

# Estimation of suspended sediment and dissolved solid load in a Mediterranean semiarid karst stream using log-linear models

Alberto Martínez-Salvador and Carmelo Conesa-García

## ABSTRACT

Many models have been developed to predict the sediment transport in watercourses. This paper attempts to test the effectiveness of log-linear models (LLM) to estimate the suspended (S-LLM), dissolved (D-LLM), and total suspended (T-LLM) load into a Mediterranean semiarid karst stream (the Argos River basin, in southeast Spain). An assessment of the supposed validity of each model and a leave-one-out cross-validation were carried out to determine their degree of statistical robustness. The T-LLM model showed higher prediction accuracy ( $R^2 = 0.98$ ,  $RMSE = 0.15$ , and  $PE = \pm 5.4\text{--}6.6\%$ ) than the D-LLM model ( $R^2 = 0.97$ ,  $RMSE = 0.16$ , and  $PE = \pm 5.5\text{--}6.8\%$ ) or the S-LLM model ( $R^2 = 0.77$ ,  $RMSE = 0.71$ , and  $PE = \pm 101\text{--}493\%$ ). In addition, different model variants, according to two flow patterns (FP1 = base flow and FP2 = rising water level), were developed. The FP2-SLLM model provided a very good fit ( $R^2 = 0.94$ ,  $RMSE = 0.34$ , and  $PE = \pm 25.3\text{--}61.5\%$ ), substantially improving the results of the S-LLM model.

**Key words** | discharge, dissolved solid load, log-linear models, Mediterranean semiarid karst stream, Southeast Spain, suspended sediment load

**Alberto Martínez-Salvador**

Research Group on Erosion and Desertification in Mediterranean Environments, University of Murcia, C/Santo Cristo, s/n, 30001 Murcia, Spain

**Carmelo Conesa-García** (corresponding author)

Department of Physical Geography, University of Murcia, Campus de La Merced, 30001 Murcia, Spain  
E-mail: cconesa@um.es

## INTRODUCTION

For most hydrometric stations located on Mediterranean semiarid streams, there are no long-term sediments data and the available data have been recorded as discharge ( $m^3/s$ ) and sediment load ( $g/s$ ) for short periods or individual events. In such cases, different methods based on direct measurement and statistical analysis have been used, mainly including regression techniques such as the sediment rating curve (SRC) method, generalized linear models (GLMs), non-parametric regression using Random Forests (RF) and Quantile Regression Forests (QRF) (Francke *et al.* 2008), log-linear models (LLM) (Walling 1977; Cohn 1995; Roman *et al.* 2012; Heng & Suetsugi 2015), and dynamic linear models (Ahn *et al.* 2017). In addition, many other alternative approaches have been applied, such as weighted regression (Abraham 1969), different variants of the SRC

concerning seasonal variations and hysteresis related to rising and falling stages (Walling 1977; Horowitz 2003, 2008), mixed-effects linear models (Araujo *et al.* 2012), optimal estimation techniques (Holtschlag 2001), and event-based suspended sediment load (SSL) models (Moliere *et al.* 2004). Also recently, artificial neural networks (ANNs) (Nourani 2009) and wavelet-based ANN (WANN) models (Gorgij *et al.* 2017; Kim *et al.* 2017), neuro-fuzzy techniques (Rajaei *et al.* 2009), such as Fuzzy Differential Evolution (Kisi *et al.* 2009) and Neural Differential Evolution (Kisi 2010), Linear Genetic Programming (Güven & Kisi 2011), decision tree algorithms (Senthil Kumar *et al.* 2012; Bharti *et al.* 2017), and WLSSVM (Wavelet-based Least Square Support Vector Machine) (Nourani & Andalib 2015) have been used to predict daily and monthly SSLs.

LLM are more flexible and interpretable with respect to the estimation of sediment load, and they have two advantages over linear models: (1) predictions are always positive and (2) the residuals are often more homoscedastic. From their graph structure, it is easy to read off the conditional independence relationships; and graph-based algorithms usually provide efficient computational algorithms for parameter estimation and model selection (Gauraha 2017). Often, it is of interest to predict the response variable or to estimate the mean of the response variable at the original scale for a new set of covariate values. In particular, log-normal linear models are widely used in applications in which linear models need to be fitted to logarithmically transformed response variables. Log-normal linear models have been applied to a wide range of studies, from water quality control (Gilliom & Helsel 1986), insurance reserves estimation (Doray 1996), and mining (Marcotte & Groleau 1997), to monitoring of air pollutant concentrations (Holland *et al.* 2000) and sediment discharge estimation (Cohn 1995; Elliott & Anders 2005).

The main purpose of this study was to evaluate the effectiveness of regression LLMs with regard to estimating the SSL, dissolved solid load (DSL), and total suspended solid load (TSL) in a Mediterranean semiarid karst stream. The environmental conditions of the Argos River basin (South-east Spain) and the availability of water and sediment discharge measurements at the entrance of the Argos reservoir, and detailed sedimentological reports for the reservoir, justify the choice of this stream and its watershed as a study area. In many cases, the flow and sediment gauging data of reservoirs are an important source of information for the validation of these models in headwater streams. Specific objectives were: (1) to develop techniques to construct a complete total suspended sediment time series from incomplete datasets of discharge and sediments, in order to use in the calibration and validation of erosion models (e.g., SWAT, WEPP) and (2) to estimate the uncertainty in the prediction of sediment loads, so the model can be extrapolated to other areas with similar characteristics knowing the reliability of the predictions obtained.

In addition, their usefulness has already been sufficiently proven in cases when measured SSL and DSL data are not available (Shen & Zhu 2008), as frequently occurs in river-flow stations and reservoirs in semiarid

Mediterranean areas. Such a modeling approach is appropriate when the objective is to predict the dependent variable (Troutman & Williams 1987), in our case the sediment transport rates based on the measured discharge series. In this work, sediment transport curves have been developed from logarithmically transformed data, with water discharge as the independent variable and either SSL, DSL, or both (suspended and dissolved load (TSL) as the dependent variable. This procedure has already been adopted by Nearing *et al.* (2007) and Polyakov *et al.* (2010), who did not find a significant improvement in the prediction accuracy of multiple regression models, compared to individual regression models that used only the relationships between flow and sediment transport.

In accordance with this approach, we propose here three transport LLMs based on all the available gauging records: (1) suspended sediment load (S-LLM), (2) dissolved solid load (D-LLM), and (3) total suspended load (T-LLM). These were statistically tested for the study stream. In addition, different model variants, according to two flow patterns (base flow and rising water level), were developed. A previous univariate analysis, the logarithmic transformation of the variables (discharge and sediment load), the evaluation of the validity assumptions for each model, and LOOCV (leave-one-out cross-validation) were carried out in all cases. Finally, the estimated values were compared with recorded data and the performance of these models was evaluated using analysis of variance (ANOVA) and statistical estimators of fit, such as RMSE (root-mean-square error), MSE (mean squared error), PE (percentage error), MAPE (mean absolute percentage error) and  $R^2$  (coefficient of determination). Such procedure could be used to provide reliable estimations of sediment loads in watercourses that have complex discharge–sediment relationships, as is the case with semiarid karst streams. Together, the above models would also be an appropriate tool to monitor and quantify the effects of changes in climate and land-use on the sediment yield in this type of watershed.

## STUDY AREA

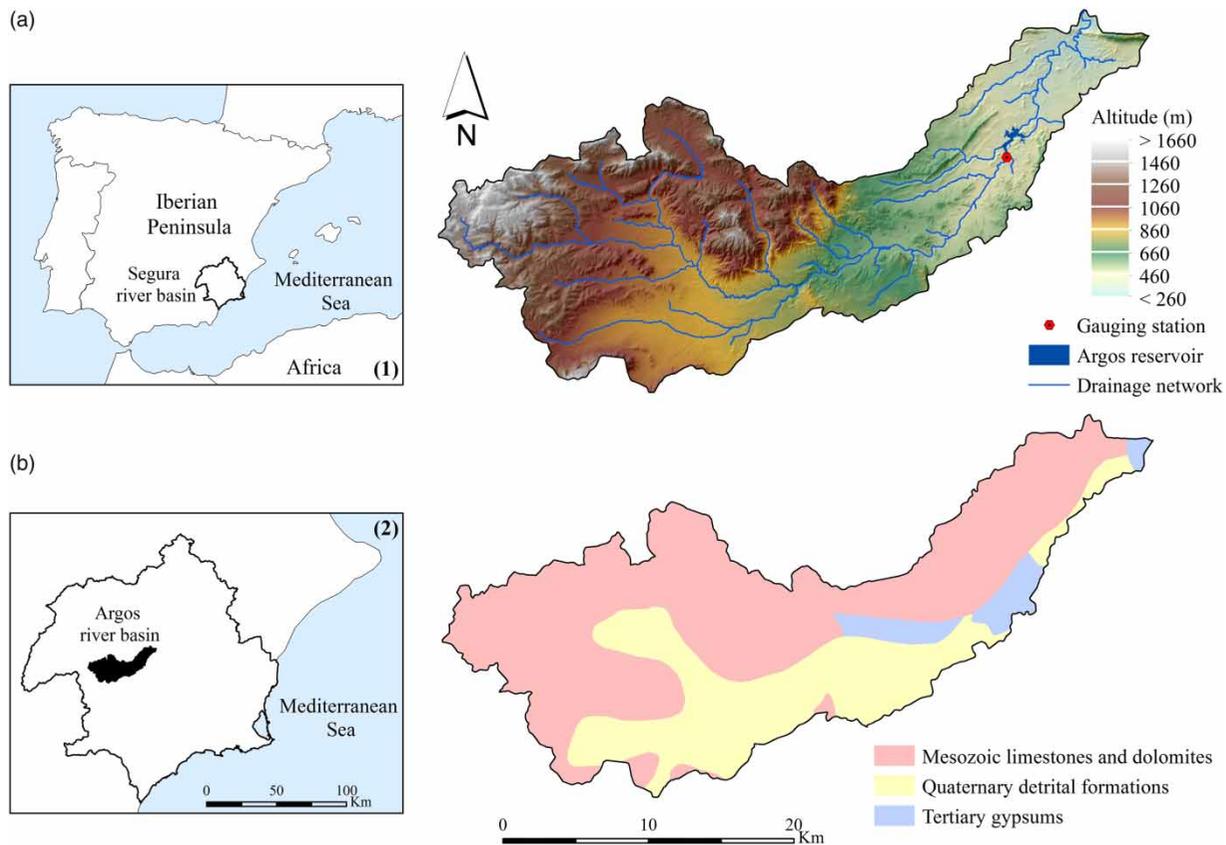
The Argos reservoir watershed (510 km<sup>2</sup>) (Segura River basin, in southeast Spain) was chosen as the study area

(Figure 1). The Argos River arises in the mountain ranges of Sierra del Gavilán and Sierra de Villafuerte, from the confluence of several gullies. As a whole, the relief is composed of a mountainous headwater and lower plains and small hills dissected by ephemeral channels that drain into the Argos River. The relief is characterized by the alternation of mountainous formations arranged in a SW–NE direction, constituted by predominantly limestone-dolomitic materials, on which a karstic relief develops. The average altitude of the basin (925 m) masks a significant contrast between the highlands and lowlands, with maximum and minimum altitude of 1,713 m and 256 m, respectively.

The Argos River basin has semiarid climatic features, locally conditioned by the relief. These include average monthly temperatures between 11 and 16 °C, an absolute maximum above 40 °C in August, and an annual temperature range around 13 °C. The precipitation is scarce

(annual average of 345 mm) and irregular, which corresponds to a typical Mediterranean regime, and is characterized by high seasonal variations. Therefore, a large part of the basin shows a deficit water balance, especially in the central-eastern zone, where the average annual evapotranspiration exceeds 800 mm. Long periods of drought alternate with isolated heavy rainfall events (100-year return period daily rainfall of 125–145 mm).

There is also a total of 11 aquifers, whose recharge is produced mainly by infiltration of rainwater and, to a lesser extent, by lateral contributions and irrigation returns. Due to the aridity, the plant cover is very low, sparse, and xerophytic. Only in the mountain areas does the typical Mediterranean scrub associated with *Pinus halepensis* forests appear, while in the riverside stretches tamarind (*Tamarix canariensis*) and reed (*Phragmites australis*) abound.



**Figure 1** | Digital elevation model (a) and lithological map (b) of the study area. The left-hand maps show (1) the location of the Segura river basin in southeast Spain and (2) the Argos catchment.

## MATERIAL AND METHODS

To estimate different transport modalities (SSL, DSL, and TSL) in the Argos River at the entrance to its reservoir, various log-linear prediction models were developed. First, and in accordance with the suggestions of other researchers (Walling 1977; Ferguson 1987; Cohn 1995; Elliott & Anders 2005; Rovira *et al.* 2005; Amin & Jacobs 2007; Lobera *et al.* 2016), a univariate analysis of the variables discharge ( $\text{m}^3/\text{s}$ ), dissolved solid load ( $\text{g}/\text{s}$ ), suspended sediment load ( $\text{g}/\text{s}$ ), and total suspended load ( $\text{g}/\text{s}$ ) (dissolved and suspended material) was performed.

As part of this analysis, it was tested whether all the variables followed a normal distribution. In the cases where this assumption was violated, a logarithmic transformation was performed (Figure 2). Then, the compliance with the assumption of normality was tested again and three LLM were implemented, as described below.

Finally, the validity assumptions of each linear regression model were evaluated, and, once these were met, a cross-validation process (LOOCV) was carried out. In fact, the model produced to estimate SSL breached an assumption of validity and the Box–Cox procedure had to be applied to achieve compliance.

### Input data

The input data for the statistical models were obtained from the gauging performed to measure river sediment discharge between 1983 and 1994, by the Center for Public Works Studies and Experimentation (CEDEX). In total, 44 records

of the daily SSL, DSL, total suspended load (TSL), and discharge ( $Q$ ) in the Argos River at the tail end of its reservoir were used. In order to group these measurements we considered two main types of flow associated with the occurrence and amount of precipitation: (1) base flow conditions and rainfall in the basin at least one month prior to gauging (flow pattern 1 – FP1) and (2) rising flow during or immediately after significant precipitation events (flow pattern 2 – FP2). In this way, different LLM have been developed under three premises: (1) considering all available registration data (LLM), (2) from the cases included in the pattern FP1 (FP1-LLM), and (3) from gauging data concerning FP2 flows (FP2-LLM). The two flow patterns represent different hydrological behaviors: FP1 can be considered a flow regulation pattern controlled by karst features and groundwater circulation, and FP2 is a flow pattern based on a quick rainfall–runoff response, as is often the case in Mediterranean semiarid streams.

### Univariate analysis

To detect the presence of possible errors in the introduction of the data, as well as atypical or omitted values, a descriptive analysis of the variables was carried out, focusing on the trend, distribution, and dispersion of the data. To determine the types of trend, statistics such as the mean and median were applied, the dispersion was defined according to the standard deviation and the coefficient of variation, and the distribution of the data using the standardized asymmetry bias and standardized kurtosis (Table 1).

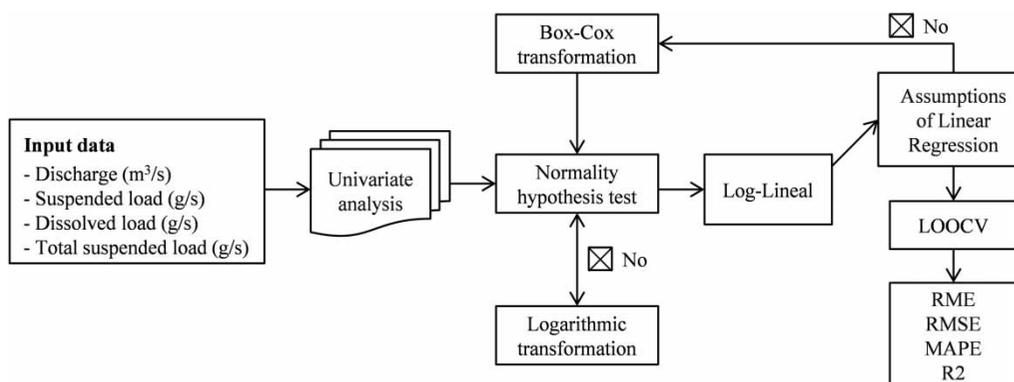


Figure 2 | Methodological scheme for the calculation of the SSL, dissolved solids load (DSL), and total suspended load (TSL).

**Table 1** | Descriptive statistics of the study variables

	Discharge (m <sup>3</sup> /s)	SSL (g/s)	DSL (g/s)	TSL (g/s)
Min	0.0273	0.3337	56.702	57.755
Median	0.212	4.161	336.7	339.4
Mean	0.321	26.362	415.2	441.5
Max	1.611	510.96	1,354	1,865
St. dev.	0.315	79.866	328.8	380.7
CV %	97.61	302.95	78.23	86.23
Standardized bias	5.532	15.206	2.971	4.495
Standardized kurtosis	8.022	46.423	1.209	4.983

The Shapiro–Wilk test was adopted to check whether the data followed a normal distribution. In those cases in which the assumption of normality was not met, the logarithmic transformation of the variables was performed.

### Log-linear model

Because the variables did not conform to a normal distribution, a transformation was carried out for all of them (discharge, dissolved solids, suspended sediments, and total SSL) as the breach of this assumption leads to invalid conclusions. When the relationship between these variables is power, as happened in all cases, they can be linearized using the least-squares function, applying a logarithmic transformation to both the independent and dependent variable, according to the expression:

$$\log Y = \beta_0 + \beta_1 \log X + e \quad (1)$$

where  $\beta_0$  and  $\beta_1$  are the regression coefficients and  $e$  is the perturbation or random error that represents model uncertainty. The least-squares function (Equation (2)) is used to minimize the mean square error in the distances between the observed and predicted values, by the linear regression line:

$$S(b) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 \log X)]^2 \quad (2)$$

where  $S(b)$  is the quadratic function of a Hessian matrix and  $y_i - \hat{y}_i$  is the residual error for the  $i$ -th observation. To

predict the response variable in its original scale an inverse transformation was carried out, which allows the actual values and those predicted by the model to be represented (Equation (3)):

$$Y_o = \text{Exp}(\beta_0 + \beta_1 \log X + e) \quad (3)$$

As the proposed models require an inverse logarithmic transformation, it was necessary to take into account the additive error or bias generated in the linear model, since it becomes multiplicative in each transformation (Ferguson 1987; Cohn & Gilroy 1991; Amin & Jacobs 2007). For this purpose, a correction factor was calculated from the standard estimation error according to Equations (4) and (5) (Sprugel 1983):

$$SEE = \sqrt{\frac{\sum (LnY_i - \overline{LnY_i})^2}{N - K - 1}} \quad (4)$$

$$CF = e^{-\frac{SEE^2}{2}} \quad (5)$$

where  $N$  is the sample size;  $K$ , the number of parameters in the model;  $LnY_i$ , the observed value; and  $Ln\hat{Y}_i$ , the value predicted by the model.

### Evaluation and validation of the log-linear regression model

The goodness of fit for each log-linear model was evaluated using ANOVA and the statistics  $R^2$  (coefficient of determination) (Equation (6)), RMSE (Equation (7)), MSE (Equation (8)), PE (Equation (9)), and MAPE (Equation (10)):

$$R^2 = \frac{(\sum_{i=1}^n (Y_i^{obs} - Y_{obs}^{mean})(Y_i^{sim} - Y_{model}^{mean}))^2}{\sum_{i=1}^n (Y_i^{obs} - Y_{obs}^{mean})^2 \sum_{i=1}^n (Y_i^{sim} - Y_{model}^{mean})^2} \quad (6)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i^{obs} - Y_i^{model})^2 \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{model})^2}{n}} \quad (8)$$

$$PE = \frac{Y_i^{obs} - Y_i^{model}}{Y_i^{obs}} \cdot 100 \tag{9}$$

$$MAPE = \frac{\sum_{i=1}^n \left( \frac{Y_i^{obs} - Y_i^{model}}{Y_i^{obs}} \right) \cdot 100}{n} \tag{10}$$

In addition, in order to guarantee the validity of the formulated models, the following basic assumptions were tested: normality (Shapiro–Wilk test), homoscedasticity (Breusch–Pagan test), no auto-correlation (Durbin–Watson test), and linearity (RESET test) (Quinn & Keough 2002). In accordance with the Gauss–Markov theorem, the fulfillment of these assumptions meant that the least-squares method would result in an optimal fit of the LLM. Since the linearity assumption was not fulfilled in the case of the model proposed for estimation of suspended sediment transport, a Box & Cox (1964) transformation was applied. This procedure determines the best power transformation by finding the value of  $\lambda$  that minimizes the standard deviation of the observations thus transformed (Statgraphics 2006):

$$\text{if } \lambda_1 \neq 0 \quad Y = 1 + \frac{(X + \lambda_2)^{\lambda_1} - 1}{\lambda_1 g^{\lambda_1 - 1}} \tag{11}$$

$$\text{if } \lambda_1 = 0 \quad Y = 1 + g \log (X + \lambda_2) \tag{12}$$

where  $g$  is the geometric mean of the observations after adding  $\lambda_2$ :

$$g = \left( \prod_{i=1}^n (X_i + \lambda_2) \right)^{1/n} \tag{13}$$

The parameter  $\lambda_2$  is set to 0 unless a different value is specified. At the center of the previous transformations is the power to which values are raised,  $\lambda_1$ . Frequently, a power between  $-2$  and  $+2$  will give the data a normal distribution.

Finally, the validation method LOOCV (leave-one-out cross-validation) (Stone 1974) was applied to prevent the error of the test data from being higher than those of the source of the model, thereby avoiding an overestimation of the errors. In this case, it allowed us to evaluate the goodness of fit using a set of independent data, the error of the

cross-validation being computed from the average of the errors produced in all the iterations. The evaluation of the cross-validation is represented by the arithmetic mean of the  $k$  error values obtained in each iteration.

## RESULTS

### LLM to estimate suspended sediment and DSL using all available gauge records

The S-LLM model shows the relationship between SSL and discharge (Figure 3(a)). To generate this model it was necessary to perform a logarithmic transformation of the power function (Box–Cox) in the discharge variable, since the linearity assumption was not fulfilled. The model presents a good fit between the two variables, with a coefficient of determination ( $R^2$ ) of 0.77, situating most of the point cloud within the 95% confidence interval (95% CI). The results indicate that for a 1% increase in the sediment load to occur, the discharge must increase by 6.2%.

A simple linear regression model obtained by logarithmic transformation of the variables discharge and DSL is shown in Figure 3(b). As can be seen, the model fit is optimal, with an  $R^2$  of 0.97, almost all residues within the 95% CI, and a homogeneous dispersion throughout the trend line. This shows that the discharge observations are representative with respect to the dissolved load. The model reveals that each 1% increase in the DSL implies an increase

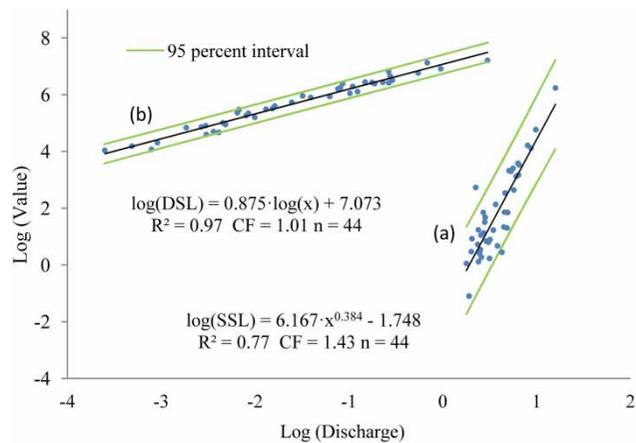


Figure 3 | Regression models for suspended sediment (S-LLM) (a) and dissolved solids (D-LLM) (b).

of 0.87% in the discharge, the standard error of the residuals being around 0.16%.

In the case of the SSL, a relatively higher dispersion is appreciated during medium discharges, which implies relatively greater uncertainty in the prediction of the sediment transport in this type of flow. This could be explained by the fact that the water samples are not very representative during such discharges with respect to the suspended sediment concentration.

There are two points outside the 95% CI, so basic diagnostic graphs were used to check if they are atypical values. In the Residuals vs Leverage graph in Figure 4, it can be seen that observation '38' deviates from the trace and '22' results in a possible dispersion of the model towards that observation. Besides, the graphs Residual vs Fitted and Normal Q-Q show that observation '38' presents an extreme value of variance and dispersion into the normal distribution, and also in the Scale-Locations graph can be seen a moderate standardized residual value. The Bonferroni test and the

Cook distance were applied to analyze if they were atypical and influential values. The Bonferroni test did not consider observation '38' atypical, although its standardized residual is above 3.5 and it is unlikely that such a high value would occur by chance.

Figure 5 shows the analysis of the Cook distances. It can be seen again that the most influential measurements are observations '38' and '22', but also that in no case is there a distance greater than 1. In fact, the highest value is 0.29, which shows that neither of these observations is an atypical case and therefore it is not necessary to eliminate them.

Figure 6 illustrates the relationship between the total SSL (dissolved and suspended load) and discharge, once both variables had been transformed logarithmically and linearized through a least-squares adjustment. The coefficient of determination indicates that the model explains 97.17% of TSL variability. The standard error of the residuals is 0.15%, which means that for each increase of 1% in the TSL an increase of 0.90% in the discharge is necessary.

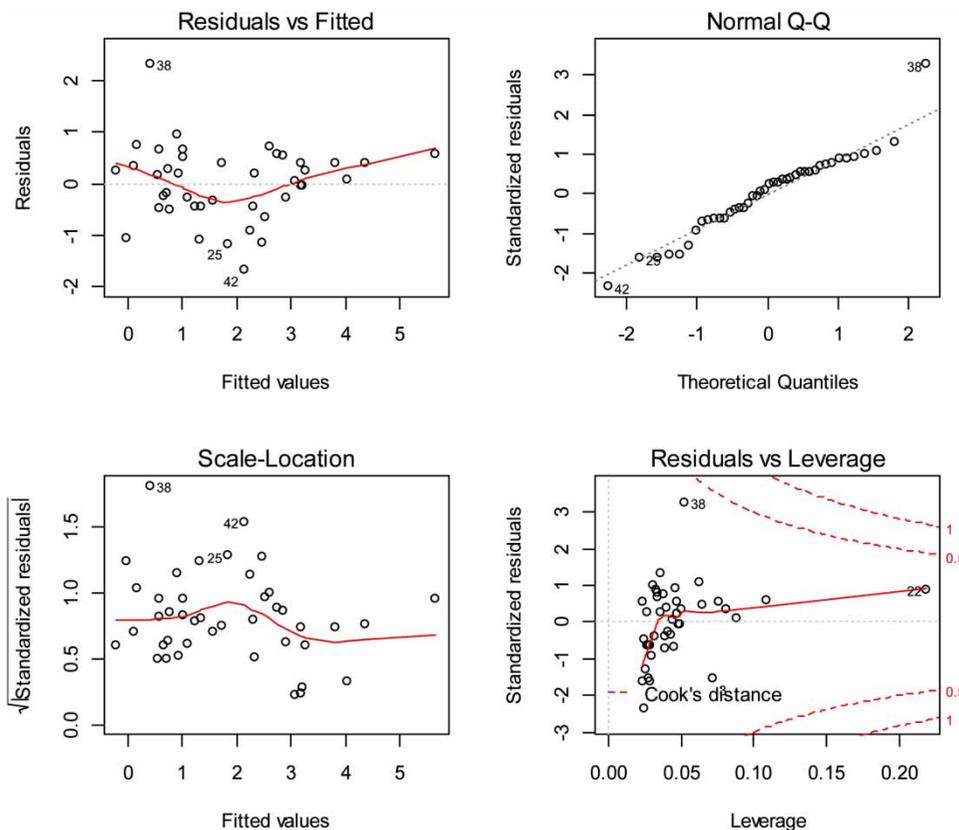


Figure 4 | Graphs of diagnosis applied to the model of SSL.

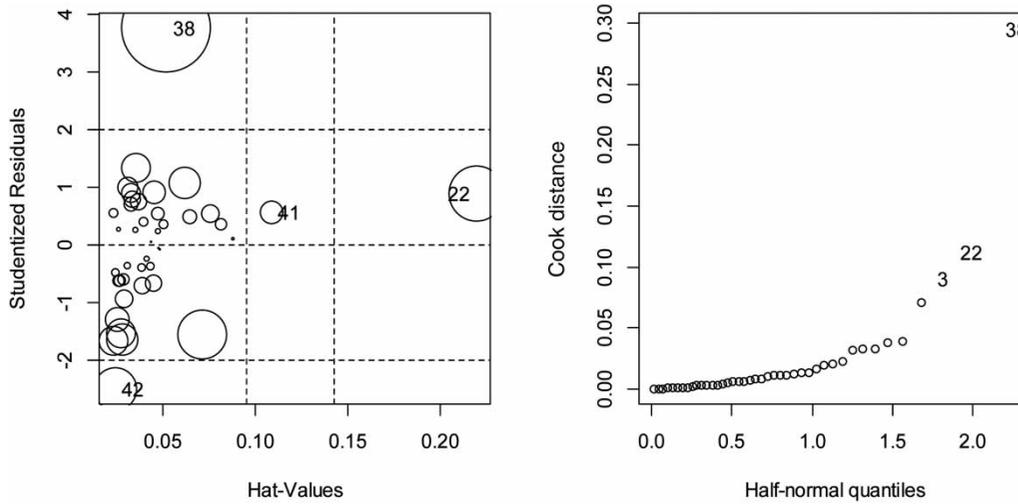


Figure 5 | Analysis of Cook distances applied to the model of SSL.

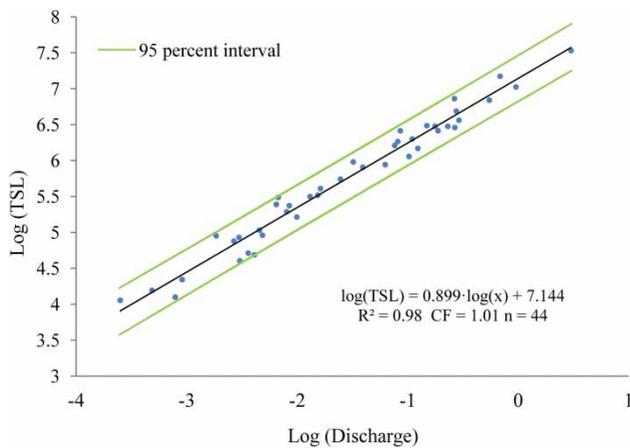


Figure 6 | Regression model for total suspended sediment load (T-LLM).

As can be seen in Figure 6, the point cloud is within the limits of the 95% confidence interval (95% CI). The homogeneous dispersion of the points around the line of linear adjustment shows a low uncertainty in the prediction of TSL for the different discharges.

To ensure that the basic assumptions were met by the linear regression model, the hypothesis tests were applied, the results of which were satisfactory for all models (Table 2).

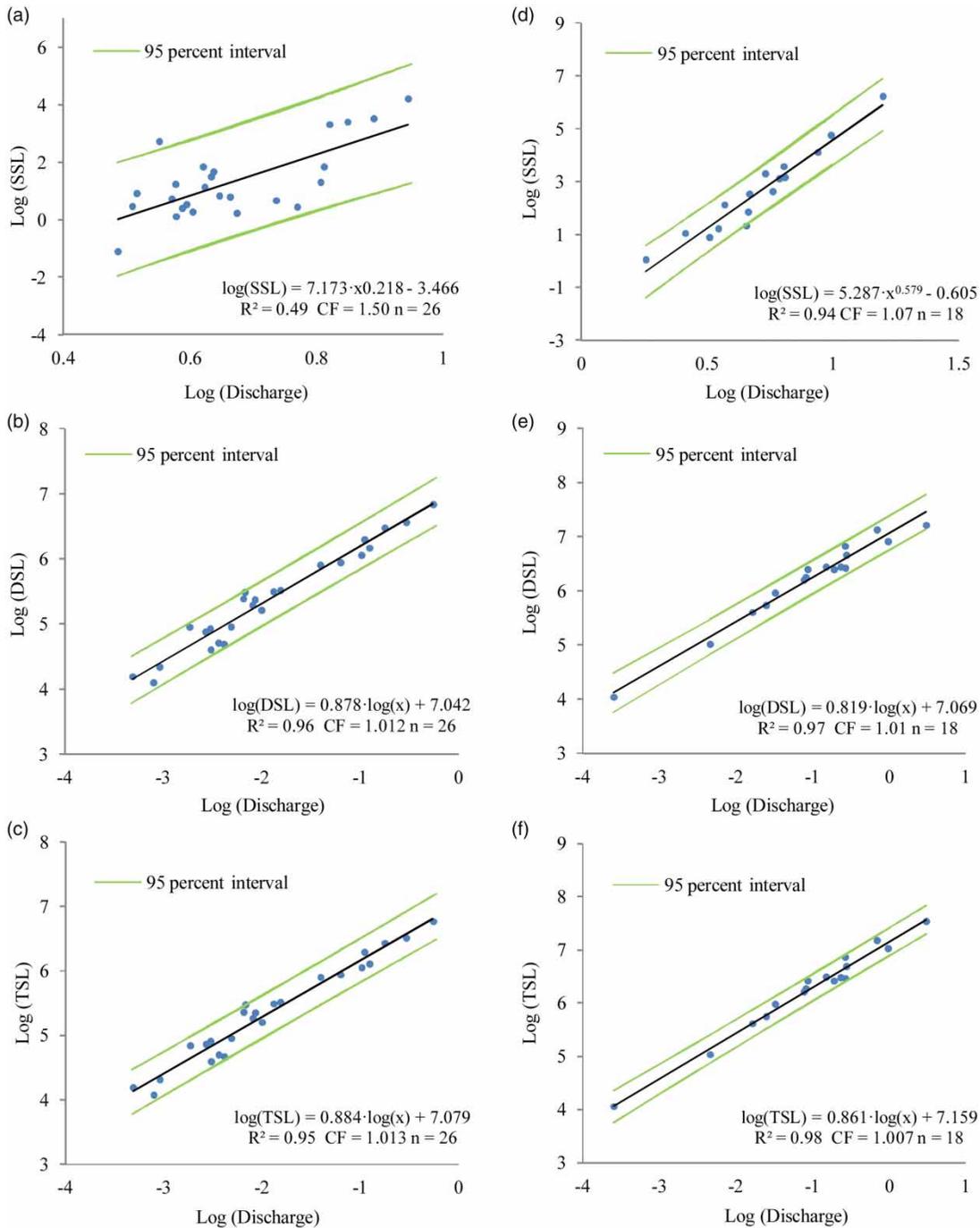
### LLM to predict suspended sediment and DSL from different flow pattern

In addition, in accordance with the same methodological approach described above, the influence of two flow patterns associated with the occurrence of rainfall prior to the discharge and sediment gauging dates was analyzed. The first pattern (FP1) represents the conditions of base flow, fed by waters of rainfall events far from the sampling dates (more than one month earlier). This base flow is directly related to the regulatory function exerted by the karstic terrains on the flow regime. In our case, the FP1-SLLM model (Figure 7(a)) gives a poor fit between SSL and base flow ( $R^2 = 0.49$ ). A wide dispersion was observed

Table 2 | Diagnosis of the LLM regression models using different hypothesis tests

LLM	NLRT		SWT		BPT		DWT	
	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
S-LLM	2.2808	0.1160	0.9523	0.0778	2.3693	0.1237	1.6570	0.1187
D-LLM	3.1287	0.0550	0.9749	0.4758	0.0435	0.8349	1.6177	0.0921
T-LLM	2.2808	0.1160	0.9523	0.0778	2.3693	0.1237	1.6570	0.1187

NLRT, 'Non-linearity' RESET test; SWT, Shapiro-Wilk test; BPT, Breusch-Pagan test; DWT, Durbin-Watson test; P-value, probability value.



**Figure 7** | Regression models for SSL, DSL, and TSL under the flow patterns FP1 (left-hand graphs) and FP2 (right-hand graphs). Models: (a) FP1-SLLM; (b) FP1-DLLM; (c) FP1-TLLM; (d) FP2-SLLM; (e) FP2-DLLM; (f) FP2-TLLM.

in the point cloud as well as a large amplitude of the 95% CI, which shows great uncertainty in the prediction of suspended sediment transport under these conditions (79% of the observations are below the average annual SSL of

the Argos River). The concentration of the errors has a very wide range of variation ( $\pm 101$ – $493\%$ ). This indicates that this type of sedimentary load is extremely sensitive to variations in discharge; therefore, for subcritical flows

with very low turbulence, such as those subjected to a karstic regulation in the Argos River, it is difficult to establish a clear causal relationship between the two variables. In addition, the slope of the fit in this model is a little lower than in S-LLM, each increment of 1% in the SSL requiring an increase of 7.17% in the discharge.

On the other hand, the FP1-DLLM and FP1-TLLM models (Figure 7(b) and 7(c)) exhibit a very good fit ( $R^2 = 0.96$  and  $0.95$ , respectively), although they do not improve on the results of the D-LLM and T-LLM models generated from the totality of the observations ( $R^2 \geq 0.97$ ). The concentration of the errors in these FP1 models ranges between  $\pm 5.9\%$  and  $\pm 6.2\%$ , also demonstrating a high degree of reliability in the prediction of both transport modes for currents regulated by groundwater inputs. This degree of similarity between the two types of models, considering conditions of variable discharge and base flow, is explained by the fact that the total suspended load is mainly composed of material in solution (96%).

The second flow pattern (FP2) represents flood periods of the river, produced during or immediately after rain events. The suspension sediment transport model generated from this pattern (FP2-SLLM) (Figure 7(d)) provides a very good fit ( $R^2 = 0.94$ ), substantially improving on the results of the S-LLM model ( $R^2 = 0.77$ ). The variability of the errors is within the range  $\pm 25.3$ – $61.5\%$ , which can be considered acceptable according to the criterion of Gray & Simões (2008). This better goodness of fit can be explained by the characteristics of FP2, in which the sediment load measurements were made. These more energetic currents, preceded by storm events or moderate to intense rainfalls,

favor the processes of superficial soil washing and the production of fine sediments.

On the other hand, the grouping of the DSL and TSL measurements under the FP2 pattern did not yield any improvement regarding the complete series of records. In fact, FP2-DLLM and FP2-TLLM (Figure 7(e) and 7(f)) give fits identical to those of their respective LLMs, with  $R^2$  values of 0.97 and 0.98, respectively, and prediction errors between  $\pm 2.9$  and  $\pm 3.45\%$ . This seems to be due to the importance acquired by the dissolution processes in a large part of the basin, regardless of the type of flow (base or rising flow), and, as in the case of FP1, to the high proportion of the dissolved load in the total suspended load.

To ensure compliance with the basic regression assumptions, the quantitative hypothesis tests described above were applied, obtaining a satisfactory result for all the models created, with both current patterns.

### Validation of LLM of sediment load

Table 3 shows the ANOVA results for the S-LLM, D-LLM, and T-LLM models. These verify that the variability of these models and the residuals that explain it are not random – which implies their acceptance and validity, since there is a statistically significant relationship between the variables.

Table 4 shows the values of the RME, RMSE, MAPE, and  $R^2$  statistics obtained in the cross-validation (LOOCV) for each of the models, calculated on the basis of the resampling of the dataset. These data have been used to select the model with the smallest errors and best fit, based on its

**Table 3** | Results of the ANOVA applied to the LLM

Model	Source	Sum of squares	df	Mean Square	F-value	P-value
S-LLM	Regression	73.6681	1	73.6681	139.11	0.000
	Residual	21.1835	40	0.5296		
	Total	94.8516	41	2.3134		
D-LLM	Regression	30.0672	1	30.0672	1,197.99	0.000
	Residual	1.0039	40	0.0251		
	Total	31.0711	41	0.7578		
T-LLM	Regression	31.6950	1	31.6953	1,375.43	0.000
	Residual	0.9218	40	0.0230		
	Total	32.6170	41	0.7955		

**Table 4** | Cross-validation of log-linear sediment load models based on goodness-of-fit statistics

Goodness-of-fit statistics	S-LLM		D-LLM		T-LLM	
	Model	Cross-validation	Model	Cross-validation	Model	Cross-validation
RME	0.504	0.766	0.024	0.027	0.022	0.024
RMSE	0.710	0.875	0.155	0.164	0.148	0.156
MAPE	79.15	125.0	2.318	2.450	2.195	2.310
R <sup>2</sup>	0.771	0.722	0.968	0.956	0.971	0.965

coefficient of determination. In general, the results show a similarity to those generated in the original model, those obtained in the validation process being slightly worse.

The values of RME, RMSE, MAPE, and R<sup>2</sup> in the models D-LLM and T-LLM are very similar; the latter model gives the best goodness of fit and, therefore, very satisfactory results in the prediction of the total suspended load. It is of note that the biggest errors are concentrated in the S-LLM model (e.g., RMSE = 0.71, MAPE = 79.15%), which makes it a less reliable predictive model than the previous ones.

Disaggregation of the dataset according to the FP1 and FP2 patterns (Tables 5 and 6) yielded satisfactory values of RME, RMSE, MAPE, and R<sup>2</sup> for the FP1-DLLM, FP2-DLLM, FP1-TLLM, and FP2-TLLM models, with a slight predictive improvement in the FP2-DLLM and FP2-TLLM models, with respect to D-LLM and T-LLM. It is worth highlighting a significant improvement in the goodness of fit provided by the FP2-SLLM model (RMSE = 0.343, MAPE = 14.41%), which makes it the most appropriate model for the prediction of the SSL when the water levels are high, above the annual average discharge. In contrast, the results of the cross-validation for the FP1-SLLM model are very unsatisfactory and discourage its application in low-regime flows.

## DISCUSSION

In the analysis of the LLM, the total suspended load model (R<sup>2</sup> = 0.98) gave the best results, followed by the dissolved solids load model (R<sup>2</sup> = 0.96). The difference in the degree of uncertainty between the S-LLM and D-LLM models is related to the influence of the two main factors that control the hydrological regime of the Argos River: the semiarid rainfall conditions and the predominance of karstic forms. The fact that the worst results were obtained with the SSL model is probably associated with the effects of the high variability in rainfall, both seasonal and event-scale, typical of the semiarid Mediterranean climate (long periods of drought are punctuated by torrential rain events). On the other hand, the dissolved solids load is particularly stable and less sensitive to these extreme phenomena, due to the hydrological regulation exerted by the karstic relief. The dissolved solids concentration increases linearly according to the magnitude of the eventual rises of the Argos River, which increase the volume of the underground flow, while the incorporation of the groundwater into the channel is produced with certain delay.

The separation of the data into different sediment transport modalities did not improve the sediment load

**Table 5** | Cross-validation of the log-linear sediment transport models for the FP1 flow pattern

Goodness-of-fit statistics	FP1-SLLM		FP1-DLLM		FP1-TLLM	
	Model	Cross-validation	Model	Cross-validation	Model	Cross-validation
RME	0.749	0.810	0.023	0.026	0.024	0.028
RMSE	0.865	0.890	0.150	0.161	0.155	0.166
MAPE	105.83	104.17	2.466	2.660	2.468	2.654
R <sup>2</sup>	0.490	0.485	0.956	0.948	0.954	0.950

**Table 6** | Cross-validation of the log-linear sediment transport models for the FP2 flow pattern

Goodness-of-fit statistics	FP2-SLLM		FP2-DLLM		FP2-TLLM	
	Model	Cross-validation	Model	Cross-validation	Model	Cross-validation
RME	0.118	0.139	0.017	0.023	0.012	0.014
RMSE	0.343	0.373	0.132	0.153	0.109	0.118
MAPE	14.414	15.877	1.756	2.063	1.316	1.445
R <sup>2</sup>	0.942	0.938	0.967	0.958	0.979	0.971

estimations, especially in the case of SSL. In fact, the estimation of SSL and DSL was less accurate than that of TSL, for which an R<sup>2</sup> value of 0.98 and low prediction errors, concentrated between  $\pm 5.5$  and 6.6%, were obtained.

However, the prediction of sediment transport by LLMs may improve or worsen depending on the flow pattern adopted (models FP1-LLM and FP2-LLM). The results show that DSL remains uniform over time, while SSL is more significant in high flows. In fact, the FP2-SLLM model improves on the estimation of S-LLM, with a range of error between  $\pm 25.3$  and 61.5%, and an R<sup>2</sup> of 0.94, which resemble the fit values obtained by [Achite & Ouillon \(2007\)](#) and [Bezak et al. \(2017\)](#) for this mode of transportation.

The uncertainty in the prediction of the sediment load has been widely discussed and, depending on the characteristics of the basin, good predictions with concentrated errors of  $\pm 15$ –20% have been obtained ([Horowitz 2008](#)). [Gray & Simões \(2008\)](#) considered  $\pm 30$ –50% as an acceptable error. However, in some studies the uncertainty was considerable. For example, [Smith & Croke \(2005\)](#) reported that the flow explained only a quarter of the variability in the concentration data. [Walling & Webb \(1996\)](#) also found a poor predictive capacity and suggested that seasonal differences in the relationship between the discharge (Q) and the concentration of the sediment components, the non-simultaneity of the runoff and the concentration peaks during storms, hysteresis, and the effects of exhaustion were the most important causes of inaccuracy.

[Campbell & Bauder \(1940\)](#) found a moderately close relationship between the estimations derived with the SRC method and the values observed in the Red River (Texas, USA), with a variation of the errors of  $\pm 14.8$ –20.0%. Similarly, [Kellerhals et al. \(1974\)](#) carried out a similar analysis

for four rivers located in Alberta and West Saskatchewan (Canada), with an average error of 22%, where underestimation predominated and errors greater than  $-50\%$  appeared. [Achite & Ouillon \(2007\)](#) quantified the fine sediment load of the Wadi River (Wadi Basin, Algeria) from the relationships between the daily discharge and the daily suspended sediment concentration, finding overestimations of 20–25% when extrapolating the suspended sediment discharge by the SRC method. [Harrington & Harrington \(2013\)](#) estimated the SSL for the Bandon and Owenabue rivers (Ireland) with the SRC method, obtaining monthly errors of  $\pm 66$ –76% in the Bandon river against errors of  $\pm 65$ –359% in the Owenabue River, underestimating the loads in the former and overestimating them in the latter.

In contrast, [Lewis \(1996\)](#) found that power functions based on the SRC method produced greater extrapolation errors than those based on LLM. Therefore, he investigated the errors associated with the application of LLM and linear regression models for the estimation of the suspended sediments concentration (SSC), obtaining errors of  $\pm 5.2$ –7.9% and  $\pm 5.8$ –8.3%, respectively.

Recent improvements in computer techniques have allowed the application of non-linear statistical models, based on ANNs, which provide a new alternative to logistic regression. [Olyaie et al. \(2015\)](#) compared the accuracy of three different models based on these techniques – an ANN, an adaptive neuro-fuzzy inference system (ANFIS), and a couplet wavelet and neural network (WANN) – with that of the conventional SRC method, for the estimation of SSL. The results show that the best model was the WANN, with a coefficient of determination of 0.91, while it was 0.65, 0.75, and 0.481 for the ANN, ANFIS, and SRC models, respectively. [Liu et al. \(2013\)](#) applied ANN, WANN, and SRC models in the Kuye River (Loess Plateau,

China), in order to relate the SSC and Q. The WANN model showed higher prediction accuracy ( $R^2 = 0.846$  and RMSE = 29.82) than the SRC model ( $R^2 = 0.537$  and RMSE = 55.40) or the ANN model ( $R^2 = 0.664$  and RMSE = 43.13) and gave satisfactory estimates of non-linear and non-stationary time series.

Neural networks offer a series of advantages, including the requirement of less statistical knowledge, the ability to implicitly detect complex non-linear relationships between dependent and independent variables, and the availability of multiple training algorithms. However, their main disadvantages are their functioning as a 'black box', their propensity for overfitting, and the empirical nature of the model; so, their use is recommended when they cannot be described in a parametric way. In addition, neural networks use non-parametric models, which eliminate the error in the estimation of the parameters and do not allow extrapolation of the results.

The results shown here demonstrate the importance of dissolved solids transport in the Argos River (representing 88–94% of the average total suspended solid load), due to the predominance of limestone materials and the intensity of the Karstification processes in the headwater area. Karstic erosion is one of the main causes of desertification and one of the most serious environmental problems in the semiarid regions of the Mediterranean basin (Nerantzaki *et al.* 2015; Malagò *et al.* 2016). In this context, Estrany *et al.* (2009) found, under environmental conditions similar to those of our study, that the SSC in the Na Borges River (Majorca Island) decreased in base flow conditions (underground supply), registering strong seasonal contrasts related with the effects of dilution in different stages of the interaction between groundwater and surface water. Hutchison (2010) showed, in the James River (Ozarks Basin, Missouri, USA), that DSL was much higher than SSL due to the entry of soluble material derived from the chemical erosion of limestone rocks in the Ozarks Basin.

Soil erosion in karst reliefs is a quite complex phenomenon because soil losses are caused not only by surface runoff (wash load), but also by underground dissolution processes that progress through the fissures and pores of the rocks (Chen *et al.* 2011; Zhang *et al.* 2011; Farines *et al.* 2015). This corroborates the sedimentological model defined by CEDEX (1994) for the Argos reservoir. This model, based on

the analysis of sediment samples obtained from the reservoir, showed a high percentage of carbonates, whose origin was attributed to allogenic and endogenous contributions – associated with the semiarid conditions of the area, the scarce depth of the surface water, and the underlying karst formations, which favor the precipitation of calcium carbonate from the bicarbonates dissolved in the water and the crystallization of the interstitial water trapped in the sediment.

In general, studies in karst areas have been limited mainly to the analysis of three types of processes: geomorphological (Ford & Williams 1989), hydrological (Bögli 1980), and vulnerability of ecosystems (Ehret 2000). When the issue of water erosion in such areas has been addressed, it has been done mainly in relation to surface erosion processes (Bou Kheir *et al.* 2008; Chen *et al.* 2011; Li *et al.* 2011; Yang *et al.* 2011; Feng *et al.* 2016; Huang *et al.* 2016), considering the internal transport of solutes to a lesser degree. This makes it difficult to understand, in an integrated way, all the processes involved in the sediment transport within this type of system. However, some work has been carried out in areas of similar environmental conditions, with a predominance of karst and semiarid forms, in which annual rates of sedimentary contributions were obtained, using LLM very similar to those of the present study (Estrany *et al.* 2009; Tena *et al.* 2011; Gamvroudis *et al.* 2015; Lobera *et al.* 2016). In these courses of irregular hydrological regime with karstic influence, in which it is very complicated to discriminate the effects of the multiple environmental processes that take part, log-linear regression models constitute an adequate option, despite their apparent simplicity. The statistical evaluation of these models provides, in terms of robustness, a generally satisfactory and realistic result in relation to the available records of water and sediment discharge at the entrance to the reservoir. Such a degree of statistical reliability means that the LLMs described here can be used to predict the transport of sediments, both suspended and dissolved, from daily discharge data in other basins with similar environmental characteristics.

## CONCLUSIONS

There is a certain lack of knowledge regarding the sedimentation rates in reservoirs, which, in many cases, leads to the

situation where the real state of the useful capacity is not known. Therefore, the study of sediment transport dynamics – through solid material gauging, statistical analysis, and application of sediment-load models – constitutes a reinforcement to determine reservoir sedimentation, until now mainly dependent on bathymetric reports.

The regression models are free of the limitations that affect physical-base models (high data requirement) and empirical equations (specific situation and localization requirement). Physical-base models are supported by simplified partial differential equations of flow and sediment, whose assumptions of simplification are often unrealistic with regard to determining the erosive effects of rainfall and runoff. The difficulty is even greater when it comes to estimating sediment transport in rivers of irregular regime, such as those of Mediterranean semiarid areas and/or subject to the combined action of multiple environmental factors. Under such circumstances, including the effects of regimes modified by a karst influence, the use of linear regression models to predict the suspended sediment and DSLs can be quite useful. In our case study, the log-linear modeling of sediment transport carried out for the Argos River, in the southeast of Spain, has provided satisfactory results, especially the prediction model for the total suspended load.

This study assumes that the errors in the LLM have a normal distribution, and are independent and homoscedastic. In practice, these assumptions can be violated and require the adoption of approaches similar to those already described in order to achieve normality in the distribution of residual errors and homogeneity in the variance of errors. Also, taking into account the principle of parsimony, the complexity of the transport processes in this type of stream, and the quality of the registered data available, it is more practical and effective to implement linear prediction models instead of non-parametric models. Despite the limitations found, the LLMs generated reached a higher level of statistical significance and improved the prediction by considering the patterns of rainfall and hydrological behavior prior to the dates of discharge measurement and sediment gauging. The D-LLM model hardly improved the estimate based on the conditions of base flow and the rainfall that occurred a month (or more) before the date of gauging, thus reflecting the

influence of a flow regulation pattern in karst terrains. In contrast, S-LLM achieved a better fit for cases in which the sediment load and discharge were measured during or immediately after heavy rain, involving a significant increase in the flow and sediment yield (flow pattern based on a quick rainfall–runoff response).

## ACKNOWLEDGEMENTS

This work has been financed by ERDF/Spanish Ministry of Science, Innovation and Universities – State Research Agency/Project CGL2017-84625-C2-1-R/CCAMICEM); Projects I+D+i, State Program for Research, Development and Innovation focused on the Challenges of Society. In addition, we extend our thanks to the Center for Public Works Studies and Experimentation (CEDEX), Ministry of Development, Spain, for providing the water and sediment discharge records.

## REFERENCES

- Abraham, C. E. 1969 *Suspended Sediment Discharges in Streams*. Technical Paper 19, Hydrologic Engineering Center (HEC), US Army Corps of Engineers, Davis, CA, USA.
- Achite, M. & Ouillon, S. 2007 *Suspended sediment transport in a semiarid watershed, Wadi Abd, Algeria (1973–1995)*. *Journal of Hydrology* **343** (3–4), 187–202.
- Ahn, K. H., Yellen, B. & Steinschneider, S. 2017 *Dynamic linear models to explore time-varying suspended sediment-discharge rating curves*. *Water Resources Research* **53** (6), 4802–4820.
- Amin, I. E. & Jacobs, A. M. 2007 *Accounting for sediment sources and sinks in the regression analysis of the suspended sediment load of streams: the Río Puerco, New Mexico as an example*. *Environmental Geosciences* **14** (1), 1–14.
- Araujo, H. A., Cooper, A. B., Hassan, M. A. & Venditti, J. 2012 *Estimating suspended sediment concentrations in areas with limited hydrological data using a mixed-effects model*. *Hydrological Processes* **26**, 3678–3688.
- Bezjak, N., Rusjan, S., Kramar Fijavž, M., Mikoš, M. & Šraj, M. 2017 *Estimation of suspended sediment loads using copula functions*. *Water* **9**, 628.
- Bharti, B., Pandey, A., Tripathi, S. K. & Kumar, D. 2017 *Modelling of runoff and sediment yield using ANN, LS-SVR, REPTree and M5 models*. *Hydrology Research* **48** (6), 1489–1507.
- Bögli, A. 1980 *Karst Hydrology and Physical Speleology*. Springer, New York.

- Bou Kheir, R., Abdallah, C. & Khawlie, M. 2008 [Assessing soil erosion in Mediterranean karst landscapes of Lebanon using remote sensing and GIS](#). *Engineering Geology* **99** (3–4), 239–254.
- Box, G. E. P. & Cox, D. R. 1964 [An analysis of transformations](#). *Journal of the Royal Statistical Society. Series B (Methodological)* **26** (2), 211–252.
- Campbell, F. B. & Bauder, H. A. 1940 [A rating-curve method for determining silt-discharge of streams](#). *Eos Transactions of the American Geophysical Union* **21** (2), 603–607.
- CEDEX (Centro de Estudios Hidrográficos) 1994 [Reconocimiento sedimentológico de embalses](#). *Embalse de Argos* Dirección General de Obras Hidráulicas, Madrid, 45 pp.
- Chen, H. S., Liu, J. W., Wang, K. L. & Zhang, W. 2011 [Spatial distribution of rock fragments on steep hillslopes in karst region of northwest Guangxi, China](#). *Catena* **84** (1–2), 21–28.
- Cohn, T. A. 1995 [Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers](#). *Reviews of Geophysics* **33** (S1), 1117–1123.
- Cohn, T. A. & Gilroy, E. J. 1991 [Estimating Loads From Periodic Records](#). Technical Memo. US Geological Survey, Branch of Systems Analysis, Reston, VA.
- Doray, L. G. 1996 [UMVUE of the IBNR reserve in a lognormal linear regression model](#). *Insurance: Mathematics and Economics* **18**, 43–57.
- Ehrelfeld, J. G. 2000 [Defining the goals of restoration: honesty is the best policy](#). *Restoration Ecology* **8**, 2–9.
- Elliott, J. G. & Anders, S. P. 2005 [Summary of Sediment Data From the Yampa River and Upper Green River Basins, Colorado and Utah, 1993–2002](#). US Geological Survey Scientific Investigations Report 2004-5242, USGS, Reston, VA, USA, 35 pp.
- Estrany, J., Garcia, C. & Batalla, R. J. 2009 [Suspended sediment transport in a small Mediterranean agricultural catchment](#). *Earth Surface Processes Landforms* **34**, 929–940.
- Farines, B., Calvet, M. & Gunnell, Y. 2015 [The summit erosion surfaces of the inner Betic Cordillera: their value as tools for reconstructing the chronology of topographic growth in southern Spain. Patterns and rates of Cenozoic landscape change in orogenic and post-orogenic settings](#). *Geomorphology* **233**, 92–111.
- Feng, T., Chen, H., Polyakov, V. O., Wang, K., Zhang, X. & Zhang, W. 2016 [Soil erosion rates in two karst peak-cluster depression basins of northwest Guangxi, China: comparison of the RUSLE model with 137cs measurements](#). *Geomorphology* **253**, 217–224.
- Ferguson, R. I. 1987 [Accuracy and precision of methods for estimating river loads](#). *Earth Surface Processes and Landforms* **12** (1), 95–104.
- Ford, D. C. & Williams, P. W. 1989 [Karst Geomorphology and Hydrology](#). Chapman & Hall, London.
- Francke, T., López-Tarazón, J. A. & Schroder, B. 2008 [Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests](#). *Hydrological Processes* **22**, 4892–4904.
- Gamvroudis, C., Nikolaidis, N. P., Tzoraki, O., Papadoulakis, V. & Karalemas, N. 2015 [Water and sediment transport modeling of a large temporary river basin in Greece](#). *Science of the Total Environment* **508**, 354–365.
- Gauraha, N. 2017 [Graphical log-linear models: fundamental concepts and applications](#). *Journal of Modern Applied Statistical Methods* **16** (1), 545–577.
- Gilliom, R. J. & Helsel, D. R. 1986 [Estimation of distributional parameters for censored trace level water quality data 1. Estimation techniques](#). *Water Resources Research* **22**, 135–146.
- Gorgji, A. D., Kisi, O. & Moghaddam, A. A. 2017 [Groundwater budget forecasting, using hybrid wavelet-ANN-GP modelling: a case study of Azarshahr Plain, East Azerbaijan, Iran](#). *Hydrology Research* **48** (2), 455–467.
- Gray, J. & Simões, F. 2008 [Estimating sediment discharge](#). In: *Sedimentation Engineering: Processes, Measurements, Modeling and Practice* (M. H. Gracia, ed.). American Society of Civil Engineers, Reston, VA, USA, pp. 1067–1088.
- Güven, A. & Kisi, Ö. 2011 [Daily pan evaporation modeling using linear genetic programming technique](#). *Irrigation Science* **29**, 135–145.
- Harrington, S. T. & Harrington, J. R. 2013 [An assessment of the suspended sediment rating curve approach for load estimation on the Rivers Bandon and Owenabue, Ireland](#). *Geomorphology* **185**, 27–38.
- Heng, S. & Suetsugi, T. 2015 [Regionalization of sediment rating curve for sediment yield prediction in ungauged catchments](#). *Hydrology Research* **46**, 2–38.
- Holland, D. M., De Oliveira, V., Cox, L. H. & Smith, R. L. 2000 [Estimation of regional trends in sulfur dioxide over the eastern United States](#). *Environmetrics* **11** (4), 373–393.
- Holtschlag, D. J. 2001 [Optimal estimation of suspended-sediment concentrations in streams](#). *Hydrological Processes* **15** (7), 1133–1155.
- Horowitz, A. 2003 [An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations](#). *Hydrological Processes* **17**, 3387–3409.
- Horowitz, A. 2008 [Determining annual suspended sediment and sediment-associated trace element and nutrient fluxes](#). *Science of the Total Environment* **400**, 315–343.
- Huang, W., Ho, H. C., Peng, Y. & Li, L. 2016 [Qualitative risk assessment of soil erosion for karst landforms in Chahe town, Southwest China: a hazard index approach](#). *Catena* **144**, 184–193.
- Hutchison, D. C. 2010 [Mass Transport of Suspended Sediment, Dissolved Solids, Nutrients, and Anions in the James River, SW Missouri](#). Master's thesis, Missouri State University.
- Kellerhals, R., Abrahams, A. D. & Giza von, H. 1974 [Possibilities for Using Suspended Sediment Rating Curves in the Canadian Sediment Survey Programme](#). Report REH/74, University of Alberta, Department of Civil Engineering.
- Kim, S., Kisi, O., Seo, Y., Singh, V. P. & Lee, C. J. 2017 [Assessment of rainfall aggregation and disaggregation using data-driven](#)

- models and wavelet decomposition. *Hydrology Research* **48** (1), 99–116.
- Kisi, Ö. 2010 River suspended sediment concentration modeling using a neural differential evolution approach. *Journal of Hydrology* **389**, 227–235.
- Kisi, O., Haktanir, T., Ardiclioglu, M., Ozturk, O., Yalcin, E. & Uludag, S. 2009 Adaptive neuro-fuzzy computing technique for suspended sediment estimation. *Advances in Engineering Software* **40** (6), 438–444.
- Lewis, J. 1996 Turbidity-controlled suspended sediment sampling for runoff-event load estimation. *Water Resources Research* **32** (7), 2299–2310.
- Li, X. Y., Contreras, S., Solé-Benet, A., Cantón, Y., Domingo, F., Lázaro, R., Lin, H., Wesemael, B. V. & Puigdefábregas, J. 2011 Controls of infiltration–runoff processes in Mediterranean karst rangelands in SE Spain. *Catena* **86** (2), 98–109.
- Liu, Q.-J., Shi, Z.-H., Fang, N.-F., Zhu, H.-D. & Ai, L. 2013 Modeling the daily suspended sediment concentration in a hyperconcentrated river on the Loess Plateau, China, using the Wavelet–ANN approach. *Geomorphology* **186**, 181–190.
- Lobera, G., Batalla, R. J., Vericat, D., López-Tarazón, J. A. & Tena, A. 2016 Sediment transport in two Mediterranean regulated rivers. *Science of the Total Environment* **540**, 101–113.
- Malagò, A., Efstathiou, D., Bouraoui, F., Nikolaidis, N. P., Franchini, M., Bidoglio, G. & Kritsotakis, M. 2016 Regional scale hydrologic modeling of a karst-dominant geomorphology: the case study of the Island of Crete. *Journal of Hydrology* **540**, 64–81.
- Marcotte, D. & Groleau, P. 1997 A simple and robust lognormal estimator. *Mathematical Geology* **29** (8), 993–1009.
- Moliere, D. R., Evans, K. G., Saynor, M. J. & Erskine, W. D. 2004 Estimation of suspended sediment loads in a seasonal stream in the wet-dry tropics, Northern Territory, Australia. *Hydrological Processes* **18**, 531–544.
- Nearing, M. A., Nichols, M. H., Stone, J. J., Renard, K. G. & Simanton, J. R. 2007 Sediment yields from unit source semiarid watersheds at Walnut Gulch. *Water Resources Research* **43**, 1–10.
- Nerantzaki, S. D., Giannakis, G. V., Efstathiou, D., Nikolaidis, N. P., Sibetheros, I. A., Karatzas, G. P. & Zacharias, I. 2015 Modeling suspended sediment transport and assessing the impacts of climate change in a karstic Mediterranean watershed. *Science of the Total Environment* **538**, 288–297.
- Nourani, V. 2009 Using artificial neural networks (ANNs) for sediment load forecasting of Talkherood River mouth. *Journal of Urban and Environmental Engineering* **3**, 1–6.
- Nourani, V. & Andalib, G. 2015 Daily and monthly suspended sediment load predictions using wavelet based artificial intelligence approaches. *Journal of Mountain Science* **12** (1), 85–100.
- Olyaie, E., Banejad, H., Chau, K. W. & Melesse, A. M. 2015 A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States. *Environ. Monit. Assess.* **187** (4), 189.
- Polyakov, V., Nearing, M. A., Nichols, M. H., Scott, R. L., Stone, J. J. & McClaran, M. 2010 Long-term runoff and sediment yields from small semi-arid watersheds in southern Arizona. *Water Resources Research* **46** (9), 1–12.
- Quinn, G. P. & Keough, M. J. 2002 *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, New York.
- Rajae, R., Mirbagheri, S. A., Zounemar-Kermani, M. & Nourani, V. 2009 Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. *Science of the Total Environment* **12**, 4916–4927.
- Roman, D. C., Vogel, R. M. & Schwarz, G. E. 2012 Regional regression models of watershed suspended-sediment discharge for the eastern United States. *Journal of Hydrology* **472–473**, 53–62.
- Rovira, A., Batalla, R. J. & Sala, M. 2005 Response of a river sediment budget after historical gravel mining (the Lower Tordera, NE Spain). In: *River Research and Applications* (P. Wood, ed.). John Wiley & Sons, Chichester, UK.
- Senthil Kumar, A. R., Ojha, C. S. P., Goyal, M. K., Singh, R. D. & Swamee, P. K. 2012 Modeling of suspended sediment concentration at Kasol in India using ANN, Fuzzy Logic, and Decision Tree Algorithms. *Journal of Hydrologic Engineering* **17**, 394–404.
- Shen, H. & Zhu, Z. 2008 Efficient mean estimation in log-normal linear models. *Journal of Statistical Planning and Inference* **138**, 552–567.
- Smith, C. & Croke, B. 2005 Sources of uncertainty in estimating suspended sediment load. *IAHS-AISH Publication* **292**, 136–143.
- Sprugel, D. G. 1983 Correcting for bias in Log-Transformed allometric equations. *Ecology* **64** (1), 209–210.
- Statgraphics 2006 Capítulo 5. Transformaciones de Box-Cox. In: *Manual de Statgraphics*. Available from [www.statgraphics.com](http://www.statgraphics.com). Statpoint Technologies, Inc., USA.
- Stone, M. 1974 Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society Series B* **36** (1), 111–147.
- Tena, A., Batalla, R. J., Vericat, D. & Lopez-Tarazón, J. A. 2011 Suspended sediment dynamics in a large regulated river over a 10-year period (the lower Ebro, NE Iberian Peninsula). *Geomorphology* **125**, 73–84.
- Troutman, B. M. & Williams, G. P. 1987 Fitting straight lines in the earth sciences. In: *International Association for Mathematical Geology Studies in Mathematical Geology – Use and Abuse of Statistical Methods in the Earth Sciences* (W. B. Size, ed.). Oxford University Press, Oxford, pp. 107–128.
- Walling, D. E. 1977 Limitations of the rating curve technique for estimating suspended sediment loads, with particular reference to British rivers. In: *Erosion and Solid Matter Transport in Inland Waters (Proceedings of the Paris*

- Symposium, July 1977*). IAHS Publication, Wallingford, UK, 122, 34–48.
- Walling, D. E. & Webb, B. W. 1996 Erosion and sediment yield: a global overview. *IAHS-AISH Publication* **236**, 3–19.
- Yang, P., Tang, Y. & Zhou, N. 2011 Characteristics of red clay creep in karst caves and loss leakage of soil in the karst rocky desertification area of Puding County, Guizhou, China. *Environmental Earth Sciences* **63**, 543–549.
- Zhang, X., Bai, X. & He, X. 2011 Soil creeping in the weathering crust of carbonate rocks and underground soil losses in the karst mountain areas of Southwest China. *Carbonates Evaporites* **26**, 149–153.

First received 4 May 2018; accepted in revised form 5 July 2018. Available online 23 July 2018