. 1030-1044, 2006.

[10] R. M. Goncalves, J. L. Awange, C. P. Krueger, B. Heck and L. dos Santos Coelho, "A comparison between three short-term shoreline prediction models", Ocean & Coastal Management, vol. 69, pp. 102–110, 2012.

[11] O. Giustolisi and D. A. Savic, "Advances in data-driven analyses and modelling using EPR-MOGA", Journal of Hydroinformatics, vol. 11, no. 3–4, pp. 225–236, 2009.

[12] M. Parise, "Surface and subsurface karst geomorphology in the Murge (Apulia, Southern Italy)", Acta Carsologica, vol. 1, no. 80, pp. 40, 2011.

[13] L. Ferranti and P. Viterbo, "The European summer of 2003: Sensitivity to soil water initial conditions", Journal of Climate, vol. 19, no. 15, pp. 3659–3680, 2006.

[14] C. J. Willmott and K. Matsuura, "Advantages of the Mean Absolute Error (MAE) Over the Root Mean Square Error (RMSE) in Assessing Average Model Performance", Climate Research, vol. 30, no. 1, pp. 79–82, 2005.

[15] E. Barca, E. Bruno, D. E. Bruno and G. Passarella, "GTest: a software tool for graphical assessment of empirical distributions' Gaussianity", Environmental monitoring and assessment, vol. 188, no. 3, pp. 1–12, 2016.

[16] I. Lawrence and K. Lin, "A concordance correlation coefficient to evaluate reproducibility", Biometrics, pp. 255–268, 1989.