Specific sediment yield model for reservoirs with medium-sized basins in Spain: An empirical and statistical approach

Alberto Rodríguez Bachiller , José Luis García Rodríguez , José Carlos Robredo Sánchez , David López Gómez

HIGHLIGHTS

- World reservoirs capacity loss, 0.5–1% per year, due to sediment deposition
- New regression model to predict specific sediment yield (SSY) in Mediterranean medium-sized basin reservoirs
- Complete characterization of 26 Mediterranean medium-sized basin reservoirs
- Sedimentation estimators: soil cover, slope, rainfall, drainage and reservoir depth

GRAPHICAL ABSTRACT

ABSTRACT

The world's reservoirs are losing capacity at a rate of 0.5–1% a year due to sedimentation processes. The strategies to reverse this trend must include an accurate estimation of sedimentation rates in terms of Specific Sediment Yield (SSY).

This research develops an empirical and statistical model based on data from the CEDEX (Spanish Studies and Experimentation Centre for Public Works). From an initial number of 131 reservoirs studied in the period 19,672,004, a group of 26 reservoirs with medium-sized basins (750 to 1750 km²) was selected for analysis. Reservoir catchments were described with 11 explanatory variables, representing the production, transport and deposition of sediment, although the calibrated model considers only six of these variables: Fournier Index (Rainfall Tormentality), Drainage Length, Reservoir Coefficient (relation between reservoir capacity and area), C (USLE Land Cover Factor), Yearly Average Rainfall and Slope. SSY and the explanatory variables were transformed during the calibration process, and the resulting model shows a non-linear relation between them. Compared to other models calculated with CEDEX data with a determination coefficient of between 17% and 80%, this model has a determination coefficient of 84%, is statistically consistent, validated by means of a jackknife analysis and contrasted with other models. However, the model is not cross validated with information on additional reservoirs, and shows substantial uncertainty and instability deriving from the definition of the explanatory variables and the quality of the data set, so extrapolation to other reservoirs is only possible under supervision and

1. Introduction

Reservoirs are built for various purposes, including flood control, water supply, crop and field irrigation, recreation, navigation, and the generation of hydroelectric power (Liu et al., 2018; Vörösmarty et al., 2010; WCD, 2017). There has been an increase in the number and storage volume of reservoirs, particularly over the last 70 years and there are today about 58,500 large dams (over 15 m in height or with a capacity of over 3 million cubic metres) in operation worldwide (WCD, 2017). These reservoirs are estimated to have a wide-ranging cumulative storage capacity of between 6,197 km³ (Lehner et al., 2011) and 16,000 km³ (WCD, 2017).

Although nearly 70% of global freshwater is stored in ice caps, glaciers, permanent snow and groundwater with relatively long retention periods (USGS, 2014), dams and reservoirs play a key role in the control and management of water resources. There is no question that mitigating floods, securing water supply, and providing hydropower have benefited human societies in many highly effective ways, allowing the improvement of human health, the expansion of food production, and economic growth (Lehner et al., 2011); large dams are estimated to contribute directly to 12–16% of global food production (WCD, 2000). Recent projections suggest that 70% more food will be needed by 2050 (nearly 100% in developing countries) to cope with a 40% increase in world population and to accommodate expected shifts in global dietary patterns (Bruinsma, 2009); part of that additional food will be produced on irrigated lands that will require 11% more water, much of which is likely to come from storage reservoirs. Hydropower provides about 15.8% of the world's electricity supply (IEA-ETSAP and IRENA, 2015) and is used in >150 countries (WCD, 2000). Climate change is also a factor for concern in terms of the supply of water resources (Foley et al., 2005), and on the Iberian Peninsula – where water demand is between 4% and 224% in some watersheds (Sabater, 2008) – the scenarios analysed predict severe droughts (García-Ruiz et al., 2011; López-Moreno et al., 2010).

However, due to sedimentation, the world's reservoirs are currently losing their storage capacity at an estimated yearly rate of 0.5–1% (Merina et al., 2016; Sumi et al., 2004; WCD, 2000). Similar rates were reported for various reservoirs in Spain (Batalla and Vericat, 2010) and even 5% under specific condition (Cobo, 2008). Syvitski et al. (2005) estimated that catchment-level human disturbances have increased the erosion of sediment from uplands and its delivery to rivers by about 2.3 billion tons per year, but that the net effect has been an estimated 1.4 billion-ton reduction in the sediment loads of rivers due to sediment trapping in reservoirs. Vörösmarty et al. (2003) extrapolated estimates from 633 large reservoirs to over 44,000 smaller reservoirs and concluded that >53% of the global sediment flux in regulated basins is potentially trapped in reservoirs, or 28% of all river basins, for a total trapping of 4–5 billion tons per year. This implies that without the necessary measures to reduce sediment deposition, around 25% of the world's current water storage capacity may be lost in the coming 25 to 50 years, reducing the supply of fresh and irrigation water, hydropower production and the effectiveness of flood control schemes. Sediment deposition not only affects a reservoir's storage capacity but also the way its different services are provided. The increase in delta depositions at the upstream end of the reservoir may lead to increased flooding or a possible rise in groundwater levels in the backwater upstream. The combination of sediment trapping and flow regulation also has a dramatic impact on the ecology, water transparency, sediment balance, nutrient budgets and river morphology downstream from the reservoir

(Wang et al., 2005). Outlets, pumps and turbines in hydropower facilities are affected (Morris and Fan, 1998). Reservoir sediments could also affect climate change mitigation strategies since they play different roles as a sink (Cole et al., 2007; Dean and Gorham, 1998; McCarty and Ritchie, 2002; Stallard, 1998) and source (Battin et al., 2009; Jacinthe et al., 2012) of carbon dioxide and other greenhouse gases.

Sediment deposition in reservoirs is most probably the main problem hindering the sustainable use of existing water resources and those that may be developed in the future (Wang et al., 2005). It is therefore crucial to quantify the exploitable reservoir capacity for the strategic planning of water, energy and food supplies in the coming decades, particularly in view of the predicted population growth and the exacerbation of impacts on hydrological variability due to climate change (Boehlert et al., 2015; Kundzewicz and Stakhiv, 2010; Soundharajan et al., 2016; Stillwell and Webber, 2013; Vörösmarty et al., 2000).

There are many approaches and models available for the estimation of reservoir sedimentation. However, they differ widely in terms of their complexity, inputs and other requirements. (de Vente et al., 2011; Yang and Lu, 2014). The first step is to determine the capacity of a watershed to produce sediments and conduct them to a specific point, and this is usually solved by calculating the SSY (Specific Sediment Yield).

The SSY is formally defined (Vanoni, 1975) as the total sediment outflow from a drainage basin or catchment area, measurable at a reference cross-section in a specific period of time. It is normally expressed in mass per unit area and unit time ($t \text{ km}^{-2} \text{ y}^{-1}$).

The SSY of a watershed can be determined by different methods (de Vente et al., 2013) which can generally be grouped into two kinds: empirical models, usually defined as a statistical regression using observed data; and parametric or physically based models, which are defined and based on hydrological and sedimentological processes which tend to require detailed and extensive data input.

In the first group, the most widely accepted method considers the sediment that has been deposited in a reservoir over a period of time (de Vente et al., 2011; ICOLD, 1989) by relating SSY and a predictive variable, usually the drainage area, and defining a curve. Many studies have reported a decreasing area-specific sediment yield (SSY; $t \text{ km}^{-2} \text{ y}^{-1}$) with increasing drainage basin area (A ; km²) (Dendy and Bolton, 1976; Einsele and Hinderer, 1997; Milliman and Meade, 1983; Milliman and Syvitski, 1992; Radoane and Radoane, 2005; Renwick et al., 2005; Summerfield and Hulton, 1994). This is often explained by the fact that, during transport through a river basin, progressively more sediment is trapped in foot slopes, concavities, alluvial plains and other sinks, while erosion rates do not increase and may even decrease due to decreasing hillslope gradients downstream (de Vente et al., 2007). This negative relation between A and SSY has often been extrapolated to make predictions of sediment yield for ungauged basins at the regional or even at the global scale (Milliman and Meade, 1983; Milliman and Syvitski, 1992; Renwick et al., 2005; Syvitski et al., 2003, 2005), often using a potential regression model (de Vente et al., 2007) following the expression $SSY = \alpha A^\beta$, where A stands for drainage basin area (km²), and α and β are empirical parameters. However, other studies have revealed either a very poor relation, a positive relation or a combination of positive and negative relation between A and SSY (Church and Slaymaker, 1989; de Araújo and Knight, 2005; de Vente et al., 2006; de Vente and Poesen, 2005; Dedkov, 2004; García Ruiz et al., 2004; Gögüs and Yener, 1997; Haregeweyn et al., 2005; Jiongxin and Yunxia, 2005; Krishnaswamy et al., 2001; Lane et al., 1997; Restrepo et al., 2006; Schiefer et al., 2001). The reasons for these

results are diverse, but they show that in some cases the drainage area is not a significant predictive variable of SSY and suggest that other variables such as topography, land cover, soil properties, climate characteristics, water discharge or gully effect should be considered (Bellin et al., 2011; de Araújo and Knight, 2005; de Vente et al., 2011; Duvert et al., 2012; García Ruiz et al., 2004; Grauso et al., 2008; López-Vicente et al., 2011; PSIAC, 1968; Tamene et al., 2006). Topographic variables that affect soil erosion capacity and hence reservoir sedimentation rates include slope angle and slope length. The higher the slope angle and the longer the slope length, the greater the likelihood of erosion (Bonan, 2008; Moore and Burch, 1986). Soil characteristics that affect soil erosion include its texture, saturated hydraulic conductivity and soil organic matter, among others (Rorke, 2000). Land cover plays a key role in water erosion by reducing surface water flow rate, while the roots keep the soil in place. However, land management practices such as conventional tillage for crop production, deforestation and overgrazing can accelerate the rate of erosion by disturbing soil and reducing vegetation cover (Cooper et al., 2013; Nearing et al., 2005). Rainfall is the major driver for sheet and rill erosion, in which total rainfall and its intensity are important variables. Intense storms increase the possibility of soil particles becoming detached and generating more surface runoff to transport detached particles (Guzzetti et al., 2007). Both the transport and deposit of sediments must be considered when addressing the issue of sedimentation. The properties of the drainage network and the reservoir, including the dam operating procedures, play an important role in the sedimentation process (Hauer et al., 2018; Merritt et al., 2003).

Physically based models tend to be used in cases where runoff and sediment load data records are unavailable. These models are of two types, namely single storm models and continuous simulation models. The Areal Non-Point Source Watershed Environment Response Simulation (ANSWERS) (Beasley et al., 1980) and the European Soil Erosion Model (EUROSEM) (Morgan et al., 1998) are examples of the first kind, while the second include the spatially distributed erosion and sediment yield component, chemicals, runoff and erosion from agricultural management systems or CREAMS (Science and Education Administration; Department of Agriculture, Washington, WA, USA) (Knisel, 1980); the SHESD model (the hydrological and sediment transportation model for hydrological modelling systems (SHE)) (Wicks and Bathurst, 1996); and the soil and water assessment model (SWAT) (Arnold et al., 1998a, 1998b). The SWAT model is the most commonly used, so a number of researchers have modified this model for different purposes (Easton et al., 2010; Swami and Kulkarni, 2016).

In the period between 1967 and 2004, the CEDEX (Centro de Estudios y Experimentación de Obras Públicas – Studies and Experimentation Centre for Public Works) in Spain conducted sedimentology and bathymetric surveys on 131 reservoirs where the basin characteristics (soil type, land uses, meteorological events, predominance of steep slopes, etc.) were potentially susceptible to the development of erosion processes that could affect their useful lifespan. The CEDEX project covered a total reservoir capacity of 20,697 hm³ (36.98% Spain's reservoir capacity (MAPAMA, 2017)).

The first study based on a partial data set from these works was published in 1997 (Avendaño Salas et al., 1997). It analysed 60 Spanish reservoirs and defined three groups of reservoirs as a result: low SSY (<150 t km⁻² y⁻¹), intermediate SSY (150–1000 t km⁻² y⁻¹) and high SSY (>1000 t km⁻² y⁻¹). A regression curve was estimated in each group for the relation SY(t y⁻¹) – Area (km⁻²). Considering SSY (t y⁻¹ km⁻²) as the dependant variable, Area (km⁻²) only explains 17% of the observed variability in SSY ($R^2 = 0.17$ with a model significance of <1%), showing that a large part of the variation in sediment yield is not explained by basin area.

Based on the data set considered in this study, other researches have included descriptive variables of the watersheds and obtained improved fits. For example, the Factorial Scoring Model (FSM) (de Vente et al., 2005; Verstraeten et al., 2003) explains 69% of the variability found in reservoir sedimentation rates, while the PSIAC model (PSIAC,

1968) explains 61% (de Vente et al., 2005), the WATEM-SEDEM model (Van Oost et al., 2000; Van Rompaey et al., 2001; Verstraeten et al., 2002) 48% (de Vente et al., 2008), and the SPADS model 67% (de Vente et al., 2008); and two statistical approaches 58% (de Vente et al., 2011) and 39% (Buendía et al., 2016).

The present research considers the complete collection of CEDEX works consisting of 272 reports (181 bathymetric studies and 91 sedimentology studies) compiled by the CEDEX in the period 1967–2004, in 131 reservoirs in Spain. This comprehensive material has been analysed to extract the data on reservoirs with medium-sized basins (750 to 1750 km²) and to define the relations between the main characteristics of this group of reservoir basins and the sedimentation process, expressed as SSY.

2. Material and methods

The analysis was carried out in three different steps. First, the CEDEX data was studied to select the group of reservoirs used to characterize and calculate the model. These reservoirs and their watersheds were then described according to ten parameters (explanatory variables), selected for their relation with the sediment production, transport and deposit processes in a reservoir watershed. Finally, the database containing this information was statistically analysed to define a model for the estimation of SSY in terms of the “sedimentary” characteristics of the reservoirs and their watersheds. These methodological steps are described in the following sections.

2.1. Reservoir screening

In the first stage the CEDEX data was analysed and filtered according to five criteria:

- **Information validity:** some results in the CEDEX reports cannot be considered, mainly because they show a net increase in the reservoir capacity between commissioning and the date when the data was gathered.
- **SSY calculation/sediment density:** a complete characterization of the sediment, including density, was made in most, although not all, the reservoirs studied by the CEDEX. Given that the density is required to transform the capacity loss (hm³) to sediment yield (t), these data were also excluded from the analysis.
- **Sample representativeness:** some reservoirs were studied by the CEDEX on several occasions, which is why there are 272 reports on 131 reservoirs. In order to maintain the statistical consistency of the data, and the optimal representativeness of the sedimentation processes, only the last study on each reservoir has been included in the analysis.
- **Statistical outliers:** after applying the above criteria, the SSY variable was analysed in SPSS PASW Statistics 17 (SPSS Inc. Released 2008. SPSS Statistics for Windows, Version 17.0. Chicago: SPSS Inc.) to define the outliers (observations outside the inner fences defined as $1.5 \times \text{IQR InterQuartile Range} = Q3 (75\% \text{ of observations}) - Q1 (25\% \text{ of observations})$).
- **Medium-sized basin (750 to 1750 km²):** some sedimentation models are conditioned by the size of the basins used to define them (de Vente et al., 2013). The analysis therefore focused on a specific range of basin sizes: 750 km² to 1750 km². This range, according to the reservoir data set resulting from the four screening criteria described above, includes the average basin size of the reservoirs analysed (950 km²) and approximately half the total area of the reservoir basins, and can be considered representative of medium-sized basins in Spain.

As a final result of the screening process, 26 reservoirs were selected for the analysis (Fig. 1).

Fig. 1. Selected reservoirs.

2.2. Reservoir basin characterization

There are many factors that explain sediment production (erosion), transport and deposition in a reservoir. This study considers ten characteristics, most of which are commonly used in hydrological and erosion studies, so they are available or can be easily estimated. The description of each reservoir is completed with the basin area (A) (km^2).

- **Soil erosivity (USLE K factor):** soil properties related to erosion have been described with the soil erodibility factor (K) in the USDA SCS USLE (Universal Soil Loss Equation) (USDA SCS, 1965). K values have been defined from 3747 soil sample analyses distributed all over Spain and used to prepare the Erosion Situation Map of Spain (Ministerio de Agricultura, Pesca y Alimentación. Instituto Nacional para la Conservación de la Naturaleza, 1987–2001) and weighted in every basin using Thiessen polygons (Thiessen and Alter, 1911) (Fig. 2). In each reservoir K was estimated by intersecting the network of Thiessen polygons, each characterized by a K value, and the reservoir's watershed, then calculating the weighted mean of K from the area of the Thiessen polygons inside the watershed.
- **Rainfall erosivity (USLE R factor):** this is the kinetic energy of rainfall, multiplied by 130 (maximum intensity of rain in 30 min expressed in cm hour^{-1}). This index corresponds to the potential erosion risk in a given region where sheet erosion appears on a bare plot with a 9% slope. In this study, it has been estimated as $h_j \text{ cm m}^{-2} \text{ h}^{-1}$, using regression models that relate annual R and yearly average rainfall (Roldán Soriano et al., 2009), described later in this section.
- **Soil cover and management (USLE C factor):** this is a simple relation between erosion on bare soil and erosion observed under a cropping system. The C factor combines plant cover, its production

level and the associated cropping techniques and varies from 1 on bare soil to 1/1000 under forest, 1/100 under grassland and cover plants, and 1 to 9/10 under root and tuber crops. This study uses the Forestry Map of Spain 1:50,000 (Ministerio de Medio Ambiente. Banco de Datos de la Naturaleza, 1998–2007) and the study of vegetation cover in Spain (Ministerio de Medio Ambiente. ICONA, 1982) to define the C factor in each basin (Fig. 3).

- **Slope (S):** this factor represents the role of topography in the sediment production and deposition process, and is also estimated using a Digital Terrain Model (Ministerio de Fomento. Dirección General del Instituto Geográfico Nacional, 2017). The slope in each pixel of the reservoir basin has been reduced to a single average value for each basin and expressed as a percentage (Fig. 4).
- **Runoff threshold (P_0):** this value represents the rainfall needed to start a runoff process and depends on the land cover (Témez, 1987). The concept is derived from the SCS curve number method (USDA SCS, 1985) and adapted to Spanish conditions. Runoff threshold and runoff have a direct negative relationship so that the higher the runoff threshold, the lower the runoff and hence the erosion it produces. These factors have been extracted from the runoff threshold map (Ministerio de Agricultura, Pesca, Alimentación y Medio ambiente, 2003) and as in the previous case, the different values in each basin have been summarized as a single average value (Fig. 5).
- **Yearly average rainfall (YAR):** this study characterizes climate using the yearly average rainfall in each basin, defined by the meteorological data from the MOPREDAS project (González-Hidalgo et al., 2011). This project develops a regular network of points, characterized with information, including monthly rainfall, from 1950 to 2010 from the Spanish meteorological stations in the National Meteorological Agency (Ministerio de Agricultura, Pesca, Alimentación y Medio Ambiente) (www.aemet.es). The rainfall information for

Fig. 2. Factor K estimation. Dots represent the points where the soil K factor was calculated to elaborate the Mapa de Estados Erosivos de España (Erosion Situation Map of Spain) and polygons are their corresponding Thiessen polygons, for the Peñarroya reservoir (Guadiana).

each point has been weighted for the reservoir basins considered using Thiessen polygons, and a single yearly average rainfall value has been estimated for each basin (Fig. 6).

- **Fournier index (FI). Rainfall aggressiveness:** another climate factor selected to describe basins is rainfall aggressiveness, defined according to the Fournier index as P/P_m , where P is the monthly average rainfall and P_m is the maximum monthly average rainfall (Fournier, 1960). After calculating this index for each MOPREDAS point, a single average value is assigned to each basin.
- **Basin Gravelius coefficient (GC)** (Gravelius, 1914): this is defined as the relation between the perimeter of the watershed and that of

a circle with the same area as the watershed; that is: $0.28 P/\sqrt{A}$, with P : perimeter (km) and A : area (km^2). This coefficient provides information about the basin shape that can be related with the sediment transport process since in elongated basins ($GC \approx 2$) the distance from the head of the basin to the reservoir is greater than in basins with a rounded shape ($GC \approx 1$).

- **Drainage length (DL):** is the sum of the entire drainage basin network, estimated using the analysis of the digital terrain model and considering a minimum watershed size of 200 ha and a length of 50 m (Fig. 7). This descriptor concerns sediment transport, since the drainage length represents the distance that sediments have to

Fig. 3. Factor C estimation. Forestry map polygons and assigned C factors, for the Forata reservoir (Júcar).

Fig. 4. Slope estimation. Using a 20×20 m DTM, for the Guadalquivir reservoir (Gudalquivir).

travel from the basin to the reservoir, and the greater the distance, the smaller the sediment load reaching the reservoir (Kirkby, 1991; Parsons et al., 1998).

- **Reservoir coefficient (RC)**: this factor has been defined as the relation between the initial reservoir capacity (hm^3) and its area (hectares); it is therefore a measure of its mean depth and has been included as a descriptor of the topography of the reservoir banks, since great depths are related with steep banks.

These characterization variables have been calculated and assigned to each basin using QGIS 2.14 (QGIS 2.14.12-Essen. GNU General Public

License. Version 2, June 1991. Copyright (C) 1989, 1991 Free Software Foundation, Inc.) Geographically distributed variables (for example R, C, S, etc.) have been reduced to an average value for each basin, using the area of each value inside the watershed as a weighting factor.

2.3. Reservoir specific sediment yield (SSY)

The ultimate objective of the CEDEX bathymetric studies was to estimate the capacity loss in each reservoir in terms of volume (hm^3) while the sedimentology studies describe the sediment characteristics, including density. Using this information, SSY values have been

Fig. 5. P_0 Runoff threshold estimation. Guadalquivir reservoir (Mediterránea Andaluza).

Fig. 6. MOPREDAS Meteorologic Data Points and their corresponding Thiessen polygons, for the Puentes reservoir (Segura).

calculated in each reservoir simply by considering the number of years since the dam was commissioned until the bathymetric study and the basin size, as follows:

$$SSY \left(t \, km^{-2} \, year^{-1} \right) = \frac{\text{Capacity loss} \left(hm^3 \right) \times \text{density} \left(t \, hm^{-3} \right)}{\text{Years in operation} \left(year \right) \times \text{basin size} \left(km^2 \right)} \quad (1)$$

The SSY definition considers that all the incoming sediments that reach a reservoir will remain trapped in it, although some sediment may escape from the reservoir through spillways, hood inlets, and other outlets depending on the hydraulic conditions in each case.

This assumption is based on the trapping efficiency (TE) calculated by CEDEX in some of the reservoirs using the basin area, reservoir capacity and operating regime (Brown, 1943), and whose results are in the range of 99.5% to 97%. This coefficient was not calculated in many other cases due to lack of information. The research therefore

considered a $TE = 100\%$, so SSY values refer strictly to the sediments deposited in the reservoirs. In any case, the deviation produced by this assumption is balanced with the uncertainty of the bathymetric surveys, conducted over a period of 37 years with different technologies and personnel.

Both the SSY calculation and the reservoir characterization consider only the basin area of each reservoir and exclude the basin area of any possible reservoir located upstream, thus removing their influence from the analysis.

2.4. Statistical analysis

Regression analysis is a statistical method that allows the examination of the relationship between two or more variables. In multiple linear regression analysis, as used in this research, a single response variable is modelled with several independent variables (Weisberg, 1980). In this study, regression is applied in an iterative way until the optimization of the explanatory variables considered in the model, and those without a significant correlation with the dependant variable, SSY, are removed.

In the first step the whole set of descriptors (11 variables, considering the basin area) is used to fit a SSY model, and any variables with a non-significant relation with SSY or with signs of collinearity are then removed from the analysis. This screening contributes to the model quality, which is inversely related with the number of independent variables.

The data on reservoirs were analysed using SPSS PASW Statistics 17. This statistical software offers multiple statistical analysis tools, including a multilinear regression analysis (Norušis and SPSS Inc, 2008).

The regressions used the "Enter" method, so all the variables must be in the model. In this method the inclusion of the variables in the model is not automatically decided by the regression process but by the users after analysing each iterative regression. The model must also include a constant to ensure the residuals have a mean of zero. The level of statistical significance was set at 0.05.

SPSS PASW Statistics 17 was configured in each regression to calculate different coefficients, statistics and indicators of the model's quality

Fig. 7. Drainage Network of La Tranquera reservoir (Ebro).

(Montgomery et al., 2012). Specifically, these tests are conceived to check the model's data correlation (determination coefficient R^2 and ANOVA significance) (Bentler and Bonett, 1980), the correlation of the explanatory variables with SSY (two-tailed Student's t significance of each variable, both unstandardized and standardized coefficients β) and the collinearity of the explanatory variables (VIF, Durbin-Watson and condition index tests of collinearity) (Graham, 2003).

The determination coefficient, R^2 , is the proportion of explained variance and is between 0 (no correlation) and 1 (perfect correlation). This estimator is highly intuitive since its value can be directly interpreted as the percentage of values that can be predicted by the estimated model. The one-way ANOVA (analysis of variance) F statistic, commonly known as the F -test, is used to determine whether the means of a variable differ between the levels or groups of another variable. That is, it compares the means of the dependent variable between the groups or categories of the independent variable. If the means of the dependent variable are different in each group or category of the independent variable, the groups differ in the dependent variable, and there is therefore a relationship between the variables. The F -test is obtained by estimating the variation of the means between the groups in the independent variable and dividing it by the estimation of the variation of the means within the groups, so by dividing the variation between the groups by the variation within the groups. If the means vary widely between the groups and the mean within a group varies slightly – that is, the groups are all heterogeneous but similar internally, the value of F will be higher, and the variables will therefore be related. In conclusion, the more the means of the dependent variable differ between the groups of the independent variable, the higher the value of F .

The significance of the F -test is assessed by calculating the p -value (probability value), which is defined as the probability of the value of F when the null hypothesis (H_0) is true. In this case, the null hypothesis is that there are no relations between dependent and independent variables, so the model has no independent variables and is a simple intercept-only model. The alternative hypothesis (H_1) is the calculated model. Following a confidence level of 95%, when the significance of F is <0.05 , the null hypothesis is rejected and it is statistically accepted that the variables are related, so the model fits the data. The two-tailed Student's t -test, also known as the t -test, is any test in which the statistic used has a Student's t distribution (Student, 1908) if the null hypothesis is true. It is applied when the population follows a normal distribution but the sample size is too small for the statistic on which the inference is based to be normally distributed, using an estimate of the standard deviation instead of the actual value. In linear regression, the statistic used is the difference between the calculated coefficient (alternative hypothesis H_1) and zero (null hypothesis H_0) divided by the standard error of least squares estimators. If this statistic has a t distribution with $n-2$ degrees of freedom (n = sample size), the null hypothesis is true, so the coefficient of relation between the independent and dependent variables is zero and there is consequently no relation.

The significance of the t -test, as described above with the F -test, is also evaluated by the calculation of the p -value (probability value) and the interpretation is the same, so when the significance of the t -test is <0.05 , the null hypothesis is rejected and it is statistically accepted that the variables are related by the calculated coefficient.

The unstandardized and standardized coefficients (β) are the factors of each independent variable used to build the model, but expressed in different units. While unstandardized coefficients are expressed in their original units, standardized coefficients are expressed in the same normalized units. The unstandardized coefficients can thus be directly used as coefficients of the independent variables in the estimated model, and the standardized coefficients can be used to interpret and compare the effect of each independent variable on the dependent variable.

The VIF (Variance Inflation Factor) quantifies, for each coefficient, the variance inflation that is considered a sign of collinearity. A VIF of

1 means there is no correlation between the predictor and the remaining predictor variables. The general rule of thumb is that VIFs exceeding 4 warrant further investigation, while VIFs exceeding 10 are signs of serious multicollinearity requiring correction (Hair et al., 2010). The Durbin-Watson statistic tests the collinearity with the residuals, taking values in the range of 0 to 4, and 2 when the residuals are completely independent. Values >2 indicate positive autocorrelation and <2 denote negative autocorrelation. It is usually assumed that the residuals are independent if the Durbin-Watson statistic is between 1.5 and 2.5. Finally, the condition indexes are derived from the breakdown of the correlation matrix into linear combinations of variables, chosen so that the first combination (first level or dimension) has the largest possible variance; the second combination (second level or dimension) has the next largest variance, assuming it is uncorrelated with the first; the third has the largest possible variance, subject to being uncorrelated with the first and second; and so forth. The variance of each of these linear combinations is known as an eigenvalue, and the condition index is calculated as the square root of the ratio between the largest eigenvalue and the eigenvalue of each level or dimension. In conditions of non-collinearity, these indexes should not exceed 15. Indexes over 15 indicate a possible problem, and indexes over 30 point to a serious problem of collinearity.

The three basic conditions of multiple linear regressions (Cohen et al., 1983) were also tested to guarantee the statistical coherence of the model: normally distributed residues, homoscedasticity and no multicollinearity.

If the relationship between independent and dependent variables is not linear, the data form a curved pattern, which could be better modelled by a nonlinear function such as a quadratic or cubic function, or transformed to make it linear. The statistical methodology of this research therefore includes the transformation of the selected explanatory variables using the Box Cox technique (Box and Cox, 1964) in addition to the multiple linear regression analysis. This method uses a power expression to transform the variables: $F(\text{Variable}) = \text{Variable}^\lambda$. The power exponent (λ) was estimated using the Box and Tidwell methodology (Box and Tidwell, 1972). The model fit is thus improved and potential collinearity issues are reduced.

The transformation of the independent variables allows the exploration of both linear and non-linear correlation between observations and explanatory variables using the linear regression methodology.

Finally, the estimated model was resampled using the jackknife procedure (Shao and Tu, 1995) based on estimating n models with $n-1$ reservoirs. The estimated models are then used to calculate the SSY in the remaining reservoir n times, comparing their values with the SSY observed. In each model, both the determination factor R^2 and the ANOVA significance are calculated to evaluate their performance. The jackknife method is a useful technique for correcting estimation bias and variance and is used in this study to detect the potential bias of the sample and correct it if necessary.

2.5. Model performance test

As an additional test for the model, the SSY values calculated with the calibrated model were correlated with the SSY values observed and compared with the results yielded by other previously developed prediction equations. Since this study focuses on reservoirs in the Mediterranean area, three models matching this requirement were selected: the global Area Relief Temperature (ART) (Syvitski et al., 2003, 2005) model and two models calibrated with reservoirs in Spain. The performance test was based on the correlation of the SSY values predicted by these models and the SSY values observed and estimated by the calibrated model, whose significance has been evaluated by the Pearson's correlation coefficient (r) (Pearson, 1895), defined as the covariance of the two variables divided by the product of their standard deviations. This coefficient takes values in the range of -1 to 1 , where -1 means negative perfect correlation, 1 positive perfect correlation and 0 no

correlation. The two-tailed significance of r (significant if it is <0.05) has been also evaluated.

The ART sediment delivery model (Syvitski et al., 2003, 2005) is a nonlinear regression model that was estimated using 340 rivers in different regions and climate zones all over the world. It consists of five regression equations for five climate zones that estimate the sediment load (Q_s , expressed in kg s^{-1} , so it must be transformed to SSY units $\text{t km}^{-2} \text{y}^{-1}$) and uses watershed area (A), large-scale relief (R), and mean annual temperature (T_m). For the Mediterranean region (temperate regions and latitude $<30^\circ$) the expression of this model is:

$$Q_s = 6.1 \cdot 10^{-5} \cdot A^{0.55} \cdot R^{1.12} \cdot e^{0.07 \cdot T_m} \quad (2)$$

The FSM Model (Factorial Scoring Model) is an expert model (de Vente et al., 2005; Verstraeten et al., 2003) that uses the catchment area (A) and five factors (topography, lithology, vegetation cover, gullies, catchment shape) to describe the watershed in the areas near (5 km) the reservoir and main tributaries. These factors are given a score between 1 and 3 (from low to high contribution to soil erosion, sediment production and delivery to the reservoir) and an index ($\text{FSM}_{\text{index}}$) is calculated by multiplying the scores. Both A and $\text{FSM}_{\text{index}}$ define the following regression equation for the estimation of SSY.

$$\text{SSY} = 4139 \cdot A^{-0.44} + 7.77 \cdot \text{FSM}_{\text{index}} - 310.99 \quad (3)$$

Finally, this study also uses a multiple regression model calibrated with 61 Spanish reservoirs (de Vente et al., 2011), attaining a correlation of $R^2 = 0.58$. It is based on slope gradient (S), relief ratio (RR), precipitation concentration index (PCI), percentage of matorral land cover (Mat) and several soil composition rates (percentage of fluvisol soils (Flu), limestone ($Lime$) and acid metamorphic rock ($Meta$)). The model has the following expression:

$$\ln \text{SSY} = 1.35 + 0.07 \cdot S + 0.11 \cdot RR + 0.34 \cdot PCI - 0.03 \cdot Mat - 0.01 \cdot Lime - 0.04 \cdot Meta + 0.04 \cdot Flu \quad (4)$$

The model performance will also be analysed through a sensibility analysis. This test is based on the comparison of the SSY predicted with the model using both the values estimated for each predictive variable and the modified values of these variables. These modifications have been defined by increasing and decreasing the explanatory variables by $\pm 1\%$, $\pm 10\%$ and $\pm 25\%$, and the results will be given as the average modification of the output values of SSY for each variable and modification percentage.

2.6. Model validation

Validation by splitting the data set in two parts, one for model development and one for model validation (cross validation) was rejected due to the low number of cases (26) resulting from the reservoir screening. The sample size is a frequent subject of debate in linear regression (Lawrence and Kerry, 1989). In this case, the sample is representative of the total set of reservoirs with medium-sized basins, since almost 50% (MAPAMA, 2017) of the reservoir capacity of the Mediterranean area of Spain was studied by CEDEX. However, as an absolute value, it is not an optimal sized sample for the number of independent variables considered (Cohen, 1988); in addition, the reduction of the sample size for the regression analysis could produce determination coefficients that are very significant but not consistent. Cross validation, the most accepted validation method, cannot therefore be used. However, some studies use other validation techniques when the data set is too small to be divided into two parts.

Some authors (Bazoffi et al., 1996; Buendía et al., 2016; de Vente et al., 2011; Verstraeten and Poesen, 2001) therefore use the jackknife resampling method to analyse the results of the n models obtained from $n-1$ data (see the description of the jackknife methodology), and

they are compared with the results of the model to be analysed. If the predictions of the resampled models are close to those of the model, it can be assumed to be validated.

3. Results

3.1. Reservoir database

As described above, the preparation of the database starts with the selection of the reservoirs to be analysed. In this process, many of the reservoirs in the original CEDEX information were eliminated following the criteria of information validity, sediment density, statistical outliers and medium basin size. Thus from an initial number of 131 reservoirs, the model fitting was reduced to 26 reservoirs (Fig. 1). Reservoirs such as Cijara (whose basin is seven times the size of the second largest basin), Mequinenza (1530 hm^3), Alarcón (1117 hm^3), Iznajar (980 hm^3), Entrepeñas (891 hm^3), Guadalcaçín (800 hm^3) and Ebro (540 hm^3), among others, were removed.

After selection, each reservoir was characterized according to the basin descriptors above, and this information, together with the values of the basin area, SY and SSY, provided by the CEDEX reports, was used to compile the database (Table 1). As previously stated, the reservoir basins range from 750 to 1750 km^2 . Eleven basins have an area of $<1000 \text{ km}^2$ while the average basin size is 1128 km^2 .

The reservoir capacities range from 13.7 to 347.4 hm^3 . Six basins are below 25 hm^3 so can be considered as small in size, while nine are over 150 hm^3 .

An analysis of SSY values shows that the average is $187 \text{ t km}^{-2} \text{ year}^{-1}$, with a significant variety represented by one reservoir with nearly $949 \text{ t km}^{-2} \text{ year}^{-1}$ and three reservoirs below $10 \text{ t km}^{-2} \text{ year}^{-1}$. If these values are compared with others cited in the bibliography, it can be concluded that the range of specific sedimentation measured in these 26 reservoirs is among the lowest in the world.

A bibliographical review of SSY in small basins ($<1000 \text{ km}^2$), the case of 42 reservoirs in this study, shows a broad range of erosion rates from 7.3 to $17,000 \text{ t km}^{-2} \text{ year}^{-1}$ (Laignel et al., 2006). The most frequently cited study is the synthesis of 280 rivers with a wide variety of basin sizes, climates and geomorphologies, on a global scale (Milliman and Syvitski, 1992) and with an erosion rate between 0.3 and $36,000 \text{ t km}^{-2}$.

The reservoir characteristics also vary widely. K values are between 0.16 and 0.47 , with all except three values, in the 0.25 to 0.4 range, with a high variability of soils (Romkens et al., 1989). R values are from 44 to 176 , with a mean value of 94 , indicating low-medium rainfall erosivity (ICONA, 1988). The other USLE factor, C , is between 0.05 and 0.13 and has an average value of 0.08 , with low C values or good soil coverage (Folly et al., 1996). Average S is 10.7% , and varies from 2.59% to 24.1% . The S in 15 reservoirs is below 10% . The reservoirs also show a broad range of rainfall values (YAR) from 341 to 950 mm , with an average value of 590 mm , mainly representing a Mediterranean climate (Lionello et al., 2006). The FI , in the 22 to 77 range and with an average value of 37 , reveals predominantly very low to low climate aggressiveness (García-Barrón et al., 2015). The GC (or compacity coefficient) is between 1.55 and 2.78 , with an average of 2.1 representing elongated basin shapes (Bardossy and Schmidt, 2002). The DL is also very variable, from 355 to 803 km , and is directly related with the basin area. Finally, the RC is between 0.07 and 0.31 , showing different reservoir depths from shallow water bodies to others located in a narrow section of the river course, producing deeper reservoirs.

3.2. Model fitting: selection of explanatory variables and calibration

The regression model obtained from the SSY values of the 26 reservoirs as dependent variables and the set of 11 descriptive variables (including basin area) as independent variables, has a determination factor of $R^2 = 0.80$. Although this value points to the model's high capacity to explain the SSY values, the statistical analysis of the estimators of fit

Table 1
Analysis data base, including SY and SSY, together with the basin explanatory variables.

Reservoir	Hydrographical confederation	Commissioning year	Initial capacity	Basin area	SY	SSY	K (USLE)	R (USLE)	C (USLE)	Slope (S)	P ₀	YAR	Four. Index (FI)	Grav. Coef. (CG)	Drainage length (DL)	Reserv. Coef. (RC)
			(hm ³)	(km ²)	(t year ⁻¹)	(t km ⁻² year ⁻¹)				(%)	(mm)	(mm)			(km)	
Linares de Arroyo	Duero	1953	54.55	767.40	3029.02	3.95	0.3768	86.10	0.078	6.48	19.99	557.27	23.39	1.85	421.81	0.12
Burgomillodo	Duero	1928	13.70	822.14	28,726.81	34.94	0.3179	91.14	0.085	6.78	23.62	581.43	24.21	1.87	429.59	0.10
Águeda	Duero	1931	22.40	812.97	170,938.75	210.26	0.4005	144.83	0.077	8.81	33.95	821.54	57.83	2.17	385.19	0.13
San Juan	Tajo	1955	162.00	874.88	830,741.08	949.55	0.1609	105.82	0.116	13.09	31.31	650.01	41.10	1.89	416.58	0.25
Peñarroya	Guadiana	1959	50.32	955.33	304.82	0.32	0.2807	74.18	0.098	2.59	21.17	498.64	25.36	1.94	476.12	0.12
Torre de Abraham	Guadiana	1974	183.40	762.09	274,331.25	359.97	0.4749	93.74	0.117	8.92	21.73	593.77	40.19	1.88	431.23	0.10
Villar del Rey	Guadiana	1987	133.00	1022.50	224,602.01	219.66	0.3329	83.19	0.065	5.29	24.80	543.17	43.01	2.03	485.86	0.10
Valuengo	Guadiana	1959	19.30	1244.86	62,582.30	50.27	0.3193	107.12	0.087	5.93	14.82	655.95	50.07	1.88	592.06	0.13
Guadalen	Guadalquivir	1954	173.00	1370.00	125,312.72	91.47	0.3289	83.67	0.108	8.72	18.75	545.47	33.98	2.21	667.94	0.13
Bembezar	Guadalquivir	1963	347.38	1713.09	172,235.50	100.54	0.3971	105.42	0.064	10.96	31.68	648.14	55.13	2.05	803.66	0.28
Bornos	Guadalete y Barbate	1961	215.40	1361.43	568,927.31	417.89	0.2512	154.88	0.078	11.96	15.23	863.70	77.88	2.24	636.43	0.09
Puente Nuevo	Guadalquivir	1972	286.76	996.23	285,566.81	286.65	0.3467	93.21	0.100	7.77	27.26	591.26	45.84	2.22	467.95	0.14
Guadalmellato	Guadalquivir	1928	162.00	1205.12	231,102.48	191.77	0.2902	96.86	0.055	9.90	31.57	608.48	49.32	1.87	564.72	0.21
Guadalhorce	Mediterránea Andaluza	1973	134.40	1194.26	244,437.80	204.68	0.2783	85.51	0.091	7.77	22.24	554.43	44.53	2.22	616.21	0.17
Fuensanta	Segura	1933	235.00	975.53	455,853.10	467.29	0.2995	83.00	0.094	20.90	24.57	542.21	29.23	2.59	457.57	0.26
Alfonso XIII	Segura	1915	42.80	854.32	259,640.10	303.91	0.3860	56.23	0.059	7.89	19.77	405.51	26.19	2.78	429.46	0.19
Puentes	Segura	1884	31.56	1001.73	190,956.99	190.63	0.3760	44.63	0.070	10.99	20.11	341.29	26.69	2.47	507.45	0.07
Talave	Segura	1918	47.20	774.40	184,007.89	237.61	0.2543	82.46	0.055	17.63	29.74	539.62	29.05	2.29	355.86	0.16
Camarillas	Segura	1960	35.84	1459.97	16,104.87	11.03	0.3180	51.26	0.071	6.48	25.82	378.45	23.34	2.40	617.78	0.11
Arquillo de San Blas	Júcar	1962	22.00	804.48	39,000.10	48.48	0.2004	93.62	0.065	11.21	33.43	593.20	26.88	2.08	389.16	0.15
Forata	Júcar	1969	39.00	1070.20	61,843.87	57.79	0.3220	71.03	0.072	7.35	19.89	482.72	28.96	2.36	540.12	0.17
María Cristina	Júcar	1920	23.28	1356.87	61,773.75	45.53	0.2668	85.50	0.091	13.73	28.28	554.36	44.84	2.23	665.10	0.09
La Tranquera	Ebro	1959	84.38	1478.82	4267.71	2.89	0.3971	70.90	0.066	7.29	19.00	482.10	22.20	1.98	728.62	0.16
Santolea	Ebro	1932	48.85	1232.33	21,561.86	17.50	0.2852	72.87	0.108	14.47	26.69	492.04	25.35	2.08	599.35	0.13
La Peña	Ebro	1913	25.00	1685.49	175,505.76	104.13	0.2977	176.06	0.136	24.10	22.22	950.38	39.52	2.32	802.10	0.08
Sau	C. I. Cataluña	1964	177.00	1537.74	345,946.07	224.97	0.2910	159.24	0.094	21.22	23.85	881.79	34.50	2.21	722.29	0.31

shows that there are explanatory variables with a non-significant correlation. The estimators of regression fit indicate that R directly ($\beta = 9.299$) and YAR inversely ($\beta = -9.932$) are the main predictors for SSY estimation, although the multicollinearity analysis points to an important correlation between both factors. The test of significance indicates a better correlation of the variables FI (significance = 0.000) and RC (0.001), while GC, K, P₀ and DL, and particularly A, have a limited linear relation with SSY according to the results of the significance tests (significance > 0.05).

The variables with the lowest levels of significance were removed from the analysis following an iterative process and the R variable was also removed due to its linear correlation with YAR (R is more difficult to obtain and its correlation significance is lower). According to this criterion, the explanatory variables were reduced to six: C, S, YAR, FI, DL and RC. The multilinear regression model calculated with these variables has a determination coefficient of $R^2 = 0.69$ and the main statistical indicators show that each variable has a high level of significance in the regression, and the effect of all the variables in the estimation of SSY is well balanced.

Following the methodology described, the variables have been transformed using a Box Cox power transformation where the potential exponents were calculated through the Box and Tidwell methodology (Table 2).

The calibrated regression model is the following:

$$\text{SSY} \left(\text{t year}^{-1} \text{ km}^{-2} \right) = (-2.57 + 447,522.26 C^5 - 29.74 S^{-1} + 7900.97 \text{YAR}^{-1} + 0.00052 \text{FI}^{2.5} - 0.0000011 \text{DL}^{2.5} + 645.60 \text{RC}^3)^2 \quad (5)$$

With a determination coefficient of $R^2 = 0.84$.

The main statistical indicators have also been calculated in this case (Table 3) for each transformed variable and show that all the variables have very high levels of significance, in some cases maximum, and their participation in the SSY estimations is well balanced. The significance value of the F = 16.22 estimator in the ANOVA is 0.000, so the statistical significance of the model is again verified. Durbin Watson coefficient is 2.37 so there is no significance correlation among residues.

According to the statistical assumptions of multiple linear regression, there are three conditions (Cohen et al., 1983) that must be met: multivariate normality, homocedasticity (the variance of error terms is similar across the values of the independent variables), and no multicollinearity. Normality was confirmed with the P–P plot of the standardized regression residuals. Homocedasticity was also confirmed with the graph of standardized predictions vs standardized residual values, which shows a random scatter plot without any distribution pattern. Finally, the no multicollinearity condition was analysed through the analysis of VIF values (below 2.5) (Table 4) and condition index (24.8 in dimension 7) (Table 5). Although no condition index is over 30, the variance analysis detects collinearity between FI and YAR. At

Table 2

Box Cox transformation coefficients for dependant and predictive variables.

Variable	λ
Dependent variable	
SSY	0.5
Predictive variable	
C	5
S	-1
YAR	-1
FI	2.5
DL	2.5
RC	3

Table 3

Model calibration. Unstandardized and standardized coefficients β , t Student and two-tailed significance level of t (confidence interval 95%).

Variable	Unstandardized coefficients		Standardized coefficients		t	Significance
	Coef	Std error	β			
Constant	-2.568	5.888			-0.436	0.668
F(C)	447,522.256	89,388.725	0.626		5.006	0.000
F(S)	-29.738	11.044	-0.279		-2.693	0.014
F(YAR)	7900.973	2391.875	0.479		3.303	0.004
F(FI)	0.00052	0.000	0.786		5.970	0.000
F(DL)	0.0000011	0.000	-0.739		-6.858	0.000
F(RC)	645.598	112.603	0.671		5.733	0.000

Where F(Variable) are the Box Cox transformed variables. F(C) = C^5 . F(Slope) = Slope^{-1} . F(YAR) = YAR^{-1} . F(FI) = $\text{FI}^{2.5}$. F(DL) = $\text{DL}^{2.5}$. F(RC) = RC^3 .

this point, it should be noted that the calculation of FI depends on YAR and maximum monthly rainfall, so correlation should be expected. However, both FI and YAR are considered as important explanatory variables in the regression model: FI defines how it rains and YAR defines how much it rains. The implementation of both in the model can effectively explain the rainfall characteristics in each reservoir basin, which neither can do alone. The bicornelation graphs (Fig. 8) do not show any linear pattern or bicornelations between the explanatory variables.

There is strong correlation between the SSY values calculated with the calibrated model and the observed SSY values (Fig. 9). The Pearson correlation coefficient is high ($r = 0.95$) and significant (r two-tailed significance = 0.00), as expected after the calibration of the model and the high value of R^2 obtained.

3.3. Jackknife analysis

The estimated model (5) was analysed using the jackknife method. When resampled for all 26 reservoirs, both R^2 and ANOVA significance are very similar to the estimated model (jackknife estimators: $R^2 = 0.84$ and the significance value of F = 15.56 estimator in the ANOVA is 0.000). The analysis of the dispersion graph, defined by the pairs of observed SSY and the jackknife-calculated SSY in each reservoir, shows an important scatter around the line 1:1, with a slight trend towards the overestimation of the model. The Pearson correlation between the jackknife model predicted values and observed values is very high ($r = 0.89$) and very significant (r significance test = 0.00). The correlation between the jackknife model predicted values and calibrated model predicted values is even higher ($r = 0.97$) and consequently significant (r significance test = 0.00) (Fig. 10).

The individual analysis of the 26 regressions indicates specific effects in some reservoirs. Thus, the model calibrated excluding San Juan has a $R^2 = 0.78$, a significant decrease in correlation compared to the calibrated model (5), meaning that an important part of the model

Table 4

Collinearity VIF statistics of the calibrated model.

Variable	Collinearity statistics	
	Tolerance	VIF
F(C)	0.550	1.818
F(S)	0.799	1.251
F(YAR)	0.409	2.443
F(FI)	0.496	2.018
F(DL)	0.741	1.349
F(RC)	0.627	1.595

Where F(Variable) are the Box Cox transformed variables. F(C) = C^5 . F(Slope) = Slope^{-1} . F(YAR) = YAR^{-1} . F(FI) = $\text{FI}^{2.5}$. F(DL) = $\text{DL}^{2.5}$. F(RC) = RC^3 .

Table 5
Collinearity condition index statistics of the calibrated model.

Collinearity diagnostics										
Dimension	Eigen value	Condition index	Variance proportions							
			Constant	F(C)	F(S)	F(YAR)	F(FI)	F(DL)	F(RC)	
1	4.852	1.000	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.01
2	0.687	2.657	0.00	0.10	0.02	0.00	0.00	0.00	0.00	0.41
3	0.660	2.711	0.00	0.32	0.05	0.00	0.01	0.01	0.01	0.04
4	0.517	3.064	0.00	0.00	0.03	0.00	0.38	0.00	0.00	0.04
5	0.165	5.429	0.00	0.18	0.13	0.00	0.07	0.75	0.21	0.21
6	0.111	6.599	0.02	0.00	0.70	0.06	0.00	0.22	0.02	0.02
7	0.008	24.825	0.98	0.39	0.06	0.93	0.53	0.00	0.28	0.28

Where F(Variable) are the Box Cox transformed variables. $F(C) = C^5$, $F(\text{Slope}) = \text{Slope}^{-1}$, $F(YAR) = YAR^{-1}$, $F(FI) = FI^{2.5}$, $F(DL) = DL^{2.5}$, $F(RC) = RC^3$.

correlation relies on this reservoir. The opposite cases are the models calculated excluding Villar del Rey and Santolea, with a $R^2 = 0.88$ and 0.87 respectively, so both reservoirs could also be considered outliers for the model estimation, in this case due to their negative effect on the correlation. However, considering that a further reduction in the number of reservoirs would decrease the representativeness of the study and the quality of the statistical analysis, they have all been retained after the jackknife analysis.

3.4. Comparison with other models

The model was compared with other models using the same data set to assess the correlation of the SSY values observed and predicted by the calibrated model and the contrast models.

In the case of the ART model, the correlation between the predicted and observed values of SSY is poor. The Pearson correlation coefficient is

low ($r = 0.11$) and not significant (r-test significance = 0.56). The correlation with the calibrated model results shows an even poorer r coefficient ($r = 0.03$ and r-test significance = 0.88). This model tends to overestimate the lower SSY values (de Vente et al., 2011) because it is adapted to large basins (Syvitski et al., 2005). In this case only the smallest SSY values are overestimated while the rest are largely below the observed values. ART does not therefore offer valid results to compare with the model (5) since it does not fit either the observed SSY values or the calibrated model. The multiple linear regression model defined by de Vente et al. (2011) produces similar results to the previous one. There seems to be no correlation between the observed and calculated SSY, as indicated by the Pearson correlation coefficient ($r = 0.25$) and the significance test (r-test significance = 0.22), although the results are slightly better than with the ART model. Similarly, the correlation between this model and the calibrated model is very low ($r = 0.18$ and r-test significance = 0.37). Thus although this model was calibrated

Bi-Correlations Matrix

Fig. 8. Explanatory variables. Bicorrelation plots.

Fig. 9. Predicted SSY vs Observed SSY scatter plot. Reservoirs with the highest deviations between predictions and observations are highlighted with black circles and names.

from many of the reservoirs that are part of the data set in this study, the conclusions are the same as in the case above.

The FSM model predicts the observed SSY values quite well. This model was applied using the same FSM_{Index} values originally defined in the calibration of the model (de Vente et al., 2005; Verstraeten et al., 2003), since many of the reservoirs in its data set coincide with those of this study (22 of 26). The correlation between the SSY resulting from this model and the observed SSY is strong (Fig. 11). The Pearson coefficient is high ($r = 0.86$) and the correlation is significant (r -test significance = 0.00). The correlation with the SSY values predicted by the calibrated model is also high ($r = 0.83$ and r -test significance = 0.00).

4. Discussion

4.1. Model performance

The statistical fit of the model (5) to estimate SSY in reservoirs with medium-sized basins (750 to 1750 km²) is statistically significant and explains 84% of the observed variability ($R^2 = 0.84$). The model uses

predictive variables based on the characteristics of the catchment areas, which can be estimated or extracted from accessible databases.

This is an empirical model (5), statistically calculated with descriptive information on reservoirs with medium-sized basins and mostly located in the Mediterranean regions of Spain. Its use should therefore be restricted to similar reservoirs and conditions.

The models defined using the data produced by CEDEX from 1967 to 2004 have different levels of statistical fit depending on the reservoirs considered, the predictive variables included in the model and the methodology applied. There are different approaches through regression models, from a 17% prediction of variability (Avendaño Salas et al., 1997) to 58% (de Vente et al., 2011). Other models have obtained better fits, like the Factorial Scoring Model (FSM) at the level of 85% (Verstraeten et al., 2003) and 69% (de Vente et al., 2005), or the PSIAC model at the level of 61% (de Vente et al., 2005). Physically and empirically based models have been calibrated using CEDEX information, like WATEM SEDEM with a determination coefficient of 48% (de Vente et al., 2008) or the ART model with 54% (de Vente et al., 2011).

The scatter plot of the observed values vs predicted values of SSY points (Fig. 9) shows a clear approximation to fit the centred line

Fig. 10. Uncertainty analysis. Jackknife analysis.

Fig. 11. Uncertainty analysis. FSM model testing.

(1:1) that represents the perfect correlation between these values. Nevertheless, there are some reservoirs with an important error in the predicted SSY, both under and overestimated. Underestimated predictions are particularly significant in the reservoirs of San Juan (Tajo watershed), Talave (Segura watershed), Villar del Rey (Guadiana watershed) and Guadalhorce (South Mediterranean watershed). In some cases (San Juan, Talave and Guadalhorce), the difference can be explained by the capacity considered in the calculation of the observed SSY. In these reservoirs CEDEX reports a higher capacity than the correct capacity given by the rivers authorities or the Ministry of Civil Works (162 hm³ instead of 137.8 hm³ in San Juan, 47.2 hm³ instead of 34.8 hm³ in Talave, and 134.4 hm³ instead of 126 hm³ in Guadalhorce), so the observed SSY appears to be overestimated. In other cases (Villar del Rey) there are apparently active gullies, unprotected areas and even quarries close to the shores (Fig. 12) that may be active sources of sediment (Poesen et al., 2003) that are not considered in the model, so it is reasonable to find higher than estimated observed SSY values.

The model overestimates SSY in some reservoirs like Bornos (Guadalete Barbate watershed) and Santolea (Ebro watershed). In both reservoirs there is a discrepancy between the capacity reported by CEDEX (243 hm³ and 48.85 hm³ respectively) and the correct capacity (215 hm³ and 42.58 hm³ respectively), which explains the difference.

The jackknife analysis is applied to detect the potential bias of the model and reports stable behaviour ($R^2 = 0.84$), without the presence of significant outliers that may condition the model calibration.

Fig. 12. PNOA Orthophoto. Villar del Rey Reservoir. Shore zone detail, with gullies and a quarry.

The explanatory variables selected and their transformed forms are all significant for the predictions, with different contributions to the SSY estimation. In terms of magnitude, the standardized coefficients (β) that compare the strength of the effect of each individual independent variable on the dependent variable show that the first is the transformed FI (exponent of transformation $\lambda = 2.5, \beta = 0.786$), followed by DL ($\lambda = 2.5, \beta = 0.739$), RC ($\lambda = 3, \beta = 0.671$), C (C) ($\lambda = 5, \beta = 0.626$), YAR ($\lambda = -1, \beta = 0.479$) and finally S ($\lambda = -1, \beta = -0.279$). The positive or negative way each variable contributes to the model depends on the sign of the coefficient and the exponent of transformation λ . The estimated SSY therefore increases as the FI, RC, C or S increase. Conversely, the estimated SSY increases when DL or YAR decrease.

The results of the sensitivity analysis (Table 6) show that the model is more sensitive to positive variations of C, DL and FI, and negative variations of YAR and DL.

Extreme rainfall has been identified as the climate agent most responsible for triggering soil erosion processes worldwide (Hilton et al., 2008), and causes serious land degradation and environmental deterioration (Wie et al., 2009). This effect is also present in other models, both directly and as part of indices or factors that include other variables. For example, PSIAC (PSIAC, 1968) and SPADS (de Vente et al., 2008) use a factorial index that includes a climate factor referring to rainfall torrentiality; however its relative value in the model is low, since in PSIAC the index can have a value of 130 while the climate factor scores only up to 10; however in SPADS the index is calculated by multiplying six parameters with scores from 1 to 3. In WATEM-SEDEM (Van Oost et al., 2000; Van Rompaey et al., 2001; Verstraeten et al., 2002), rainfall intensity is used directly as the USLE rainfall erosivity factor (USDA SCS, 1965) along with five other variables, and other regression-based equations like the model defined by de Vente et al. (2011), which includes

Table 6
SSY variations (%) respect to predictive variable variations (%).

Predictive variable	Average value	% Variations in predictive variables					
		+1%	+10%	+25%	-1%	-10%	-25%
C	0.08	2.2%	28.2%	108.7%	-2.1%	-17.0%	-30.3%
S	10.70	0.6%	5.7%	12.8%	-0.6%	-6.8%	-19.6%
YAR	590.65	-2.9%	-25.3%	-50.8%	3.0%	35.9%	124.4%
FI	37.25	2.5%	28.2%	87.0%	-2.4%	-21.5%	-44.1%
DL	546.55	-4.3%	-40.8%	-87.1%	4.3%	43.7%	107.3%
RC	0.15	1.5%	17.4%	54.0%	-1.5%	-13.2%	-27.1%

the modified Fournier index (Arnoldus, 1977) in the prediction of SSY and a precipitation index in the prediction of SSY.

The importance of both vegetation cover (García-Ruiz, 2010; Eshghizadeh et al., 2015; Gyssel et al., 2005; Igwe et al., 2017) and slope (Barthes and Roose, 2002; Dadson et al., 2003) in erosion processes have been described frequently. Practically all the models developed in study areas comparable to the ones in the current research consider both factors (de Vente et al., 2013). In the FSM the vegetation cover is the most important variable (de Vente et al., 2005; Verstraeten et al., 2003) in the estimation of SSY.

The model suggests that the higher the yearly average rainfall (YAR), the lower the sedimentation. This effect could be related with the growth of thicker vegetation when annual precipitation increases, thus providing better soil protection (Langbein and Schumm, 1958). Other models such as WATEM-SEDEM, AGNPS (Young et al., 1989) or SWAT (Arnold et al., 1998a, 1998b) also include precipitation in the calculation of SSY, but directly related with erosion and sedimentation.

The same inverse effect on SSY is assigned for the drainage length, which appears to be a logical consequence of the transport of the eroded materials from the basin to the reservoir (Ferro and Minacapilli, 1995). However, variables related with sediment transport are not common in SSY models (de Vente et al., 2013), with some exceptions such as the SSY Index model (Delmas et al., 2009), which uses the drainage density for predictions, and the regression-based model described by Buendía et al. (2016) which also uses stream length.

In the case of Reservoir Coefficient (RC), the model shows that the deeper the reservoir, the higher the SSY. This result may be explained by the relation between the average depth of the reservoirs and the slope of the banks, since a significant part of the sediment load deposited in reservoirs comes from their direct watersheds. Some models such as FSM are based on the sedimentation from the shore area of the reservoir, including the slope gradient, so by considering the reservoir depth and the bank slope to be correlated, RC may be introducing this factor in the model. Another possibility is the accuracy of the observations, where the deeper the reservoir, the more accurate the measure. Only the model defined by Buendía et al. (2016) corrects SSY data in reservoirs located downstream from one or more reservoirs, considering the trapping efficiency of upstream dams, although the estimated equation does not include any morphological characteristic of the reservoirs. Although reservoir depth and sediment deposition may also be related to the reservoir's trapping efficiency, in this study it is assumed that 100% of the sediment remains trapped in the reservoir. Other factors may come into play such as the design of the dam and outlet and the reservoir's operating procedure (Hauer et al., 2018; Merritt et al., 2003), among others. Thus, the regression analysis shows a clear correlation between the observed values of SSY and RC ($\beta = 0.671$, $t = 5.73$ and significance = 0.000).

The catchment area (A) has been used in many models to predict SSY. It is often the determining factor and in many studies is the only explanatory variable needed to define a significant predictive model (Dendy and Bolton, 1976; Lahlou, 1988; Strand, 1975). The relation between SSY and area is usually indirect, so in these models SSY decreases with increasing basin area following a power equation, an effect described in the theory of sediment sources and sinks (Walling, 1983). This relation is not present in the data set in this study. The potential regression, the most usual expression described for this relation (de Vente et al., 2007), was tested with very poor results, with a determination coefficient of $R^2 = 0.01$ and an ANOVA F-test = 0.23, clearly non-significant. The correlation between the catchment area and SSY was also analysed with the other explanatory variables, and the results show no link between them in this data set ($t = -0.77$ and t-significance = 0.94). This could be mainly due to the variability in the climate and geomorphology of the reservoirs considered in the analysis (Verstraeten et al., 2003) and the prior selection of reservoirs according to their basin area, which reduced the possibility of explaining their variability with this factor. Several studies have determined the importance

of gullies and bank erosion in sediment production (Poesen et al., 1996, 2003), and they constitute an essential fraction of sediment yield, especially in mountainous and Mediterranean areas (de Vente et al., 2008; Vanmaercke et al., 2011). It should be noted that unlike other models used to predict SSY (Verstraeten et al., 2003), this model does not include the influence of gullies in the estimations, particularly gullies located in the reservoir shore zone where transport and sedimentation processes previous to the reservoir are minimized, thereby increasing the sediment's capacity to reach the reservoir. As described above, this could be the reason for some underpredictions in the calibrated model, although some studies suggest that models without gully-related variables compensate for this by overpredicting the hillslope erosion (de Vente et al., 2013).

4.2. Model testing and validation

SSY models based on linear regression can have high correlation values and so can apparently accurately predict the production of sediments. However, these models are often unstable and lead to important calculation errors when are extrapolated to other reservoirs. It is therefore important to adequately validate and test the model and not rely solely on correlation estimators, as is common in regression studies (Verstraeten and Poesen, 2001).

The correlation values obtained by applying the abovementioned models are very variable. Both the ART model and the regression model of de Vente et al. (2011) give some results that are very far from the SSY values observed and predicted by the calibrated model. In contrast, the FSM model produces results that are significantly correlated with the observed and predicted SSY.

The poor results of the ART model may be due to the fact that this model was originally calibrated for large basins (Syvitski et al., 2003) and is a global model that does not reflect the local peculiarities of this study. However, it has produced good results for Spanish reservoirs in other studies (de Vente et al., 2011). It also tends to overestimate sedimentation values (Syvitski et al., 2003), which may be related to the fact that it does not consider sediment trapping in upstream reservoirs (de Vente et al., 2011). However, in this validation ART predicts much lower values than those observed. The conclusion is that ART is unable to predict the existing variability in SSY values with the variables in this model, namely basin area, elevation and temperature, which offer little variability in the data set in this study. The regression model proposed by de Vente et al. (2011) offers similar results, with a very poor correlation with the observed values of SSY. In this case, this model was calibrated with a data set similar to the one used in this study, although it was broader and based on different explanatory variables (de Vente et al., 2011), which would explain some lack of correlation with the values of observed SSY. However, the prediction errors are very high in this model, so may be related with the selection of the data set.

Finally, FSM predicted SSY values that are strongly correlated with the values observed and predicted by the calibrated model. In view of the fact that the parametric models seem better calibrated for the prediction of SSY, and particularly the FSM model applied to medium-sized Spanish reservoirs (de Vente et al., 2013), this result appears to support the performance of the calibrated model. Nevertheless, FSM is an expert scoring model that is based on the assessment of the areas closest to the reservoir and includes gullies together with other topography, vegetation cover, lithology and basin shape ratios. In contrast, the model calibrated in this study is based on the analysis of the entire catchment and does not consider the contribution of gullies to the sediment yield, but rather variables for climate, vegetation, drainage network and topography. These two models are conceptually very different, although the results show accurate predictions for both.

Although validation is not a common practice (de Vente et al., 2011), the stability of other models in the literature has been tested using different methods. Some examples are a regression model that reduces its

performance from 58% to 17% after validation in Spain (de Vente et al., 2011), from 96% to 21% in Italy (Grauso et al., 2008), and from 76% to 36% in Belgium (Verstraeten and Poesen, 2001).

As described above, the sample size does not allow the implementation of a cross validation methodology, the most accepted and reliable method of model validation. An alternative validation approach is offered by the jackknife analysis (Bazoffi et al., 1996; Buendía et al., 2016; de Vente et al., 2011; Verstraeten and Poesen, 2001). The model can be considered to be validated by comparing the results obtained when applying the jackknife methodology to the sample with the values of SSY produced by the calibrated model (5).

The comparison of the calibrated model with other previously validated models (de Vente et al., 2011) provides additional information for the validation of the model. In this case, the strong correlation of the calibrated model (5) and the FSM model also supports the validation approach.

4.3. Model uncertainty

Possibly the greatest source of uncertainty in the model is the quality of the data considered in the analysis, in regard to both the observed SSY values and the variables input in the model.

The observed SSY values appear to have a high degree of uncertainty, mainly in reservoirs undergoing several surveys during the CEDEX study period. In some cases the differences between successive bathymetries point to a change in the sedimentation trend that cannot be justified by the available information. The analysis of the evolution of the successive bathymetries in the 26 reservoirs considered in the study shows that SSY varies from a significant dispersion in the first surveys to a converging trend in recent years (Fig. 13). This fact reinforces the decision to take the value for the last bathymetry in reservoirs with more than one data survey as the reference sedimentation.

Another source of data uncertainty is the lack of information on the reservoirs' operating procedures, so there is no knowledge of how dam management can condition sediment evacuation through the outlets (Hauer et al., 2018; Merritt et al., 2003). None of the models that have been mentioned (de Vente et al., 2013), nor the one calibrated in this analysis, takes this effect into account. Depending on the dam design and the availability of water resources, the sediment load in reservoirs can be reduced by flushing through low level-gates, sluicing or by the

use of by-passes. Some simulations have indicated that between 90% and 95% of the total sediment load can be eliminated from an off-stream reservoir using these methods (Morris, 2010), so it is essential to consider them when interpreting the results of bathymetric surveys.

The selection of the explanatory variables used in the model is another potential cause of uncertainty. In the current study, the variables chosen are related to the processes of production (erosion), transport and deposition of sediments, although other explanatory factors could have been selected. For example, de Vente et al. (2011) fitted a regression model based on various lithology and soil descriptors (fluvisol, limestone and acid metamorphic rock rates), topography (relief ratio and slope gradient), precipitation (precipitation concentration index) and land cover (shrub cover rate); and Grauso et al. (2008) calibrated a regression model in Italy using the catchment area and perimeter, and lithology-based erodibility. Verstraeten and Poesen (2001) developed a model using only morphological variables such as the catchment area, topography index and drainage length. Some global models like ART use only the catchment area, altitude and temperature, while the later BQART introduces annual discharge and an index related to lithology and human impact (Syvitski and Milliman, 2007). As described above, there are many models in the literature that predict sedimentation with only the catchment area. With such a wide variability of descriptors it is reasonable to assume that the selection of variables and the definition of a SSY model depend on the local scope of the study and the characteristics of both the catchment and the observations.

The model considers that the variables that characterize the watershed are representative, without any temporal variations. This is true for its morphological and topographic characteristics, drainage network (although there may be actions on its linear continuity that are not known) and even the climate if the time scale is reduced. However, there may be substantial variations in land use and vegetation cover (García-Ruiz, 2010) over short periods of time which, depending on the updating of the data source, may introduce significant errors in the model. The reason for these modifications can be diverse: forest fires (Moreira et al., 2011), changes in land use (Serra et al., 2008), the natural evolution of the vegetation (Herguido Sevillano et al., 2018) or reforestation, among others. The calibrated model is very sensitive to increases in the C factor. Since an increase in C indicates less protection against erosion (USDA SCS, 1965), the model will yield very different results if the degradation of the vegetation cover is not accurately characterized.

Fig. 13. Evolution of observed SSY values in the analysed reservoirs from 1968 to 2001.

Finally, it should be noted again that the model has not been cross validated due to the small size of the data set. Although the jackknife analysis can be used in these cases to provide a validation approach, this method is not as robust and reliable as cross validation.

4.4. Statistical aspects

The analysis of a sample using statistical techniques requires a minimum number of cases so that the results – in this case the regression equation and the associated performance estimators – are representative and the uncertainty is reduced to the confidence interval (Kish, 1965). There is a major controversy surrounding the sample sizes needed for different statistical methodologies (Russell, 2001). In the case of linear regression, the number of observations varies from a minimum of 15 (Schmidt, 1971) to 200 (Green, 1991), and a standard amount of $10 \times$ the number of independent variables is often applied in multilinear regression (Harrell, 2001). The size of the population where the sample is collected is an important factor in assessing the appropriate size. In this study, the initial data set constituted by the CEDEX bathymetric information includes 131 reservoirs that represent almost 37% of the total capacity of large Spanish reservoirs, and 50% of the capacity in the Spanish Mediterranean basin, although its representativeness decreases if the entire Mediterranean region is considered. If only the 26 selected reservoirs are considered, the sum of capacities represents almost 6% of Spain's total capacity and 9.5% of the capacity of the Spanish Mediterranean basin. Consequently, the size of the data set analysed in this study can be considered large enough in terms of the analysed population, but insufficient considering the requirements of the multilinear regression.

The models described in the literature show a wide variability in the size of the observation sample. For example, models calibrated in Spain with the CEDEX data are based on 61 (de Vente et al., 2008, 2011) or 60 (de Vente et al., 2005) reservoirs. In Italy, Grauso et al. (2008) used only 16, and 26 were used in Belgium (Verstraeten and Poesen, 2001). Some global models are based on an extensive data set, such as the ART model which is based on 327 observations (Syvitski et al., 2003) and the BQART model on 488 observations (Syvitski and Milliman, 2007).

The model calibration applied transformation methods to both the dependent and independent variables to allow the analysis of non-linear relationships between them using linear regression methodology (Box and Cox, 1964). The models defined by linear regression (de Vente et al., 2011; Grauso et al., 2008; Verstraeten and Poesen, 2001) are based on the existence of a linear relationship between the observed SSY values (in some cases a logarithmic transformation of SSY is used) and the explanatory variables, but when the relationship between independent variables and the dependent SSY variable is non-linear, the results of the regression analysis will underestimate the true relationship (Osborne and Waters, 2002). In this study, this restriction is eliminated by transforming the variables, which serves many other functions in quantitative data analysis, such as improving the normality of a distribution and equalizing variance to meet assumptions and improve effect sizes, thus constituting important aspects of data cleaning and preparing for statistical analyses (Osborne, 2010).

5. Conclusions

This study analyses the complete sedimentology and bathymetry database compiled and calculated by the CEDEX in the period 1967–2004 in 272 reports on 131 Spanish reservoirs, focusing on a group of 26 reservoirs with a basin area in the range of 750 to 1750 km², located in the area of Mediterranean influence in Spain. The basin area and reservoirs have been characterized for this group through 11 explanatory variables in order to define a regression model to estimate the Specific Sediment Yield (SSY).

After a three-step process, only variables C, Slope, Yearly Average Rainfall, Fournier Factor (Torrentiality Factor), Drainage Length and

Reservoir Coefficient were included in the final regression model, all with a very high level of correlation with SSY. The model calibration includes the transformation of the explanatory variables using a potential expression so both the linear and non-linear relation between SSY and the explanatory variables have been investigated.

The estimated regression model (5) has a determination coefficient of $R^2 = 0.84$ (able to explain 84% of the observation) with a very high level of significance of the ANOVA F-test = 0.000, and is consistent with the usual assumptions of multiple linear regression models. The graph analysis of the values of SSY estimated with the model shows a fairly close fit to the observed values. The model has high sensitivity to positive variations in C, Drainage Length and FI, and to negative variations in YAR and Drainage Length. It was resampled with a jackknife analysis ($R^2 = 0.84$) and tested with three SSY models: ART, FSM and the regression model described by de Vente et al. (2011). ART and the regression model show no correlation with the observed and predicted values of SSY, so no performance information can be derived from them. FSM shows a close correlation with both the observed and predicted SSY, so this model reinforces the performance of the model calibrated in this study. However, many causes for uncertainty and instability have been described, such as data quality (bathymetric surveys) and availability (reservoir operating procedures), and subjectivity in the selection and characterization of variables. Notwithstanding these limitations, the model offers valuable information to explain the processes involved in the estimation of SSY in reservoirs with medium-sized basins without bathymetric information or to analyse the effect of modifications in the conditions of the reservoir basins on sedimentation. Although the model has not been validated using the cross validation methodology, the main and most accepted method to validate regression equations, the jackknife methodology and the correlation with the FSM model may support the model validation. In any case, this model instability requires local calibration before its application to other reservoirs. The model also shows interesting relations between SSY and certain basin characteristics, while others do not appear to be related. For example, there is a positive relation with rainfall torrentiality, land cover and slope gradient, a negative relation with yearly rainfall and drainage length, and a lack of relation with soil erosivity (characterized with the USKE K factor) and the catchment shape and area. The negative relation with yearly rainfall is likely due to the higher soil protection afforded by the vegetation growth, and drainage length is related with the transport distance the eroded materials have to travel before reaching a reservoir. The model also reveals a strong and positive correlation between sedimentation and reservoir depth ratio, which has been explained by the possible relationship between the reservoir depth and bank topography. In contrast, there appears to be no relation between SSY and the catchment area, a factor that is broadly related with sedimentation in the literature.

This study has allowed the definition of a sedimentation model for reservoirs that can be used, after local calibration, as an indicator for planning actions in the basin or for estimating available resources. The correlations detected by the regression analysis confirm the main factors that characterize the processes of production, transport and deposition of sediments. The transformation of variables in the calibration allows the investigation of the non-linear relationships between the variables. In any case, the uncertainty of the model seems to be mainly associated with the data used, so its extrapolation to other reservoirs should be based on a comprehensive and verified data set.

Acknowledgement

This research has been possible thanks to the collaboration of CEDEX (Centro de Estudios y Experimentación de Obras Públicas. Ministerio de Fomento y Ministerio de Transición Ecológica. Spain) and the disinterested sharing of a large amount of data and reports produced by its excellent professionals. We also thank our colleagues from the Escuela Técnica Superior de Ingeniería de Montes, Forestal y del Medio Natural

de Madrid, Universidad Politécnica de Madrid (Spain) who provided insight and expertise that greatly assisted the research.

References

- Arnold, J.G., Srinivasan, R., Muttiah, R.S., 1998a. Large area hydrologic modeling and assessment. Part I: model development. *J. Am. Water Resour. Assoc.* 34, 73–89.
- Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998b. Large area hydrologic modelling and assessment part I: model development. *J. Am. Water Res. Assoc.* 34 (1), 73–89.
- Arnoldus, H.M.J., 1977. Methodology Used to Determine the Maximum Potential Average Annual Soil Loss Due to Sheet and Rill Erosion in Morocco (FAO Soils Bulletin).
- Avendaño Salas, C., Sanz Montero, E., Cobo Rayán, R., Gómez Montaña, J.L., 1997. Sediment yield at Spanish reservoirs and its relationship with the drainage basin area. *Commission Internationale des Grands Barrages. Dix-neuvième Congrès des Grands Barrages*, Florence, pp. 863–874.
- Bardossy, A., Schmidt, F., 2002. GIS approach to scale issues of perimeter-based shape indices for drainage basins. *Hydrol. Sci. J.* 47 (6), 931–942.
- Barthes, B., Roose, E., 2002. Aggregate stability as an indicator of soil susceptibility to runoff and erosion; validation at several levels. *CATENA* 47, 133–149.
- Batalla, R.J., Vericat, D., 2010. A review of sediment quantity issues: examples from the river Ebro and adjacent basins (northeastern Spain). *Integr. Environ. Assess. Manag.* 7, 256–268.
- Battin, T.J., et al., 2009. The boundless carbon cycle. *Nat. Geosci.* 2 (9), 598–600.
- Bazoffi, P., Baldassarre, G., Vacca, S., 1996. Validation of PISA2 model for automatic assessment of reservoir sedimentation. In: Albertson, M. (Ed.), *Proceedings of the International Conference on Reservoir Sedimentation*. Colorado State University, Fort Collins, Colorado, USA, pp. 519–528.
- Beasley, D.B., Huggins, L.F., Monke, E.J., 1980. ANSWERS: a model for watershed planning. *Trans. ASAE* 23, 938–944.
- Bellin, N., Vanacker, V., van Wesemael, B., Sole-Benet, A., Bakker, M.M., 2011. Natural and anthropogenic controls on soil erosion in the Internal Betic Cordillera (southeast Spain). *Catena* 87 (2), 190–200.
- Bentler, P.M., Bonett, D.G., 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88 (3), 588.
- Boehrlert, B., Solomon, S., Strzepek, K.M., 2015. Water under a changing and uncertain climate: lessons from climate model ensembles. *J. Clim.* 28, 9561–9582.
- Bonan, G.B., 2008. *Ecological Climatology: Concepts and Applications*. Cambridge University Press, Cambridge.
- Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. *J. R. Stat. Soc. Ser. B* 26, 211–252.
- Box, G.E.P., Tidwell, P.W., 1972. Transformation of the independent variables. *Technometrics* 4, 531–550.
- Brown, C.B., 1943. Discussion of sedimentation in reservoirs, by J. Witzig. *Transactions of the American Society of Civil Engineers* 69, 1493–1500.
- Bruinsma, J., 2009. The resource outlook to 2050: by how much do land, water use and crop yields need to increase by 2050? Expert Meeting on How to Feed the World in 2050. Rome, FAO and ESDD. Available at: <ftp://ftp.fao.org/docrep/fao/012/ak542e/ak542e06.pdf>.
- Buendía, C., Herrero, A., Sabater, S., Batalla, R.J., 2016. An appraisal of the sediment yield in western Mediterranean river basins. *Sci. Total Environ.* 572, 538–553.
- Church, M., Slaymaker, H.O., 1989. Disequilibrium of Holocene sediment yield in glaciated British Columbia. *Nature* 337, 452–454.
- Cobo, R., 2008. Los sedimentos de los embalses españoles. *Ingeniería del agua* 15, 231–241.
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Academic Press, New York, New York.
- Cohen, P., Cohen, P., West, S., Aiken, L., 1983. *Applied Multiple Regression/Correlation (Analysis for the Behavioral Sciences)*.
- Cole, J.J., et al., 2007. Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems* 10 (1), 171–184.
- Cooper, S.D., Lake, P.S., Sabater, S., et al., 2013. The effects of land use changes on streams and rivers in Mediterranean climates. *Hydrobiologia* 719, 383–425.
- Dadson, S.J., Hovius, N., Chen, H.G., Dade, W.B., et al., 2003. Links between erosion, runoff variability and seismicity in the Taiwan orogeny. *Nature* 426, 648–651.
- de Araújo, J.C., Knight, D.W., 2005. A review of the measurement of sediment yield in different scales. *REM: Revista Escola de Minas, Ouro Preto* 58, 257–265.
- de Vente, J., Poesen, J., 2005. Predicting soil erosion and sediment yield at the basin scale: scale issues and semi-quantitative models. *Earth Sci. Rev.* 71, 95–125.
- de Vente, J., Poesen, J., Verstraeten, G., 2005. The application of semi-quantitative methods and reservoir sedimentation rates for the prediction of basin sediment yield in Spain. *J. Hydrol.* 305 (1–4), 63–86.
- de Vente, J., Poesen, J., Bazoffi, P., Van Rompaey, A., Verstraeten, G., 2006. Predicting catchment sediment yield in Mediterranean environments: the importance of sediment sources and connectivity in Italian drainage basins. *Earth Surf. Process. Landf.* 31 (8), 1017–1034.
- de Vente, J., Poesen, J., Arabkhedri, M., Verstraeten, G., 2007. The sediment delivery problem revisited. *Prog. Phys. Geogr.* 31 (2), 155–178.
- de Vente, J., Poesen, J., Verstraeten, G., Van Rompaey, A., Govers, G., 2008. Spatially distributed modelling of soil erosion and sediment yield at regional scales in Spain. *Glob. Planet. Chang.* 60, 393–415.
- de Vente, J., Verduyn, R., Verstraeten, G., Vanmaercke, M., Poesen, J., 2011. Factors controlling sediment yield at the catchment scale in NW Mediterranean geoecosystems. *J. Soils Sediments* 11, 690–707.
- de Vente, J., Poesen, J., Verstraeten, G., Govers, G., Vanmaercke, M., Van Rompaey, A., Arabkhedri, M., Boix-Fayos, C., 2013. Predicting soil erosion and sediment yield at regional scales: where do we stand? *Earth Sci. Rev.* 127, 16–29.
- Dean, W.E., Gorham, E., 1998. Magnitude and significance of carbon burial in lakes, reservoirs, and peatlands. *Geology* 26 (6), 535–538.
- Dedkov, A.P., 2004. The relationship between sediment yield and drainage basin area. In: Golosov, V., Belyaev, V., Walling, D.E. (Eds.), *Sediment Transfer Through the Fluvial System*. vol. 288. IAHS publication, Moscow, pp. 197–204.
- Delmas, M., Cerdan, O., Mouchel, J.-M., Garcin, M., 2009. A method for developing a largescale sediment yield index for European river basins. *J. Soils Sediment.* 9 (6), 613–626.
- Dendy, F.E., Bolton, G.C., 1976. Sediment yield runoff–drainage area relationships in the United States. *J. Soil Water Conserv.* 31, 264–266.
- Duvert, C., Nord, G., Gratiot, N., Navratil, O., Nadal-Romero, E., Mathys, N., Némery, J., Regués, D., García-Ruiz, J.M., Gallart, F., Esteves, M., 2012. Towards prediction of suspended sediment yield from peak discharge in small erodible mountainous catchments (0.45–22km) of France, Mexico and Spain. *J. Hydrol.* 454, 42–55.
- Easton, Z.M., Fukal, D.R., White, E.D., Collick, A.S., Ashagre, B.B., McCartney, M., Awulachew, S.B., Ahmed, A.A., Steenhuis, T.S., 2010. A multi basin SWAT model analysis of runoff and sedimentation in the Blue Nile, Ethiopia. *Hydrology and Earth System Sciences* 14, 827–841.
- Einsele, G., Hinderer, M., 1997. Terrestrial sediment yield and the lifetime of reservoirs, lakes, and larger basins. *Geol. Rundsch.* 86, 288–310.
- Eshghizadeh, M., Talebi, A., Dastorani, M.T., Azimzadeh, H.R., 2015. Effect of natural land covers on runoff and soil loss at the hill-slope scale. *Global Journal of Environmental Science Management* 2 (2), 125–134.
- Ferro, V., Minacapilli, M., 1995. Sediment delivery processes at basin scale. *Hydrol. Sci. J.* 40 (6), 703–717.
- Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., et al., 2005. Global consequences of land use. *Science* 309, 570–574.
- Folly, A., Bronsvelt, M.C., Clavaux, M., 1996. A knowledge-based approach for C-factor mapping in Spain using Landsat TM and GIS. *Int. J. Remote Sens.* 17 (12), 2401–2415.
- Fournier, F., 1960. *Climat et Érosion. La Relation Entre L'érosion du sol par l'eau et les Précipitations Atmosphériques*. Presses Universitaires de France, Paris.
- García Ruiz, J.M., Lana Renault, N., Beguería, S., Valero Garcés, B.L., Lasanta, T., Arnáez, J., López Moreno, J.I., Regués, D., Martí Bono, C., 2004. Temporal and spatial interactions of slope and catchment processes in the central Spanish Pyrenees. In: Golosov, V., Belyaev, V., Walling, D.E. (Eds.), *Sediment Transfer Through the Fluvial System*. vol. 288, pp. 21–28 Moscow: IAHS publication.
- García-Barrón, L., Camarillo, J.M., Morales, J., Sousa, A., 2015. Temporal analysis (1940–2010) of rainfall aggressiveness in the Iberian Peninsula basins. *J. Hydrol.* 525, 747–759.
- García-Ruiz, J.M., 2010. The effects of land uses on soil erosion in Spain: a review. *CATENA* 81, 1–11.
- García-Ruiz, J.M., López-Moreno, J.I., Vicente-Serrano, S.M., Lasanta-Martínez, T., Beguería, S., 2011. Mediterranean water resources in a global change scenario. *Earth Sci. Rev.* 105 (3–4), 121–139.
- Gögüs, M., Yener, A.G., 1997. Estimation of Sediment Yield Rates of Reservoirs in Turkey, Dix-Neuvième Congrès des Grands Barrages. *Commission Internationale Des Grands Barrages*, Florence, pp. 1265–1276.
- González-Hidalgo, J., Brunetti, M., Stepanek, P., De Luis, M., 2011. A new tool for monthly precipitation analysis in Spain: MOPREDAS database (monthly precipitation trends December 1945 – November 2005). *Int. J. Climatol.* 31 (5), 715–731.
- Graham, M.H., 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* 84 (11), 2809–2815.
- Grauso, S., Pagano, A., Fattoruso, G., De Bonis, P., Onori, F., Regina, P., Tebano, C., 2008. Relations between climatic–geomorphological parameters and sediment yield in a Mediterranean semi-arid area (Sicily, southern Italy). *Environ. Geol.* 54 (2), 219–234.
- Gravelius, H., 1914. *Flusskunde*. Goschen Verlagshan dlug Berlin. In: Zavoianu, I. (Ed.), 1985: *Morphometry of Drainage Basins*.
- Green, S.B., 1991. How many subjects does it take to do a regression analysis. *Multivar. Behav. Res.* 26, 499–510.
- Guzzetti, F., Peruccacci, S., Rossi, M., et al., 2007. Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorog. Atmos. Phys.* 98, 239–267.
- Gyssel, G., Poesen, J., Bochet, E., Li, Y., 2005. Impact of plant roots on the resistance of soils to erosion by water: a review. *Phys. Geogr.* (2), 189–217.
- Hair, J.F., Black, B., Babin, B., Anderson, R.E., 2010. *Multivariate Data Analysis* 7th Pearson Prentice Hall. Upper Saddle River, NJ.
- Haregeweyn, N., Poesen, J., Nyssen, J., Verstraeten, G., de Vente, J., Govers, G., Deckers, S., Moeyersons, J., 2005. Specific sediment yield in Tigray-Northern Ethiopia: assessment and semi-quantitative modelling. *Geomorphology* 69, 315–331.
- Harrell, F.E., 2001. *Regression Modelling Strategies*. Springer-Verlag, New York.
- Hauer, C., Wagner, B., Aigner, J., Holzapfel, P., Flödl, P., Liedermann, M., Tritthart, M., Sindelar, C., Pulg, U., Klösch, M., Haimann, M., Donnum, B.O., Stickler, M., Habersack, H., 2018. State of the art, shortcomings and future challenges for a sustainable sediment management in hydropower: a review. *Renew. Sust. Energ. Rev.* 98, 40–55.
- Herguido Sevillano, E., Lavado Contador, J.F., Schnabel, S., Pulido, M., Ibáñez, J., 2018. Using spatial models of temporal tree dynamics to evaluate the implementation of EU afforestation policies in rangelands of SW Spain. *Land Use Policy* 78, 166–175.
- Hilton, R.G., et al., 2008. Tropical-cyclone-driven erosion of the terrestrial biosphere from mountains. *Nat. Geosci.* 1, 759–762.
- ICOLD (International Committee on Large Dams), 1989. *Sedimentation control of reservoirs. Guidelines*. Bulletin 67, 159.
- ICONA, 1988. *Agresividad de la lluvia en España*. In: Ariel, S.A. (Ed.), *Valores del Factor R de la Ecuación Universal de Pérdidas de Suelo*. Servicio de publicaciones del Ministerio de Agricultura, Pesca y Alimentación, Barcelona, Madrid (292 pp).

- IEA-ETSAP and IRENA, 2015. Technology brief E06. Hydropower.
- Igwé, P.U., Ezeukwu, J.C., Edoaka, N.E., Ejie, O.C., Ifi, G.I., 2017. A review of vegetation cover as a natural factor to soil erosion. *International journal of Horticulture, Agriculture and Food science* 1 (4), 21–28.
- Jacinto, P.A., Filippelli, G.M., Tedesco, L.P., Raftis, R., 2012. Carbon storage and greenhouse gases emission from a fluvial reservoir in an agricultural landscape. *Catena* 94, 53–63.
- Jiongxin, X., Yunxia, Y., 2005. Scale effects on specific sediment yield in the Yellow River basin and geomorphological explanations. *J. Hydrol.* 307, 219–232.
- Kirkby, M.J., 1991. Sediment travel distance as an experimental and model variable in particulate movement. *Catena Suppl.* 19, 111–128.
- Kish, L., 1965. *Survey Sampling*. Wiley, New York.
- Knisel, W.G., 1980. CREAMS: A Field-Scale Model for Chemical, Runoff and Erosion From Agricultural Management Systems. USDA Conservation Research Report, Conservation Model Development, Washington, WA, USA.
- Krishnaswamy, J., Richter, D.D., Halpin, P.N., Hofmockel, M.S., 2001. Spatial pattern of suspended sediment yields in a humid tropical watershed in Costa Rica. *Hydrol. Process.* 15, 2237–2257.
- Kundzewicz, Z.W., Stakhiv, E.Z., 2010. Are climate models 'ready for prime time' in water resources management applications, or is more research needed? *Hydrol. Sci. J.* 55, 1085–1089.
- Lahlou, A., 1988. The silting of Moroccan dams. *Sediment Budgets (Proceedings of the Porte Alegre Symposium, December)*. vol. 174. IAHS Publ, Wallingford, UK, pp. 71–77.
- Laignel, B., Dupuis, E., Durand, A., Dupont, J.P., Hauchard, E., Massei, N., 2006. Erosion balance in the watersheds of the western Paris Basin by high-frequency monitoring of the discharge and suspended sediment in surface water. *Compt. Rendus Geosci.* 338, 556–564.
- Lane, L.J., Hernandez, M., Nichols, M., 1997. Processes controlling sediment yield from watersheds as function of spatial scale. *Environ. Model. Softw.* 12, 355–369.
- Langbein, W.B., Schumm, S.A., 1958. Yield of sediment in relation to mean annual precipitation. *Transactions, American Geophysical Union*. 39.
- Lawrence, L.K., Kerry, B.H., 1989. How appropriate are popular sample size formulas? *Am. Stat.* 43 (2), 101–105.
- Lehner, B., Reidy Liermann, C., Carmen, R., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rödel, R., Sindorf, N., Wisser, D., 2011. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front. Ecol. Environ.* 9 (9), 494–502.
- Lionello, P., Malanotte-Rizzoli, P., Boscolo, R., Alpert, P., Artale, V., Li, L., Luterbacher, J., May, W., Trigo, R., Tsimilis, M., Ulbrich, U., Xoplaki, E., 2006. The Mediterranean climate: an overview of the main characteristics and issues. *Developments in Earth and Environmental Sciences* 4, 1–26.
- Liu, L., et al., 2018. Quantifying the potential for reservoirs to secure future surface water yields in the world's largest river basins. *Environ. Res. Lett.* 13.
- López-Moreno, J.I., Vicente-Serrano, S.M., Moran-Tejeda, E., Zabalza, J., Lorenzo-Lacruz, J., García-Ruiz, J.M., 2010. Impact of climate evolution and land use changes on water yield in the Ebro basin. *Hydrology and Earth Systems Science* 7, 2651–2681.
- López-Vicente, M., Lana-Renault, N., García-Ruiz, J.M., Navas, A., 2011. Assessing the potential effect of different land cover management practices on sediment yield from an abandoned farmland catchment in the Spanish Pyrenees. *J. Soils Sediments* 11, 1614–1640.
- MAPAMA, 2017. Ministerio de Agricultura, alimentación y Medio ambiente. *Boletín Hidrológico Semanal*. Julio 24–31.
- McCarty, G.W., Ritchie, J.C., 2002. Impact of soil movement on carbon sequestration in agricultural ecosystems. *Environ. Pollut.* 116 (3), 423–430.
- Merina, N.R., Sashikumar, M., Rizvana, N., Adlin, R., 2016. Sedimentation study in a reservoir using remote sensing technique. *Appl. Ecol. Environ. Res.* 14 (4), 296–304.
- Merritt, W.S., Letcher, R.A., Jakeman, A.J., 2003. A review of erosion and sediment transport models. *Environ. Model. Softw.* 18, 761–799.
- Milliman, J.D., Meade, R.H., 1983. World wide delivery of river sediment to the oceans. *The Journal of Geology* 91, 1–21.
- Milliman, J.D., Syvitski, J.P.M., 1992. Geomorphic/tectonic control of sediment discharge to the ocean: the importance of small mountainous rivers. *The Journal of Geology* 100, 325–344.
- Ministerio de Agricultura, Pesca, Alimentación y Medio ambiente, 2003. Mapa de Umbral de Escorrentía en condiciones de humedad media del suelo para el método racional modificado obtenido a partir de los usos del suelo del CORINE LAND COVER 2000 y de acuerdo con la metodología expuesta en la Publicación "Análisis de nuevas fuentes de datos para la estimación del parámetro número de curva: perfiles de suelos y teledetección". CEDEX 2003, 346 pp.
- Ministerio de Agricultura, Pesca y Alimentación. Instituto Nacional para la Conservación de la Naturaleza. 1987–2001. Mapa de Estados Erosivos de España.
- Ministerio de Medio Ambiente. ICONA, 1982. Paisajes erosivos en el sureste español. Ensayo de metodología para el estudio de su cualificación y cuantificación. MAPA. Proyecto Lucdeme, Madrid.
- Montgomery, D.C., Peck, E.A., Vining, G.G., 2012. *Introduction to Linear Regression Analysis*. vol. 821. John Wiley & Sons.
- Moore, I.D., Burch, G.J., 1986. Physical basis of the length-slope factor in the universal soil loss equation 1. *Soil Sci. Soc. Am. J.* 50, 1294–1298.
- Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., Barbati, A., Corona, P., Vaz, P., Xanthopoulos, G., Mouillot, F., Bilgili, E., 2011. Landscape – wildfire interactions in southern Europe: implications for landscape management. *J. Environ. Manag.* 92 (10), 2389–2402.
- Morgan, R.P.C., Quinton, J.N., Smith, R.E., Govers, G., Poesen, J.W., Auerswald, K., Chisci, G., Torri, D., Styczen, M.E., Folly, A.J., 1998. The European soil erosion model (EUROSEM): a dynamic approach for predicting sediment transport from fields and small catchments. *Earth Surface Processes and Landform* 23, 527–544.
- Morris, G.L., 2010. Offstream reservoirs for sustainable water supply in Puerto Rico. *Am. Water Resource Assn., Summer Specialty Conf.*, 30 Aug.–1 Sept., San Juan, Puerto Rico.
- Morris, G.L., Fan, J., 1998. *Reservoir Sedimentation Handbook*. McGraw-Hill Book Co, New York (848 pp).
- Nearing, M.A., Jetten, V., Baffaut, C., Cerdan, O., Couturier, A., Hernandez, M., Le Bissonnais, Y., Nichols, M.H., Nunes, J.P., Renschler, C.S., Souchère, V., van Oost, K., 2005. Modeling response of soil erosion and runoff to changes in precipitation and cover. *CATENA* 61 (2–3), 131–154.
- Norušis, Marija J., SPSS Inc, 2008. *SPSS Statistics 17.0 Guide to Data Analysis*. Prentice Hall, Upper Saddle River, N. J.
- Osborne, J.W., 2010. Improving your data transformations: applying the Box-Cox transformation. *Pract. Assess. Res. Eval.* 15 (12), 1–9.
- Osborne, J.W., Waters, E., 2002. Four assumptions of multiple regression that researchers should always test. *Pract. Assess. Res. Eval.* 8 (2).
- Parsons, A. J., Stromberg, S. G., L., Greener, M., 1998. Sediment transport competence of rain-impacted interrill overland flow, *Earth Surface Processes and Landforms* 23, pp 365–375.
- Pearson, K., 1895. Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London* 58, 240–242.
- Poesen, J., Vandaele, K., van Wesemael, B., 1996. Contribution of gully erosion to sediment production on cultivated lands and rangelands. In: Walling, D.E., Webb, B.W. (Eds.), *Erosion and Sediment Yield: Global and Regional Perspectives*. IAHS Symposium 236. Exeter, pp. 251–266.
- Poesen, J., Nachtergaele, J., Verstraeten, G., Valentin, C., 2003. Gully erosion and environmental change: importance and research needs. *CATENA* 50 (2–4), 91–133.
- PSIAC, 1968. Pacific Southwest Inter-Agency Committee (Report of the water management subcommittee on factors affecting sediment yield in the Pacific southwest area and selection and evaluation of measures for reduction of erosion and sediment yield).
- Radoane, M., Radoane, N., 2005. Dams, sediment sources and reservoir silting in Romania. *Geomorphology* 71, 112–125.
- Renwick, W.H., Smith, S.V., Bartley, J.D., Buddemeier, R.W., 2005. The role of impoundments in the sediment budget of the conterminous United States. *Geomorphology* 71, 99–111.
- Restrepo, J.D., Kjerfve, B., Hermelin, M., Restrepo, J.C., 2006. Factors controlling sediment yield in a major South American drainage basin: the Magdalena River, Colombia. *J. Hydrol.* 316, 213–232.
- Roldán Soriano, M., Carrero Díez, L., Gómez Sanz, V., 2009. Estimación del factor R de erosividad de la lluvia mensual y anual con datos de precipitación mensual y anual respectivamente. IV Simposio sobre control de la degradación de los suelos y cambio global. Valencia, España. Sept. 8–11, pp 77–78.
- Romkens, M.J.M., Young, R.A., Poessen, J.W.A., El-Swaify, S.A., Bradford, J.M., 1989. The USLE soil erodibility factor. The Revised Universal Soil Loss Equation, Chap. 3. USDA-ARS, Tucson, Arizona.
- Rorke, B.B., 2000. Soil erodibility and processes of water erosion on hillslope. *Geomorphology* 32 (3–4), 385–415.
- Russell, V.L., 2001. Some practical guidelines for effective sample size determination. *Am. Stat.* 55 (3), 187–193.
- Sabater, S., 2008. Alterations of the global water cycle and their effects on river structure, function and services. *Fr. Rev.* 1 (1), 75–88.
- Schiefer, E., Slaymaker, H.O., Klinkenberg, B., 2001. Physiographically controlled allometry of specific sediment yield in the Canadian cordillera: a lake sediment-based approach. *Geogr. Ann.* 83A, 55–65.
- Schmidt, F.L., 1971. The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educ. Psychol. Meas.* 31 (3), 699–714.
- Serra, P., Pons, X., Sauri, D., 2008. Land-cover and land-use change in a Mediterranean landscape: a spatial analysis of driving forces integrating biophysical and human factors. *Appl. Geogr.* 28 (3), 189–209.
- Shao, J., Tu, D., 1995. *The Jackknife and Bootstrap*. Springer-Verlag, Inc., New York.
- Soundharajan, B.S., Adeloje, A.J., Remesan, R., 2016. Evaluating the variability in surface water reservoir planning characteristics during climate change impacts assessment. *J. Hydrol.* 39, 538–625.
- Stallard, R.F., 1998. Terrestrial sedimentation and the carbon cycle: coupling weathering and erosion to carbon burial. *Glob. Biogeochem. Cycles* 12 (2), 231–257.
- Stillwell, A.S., Webber, M.E., 2013. Evaluation of power generation operations in response to changes in surface water reservoir storage. *Environmental Research Letter* 8.
- Strand, R.L., 1975. Bureau of Reclamation procedures for predicting sediment yield. Present and Prospective Technology for Predicting Sediment Yield and Sources. Proceedings of the Sediment Yield Workshop, USDA Sedimentation Laboratory, Oxford Mississippi, November 28–30, 1972, ARS-S-40. USDA-Agricultural Research Service, Washington, DC, pp. 10–15.
- Student, 1908. The probable error of mean. *Biometrika*, vol VI, pp 1–25.
- Sumi, T., Okano, M., Takata, Y., 2004. Reservoir sedimentation management with bypass tunnels in Japan. Proceedings of 9th International Symposium on River Sedimentation, ii, pp 1036. Yichang, China, p. 1043.
- Summerfield, M.A., Hulton, N.J., 1994. Natural controls of fluvial denudation rates in major world drainage basins. *J. Geophys. Res.* 99 (B7), 13871–13883.
- Swami, V.A., Kulkarni, S.S., 2016. Simulation of runoff and sediment yield for a Kaneri watershed using SWAT model. *Journal of Geoscience and Environment Protection* 4, 1–15.
- Syvitski, J.P.M., Milliman, J.D., 2007. Geology, geography, and humans battle for dominance over the delivery of fluvial sediment to the coastal ocean. *J. Geol.* 115 (1), 1–19.
- Syvitski, J.P.M., Peckham, S.D., Hilberman, R., Mulder, T., 2003. Predicting the terrestrial flux of sediment to the global ocean: a planetary perspective. *Sediment. Geol.* 162, 5–24.

- Syvitski, J.P.M., Vörösmarty, C.J., Kettner, A.J., Green, P., 2005. Impact of humans on the flux of terrestrial sediment to the global coastal ocean. *Science* 308, 376–380.
- Tamene, L., Park, S.J., Dikau, R., Vlek, P.L.G., 2006. Analysis of factors determining sediment yield variability in the highlands of northern Ethiopia. *Geomorphology* 76 (1–2), 76–91.
- Témez, J.R., 1987. Cálculo hidrometeorológico de caudales máximos en pequeñas cuencas naturales, MOPU. Dirección General de Carreteras.
- Thiessen, A.J., Alter, J.C., 1911. Precipitation averages for large areas. *Mon. Weather Rev.* 39, 1082–1084.
- USDA SCS, 1965. *USDA Agriculture Handbook* 282.
- USDA SCS, 1985. *National engineering handbook. Section 4- Hydrology*. DC, Washington.
- USGS, 2014. One estimate of global water distribution. *The World's Water*. Geological Survey, U.S.
- Van Oost, K., Govers, G., Desmet, P.J.J., 2000. Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. *Landsc. Ecol.* 15, 577–589.
- Van Rompaey, A.J.J., Verstraeten, G., Van Oost, K., Govers, G., Poesen, J., 2001. Modelling mean annual sediment yield using a distributed approach. *Earth Surf. Process. Landf.* 26 (11), 1221–1236.
- Vanmaercke, M., Poesen, J., Maetens, W., de Vente, J., Verstraeten, G., 2011. Sediment yield as a desertification risk indicator. *Sci. Total Environ.* 409 (9), 1715–1725.
- Vanoni, V.A., 1975. *Sediment Engineering. Reports in Engineering Practice*. vol. 54. A.S.C.E, New York, p. 745.
- Verstraeten, G., Poesen, J., 2001. Factors controlling sediment yield from small intensively cultivated catchments in a temperate humid climate. *Geomorphology* 40, 123–144.
- Verstraeten, G., Van Oost, K., Van Rompaey, A., Poesen, J., Govers, G., 2002. Evaluating an integrated approach to catchment management to reduce soil loss and sediment pollution through modelling. *Soil Use Manag.* 19, 386–394.
- Verstraeten, G., Poesen, J., de Vente, J., Koninckx, X., 2003. Sediment yield variability in Spain: a quantitative and semiquantitative analysis using reservoir sedimentation rates. *Geomorphology* 50 (4), 327–348.
- Vörösmarty, C.J., Green, P., Salisbury, J., Lammers, R., 2000. Global water resources: vulnerability from climate change and population growth. *Science* 289, 284–288.
- Vörösmarty, C.J., Meybeck, M., Fekete, B., Sharma, K., Green, P., Syvitski, J.P.M., 2003. Anthropogenic sediment retention: major global impact from registered river impoundments. *Glob. Planet. Chang.* 39, 169–190.
- Vörösmarty, C.J., Lévêque, C., Revenga, C., 2010. *Millennium Ecosystem Assessment – Ecosystems and Human Well-Being: Current State and Trends*.
- Walling, D.E., 1983. The sediment delivery problem. *J. Hydrol.* 65, 209–237.
- Wang, G., Wu, B., Wang, Z.-Y., 2005. Sedimentation problems and management strategies of Sanmenxia Reservoir, Yellow River, China. *Water Resour. Res.* 41 (9).
- WCD (World Commission of Dams), 2000. *Dams and Development: A New Framework for Decision-Making*. Earthscan Publications Ltd, London (356 pp).
- WCD (World Commission of Dams), 2017. *World register of dams*. <http://www.icold-cigb.org>.
- Weisberg, S., 1980. *Applied Linear Regression*. John Wiley & Sons, Chichester.
- Wicks, J.M., Bathurst, J.C., 1996. SHESED: a physically based, distributed erosion and sediment yield component for the SHE hydrological modelling system. *J. Hydrol.* 175, 213–238.
- Wie, W., Chen, L., Fu, B., 2009. Effects of rainfall change on water erosion processes in terrestrial ecosystems: a review. *Prog. Phys. Geogr.* 33, 307–318.
- Yang, X., Lu, X.X., 2014. Estimate of cumulative sediment trapping by multiple reservoirs in large river basins: an example of the Yangtze River basin. *Geomorphology* 227, 49–59.
- Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P., 1989. AGNPS: a nonpoint-source pollution model for evaluating agricultural watersheds. *J. Soil Water Conserv.* 44 (2), 168–173.