Development of a statistical tool for the estimation of riverbank erosion probability

E. A. Varouchakis[1], G. V. Giannakis[1], M. A. Lilli[1], E. Ioannidou[1], N. P. Nikolaidis[1] and G. P. Karatzas[1].

[1]{School of Environmental Engineering, Technical University of Crete, Chania, Greece}

Correspondence to: E. A. Varouchakis (varuhaki@mred.tuc.gr)

Abstract

Riverbank erosion affects river morphology and local habitat, and results in riparian land loss, property and infrastructure damage, and ultimately flood defence weakening. An important issue concerning riverbank erosion is the identification of the vulnerable areas in order to predict river changes and assist stream management/restoration. An approach to predict areas vulnerable to erosion is to quantify the erosion probability by identifying the underlying relations between riverbank erosion and geomorphological or hydrological variables that prevent or stimulate erosion. In the present work, a statistical methodology is proposed to predict the probability of presence or absence of erosion in a river section. A physically based model determines the locations vulnerable to erosion by quantifying the potential eroded area. The derived results are used to determine validation locations for the evaluation of the statistical tool performance. The statistical tool is based on a series of independent local variables and employs the Logistic Regression methodology. It is developed in two forms, Logistic Regression and Locally Weighted Logistic Regression, which both deliver useful and accurate results. The second form though, provides the most accurate results as it validates the presence or absence of erosion at all validation locations. The proposed tool is easy to use, accurate and can be applied to any region and river.

1 Introduction

Riverbank erosion is a complex phenomenon resulting from various factors, which affect the balance of ecosystems. It is also important from the geomorphological aspect as it also induces changes in the river channel course and in the development of the floodplain (Hooke, 1979; Bridge, 2003). Mass-failure processes constitute a significant source of sediment in disturbed streams, which occur due to a combination of hydraulic and geotechnical processes that undercut bank toes and cause bank collapse (Simon et al., 2009). Riverbank erosion is a natural geomorphologic process that affects the fluvial environment in many aspects; physical, ecological and socio-economical. It is the result of a complex interaction between the channel hydraulic conditions and the physical characteristics of the banks, both of which are highly variable in nature. Bank retreat affects the riverbed structure and morphology as well as the floodplain morphology and the physical habitat. In addition, riparian land losses and damage to human property and infrastructures, lead to direct financial consequences. Moreover, turbidity increase, sediment and debris transport, and flood defense weakening, reveal a complex combination of arising issues due to riverbank erosion. According to Atkinson et al. (2003), significant parameters affecting erosion are vegetation index (stability), the presence or absence of meanders, bank material (classification) and stream power. Also other factors such as bank height, riverbank slope, river cross section width, riverbed slope and water velocity have been reported to affect the erosion rate (Hooke, 1979; Abam, 1993; Winterbottom and Gilvear, 2000; Rinaldi et al., 2008; Luppi et al., 2009). Therefore, the identification of riverbanks which are vulnerable to erosion is of utmost importance, either for their protection, or restoration.

On the other hand, riverbank erosion constitutes a significant factor to the functioning of river dependent ecosystems and provides a sediment source that creates riparian habitat. Bank erosion is a key geomorphological mechanism for the fluvial ecosystems, since it regulates the diversity of habitats, species and vegetal units. The process provides riparian vegetation succession and develops dynamic habitats, vital for fluvial plants and animals. For small scale bank erosion or for local extent, there is no significant influence on the aquatic ecosystem and it is contributing to the ecosystem sustainability. In the opposite case, the ecosystem is significantly affected, while riparian land losses and damages are caused providing areas vulnerable to flooding (Piégay et al., 1997; Piégay et al., 2005; Florsheim, 2008).

The bank erosion process is closely related to soil composition of the riverbanks, and the erodibility factor is affected by the composition of sand, silt and clay. A high content of sand and silt leads to easily eroded soils since they are fine in size and can be carried away by river flow. The most common type of bank structure is a stratified or interbedded bank of cohesive or non-cohesive layers. Riverbanks made up of non-cohesive soil are very erodible due to the low clay-content and the weak erosion-resistant strength of the bank soil. Instead cohesive soils have increased clay or clayey silt content and are more resistant to erosion. Non-cohesive soils erode as individual grains, while cohesive soils erode as aggregates. On the other hand a bedrock bank is usually very stable and will only experience gradual erosion (Raudkivi 1998, Roslan et al., 2013).

Although riverbank erosion is a common phenomenon, the prediction of the location and of the extent of riverbank erosion is difficult. Therefore, a range of approaches and methods have been developed and tested. The most important issue concerning riverbank erosion is the identification of the areas vulnerable to bank erosion, in order to predict changes in the river channel form and assist stream management/restoration options. Different methods have been used to predict erodibility, such as analyses of historical maps and the use of sequential aerial photographs based on GIS technology. However, riverbank erosion is usually approached by using a combination of bank stability methods and hydrodynamic models to predict the vulnerable areas and estimate the erosion rate (Nardi et al., 2013). Of these two methods, the former has a relatively high degree of inaccuracy, while the latter is too complex to be applied, as it requires significant number of data variables.

Herein, a statistical tool is proposed using the Logistic Regression (LR) technique, for the determination of riverbank erosion probability. This technique was selected due to its ability to link related dependent and independent variables, by converting their relationship to a probability of presence or absence of the dependent variable. In addition, it can be extended to account for locally spatial correlated independent variables. The suggested statistical model entitled Locally Weighted Logistic Regression (LWLR), combines LR and Locally Weighted Regression (LWR) principles to create a local model that calculates the probability of erosion to occur based on spatially correlated secondary information (e.g. bank slope, river cross section). Therefore, the accuracy of the predictions is expected to improve compared to the global regression model LR.

The proposed statistical model identifies the underlying relations between riverbank erosion and the geomorphological or hydrological variables that prevent or stimulate erosion. It utilises the available data to detect areas vulnerable to erosion. In addition, the erosion occurrence probability can be calculated in conjunction with the model deviance for each independent variable or model form tested. A similar method was introduced and applied successfully to a river in North Wales (Atkinson et al., 2003), for the estimation of the variables that mostly affect riverbank erosion. The simple Logistic Regression was applied.

This work also involves the application of the Bank-Stability and Toe-Erosion Model (BSTEM 5.2) in order to predict eroded or not riverbank areas, for the validation of the proposed tool. The BSTEM model is a physically-based model, developed by the National Sedimentation Laboratory in Oxford, Mississippi, USA (Simon et al., 2000), and it has been used to simulate the hydraulic and geotechnical processes responsible for mass failure. It represents two distinct processes namely, the failure by shearing of a soil block of variable geometry and the erosion by flow of bank and bank toe material. The BSTEM has been successfully applied in diverse alluvial environments (e.g., Simon et al., 2000; Simon et al., 2002; Simon and Thomas, 2002; Pollen and Simon, 2005; Pollen-Bankhead and Simon, 2009; Simon et al., 2011). It was used to simulate the effects of enhanced matric suction from evapotranspiration and decreased soil erodibility driven by the presence of plant roots, quantifying the effects on streambank factor of safety and comparing with the effects of mechanical root-reinforcement (Pollen-Bankhead and Simon, 2010). BSTEM was also used to quantify bank retreat, which ranged from 7.8 to 20.9 m along 100 m of riverbank at the Barren Fork Creek site (Midgley et al., 2012). In addition, it was also used to quantify the reductions of mass failure frequency and sediment loading from streambanks in the Lake Tahoe in United States (Simon et al., 2009).

The developed methodology was applied to Koiliaris River Basin at the island of Crete, Greece. The overall concept of this work is to provide estimates of the erosion probability at specific ungauged riverbank locations, based on independent secondary explanatory information in terms of LWLR methodology. BSTEM has an auxiliary role to estimate/validate potential eroded riverbank locations by calculating the potential eroded area, using field measurements of hydraulic, hydrologic and geomorphologic variables. These estimations (dependent variables) are then used to set up and validate the statistical model. To the best of our knowledge, the combination of deterministic and stochastic models to predict river bank erosion appears for the first time in the scientific literature.

2 Case Study

The Koiliaris River Basin is situated 25 km east of Chania (005-12-489E, 039-22-112N) and occupies an area of about 130 km^2. Watershed elevation ranges from 0 to 2041 m.a.s.l. with slopes ranging from 1-2% at low elevations up to 43% (high elevations). The total length of it's hydrographic network is 36 km (Moraetis et al., 2010). The area has been studied extensively in the last ten years and especially since 2009 as part of the European network of Critical Zone Observatories (Koiliaris CZO). The Koiliaris River Basin, as a typical Mediterranean watershed, is characterized by varying spatial and temporal hydrologic and geochemical processes. Lithology and geomorphology as well as the climatic conditions in the area, have major influence on the hydrologic characteristics of the Koiliaris CZO (Moraetis et al., 2014). The river is mainly fed by the Stylos karstic springs with water originating from the White Mountains and traveling through an extensive karstic system, which drains the rain and snow melt at high elevations. It is also fed temporarily, during the rain period (October to April), by the Keramianos tributary stream. Keramianos is the main temporary tributary, which drains a watershed sub-catchment characterized by steep slopes, schist geologic formation and degraded erodible soils. As a result, when high rainfall intensities fall upon this area, especially after the dry summer period, surface runoff is induced, transferring large quantities of sediments to the Koiliaris River (flush floods) (Moraetis et al., 2010). During these events, river flow conditions change dramatically with a rapid increase in water level and high flow velocities, affecting riverbank erodibility up to causing bank failure. Such events occur two to three times a year during the rainy period affecting the riparian area and enhancing soil losses through riverbank erosion. The current study focuses on the downstream section of River Koiliaris (Fig. 1). A hydrochemical station (Gauge Station), has been strategically located at the intersection of the Koiliaris River with the Keramianos tributary, recording the water level which was used to generate the flow hydrograph (Fig. 2).

3 Methodology

The bank erosion vulnerability of the Koiliaris' riverbanks was first studied during the hydrologic period 2010-11. The downstream section of the river was divided in eight subsections of variable length, starting from the Gauge Station up to pin No 8 on the study area map (Fig. 1). In each subsection, the geomorphological characteristics of the riverbanks and the riverbed were measured at the beginning and at the end of the subsection, during the first field campaign. Channel and bank geometry characteristics, flow parameters, bank material, bank vegetation and protection parameters were identified and used as an input to the BSTEM model to calculate the riverbank eroded area (L^2).

Therefore, reach slope varied between 0.0042 and 0.11 m/m and the bank material was set, after field measurements analysis, to "fine rounded sand" with an average medium grain size of 0.3 (± 0.06) mm. The "geyer willow" was selected from the predefined list to describe the bank vegetation with the assumptions of the plants age of about 100 years and 100% contribution to assemblage. Additionally, for the locations where the bank was protected, the "boulders" choice was used to describe the bank material. Bank slope and river cross-section measurements were supplemented by a second field campaign. As far as the flow parameters, river water elevation was set to 1.27 m for a 48 h duration event, based on field data. The BSTEM model was then applied to determine the vulnerability of bank erosion at the under study river subsections. The model results, for such long distances (min = 20 m and max = 200 m), were interpreted as potential erosion vulnerability of riverbank, considering the extent of the estimated eroded area.

At the beginning of hydrological year 2013-14, a second field campaign was designed to identify this time specific locations vulnerable to erosion. Twelve riverbank locations were selected along the aforementioned eight subsections and scaled sticks were installed. Two of those locations were selected at restored parts of the river section to monitor potentially stable riverbank points. Six months later, at the end of the wet period and after three flood events (Fig. 2 – Red peaks), the erosion sticks were visually inspected, during a field trip, to identify the presence or absence of erosion. Therefore, the eroded area was roughly estimated.

The concept for this second campaign was to establish measurement points, necessary to develop and apply a statistical model that taking into account a series of explanatory variables, would determine the probability of riverbank erosion at local scale. Furthermore, a series of validation points were necessary to validate the model's efficiency. Thus, the endpoints of each subsection from the first campaign were used because an overall estimate of the riverbank vulnerability was available from the BSTEM results.

However, in order to verify the BSTEM prediction efficiency, it was decided to test the model by using the twelve locations of the second campaign. Based on the model's efficiency and the quality of estimation, the reliability of BSTEM results was evaluated at the eight subsections of the aforementioned river section. The 2nd BSTEM model application estimated the cumulative riverbank erosion effect for the three flash flood events (Fig. 2) at the twelve locations. The other parameters were similar for both model applications, since the same river section was employed.

The BSTEM model results (at the twelve locations) together with field inspection were used to setup the statistical model by interpreting the erosion existence in terms of binary data (1 = "presence of erosion" and 0 = "no erosion"). The BSTEM model, has the capacity to quantitatively calculate the eroded area (L^2). The interpretation of the significance of the estimated eroded area was determined through a statistical process that involves the 25th and 75th percentiles of the estimated values. Therefore, the eroded area can be classified to levels of significance. Under the 25th percentile the erosion is categorized as not significant (no erosion) and over 75th as significant. The in-between values are signified as erosion. The latter two, fall in the ''presence of erosion'' category.

Next, the probability of erosion at the riverbanks of River Koiliaris, was estimated considering a series of easy to determine independent geomorphological variables (bank slope, river cross section) through LR and LWLR methodologies, first at the validation points and then at ungauged riverbank locations. The methodological steps of the proposed tool and of the overall process are briefly described by a flowchart presented in Figure 3.

3.1 Logistic Regression

Riverbank erosion can be simulated by a regression model using independent variables that are considered to affect the erosion process. The impact of such variables may vary with geographical location and, therefore, a spatially non-stationary regression model is preferred instead of a stationary equivalent. Locally Weighted Regression (LWR) is proposed as a suitable choice. This method can be extended to predict the binary presence or absence of erosion based on a series of independent local variables by using the Logistic Regression (LR) model. It is referred to as Locally Weighted Logistic Regression (LWLR). The two independent variables considered herein were river cross section width and bank slope.

In statistics, LR is a type of regression analysis used for predicting the outcome of a categorical dependent variable (e.g. binary response) based on one or more predictor variables (continuous or categorical). The method can be used along with LWR to assign weights to local independent variables. LWR allows model parameters to vary over space in order to reflect spatial heterogeneity (Atkinson et al., 2003; Lall et al., 2006). The probabilities of the possible outcomes are modelled as a function of independent variables using a logistic function. LR measures the relationship between a categorical dependent variable and, usually, one or several continuous independent variables by converting the dependent variable to probability scores.

Then, a LR is formed, which predicts success or failure of a given binary variable (e.g. 1 = “presence of erosion” and 0 = “no erosion”) for any value of the independent variables.

The LR model is based on the logistic function, a common sigmoid function. The mathematical form is represented by the following equation:

$$p(\boldsymbol{x}) = \frac{1}{1+\exp(-\boldsymbol{x})}, \quad (1)$$

$$\boldsymbol{x} = \beta_0 + \sum_{k=1}^{K} \beta_k x_k,\ k = 1,\ldots,K, \quad (2)$$

where $p(\boldsymbol{x})$ is the probability of the dependent variable, $0 \leq p(\boldsymbol{x}) \leq 1$, associated with a given location, K is the number of the respective independent variables, β_0, $\beta_k, k = 1, \ldots K$ are the logistic regression coefficients estimated from n sample observations and x_k are the independent variables (Menard, 2001; Atkinson et al., 2003; Ozdemir, 2011). The regression coefficients are estimated by using maximum likelihood estimation.

The goal of LR is to derive estimates for the $K + 1$ unknown parameters, $\beta_0, \beta_1, \ldots, \beta_K$ by maximizing the likelihood function given in Eq. (3):

$$L(\boldsymbol{\beta} / y_1,\ldots,y_n) = \prod_{i=1}^{n} \left(p(\boldsymbol{x}_i)\right)^{y_i} \left(1-p(\boldsymbol{x}_i)\right)^{1-y_i}, \quad (3)$$

where n is the sample size, $\boldsymbol{x}_i$ represents the values of the independent variables for the i^{th} sample (Eq. 2), $p(\boldsymbol{x}_i)$ is determined by Eq. (1) and y_i is the value of the dependent variable for the i^{th} sample. As the equations are non-linear, the solution was numerically estimated using Newton’s method (Hosmer and Lemeshow, 2004).

LWR is an extension to the concept of general regression. The difference between LWR and Multiple Linear Regression, is that in LWR, the independent variables’ effect on the dependent one is weighted based on a weighted function in terms of their geographical location. Basically, LWR is a form of spatial data analysis that allows for the evaluation of a dependent variable, based on one or more local independent variables (Cleveland and Devlin, 1988; Brunsdon et al., 1996; Fotheringham et al., 2002; Atkinson et al., 2003; Lall et al., 2006). LWR is used to improve the results obtained with simple LR, allowing for the coefficients β_k to vary for each estimation point. In this work, the exponential (Eq. 4) and the tri-cubic (Eq. 5) weighting functions are used to assign weights to the observation points. The first was applied in a similar work (Atkinson et al., 2003), while the latter is a common, efficient weighting function that is used with LWR.

$$w(d) = exp(-d/a), \tag{4}$$

$$w(d) = \left[1 - |d/h|^3\right]^3, \; |d/h| \le 1. \tag{5}$$

In Eqs. (4) and (5) above, w denotes the weights, a and h are nonlinear parameters which determine the spatial correlation distance of measurement points with respect to the estimation point, for each function and d is the Euclidean distance between the estimation point and the measurement point.

3.2 Calculation of model deviance

The erosion occurrence probability can be calculated in conjunction with the model deviance. The reliability of both LR and LWLR is determined using the G-Statistic method. It is a simple and effective statistical approach to evaluate the model efficiency and the reliability of each of the independent variables tested. The model deviance is given by

$$D = -2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{p(\boldsymbol{x}_i)}{y_i}\right) + (1-y_i)\ln\left(\frac{1-p(\boldsymbol{x}_i)}{1-y_i}\right)\right], \tag{6}$$

where y is a binary variable that indicates the result of an experiment. The conditional probability of the effect to be present is expressed as $P = (y = 1|\boldsymbol{x}) = p(\boldsymbol{x})$. Variable $\boldsymbol{x} = (x_1, x_2, \ldots, x_K)$ denotes a series of independent variables. Probability $p(\boldsymbol{x})$ is calculated as in Eq. (1),

$$p(\boldsymbol{x}_i) = \frac{exp(\beta_0 + \beta_k x_k)}{1 + exp(\beta_0 + \beta_k x_k)}. \tag{7}$$

The G-Statistic is given by,

$$G = D_{null} - D_k, \tag{8}$$

where term D_{null} denotes the deviance when the model is applied without independent variables, i.e., when $p(\boldsymbol{x}) = [1 + \exp(\beta_0)]^{-1}$. Term D_k refers to the deviance for the model with k independent variables. The difference between these two terms is often cited as a sign of goodness of fit. The greater this difference, the more important is the influence of the estimation variables used. The optimal result for *D* is zero (Hosmer and Lemeshow, 2004). The process of the proposed statistical model described above was implemented with original code developed in the Matlab programming environment.

4 Results and Discussion

The BSTEM model was validated for the predicted erosion (m^2) after a field investigation that was performed at the end of the wet period of the hydrologic year of 2013-14. The eroded area at each location was successfully predicted based on the field observations at the affected area. However, quantified measurements at those points were not performed, but only field inspection to validate that the BSTEM results are consistent with reality. During the inspection, photographs were taken at some locations where the 50 cm scaled stick was placed to highlight the eroded area. However, only at the point with the most intense erosion, a close up photo was taken and analyzed to quantify the erosion.

The evaluation of the BSTEM model results, involved the calculation of the percentiles used to categorize the significance of the BSTEM calculated eroded area. The BSTEM model results are in very good agreement with the behavior of the banks after the flood events. Of the twelve measurement points, four were identified with no or low erosion, as the affected area was under or very close to the 25th percentile and equal to 0.52 m^2 (Table 1). In addition, the erosion sticks' inspection, showed that the observed affected area at the four locations, was limited considering that the bank form had not changed. The remaining eight points were identified as eroded and significantly eroded based on the model results for the affected area, while the bank form had changed at those locations. The affected area at the three significantly eroded locations ranged from 1.399 to 2.043 m^2, close and over the 75th percentile, which is equal to 1.38 m^2. Location (KI) (Fig. 1) presented the most significant erosion effect. The predicted eroded area was equal to 2.043 m^2 and the affected area measured at the field (and represented in the modified photo, Fig. 4) was roughly 2.08 m^2. The situation is the same for the other locations. The statistical model input considers the twelve measurement locations as eroded or not eroded based on the BSTEM results and the observed bank formation (Table 2).

The aforementioned results mean that the BSTEM outcome for the eight subsections of the first campaign can be also characterized as reliable, as they are located in-between of the twelve points that were successfully validated by the field inspection. The model outcome provided seven subsections with potential to erosion vulnerability and one not vulnerable to erosion, based on the estimated affected area in comparison to the total area of the bank at the respective river subsection. Therefore, they could be used as validation locations for the assessment of the statistical model performance.

Consequently, the twelve measurements of the 2nd field campaign were used to apply LR and LWLR, while the eight locations of the 1st campaign were employed as validation points. The first BSTEM application has provided a vulnerability assessment of the riverbank sections that these eight locations assign. The riverbank areas vulnerable to erosion, and therefore the associated locations are characterized as Unstable "U" and the non-vulnerable as Stable "S". Correspondingly to the LR and LWLR that deliver probabilities of erosion to occur, $P \geq 0.5$ is interpreted as presence of erosion and is denoted as Unstable "U" and absence of erosion $P<0.5$ as "S", Stable. Therefore, the different statistical model forms are validated based on the erosion vulnerability of the eight locations of the 1st field campaign (Tables 3 and 4).

The results derived from the application of the LR model, with uniform parameters for all estimation points, are presented in Table 3. The values of the independent variables and the BSTEM erosion estimates at the validation points are also presented in the same table. The model deviance was calculated equal to 6.14 and the G-Statistic equal to 7.23. Results for the erosion probability at different ungauged locations along the Koiliaris' riverbanks obtained with the LR model are presented in Fig. 5a. The values for the independent variables were randomly selected from locations among the measurement points based on a 3D digital model of River Koiliaris, which was developed based on a 5 m Digital Elevation Model (DEM) .

On the other hand, results for the erosion probability at the validation points derived by applying LWLR with the exponential and the tri-cubic weighting functions, are presented in Table 4. The graphical representation of the results for the erosion probability at the ungauged locations, is provided in Figs. 5b and 5c for the exponential and tri-cubic functions respectively. In the case of the exponential weighting function, the model deviance is equal to 6.27 and the G-statistic equal to 5.10, while in the case of the tri-cubic function, the model deviance is equal to 5.12 and the G-statistic equal to 6.25.

Inter-comparison of estimations of the three methods tested is possible, as the x and y axis of the plots (Fig. 5) are at the same scale. In addition, the validation points are shown on the plots for easier contrast. The produced 3D figures (Fig. 5) actually work as a probability map presenting the probability of erosion to occur (z axis) at the specific riverbank locations when a couple of independent values is met (x and y axes). In a similar work recently published

(Vozinaki et al., 2015), the simple LR model was applied on predicting crop damage curves based on measurements of river flood depth and velocity (secondary data). The secondary data required to develop the probability curves (predictions) were produced by a Monte Carlo simulation in the absence of sufficient measurement data. Herein, the selected secondary values were derived from the 3D river structure model as it was previously explained.

Both LWLR models, involve a nonlinear parameter in the weighting function that determines the correlation distance of the spatially correlated measurement points. The optimal distance in each case, was calculated using a leave-one-out cross validation analysis, involving the measurement locations. As a result, parameter a of the exponential weighting function, was set to 600 m and parameter h of tri-cubic function, was set to 400 m.

The results obtained with the LR method, were in very close agreement with those of BSTEM as the erosion presence or absence was accurately predicted at six out of the eight locations, with one of the fail locations to have a narrow deviance from the set erosion presence limit. Next, to improve predictions, the LWLR method was applied to account for the local spatial dependence of the independent variables at the measurement locations. The LWLR model with the exponential function has, overall, similar performance to the LR model. The derived results are in agreement with the BSTEM estimates at seven out of the eight validation locations and the approach fails at only one validation location. The application of the LWLR model with the tri-cubic function leads to significant improvement of the estimates and to the accurate prediction of the erosion probability at all eight validation locations. The significant result for this model was the validation of a clearly unstable point (pin no. 7) which has independent variables that should provide a stable indication (as delivered by LR). Another point with similar characteristics (pin no. 4, Fig. 1) was correctly identified as stable. Therefore, such performance is possible only when local spatial weighting functions are used.

The only validation point indicated as stable (pin no. 4, Fig. 1) belongs to the fourth river section (between pins no. 3 and 4, Fig. 1) which as a whole, was determined by BSTEM as stable. However, two out of the three local measurements in the same section (pins KB and KC in Fig. 1), showed signs of erosion after the inspection. Generally though, apart from limited locations, the banks of that section did not show erosion signs due to the presence of dense seasonal riparian vegetation. The erosion probability estimation at this point is affected significantly, at local scale, by the spatially correlated measurement points with low vulnerability to erosion. Similarly, validation points 6 and 7 are also affected by the close presence of measurement locations with low vulnerability to erosion. This explains the difficulty in predicting erosion at these points. The model results may confirm the presence or absence of erosion at the validation points, but they are quite different from the targeted values of zero for no erosion and one for erosion presence. This is expected to improve when a larger dataset with greater variability of the independent variables effect on erosion becomes available.

The graphical representation of the LWLR model results at the discretized river section (Figs. 5b and 5c) shows a significant difference in performance for the two weighting functions. The tri-cubic function (Fig. 5c) delivers more reliable results as it is clearly considers the variability of the independent variables inside the correlation distance. This can be observed through the color variability in the graph of Fig. 5c that represents the variability of the erosion occurrence probability. On the other hand, the exponential function (Fig. 5b) shows a smooth change in probability for the different pairs of independent variable values. This can be explained in terms of the function shape behavior and the correlation distance. The tri-cubic function is herein applied in a shorter correlation distance according to the cross validation results, which can capture the local dependence of the explanatory variables that in longer distances are smoothed due to the presence of more data.

The LWLR method with the tri-cubic function, yields the highest value for the G-Statistic for the selected independent variables. Therefore, it can be viewed as the optimum approach to calculate the erosion presence probability at local scale. The G-Statistic can be also used to assess the impact and importance of each independent variable on the estimates. Each variable was separately applied both in LR and LWLR. The G-Statistic obtained its highest values when the cross section width was applied. The results of the statistical term, improved by 12% and 20%, respectively, compared to the bank slope application.

The LR based models results suggest that riverbank erosion probability generally increases as the bank slope increases and the river cross section decreases. This is due to an increase of the flow velocity that removes the non-cohesive soil components from the banks. Based on field measurements analysis, the bank material at the Koiliaris River was classified as "fine rounded sand". The fine rounded material is easier removed due to its low resistance and increased flow friction. This characteristic is associated with the LR based models results, as they provide mainly favorable probabilities of riverbank erosion at the validation points. However, to connect the soil properties effect with the probability of erosion that results from geomorphological variables in detail, the LR based models should account also for soil properties, such as particle size distribution and the bulk density that consider also mechanical properties of the riverbanks. This is a task that the authors plan to address at a near future research.

The proposed statistical model is a useful, fast, efficient and fairly easy to apply tool that requires information from easy to determine geomorphological and/or hydrological variables. This tool provides a quantified measure of the erosion probability along the riverbanks, and could be used to assist managing erosion and flooding events. On the other hand, BSTEM model can be successfully applied to determine the potential riverbank eroded area (L^2). Both are useful, depending on data and software availability, in providing information regarding the vulnerability of riverbanks to erosion.

5 Conclusions

The BSTEM model set up provides reliable results regarding the potential erosion vulnerability of the riverbanks that can be used to validate the estimations of the proposed statistical model. On the other hand, the proposed LR based statistical model, estimates efficiently the erosion probability at the riverbanks, using two secondary variables that affect significantly the presence or absence of erosion. However, in LWLR, locality is of utmost importance; the location of the new couple of secondary variables was used to identify and weight the effect of spatially correlated measurement points in order to calculate the model parameters. The proposed methodology, LWLR, exploits the local information of independent variables and translates it successfully to bank erosion probability. This is not a typical regression estimation based on global parameters but herein, the model parameters are calculated iteratively for the new couples of secondary variables.

The LR method performs satisfactorily in the plain form where uniform parameters are considered for all estimation points. Difference from the BSTEM results is observed only at two of the eight validation points. The LWLR method with the exponential weighting function gives results similar to those of LR. The LWLR method with the tri-cubic function provides significantly improved estimates which coincide with the BSTEM results at all validation points. The graphical presentation of the results in the discretized river section shows that the erosion probability increases with bank slope and decreases with cross section width. This is also confirmed by the positive sign of the bank slope coefficients and the negative sign of the cross section width coefficients in all LR applications. The deviance and the G-Statistic results show that the cross section width parameter is more important than bank slope for the estimation of erosion probability at the banks of the Koiliaris River.

This work presents the framework of a methodology that can be applied in order to estimate the probability of erosion at specific riverbank locations considering explanatory and easy to determine secondary variables. Channel geomorphological characteristics, such as cross section and bank slope, are relatively easy to be determined at unmeasured locations by using a digital elevation model. On the other hand, hydrological variables or bank material requires extensive field measurements in order characteristic variables to be considered as secondary information. Such measurements did not take place during the field campaigns as it was not in the context of this work. The developed statistical tool provides an alternative proposition for the estimation of riverbank locations vulnerable to erosion, which requires limited information on explanatory variables, yet can provide vulnerable location estimates with increased reliability. It is therefore considered, as a very promising approach for the estimation of riverbank erosion probability. The tool is proposed as a supplementary solution to the riverbank erosion identification issue.

Author contribution

E. A. Varouchakis developed the statistical model, the model code and performed the simulations. Along with G. V. Giannakis, M. A. Lilli and N. P. Nikolaidis they designed and carried out the field campaigns, while with the aid of G. P. Karatzas, they analyzed the collected data and the model results. E. Ioannidou performed part of the model simulations. M. A. Lilli and N. P. Nikolaidis applied the BSTEM model. Finally, E. A. Varouchakis prepared the manuscript with the contribution of all co-authors.

Acknowledgements

This work is part of a THALES project (CYBERSENSORS - High Frequency Monitoring System for Integrated Water Resources Management of Rivers). The Project has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

In addition, the authors would like to thank the topical editor and the anonymous reviewers for their contribution to improve the manuscript.

References

Abam, T. K. S.: Factors affecting distribution of instability of river banks in the Niger delta, Eng. Geol., 35, 123-133, 1993.

Atkinson, P. M., German, S. E., Sear, D. A., and Clark, M. J.: Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression, Geogr. Anal., 35, 58-82, 2003.

Bakker, M. M., Govers, G., Jones, R. A., and Rounsevell, M. D. A.: The effect of soil erosion on Europe's crop yields, Ecosystems, 10, 1209–1219, 2007

Bridge, J. S.: Rivers and floodplains: forms, processes, and sedimentary record, Blackwell, Malden, Mass., USA, 2003.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E.: Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr. Anal., 28, 281-298, 1996.

Cleveland, W. S. and Devlin, S. J.: Locally weighted regression: an approach to regression analysis by local fitting, J. Am. Statist. Assoc., 83, 596-610, 1988.

Florsheim, J.,L., Mount, J. F., and Chin A.: Bank Erosion as a Desirable Attribute of Rivers, BioScience, 58, 519-529, 2008.

Fotheringham, A. S., Brunsdon C., and Charlton, M.: Geographically weighted regression: the analysis of spatially varying relationships, Wiley, Chichester, England, 2002.

Grimm, M., Jones, R., and Montanarella, R. J.: Soil erosion risk in Europe, European Soil Bureau, Institute for Environment & Sustainability JRC, Ispra, Italy, 44 pp., 2002.

Hooke, J. M.: An analysis of the processes of river bank erosion, J. Hydrol., 42, 39–62, 1979.

Hosmer, Jr. D. W. and Lemeshow, S.: Applied logistic regression, John Wiley & Sons, Canada, 2004.

Lall, U., Moon, Y. I., Kwon, H. H., and Bosworth, K.: Locally weighted polynomial regression: Parameter choice and application to forecasts of the Great Salt Lake, Water Resour. Res., 42, W05422, doi:10.1029/2004WR003782, 2006.

Luppi, L., Rinaldi, M., Teruggi, L. B., Darby, S. E., and Nardi, L.: Monitoring and numerical modelling of riverbank erosion processes: A case study along the Cecina River (central Italy), Earth Surf. Process. Landf., 34, 530-546, 2009.

Menard, S.: Applied logistic regression analysis (Vol. 106), Sage Publications Inc., Thousand Oaks, California, USA, 2001.

Midgley, T. L., Fox, G. A., and Heeren, D. M.: Evaluation of the bank stability and toe erosion model (BSTEM) for predicting lateral retreat on composite streambanks. Geomorphology, 145–146, 107–114, 2012.

Moraetis, D., Efstathiou, D., Stamati, F. E., Tzoraki, O., Nikolaidis, N. P., Schnoor, J. L., and Vozinakis, K.: High-frequency monitoring for the identification of hydrological and bio-geochemical processes in a Mediterranean river basin. J. Hydrol., 389, 127–136, 2010.

Moraetis, D., Paranychianakis, N. V., Nikolaidis, N. P., Banwart, S. A., Rousseva, S., Kercheva, M., Nenov, M., Shishkov, T., de Ruiter, P., Bloem, J., Blum, W. E. H., Lair, G. J., van Gaans, P., and Verheul, M.: Sediment provenance, soil development, and carbon content in fluvial and manmade terraces at Koiliaris River Critical Zone Observatory, J. Soils Sediments, 15, 347-364, 2014.

Nardi, L., Campo, L., and Rinaldi, M.: Quantification of riverbank erosion and application in risk analysis, Nat. Hazards, 69, 869-887, 2013.

Ozdemir, A.: Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey), J. Hydrol., 405, 123-136, 2011.

Piégay, H., Darby, S. E., Mosselman, E., and Surian, N.: A review of techniques available for delimiting the erodible river corridor: a sustainable approach to managing bank erosion, River Res. Appl., 21, 773-789, 2005.

Piégay, H., Cuaz, M., Javelle, E., and Mandier, P.: Bank erosion management based on geomorphological, ecological and economic criteria on the Galaure River, France, Regul. River, 13, 433-448, 1997.

Pollen, N. L. and Simon, A.: Estimating the mechanical effects of riparian vegetation on stream bank stability using a fiber bundle model, Water Resour. Res., 41, W07025, doi:10.1029/2004WR003801, 2005.

Pollen-Bankhead, N. and Simon, A.: Enhanced application of root-reinforcement algorithms for bank-stability modelling, Earth Surf. Process. Landf., 34, 471–480, 2009.

Pollen-Bankhead, N. and Simon, A.: Hydrologic and hydraulic effects of riparian root networks on stream bank stability: is mechanical root-reinforcement the whole story?, Geomorphology, 116, 353–362, 2010.

Raudkivi, A. J.: Loose boundary hydraulics. CRC Press, Netherlands, 1998.

Rinaldi, M., Mengoni, B., Luppi, L., Darby, S. E., and Mosselman, E.: Numerical simulation of hydrodynamics and bank erosion in a river bend, Water Resour. Res., 44, W09428, doi:10.1029/2008WR007008, 2008.

Roslan, Z. A., Naimah, Y., and Roseli, Z. A.: River bank erosion risk potential with regards to soil erodibility, in: River Basin Management VII, Brebbia, C. A (Ed.), Wessex Institute of Technology Press, UK, 289-297, 2013

Ruiz, J. M. G., Romero, A. N., Renault, N. L., and Begueria, S.: Erosion in Mediterranean landscapes: Changes and future challenges, Geomorphology, 198, 20–36, 2013.

Simon, A., Pollen-Bankhead, N., Mahacek, V., and Langendoen, E.: Quantifying reductions of mass-failure frequency and sediment loadings from streambanks using toe protection and other means: Lake Tahoe, United States, J. Am. Water Resour. Assoc., 45, 170–186, 2009.

Simon, A., Curini, A., Darby, S. E., and Langendoen, E. J.: Bank and near-bank processes in an incised channel, Geomorphology, 35, 183–217, 2000.

Simon, A., Pollen-Bankhead, N., and Thomas, R. E.: Development and application of a deterministic bank stability and toe erosion model for stream restoration. Stream Restoration in Dynamic Fluvial Systems: Scientific Approaches, Analyses, and Tools, American Geophysical Union, Geophysical Monograph Series, 194, 453–474, 2011.

Simon, A. and Thomas R. E.: Processes and forms of an unstable alluvial system with resistant, cohesive streambeds, Earth Surf. Process. Landf., 27, 699–718, 2002.

Simon, A., Thomas, R. E., Curini, A., and Shields, F. D.: Case Study: Channel stability of the Missouri River, Eastern Montana, J. Hydraul. Eng. ASCE, 128, 880–890, 2002.

Vozinaki, A. E. K., Karatzas, G. P., Sibetheros, I. A., and Varouchakis E. A.: An agricultural flash flood loss estimation methodology: the case study of the Koiliaris basin (Greece), February 2003 flood, Nat. Hazards, 10.1007/s11069-015-1882-8, 2015.

Winterbottom, S. J. and Gilvear, D. J.: A GIS-based approach to mapping probabilities of river bank erosion: Regulated River Tummel, Scotland, River Res. Appl., 16, 127-140, 2000.

Table 1. Amount of bank erosion at the measurement locations (Fig. 1) - Modelling results obtained by BSTEM.

	Eroded area (m^2)			
Map location	1st flood	2nd flood	3rd flood	Cumulative effect
KA	0.440	0.404	0.349	1.193
KB	0.566	0.510	0.394	1.470
KC	0.498	0.512	0.389	1.399
KD	0.411	0.410	0.328	1.149
KE	0.459	0.437	0.320	1.216
KG	0.258	0.255	0.213	0.726
KZ	0	0	0	0
KH	0.207	0.187	0.145	0.539
KJ	0.368	0.421	0.357	1.146
KI	0.741	0.728	0.574	2.043
KK	0	0	0	0
KL	0.167	0.162	0.132	0.461

Table 2. Presence (1) or absence (0) of erosion at measurement locations using a binary indication for the statistical model (LR and LWLR) set up based on inspection and BSTEM results. Columns 3 and 4 present the measured independent geomorphological variables.

Map location	Presence/Absence of erosion	Bank slope (degrees)	Cross section width (m)
KA	1	60	9.00
KB	1	75	9.25
KC	1	65	8.75
KD	1	55	9.00
KE	1	85	10.76
KG	1	60	11.55
KZ	0	75	10.00
KH	0	65	13.5
KJ	1	60	13.35
KI	1	75	7.60
KK	0	70	13.00
KL	0	50	9.00

Table 3. Result of LR application at the eight validation locations (Fig. 1). The independent variables used and the BSTEM estimates are also presented. In column 4, S denotes stable and U unstable bank locations.

Validation points	Bank slope (degrees)	Cross section width (m)	BSTEM erosion estimates	Erosion (P) LR
1	84	9.25	U	0.84
2	58	9.05	U	0.64
3	81	9.35	U	0.82
4	33	9.00	S	0.38
5	82.5	8.75	U	0.85
6	44	9.00	U	0.49
7	27	9.26	U	0.30
8	57	9.25	U	0.62

Table 4. Result of LWLR application at the eight validation locations (Fig. 1). The LR estimates, the independent variables used and the BSTEM estimates are also presented. With bold face the diverged values are indicated. In column 4, S denotes stable and U unstable bank locations.

Validation point	Bank slope (degrees)	Cross section width (m)	BSTEM Erosion estimates	Erosion (P) LR	Erosion (P) LWLR (exp. model)	Erosion (P) LWLR (tri-cubic model)
1	84	9.25	U	0.84	0.85	0.86
2	58	9.05	U	0.64	0.58	0.74
3	81	9.35	U	0.82	0.75	0.87
4	33	9.00	S	0.38	0.27	0.25
5	82.5	8.75	U	0.85	0.81	0.82
6	44	9.00	U	**0.49**	0.54	0.53
7	27	9.26	U	**0.30**	**0.21**	0.52
8	57	9.25	U	0.62	0.64	0.73

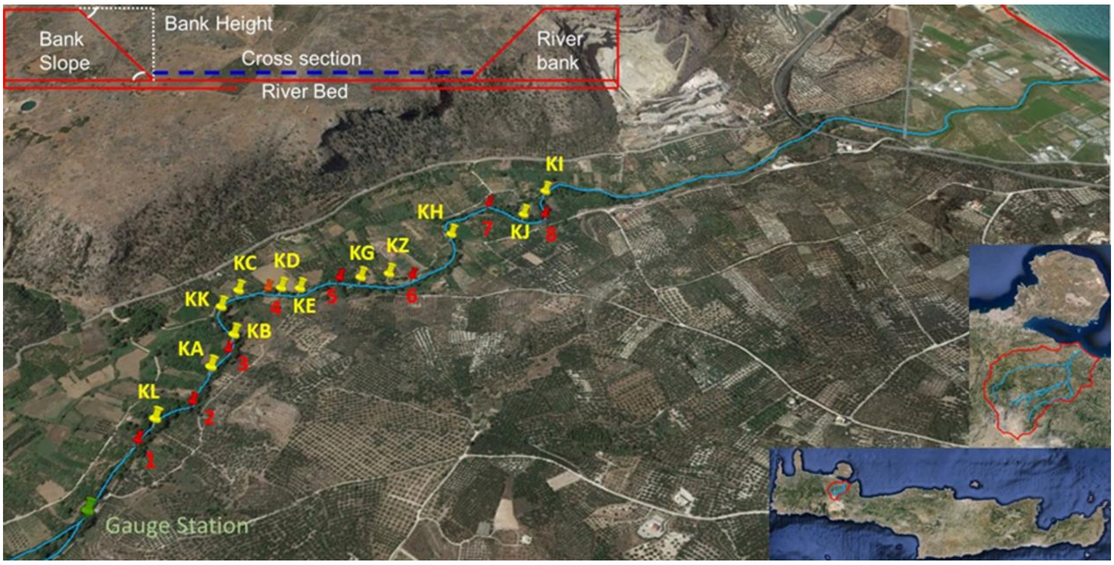

Figure 1**.** The downsream part of the Koiliaris River located in the western part of the island of Crete. The yellow pins represent the measurement locations, the red pins the validation locations and the green pin the Gauge Station located at the intersection of the the Koiliaris River with the Keramianos tributary. A representation of the measured geomorphological values is provided in the upper left corner.

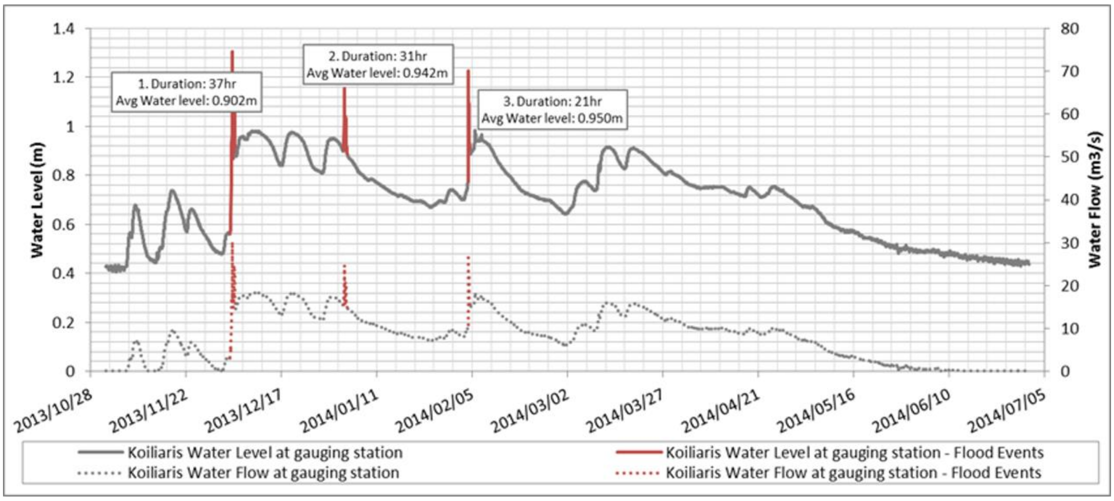

Figure 2. Typical hydrograph of the Koiliaris River at the Gauge Station (November 2013 – June 2014).

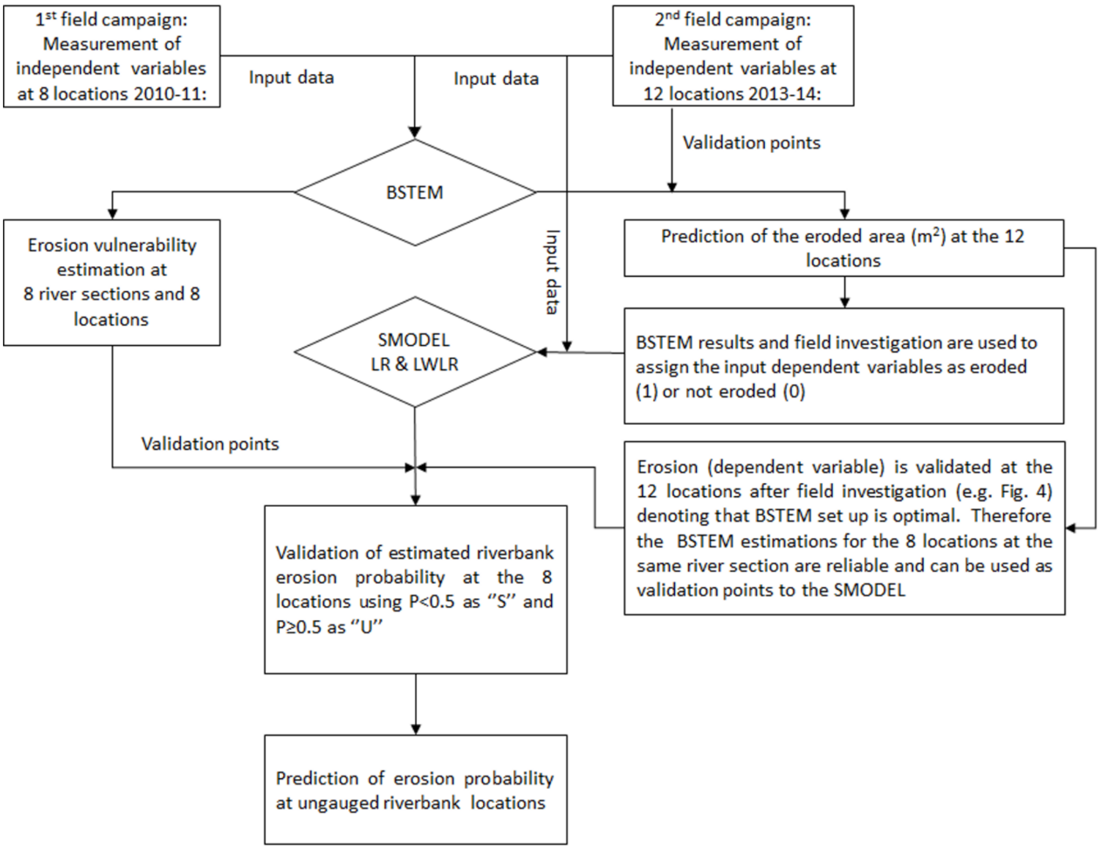

Figure 3. Process flowchart that presents the combined application of the BSTEM and of the proposed statistical model (SMODEL) based on LR principles. The notation ''S'' and ''U'' correspond to Stable and Unstable riverbanks respectively.

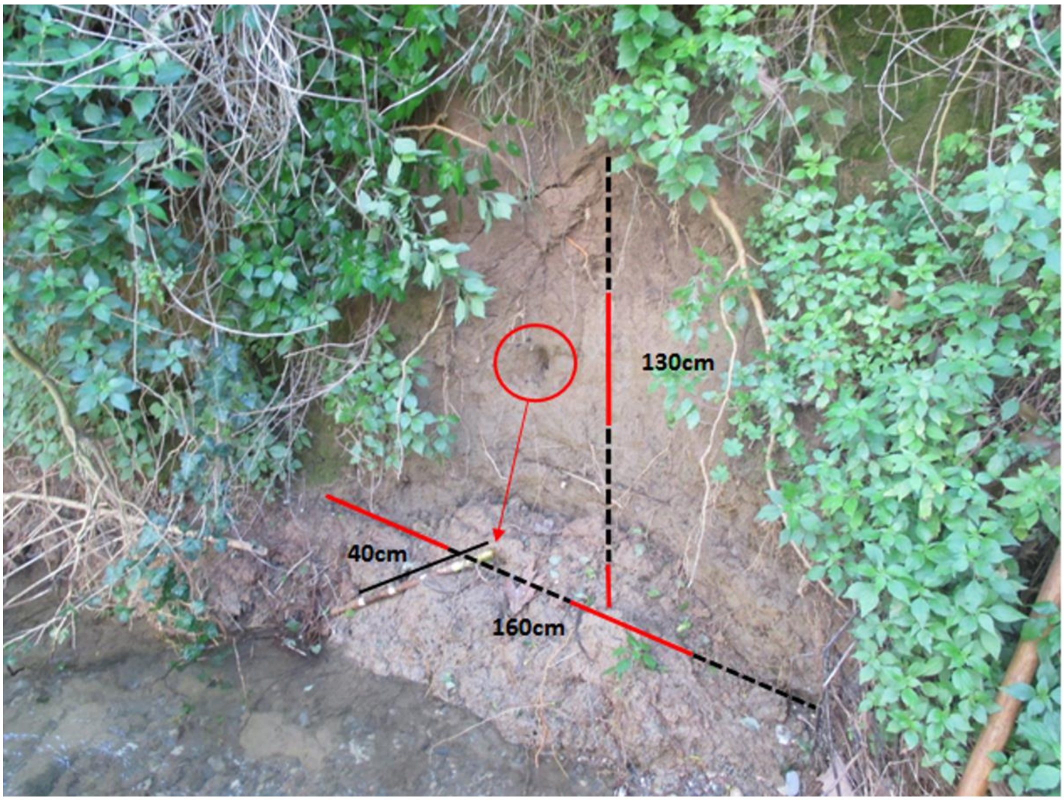

Figure 4. Photo highlight of the riverbank location (KI) with the most intense observed erosion accompanied by the appropriate scaled tools to provide a rough estimate of the eroded area.

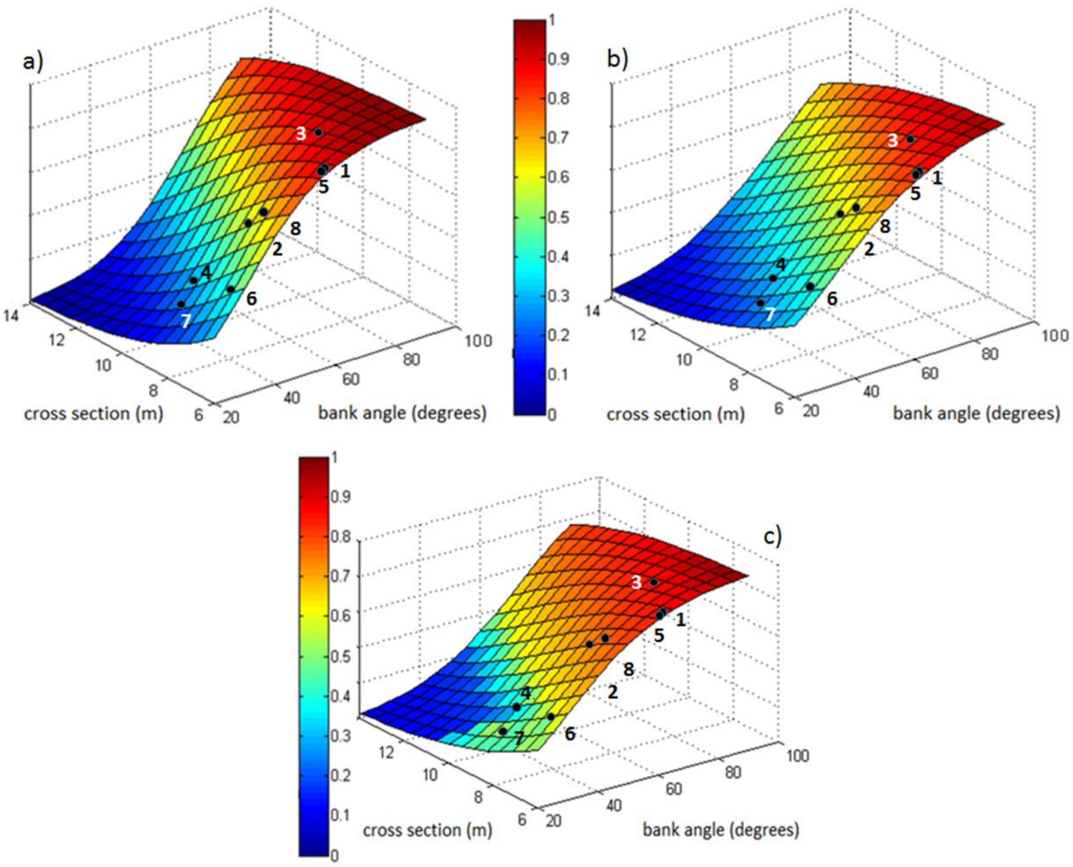

Figure 5. Erosion probability predictions using a) LR, b) LWLR with the exponential weighting function and c) LWLR with the tri-cubic weighting function versus variable independent values at random ungauged Koiliaris' riverbank locations. The black dots indicate the eight validation points.