Geomorphology 300 (2018) 128–140

Contents lists available at ScienceDirect

Geomorphology

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

Linear and evolutionary polynomial regression models to forecast coastal dynamics: Comparison and reliability assessment

Delia Evelina Bruno [a,*], Emanuele Barca [a], Rodrigo Mikosz Goncalves [b], Heithor Alexandre de Araujo Queiroz [b], Luigi Berardi [c], Giuseppe Passarella [a]

[a] Water Research Institute, National Research Council, IRSA, CNR, Viale F. De Blasio, n. 5, 70125 Bari, Italy
[b] Department of Cartography Engineering, Federal University of Pernambuco (UFPE), Geodetic Science and Technology of Geoinformation Post Graduation Program, Recife 50670-901, PE, Brazil
[c] Polytechnic University of Bari, Civil Engineering and Architecture Department, Via Edoardo Orabona, 4, 70125 Bari, Italy

ARTICLE INFO

Article history:
Received 23 September 2016
Received in revised form 26 September 2017
Accepted 15 October 2017
Available online 31 October 2017

Keywords:
Evolutionary polynomial regression
Multilinear regression
Coastal morphodynamics
Coastal erosion

ABSTRACT

In this paper, the Evolutionary Polynomial Regression data modelling strategy has been applied to study small scale, short-term coastal morphodynamics, given its capability for treating a wide database of known information, non-linearly. Simple linear and multilinear regression models were also applied to achieve a balance between the computational load and reliability of estimations of the three models. In fact, even though it is easy to imagine that the more complex the model, the more the prediction improves, sometimes a "slight" worsening of estimations can be accepted in exchange for the time saved in data organization and computational load. The models' outcomes were validated through a detailed statistical, error analysis, which revealed a slightly better estimation of the polynomial model with respect to the multilinear model, as expected. On the other hand, even though the data organization was identical for the two models, the multilinear one required a simpler simulation setting and a faster run time. Finally, the most reliable evolutionary polynomial regression model was used in order to make some conjecture about the uncertainty increase with the extension of extrapolation time of the estimation. The overlapping rate between the confidence band of the mean of the known coast position and the prediction band of the estimated position can be a good index of the weakness in producing reliable estimations when the extrapolation time increases too much. The proposed models and tests have been applied to a coastal sector located nearby Torre Colimena in the Apulia region, south Italy.

1. Introduction

The coastal management aims to define future morphodynamics in order to plan and realize defense works safeguarding such areas (Masciopinto, 2006). Nevertheless, forecasting morphodynamics can be affected by errors due to morphological instabilities over long time periods and provide incorrect representations. Sandy beaches are very dynamic systems and the relationship between hydro- and morphodynamics is a non-linear function, strongly dependent on time and spatial scales (Ranasinghe, 2016). In contrast, rocky coast evolution occurs over millennia even though some morphological evidences can be observed, (Matsumoto et al., 2016) as local rock falls and blocks on the beach or into the sea. Consequently, the assessment of shoreline evolution at different scales requires advanced methods and tools. Studies of coastal evolution have so far been based on the paleogeographic reconstruction (Rao et al., 2015) by means of stratigraphic surveys, radiometric and luminescence dating (Fruergaard et al., 2015), high-resolution topographic data (Jara-Muñoz et al., 2016), observations, and comparisons of historical maps. At the basin scale, the coastline is explained as the result of events over the Holocene (Brill et al., 2015), generally linked to river course changes, prevailing winds, marine currents (Cooper et al., 2013), tides (Davis and Hayes, 1984) and, more recently, to anthropogenic environmental changes (Hapke et al., 2013; Le Cozannet et al., 2014).

According to the Intergovernmental Panel on Climate Change (IPCC), future climate change scenarios referred to the end of the 21st century will impact mean sea level, wave conditions, storm surges, river flow and, ultimately, the coastal zone evolution (Gornitz, 1991; Solomon et al., 2007; Ford, 2013). Given the stochastic nature of the considered phenomena, some authors propose probabilistic models to handle the uncertainty associated with coastal dynamics forecasting (Cowell and Zeng, 2003) at different space and time scales. In this context, since the '90s the scientific literature reports approaches based on time-series and trend analysis (Crowell et al., 1997; Galgano and Douglas, 2000), fuzzy-logic numerical models (Altunkaynak, 2014), Bayesian networks (Gutiérrez et al., 2011) and Neural Network Approximation (NNA) (Gopinath and Dwarakish, 2015). Bheeroo et al. (2016) estimated the risk of marine erosion using the Digital Shoreline Analysis System (DSAS), a free software application developed and

* Corresponding author.
E-mail address: delia.bruno@ba.irsa.cnr.it (D.E. Bruno).

https://doi.org/10.1016/j.geomorph.2017.10.012
0169-555X/

implemented at the United States Geological Survey (USGS) (Thieler and Danforth, 1994). Scientific literature frequently proposes data-driven modelling as a reliable coastal management decision-support tool (Hsu et al., 1994).

Independently from the approach, specific attention must be paid to the timescale of the considered phenomenon. Actually, shoreline changes are triggered by natural processes characterized by different timescales, which range from episodic (hours-days) to medium-term or inter-annual (year-decade), up to long-term (decades-century) (Thom and Hall, 1991; Cowell et al., 1992; Ranasinghe, 2016). Consequently, the choice of the timescale is strictly related to the objectives of the study but strongly constrained by the available data.

In this paper, an advanced data modelling strategy, based on the evolutionary polynomial regression (EPR) approach (Giustolisi and Savic, 2006), has been applied to study small spatial scale and medium-term coastal morphodynamics. Simple linear and a multilinear regression models have also been applied, and a balance between computational load and the reliability of the outcome of the three models has been explored. Even expecting that the most advanced modelling strategy provides most reliable results, such a comparison aims to evaluate if a “slight” worsening of the estimations can be accepted in exchange for the time saved in data organization and computational load. The choice of the EPR approach has been based on our intention of testing a data-driven technique, based on evolutionary computation, to simulate the coastal dynamics. This is a hybrid approach combining numerical regression and evolutionary search of mode expressions, providing a number of reliable regression models, all capable of making good forecasts of the evolution of a given shoreline sector. Although it is not physically based, it allows the introduction of some prior insight on the phenomenon before (i.e. selection of candidate inputs) and after (i.e. selection among expressions) model exploration. Even though scientific and technical literature reports the application of various regression models to the considered problem of coastal change, it does not include the use of EPR, which instead has been applied in various environmental studies, such as groundwater quantitative and qualitative assessment, and hydrological time-series analysis (Giustolisi et al., 2008; Markus et al., 2010).

The proposed model has been applied to an about 1.2 km-long coastal sector nearby Torre Colimena in the Apulia region, south Italy, where several geomorphological and hydrogeological studies have been carried out during the last 50 years (Dai Pra and Hearty, 1989; Bruno et al., 2016).

2. Materials and methods

In this paper, the SLR and MLR have been executed by means of simple MS-Excel formulas, while the EPR modelling strategy have been applied by means of the MS-Excel add-in “EPR MOGA-XL” (Giustolisi and Savic, 2006; Giustolisi and Savic, 2009). In the following section, after a brief introduction to linear regression, a more detailed description of the EPR main theoretical and practical issues are provided.

2.1. Multivariate linear regression

The MLR is a well-known and, perhaps, the simplest method to simulate coastal morphodynamics based on the knowledge of the past shoreline evolution. This method is capable of quantifying any relationship between a dependent variable and one or more independent variables and is expressed by the linear model:

$$y = \sum_{j=1}^{m} a_j x_j + a_0 \tag{1}$$

where y is the dependent variable; x_j are the m independent variables; and a_j and a_0 are the unknown regression coefficients and the bias, respectively. The coefficient of determination R^2 and the *Lin* coefficient are used to measure the capability of the regression model to describe y. SLR refers to Eq. (1) where $j = 1$.

2.2. Evolutionary polynomial regression

The EPR is a data-driven technique based on the evolutionary computation (Giustolisi and Savic, 2006), which deals with pseudo-polynomial structures representing a true physical system. A typical compact formulation of the EPR expression is:

$$y = \sum_{j=1}^{m} F\left(X, f(X), a_j\right) + a_0 \tag{2}$$

where y is the dependent variable; a_j is a coefficient for the j^{th} term; F is a function produced by the process; X is the matrix of dependent variables; f is a function defined by the user; m is the maximum number of terms in the expression; and a_0 is an optional bias.

EPR MOGA-XL is implemented through two main stages: an evolutionary procedure, and a linear regression step. The former is based on a multi-objective genetic algorithm (GA) for searching model structures (Goldberg et al., 1989; Giustolisi and Savic, 2009), and the latter is based on a least squares (LS) technique (Giustolisi et al., 2007; Laucelli and Giustolisi, 2011) for computing model parameters. This multi-objective strategy allows searching for models achieving the best trade-off between fit to observed data, achieved by minimizing the Sum of Squared Error (SSE), and model expression complexity, as number of pseudo-polynomial terms and/or number of input variables selected in the model. Such a strategy enabled us to overcome the introduction of the penalization of complexity (PCS) fitness function applied in the earlier EPR strategy (Giustolisi and Savic, 2006). This approach allows one to evaluate what is the real contribution of adding a new term or a further variable to the prediction accuracy as the model complexity increases. In other words, considering the accuracy as a benefit and the structural complexity as a cost, EPR MOGA-XL allows one to assess which cost produces higher accuracy.

EPR MOGA-XL is implemented in a software application running in a mixed environment (Rezania et al., 2008), which takes advantage of the computational capability of MATLAB and the graphical and data manipulation facilities of MS-Excel. In order to run EPR MOGA-XL, some parameters must be assigned to drive both the evolutionary search and the linear regression steps: the general model structure (i.e. F and f(X) in Eq. (2)), the maximum number of the pseudo-polynomial terms (m), the range of exponents for variables (X) to be assigned in Eq. (2), the modelling type (e.g. for time-series or not), the coefficients estimation method (i.e. least squares), and the optimization strategy (i.e. minimizing the number of pseudo polynomial terms, the number of selected variables or both against the minimization of the SSE). A reasonable setting of such parameters would positively affect the software application runtime.

2.3. Simulations

Generally speaking, the three proposed methods share the same preliminary steps, which fundamentally consist in the data organization. Fig. 1 shows the main preliminary steps. Preliminary step 1) concerns the discretization of the considered coastline; the more transects there are, the more the discretization fits the actual coastline but the more the computational time increases. Note that transects all have the same width and are all oriented N-S in order to measure their changes of position just as upward/downward movements along the vertical axis. Setting the number and the dimension of transects is an important task; the more the number increases, the more the relative dimension decreases; nevertheless, this increases the computational load. Given a local spatial scale (hundred meters to few kilometers),