

Article

Partial Least Squares Regression for Determining the Control Factors for Runoff and Suspended Sediment Yield during Rainfall Events

Nufang Fang ^{1,2}, Zhihua Shi ^{3,*}, Fangxin Chen ³ and Yixia Wang ²

¹ State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Northwest A & F University, Yangling 712100, China; E-Mail: fnf2012@126.com

² Institute of Soil and Water Conservation of Chinese Academy of Sciences and Ministry of Water Resources, Yangling 712100, China; E-Mail: wyx19880730@163.com

³ College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China; E-Mail: chenfx_520@163.com

* Author to whom correspondence should be addressed; E-Mail: shizhijhua70@gmail.com; Tel.: +86-27-8728-8249; Fax: +86-27-8767-1035.

Academic Editor: Miklas Scholz

Received: 20 May 2015 / Accepted: 6 July 2015 / Published: 14 July 2015

Abstract: Multivariate statistics are commonly used to identify the factors that control the dynamics of runoff or sediment yields during hydrological processes. However, one issue with the use of conventional statistical methods to address relationships between variables and runoff or sediment yield is multicollinearity. The main objectives of this study were to apply a method for effectively identifying runoff and sediment control factors during hydrological processes and apply that method to a case study. The method combines the clustering approach and partial least squares regression (PLSR) models. The case study was conducted in a mountainous watershed in the Three Gorges Area. A total of 29 flood events in three hydrological years in areas with different land uses were obtained. In total, fourteen related variables were separated from hydrographs using the classical hydrograph separation method. Twenty-nine rainfall events were classified into two rainfall regimes (heavy Rainfall Regime I and moderate Rainfall Regime II) based on rainfall characteristics and K-means clustering. Four separate PLSR models were constructed to identify the main variables that control runoff and sediment yield for the two rainfall regimes. For Rainfall Regime I, the dominant first-order factors affecting the changes in sediment yield in our study were all of the four rainfall-related variables, flood peak discharge, maximum flood

suspended sediment concentration, runoff, and the percentages of forest and farmland. For Rainfall Regime II, antecedent condition-related variables have more effects on both runoff and sediment yield than in Rainfall Regime I. The results suggest that the different control factors of the two rainfall regimes are determined by the rainfall characteristics and thus different runoff mechanisms.

Keywords: hydrological response; suspended sediment; partial least squares regression; small watershed

1. Introduction

Soil erosion affects soil productivity and sustainable agriculture. Erosion by water strips the fertile topsoil on site, degrades water quality, and clogs streams, rivers, and reservoirs by transporting sediments off site [1]. Suspended sediment yields represent the sum of the erosion produced by all active sources within a watershed [2]. Analysis of the relationships between sediment transport, rainfall, and runoff characteristics can facilitate the elucidation of the factors and processes determining sediment responses [3].

Many recent studies have evaluated factors that control hydrological and sediment responses at an inter-event scale. Oeurng *et al.* [4] use a Pearson correlation matrix and factorial analysis to assess the relationships between precipitation, discharge, and suspended sediment transport to explain the hydrological and sedimentological responses in a catchment of France. López-Tarazón *et al.* [5] use multiple regression equations derived from Pearson correlation analysis to describe the relationships between rainfall, runoff, and sediment transport in a mountainous catchment. Wine *et al.* [6] apply stepwise regression to determine control factors of runoff in watersheds of the Southern Great Plains. When analyzing the relationship between rainfall, runoff, and sediment yield, multivariate statistics are commonly used to relate control factors to the dynamics of discharge and sediment yield [2,7–9].

These studies enable soil and water conservationists to understand the complexities of hydrological processes. However, these statistical approaches present particular analytical challenges despite their great potential [10]. Many control factors are highly correlated, which can result in redundancy. Thus, the application of these statistical approaches is somewhat limited if inappropriate approaches or unrepresentative variables are selected. Canonical correlation requires that the ratio of the number of predictors to the sample size be at least 0.025–0.05 [10]. Ordinary regression is hindered by limitations imposed by sample size (the number of observations). Classical multiple regression requires a large sample size relative to the number of predictors [11]. The limitations of traditional multivariate regression approaches in handling multi-collinear and noisy data can be overcome by applying techniques based on multivariate statistical projection [12]. Therefore, the influence of changes in each of the hydrologic variables on runoff and sediment yield must be investigated to enable more effective and accurate watershed management and prediction of the hydrological consequences of rainfall.

Knowledge of sediment yield from small watersheds is critical to understand the linkage between soil erosion processes on hill slopes and suspended sediment transport in large rivers [13]. In this study, data on sediment yield for a small watershed in the Three Gorges Area (TGA) were collected. The Three

Gorges Project (TGP) on the Yangtze River in China is the world's largest hydropower complex project. Following the construction of the Three Gorges Dam (1994–2009), millions of farmers resettled in the surrounding mountain areas and cultivated marginal lands, which are largely on steep slopes with soil of poor structure. The TGA, which refers to the riparian parts along the Yangtze valley between Yichang and Chongqing (Figure 1), contains a substantial amount of arable land on steep slopes known to be susceptible to soil erosion. Sediment yield in this area is estimated to be approximately $700 \text{ t}\cdot\text{km}^{-2}\cdot\text{year}^{-1}$ [14]. Soil erosion is a serious issue in this area because of long-term anthropogenic pressure, including over use and inappropriate development [15].

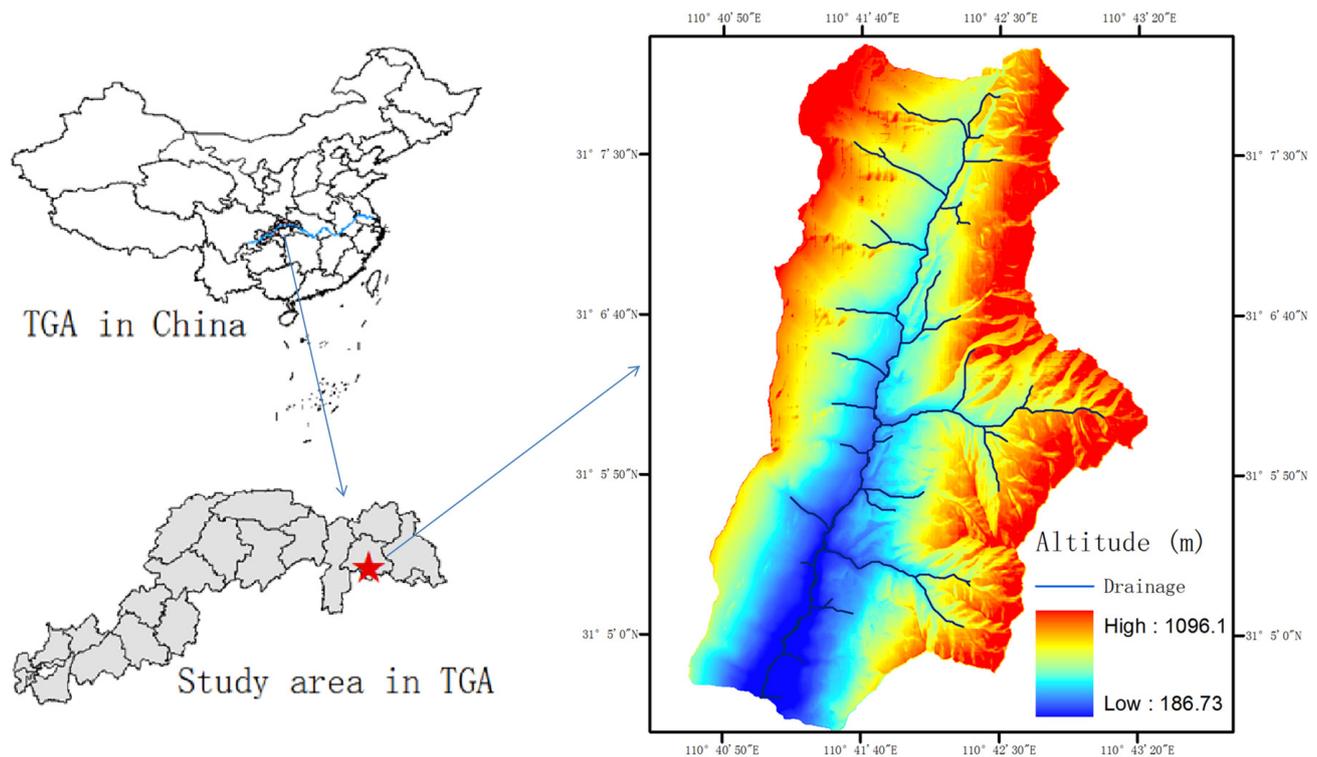


Figure 1. Location of the study watershed in the TGA of China.

The objectives of this paper are: (1) to quantify the contribution of control variables to runoff and sediment yield using partial least squares regression (PLSR); and (2) to investigate the effects of land use change on runoff and suspended sediment at an inter-event scale on a mountainous watershed in the TGA.

2. Study Area and Methods

2.1. Study Area

Integrated small watershed management (ISWM) for soil conservation has developed rapidly in the TGA since the 1990s. ISWM has been conducted in more than 5000 small watersheds with an area of $96,000 \text{ km}^2$, and the central government has invested 15.2 billion RMB (about 2.5 billion USD) in this project [16]. As a part of the ISWM program, the Wangjiaqiao watershed was selected as a monitoring site. A national Gauging station on the outlet of the watershed was constructed in 1989. Gauging records

of discharge and sediment have provided useful information for decision makers and planners since the 1990s.

The Wangjiaqiao watershed lies in Zigui County of Hubei Province, China (31°5' N–31°9' N, 110°40' E–110°43' E). The watershed is approximately 50 km northwest of the Three Gorges Dam and covers an area of 1670 ha (Figure 1). Elevations within the watershed range from 184 to 1180 m; slopes range from 2° to 58°, with an average of 23°. Two main soil great groups occur in the study watershed, namely, purple soil derived from purple sandy shale and paddy soil developed from the purple soil. According to the Soil Taxonomy of the USDA, the purple soil and paddy soil are classified into Entisols and Aquepts, respectively. The climate is subtropical, with mean temperatures between 11 and 18 °C. Annual precipitation averages 1016 mm, of which 70% occurs between May and September. Previous studies about this watershed can see Fang *et al.* [17] and Fang *et al.* [18].

2.2. Field Surveys and Land Use

Field surveys were conducted in 1995, 2000, and 2005. The watershed topographic map (scale 1:10,000) was used in combination with 1995 and 1999 aerial photographs and 2005 SPOT5 imagery. The land use types were delineated on the photographs and verified in the field. In this watershed, land use is mainly a function of elevation and topography. The remnant forest patches exist primarily on steep, inaccessible peaks and slopes. Little natural vegetation is observed, and most areas are covered by secondary vegetation under human influence. The main agricultural crops are rice (*Oryza sativa* L.), maize (*Zea mays* L.), and wheat (*Triticum aestivum* L.). The streams in the Wangjiaqiao watershed have a trellis drainage pattern, and the length of the main channel is approximately 6500 m. Due to the implementation of ISWM for soil conservation in the TGA in the 1990s, land use was altered between 1995 and 2005 in the Wangjiaqiao watershed. Table 1 provides the areas of various types of land use and the corresponding percentages. In 1995, forest covered 48.7% of the study area, whereas farmland covered 43.1%. The other land use types were relatively minor and consisted of shrub land (3.2%), rural residential land (4.3%), and water bodies (0.7%). During the 1995–2005 periods, some steep lands with slope gradients of more than 25° were converted to forest. This change was related to the implementation of ISWM for soil conservation in the TGA in the 1990s. During this period, forest increased to 56.9% in 2000 and to 66.3% in 2005, whereas farmland decreased to 34.6% and 24.1%, respectively (Figure 2).

Table 1. Changes in different land use categories as a percentage of the total watershed area.

Land Use	% of Total (16.7 km ²) Area			Change (%)		
	1995	2000	2005	1995–2000	2000–2005	1995–2005
Forest	48.7	56.9	66.3	8.2	9.4	17.6
Farmland	43.1	34.6	24.1	−8.5	−10.5	−19
Shrub land	3.2	3.1	3.0	−0.1	−0.1	−0.2
Rural residential land	4.3	4.6	5.3	0.3	0.7	1
Water body	0.7	0.8	1.3	0.1	0.5	0.6

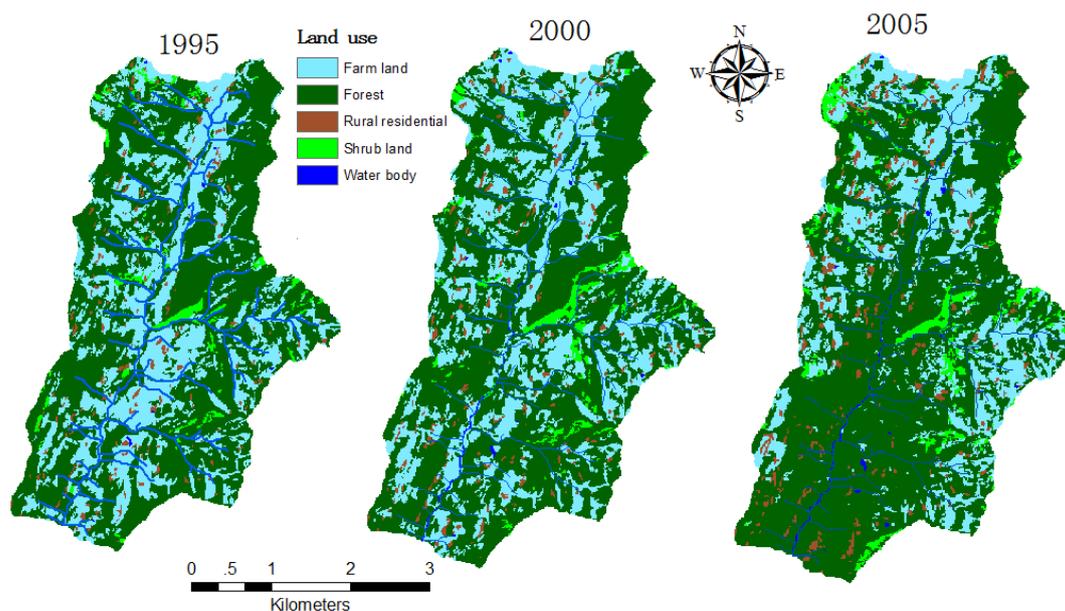


Figure 2. Land use change in the study watershed for 1995, 2000, and 2005.

2.3. Field Monitoring

A set of instruments consisting of a continuously recording rain gauge, water-level stage recorder, and silt samplers (bottle type) were used to record rainfall, stream flow, and sediment flow, respectively. The water stage was measured every 15 min and then transformed into discharge via the calibrated rating curve obtained through periodic flow measurements. Suspended sediment concentrations (SSCs) were determined by the gravimetric method. Water samples were vacuum filtered through a 0.45 μm filter, and the residue was oven dried at 105 $^{\circ}\text{C}$ for 24 h. The weight of each dried residue and the sample volume were used to determine the SSC ($\text{g}\cdot\text{m}^{-3}$). The suspended sediment yield (TL) was then calculated from the SSC and water discharge (Q) data. Watershed runoff and rainfall data have been collected since 1989.

2.4. Data Processes

In this study, runoff was separated between storm-flow and base-flow, using the classical hydrograph separation method [19]. Despite its arbitrary nature (similar to all other methods), this hydrograph separation method was used in this study to characterize the response of the catchment to a rainstorm, and no interpretation in terms of runoff processes was derived from the separation [20]. Equipment malfunctions prevented complete monitoring of all storms on several occasions. The land pattern of the study watershed varied continuously during the 1990–2005 period. Hydrograph separation was conducted on 29 events occurring in 1995, 2000, and 2004; for these events, we had reasonably completed records of sediment concentrations and conducted reconnaissance field surveys. The inner-event rainfall data for 2005 were missing, so we used data of 2004 instead of 2005. Floods were identified when the increase in stream discharge exceeded 1.5 times the base flow recorded at the beginning of the rainfall event [21]. For each rainfall-runoff event, the characteristics of individual storms were evaluated based on their erosive characteristics. The flood events were subsequently characterized using three groups of variables (Table 2).

Table 2. Flood variables and associated abbreviations used in the statistical analysis of the relationship between rainfall, runoff, and suspended sediment transport.

Variables	Variable	Abbreviation	Unit
Rainfall related variables	Total precipitation	P	mm
	Duration	D	min
	Maximum 30 min rainfall depth	I_{30}	mm
	Mean rainfall intensity	I_m	$\text{mm}\cdot\text{h}^{-1}$
	Runoff duration	D_r	mm
	Antecedent precipitation 1 day before	$P1D$	mm
	Antecedent precipitation 5 days before	$P5D$	mm
Runoff related variables	Antecedent precipitation 10 days before	$P10D$	mm
	Runoff	R	m^3
	Base flow	BF	$\text{m}^3\cdot\text{s}^{-1}$
	Total discharge	TQ	$\text{m}^3\cdot\text{s}^{-1}$
	Flood peak discharge	Q_{\max}	$\text{m}^3\cdot\text{s}^{-1}$
	Duration of runoff	Dq	s
Suspended sediment-related variables	Maximum flood suspended sediment concentration	SSC_{\max}	$\text{g}\cdot\text{m}^{-3}$
	Total suspended sediment load	TL	t

Some variables use the follow equations to calculate:

$$R = TQ - BF \times Dq \quad (1)$$

where R , TQ , BF , and Dq are the runoff, total discharge, base flow, and duration of discharge respectively.

$$TL_i = SSC_i \times Q_i \quad (2)$$

$$TL = \sum_{i=1}^n TL_i \quad (3)$$

where TL_i , SSC_i and Q_i are the suspended sediment yield, maximum flood suspended sediment concentration, and discharge, respectively, during period i .

2.5. Clustering Approach and Partial Least Squares Regression (PLSR)

Runoff and sediment generation varies considerably depending on rainfall type. Many studies have suggested that local storm patterns are important for determining runoff and sediment yield [22,23]. Such rain parameters as depth, duration, and intensity play a key role in inducing various water erosion rates [24,25]. Thus, prior to PLSR analysis, we used a clustering approach to distinguish rainfall regimes. Clustering approach was evaluated with the SPSS13.0 statistical software package.

Partial least squares regression is a robust multivariate regression method that allows users to perform a wide range of analyses [26]. Partial least squares regression provides a quick overview of the main systematic types of variation in data from complex systems and helps to identify mistakes in the input data. Applied as a multivariate calibration of one dependent variable vs. many independent variables, PLSR is suitable for selectivity enhancements of analytical instruments [26]. PLSR is a method for relating two data matrices, X and Y, by a linear multivariate model, but goes beyond traditional regression in that it models also the structure of X and Y. PLSR derives its usefulness from its ability to

analyze data with many, noisy, collinear, and even incomplete variables in both X and Y [27]. Details on the theory, principles, and application of PLSR can be found in the literature [28]. In this study, PLSR was performed with SIMCA-P (Umetrics AB, Umeå, Sweden). Four separate PLSR models were constructed to identify the main variables that control runoff and sediment yield for the two rainfall regimes. For runoff models of two rainfall regimes, runoff (R) is considered to be dependent variable and the other variables in Table 3 are considered to be independent variables. As similar, for sediment load models, total suspended sediment load (TL) are considered to be dependent variable.

Table 3. General characteristics of the analyzed rainfall-runoff events recorded in the Wangjiaqiao watershed.

Date	D (min)	P (mm)	I_{30} (mm)	I_m (mm/h)	Q_{max} (m ³ /s)	R (m ³)	TQ (m ³)	SSC_{max} (g/m ³)	TL (Kg)	PID (mm)	$P5D$ (mm)	$P10D$ (mm)	BF (m ³ /s)
1995.4.16	1380	14.6	3.8	0.63	1.16	12,407	45,507	520	1056	1.6	13.3	24.2	0.034
1995.5.12	280	19	16.8	4.07	0.97	5296	217,852	1590	1345	0	5.7	6.4	0.032
1995.5.19	1420	27.1	4.7	1.15	4.05	155,909	170,360	1310	105,357	8	8	37.2	0.036
1995.6.1	610	23.6	3.8	2.32	0.21	32,918	35,269	610	7618	0	0.1	0.1	0.028
1995.6.5	780	29.5	4.7	2.27	4.05	244,089	217,852	1120	59,781	0	32.3	32.4	0.039
1995.6.13	1790	21.9	1.4	0.73	1.58	57,296	53,469	240	5129	0	54.1	89	0.033
1995.6.20	1420	18.7	2.8	0.79	3.33	73,155	95,008	400	18,391	0	3.9	26.1	0.034
1995.6.21	825	12.1	3.5	0.88	3.41	146,189	160,937	150	23,048	18.7	22.6	44.8	0.857
1995.10.2	1250	31.4	5.8	1.51	3.13	72,058	73,796	370	22,701	24.6	24.6	41.8	0.033
1995.10.13	145	16.5	9.5	6.83	1.20	10,022	15,731	950	73	0.4	0.4	13.8	0.039
1995.10.17	480	24.1	3.2	3.01	3.37	96,509	89,487	1360	44,569	1	17.9	20.8	0.053
1995.10.19	1430	41.2	3.7	1.73	5.50	311,040	561,755	1460	180,757	3	28.1	45	0.104
1995.10.22	1035	14.8	2.3	0.86	2.87	157,378	141,538	380	19,554	0	69.2	87.1	0.083
2000.5.15	710	28.6	3.0	2.42	0.08	4717	4001	370	744	0	0	21.7	0.012
2000.6.3	565	22	2.6	2.34	0.05	2972	2624	270	299	0	5.4	26	0.006
2000.6.5	720	15.2	3.9	1.27	0.88	42,025	42,571	3020	23,263	0	22.8	48	0.011
2000.6.26	600	18.4	16.4	1.84	0.16	7857	10,200	1110	2254	1.2	37.3	50.9	0.014
2000.7.1	1420	45.2	4.1	1.91	3.16	265,766	352,487	2810	162,234	0	3.2	58.6	0.012
2000.7.29	235	25.9	13.1	6.61	0.07	3335	3271	950	917	14.3	17.8	18.6	0.011
2000.8.2	1090	52.2	3.2	2.87	2.66	210,591	328,436	2170	450	0	40.2	43.7	0.012
2000.10.21	1195	22.6	2.6	1.13	2.32	165,335	177,404	670	7449	3.7	6.4	35.4	0.088
2000.10.24	1080	23.1	4.4	1.28	2.58	136,236	141,426	620	56,253	4.8	31.1	37.3	0.475
2000.10.25	665	10.2	3.2	0.92	5.59	126,144	972,990	18200	42,791	23.1	50.5	60.4	0.635
2004.5.29	590	26	3.1	2.64	0.68	24,710	26,772	280	6053	0	40.5	45.9	0.030
2004.6.3	2005	72.7	10.1	2.18	5.44	263,002	265,446	5240	412,651	0	26	66.5	0.020
2004.8.3	300	31.1	20.1	6.22	0.68	16,278	39,391	3010	7631	0	2	2	0.010
2004.9.19	485	61	10.9	7.55	2.20	80,784	70,506	660	29,117	0	0	0.4	0.010
2004.9.24	225	39.2	3.0	10.45	2.43	177,984	125,341	490	33,074	0	63.5	63.5	0.850
2004.11.12	880	17.1	30.0	1.17	0.05	4640	4126	130	234	0	15.8	15.8	0.022

3. Results

3.1. Characteristics of the Flood Events

Table 3 summarizes the general characteristics of the rainfall, runoff, suspended sediment, and antecedent conditions associated with the observed floods and variables as analyzed by statistical analysis.

The maximum amount of precipitation for a single event was 72.7 mm (during the event on 3 June 2004); most of the events were relatively small in magnitude. Only 10 events (34%) were greater than the average rainfall value (27.8 mm), whereas the remaining events were below the average. The mean intensity varied from 0.63 to 10.45 mm·h⁻¹, and eight events were greater than the mean value (2.74 mm·h⁻¹). The maximum 30 min intensity ranged from 1.3 to 30 mm; 24% exceeded 10 mm during the 30 min interval, and eight of the events were greater than the average value (6.89 mm). The antecedent rainfall values varied considerably, ranging from 0 to 24.6 mm, 0 to 69.2 mm, and 0.1 to 89.0 mm of precipitation during the 1-, 5-, and 10-day previous periods, respectively.

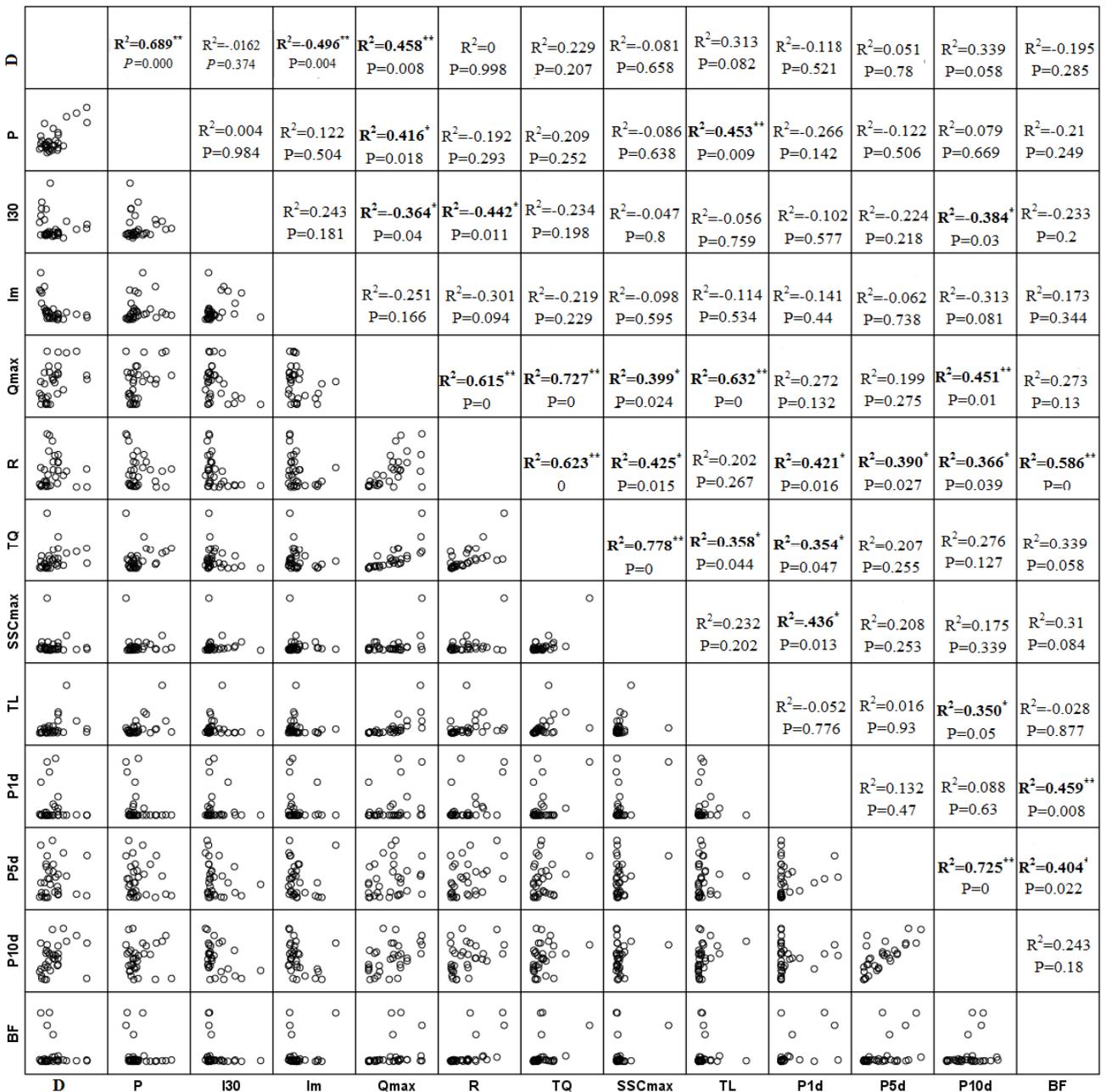


Figure 3. Bivariate scatter plot matrix of selected event characteristics. Note: ** means very significant levels ($p < 0.01$), and * means significant levels ($p < 0.05$).

The runoff generated by rainfall varied between 3674 and 309,618 m³, with a mean value of 100,459 m³. Peak discharge oscillated between 0.05 and 5.59 m³·s⁻¹. The peak was greater than the mean value (2.20 m³·s⁻¹) during 16 floods (55% of the total sample). The baseflow level fluctuated from 0.006 to 0.857 m³·s⁻¹. The maximum flood sediment concentrations varied from 130 to 18,200 g·m⁻³; with a mean concentration of 1740 g·m⁻³. The total suspended sediment load carried by 15 floods exceeded 167 t (representing a specific suspended sediment yield of 10·t·km⁻²). The maximum yield during a single flood occurred on 4 June 2004 and reached 412,651 kg. This yield was generated by 72.7 mm of precipitation, which created a flood peak discharge of 5.44 m³·s⁻¹. These values illustrate the degree of geomorphic activity of the system and confirm the high sediment contribution and transport capacity of the channels in the Wangjiaqiao watershed, which are largely related to the availability of fine materials in the TGA areas and the accumulation of these materials along the main channel [15].

We generated a Pearson correlation matrix (Figure 3). The linear correlation coefficients among rainfall-, runoff-discharge-, sediment-, and antecedent condition-related variables are described in detail. Peak flow (Q_{\max}) was significantly correlated with R , TQ , SSC_{\max} , TL , and $P10D$. The strongest correlation was between precipitation (TQ) and runoff (Q_{\max}). Runoff was significantly correlated with TQ , SSC_{\max} , $P1D$, $P5D$, $P10D$, and BF . The results confirmed that many variables were co-linear.

3.2. Results of Clustering Approach

The 29 rainfall events were divided into two groups using K-means clustering. Three rainfall variables were used during this process: the depth (P), duration (D), and maximum 30 min rainfall intensity (I_{30}) (Table 4). The general characteristics of the two rainfall regimes can be described as heavy and moderate rainfall ($p < 0.0001$). Compared with Rainfall Regime II (17 events), Rainfall Regime I (12 events) was composed of rainfall events with a high mean P and D and low I_{30} .

Table 4. Statistical features of the different rainfall regimes.

Rainfall Regimes	Variables	Mean	SD	Variation	Frequency (time)
Rainfall Regime I	P (mm)	32.1	17.6	0.55	
	D (min)	1376	288	0.21	
	I_{30} (mm)	4.1	2.2	0.55	12
	Runoff (m ³)	156,601	92,988	0.59	
	TL (kg)	82,665	121,676	1.47	
Rainfall Regime II	P (mm)	24.7	11.9	0.48	
	D (min)	535	227	0.42	
	I_{30} (mm)	8.9	8.0	0.90	17
	Runoff (m ³)	60,829	73,380	1.21	
	TL (kg)	16,636	19,140	1.15	

3.3. Results of PLSR Analysis

Many studies have demonstrated that land use type can change runoff and sediment yield [29]; thus, the percentages of forest and farmland during the study years were included in the PLSR models. In a PLSR model, the importance of a predictor for both the independent and dependent variables is given by

the variable importance for the projection (VIP) [10,12]. Terms with large VIP values are the most relevant for explaining the dependent variable. To overcome the problem of over-fitting, the appropriate number of components of each PLSR model was determined by cross-validation to achieve an optimal balance between the explained variation in the response (R^2) and the predictive ability of the model (goodness of prediction: Q^2) [10].

Table 5 provides a summary of the four PLSR models constructed separately for the runoff and sediment yield of the two rainfall regimes.

Table 5. Summaries of the partial least squares regression (PLSR) models.

Rainfall Regime	Response Variable Y	R^2	Q^2	Component	% of Explained Variability in Y	Cumulative Explained Variability in Y (%)	RMSECV ^a (m ³ or kg)	Q^2_{cum}
Rainfall Regime I	R	0.99	0.85	1	71.6	71.6	51,977	0.564
				2	22.5	94.1	25,035	0.761
				3	2.7	96.8	19,507	0.737
				4	2.0	98.8	12,801	0.803
				5	0.7	99.5	8862	0.852
				6	0	99.5	9265	0.837
Rainfall Regime I	TL	0.97	0.62	1	80.6	80.6	52,181	0.570
				2	16.2	96.8	24,088	0.624
				3	1.9	98.7	16,500	0.587
				4	0.5	99.2	13,434	0.546
Rainfall Regime II	R	0.90	0.65	1	69.6	69.6	41,776	0.444
				2	20.1	89.7	25,167	0.646
				3	3.3	93.0	21,477	0.610
Rainfall Regime II	TL	0.86	0.66	1	80.6	66.6	11,427	0.503
				2	4.3	88.6	6912	0.656
				3	1.4	92.2	5933	0.637

Notes: ^a The RMSECV (cross-validated root mean squared error), Q^2_{cum} (cross-validated goodness of prediction) per component, R^2 (goodness of fit), and Q^2 (cross-validated goodness of prediction) were calculated for the PLSR models.

For the runoff model of Rainfall Regime I, the prediction error decreased with an increasing number of components, and the minimum RMSECV and maximum Q^2 were obtained with five components. An additional increase in the number of components generated a higher prediction error, suggesting that the other components were not strongly correlated with the residuals of the predicted variable [12]. The first component explained 71.6% of the variance in the dataset in terms of the changes in runoff (Table 5). The addition of the second through fifth components to the models cumulatively explained 99.5% of the total variance. For the *TL* model of heavy rainfall, the maximum Q^2 was obtained with two components. The first component explained 80.6% of the variance in the dataset in terms of the changes in *TL*. The second component explained 16.2% of the variance. These two components explained 96.8% of the total variance. Further addition of components to the PLSR models did not substantially increase the percentage of the variance explained (Table 5). The PLSR weights could be used to describe the quantitative relationship between the predictors and results because they are linear combinations of the original variables that defined the scores [28].

For the runoff model of moderate rainfall, the optimum model had two components, with a maximum Q^2 of 0.646. The model explained 89.7% of the total variance, with 69.6% of the variance explained by the first component. The optimum model for the TL of moderate rainfall also contained two components. Those two components explained 88.6% of the total variance. The maximum Q^2 was 0.656.

The first component of the runoff model for heavy rainfall (Table 6) was dominated by P , I_m , Q_{max} , TQ , SSC_{max} , and TL , whereas the second component was dominated by Q_{max} and TQ . The third, fourth, and fifth components were dominated by many variables, mainly on the negative side. A more convenient and comprehensive expression of the relative importance of the predictors was obtained by exploring their VIP values [12]. For runoff in Rainfall Regime I, higher VIP values were obtained for changes in TQ , Q_{max} , P , I_m , SSC_{max} , and TL ($VIP > 1$), followed by the percentage of forest and farmland (0.936 and 0.941) (Figure 4). Predictors with VIP values below one are considered of minor predictive importance. For runoff in Rainfall Regime II, a higher VIP value was obtained for changes in TL , Q_{max} , BF , $P5D$, and TQ . Compared with Rainfall Regime I, runoff in Rainfall Regime II was more likely to be affected by antecedent conditions, such as BF and $P5D$.

Table 6. PLSR for runoff ^a.

Predictors	R of Rainfall Regime I					R of Rainfall Regime II			
	RCs ^b	<i>W</i> * (1)	<i>W</i> * (2)	<i>W</i> * (3)	<i>W</i> * (4)	<i>W</i> * (5)	RCs	<i>W</i> * (1)	<i>W</i> * (2)
<i>P</i>	-0.023	0.357	-0.038	-0.295	-0.346	-0.137	0.206	0.102	0.496
<i>I</i> ₃₀	-0.187	0.172	-0.397	-0.124	-0.300	-0.17	-0.114	-0.251	-0.040
<i>I</i> _m	0.005	0.344	0.161	-0.469	-0.486	0.025	0.110	0.036	0.289
<i>Q</i> _{max}	0.288	0.365	0.305	0.293	-0.031	-0.188	0.279	0.478	0.263
<i>TQ</i>	0.538	0.446	0.693	0.294	0.182	0.192	0.018	0.251	-0.251
<i>SSC</i> _{max}	0.079	0.339	-0.098	-0.031	0.022	0.025	-0.107	0.118	-0.465
<i>TL</i>	0.195	0.334	-0.024	0.278	0.212	-0.074	0.388	0.523	0.536
<i>P</i> <i>I</i> <i>D</i>	0.024	-0.129	-0.164	0.175	0.194	0.731	-0.053	0.147	-0.339
<i>P</i> 5 <i>D</i>	-0.193	-0.011	0.103	-0.620	-0.325	-0.499	0.151	0.325	0.061
<i>P</i> 10 <i>D</i>	0.326	0.099	0.138	0.218	0.927	0.414	0.063	0.258	-0.122
<i>BF</i>	0.007	-0.011	0.136	0.004	-0.081	-0.347	0.167	0.346	0.086
<i>Forest</i>	0.144	0.260	-0.150	0.063	0.394	0.382	-0.059	-0.095	-0.064
<i>Farm</i>	-0.133	-0.26	0.160	-0.054	-0.375	-0.357	0.053	0.091	-0.064
<i>D</i>	-0.279	0.061	-0.431	-0.111	-0.303	-0.644	0.081	0.128	0.088

Notes: ^a Values larger than 0.3, which are shown in bold face, indicate that the PLSR components are mainly loaded on the corresponding variables. *W** means component. ^b RCs means Regression Coefficients.

For the TL model of Rainfall Regime I, higher VIP values were obtained for changes in P , I_{30} , I_m , Q_{max} , SSC_{max} , R , percentage of forest and farmland, and D (see Figure 5). Thus, all variables have important effects on TL with the exception of antecedent conditions and TQ . For the TL model of Rainfall Regime II, high VIP values were obtained for P , Q_{max} , TQ , R , and BF . Table 7 indicates that Q_{max} was important in the first and second components of Rainfall Regimes I and II, whereas $P5D$ and $P10D$ had only minor effects on TL for either rainfall regime.

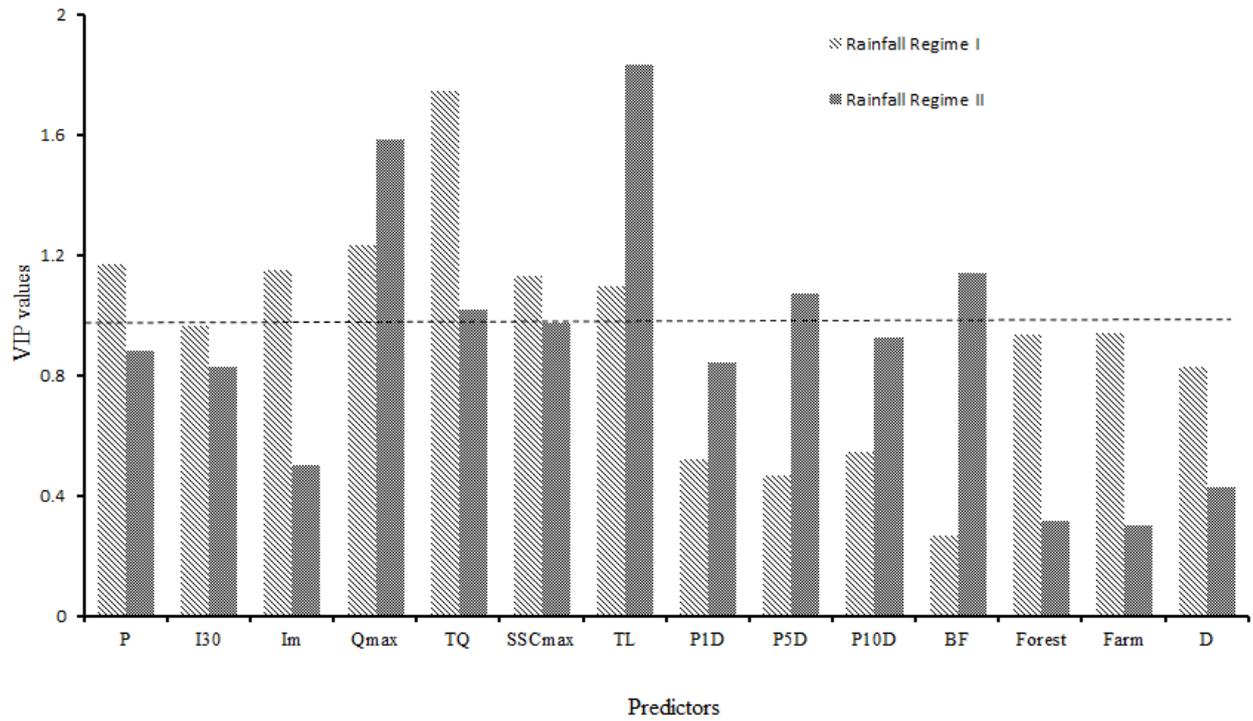


Figure 4. VIP values of each predictor for PLSR of runoff.

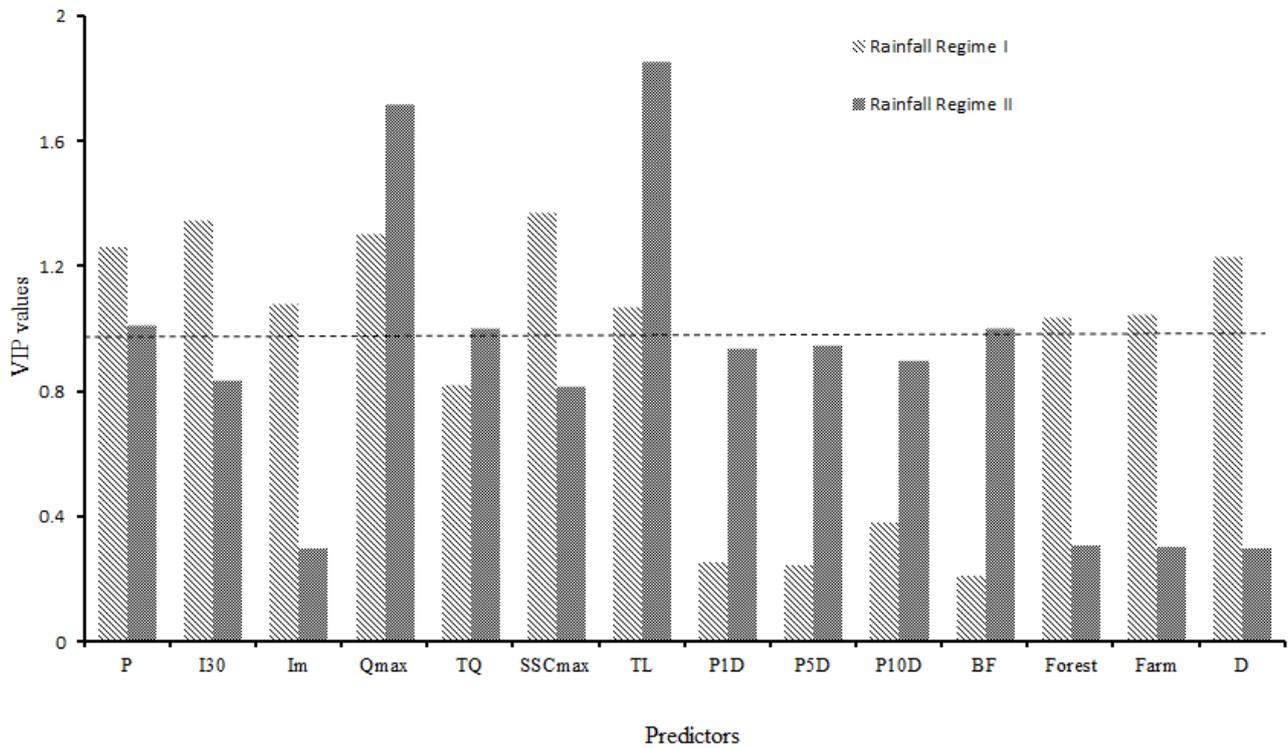


Figure 5. VIP values of each predictor for PLSR of sediment load.

Table 7. PLSR for sediment load.

Predictors	TL of Rainfall Regime I			TL of Rainfall Regime II		
	RCs ^a	<i>W</i> * (1)	<i>W</i> * (2)	RCs	<i>W</i> * (1)	<i>W</i> * (2)
<i>P</i>	0.086	0.355	−0.133	0.244	0.112	0.547
<i>I</i> ₃₀	0.266	0.367	0.383	−0.131	−0.259	−0.082
<i>I</i> _m	−0.116	0.200	−0.559	0.052	−0.014	0.156
<i>Q</i> _{max}	0.258	0.355	0.373	0.344	0.516	0.388
<i>TQ</i>	0.036	0.225	−0.141	0.084	0.304	−0.094
<i>SSC</i> _{max}	0.175	0.396	0.084	0.031	0.216	−0.148
<i>R</i>	0.093	0.305	−0.058	0.411	0.532	0.553
<i>P1D</i>	−0.016	−0.037	0.033	−0.098	0.130	−0.406
<i>P5D</i>	−0.050	−0.067	−0.074	0.092	0.281	−0.048
<i>P10D</i>	0.091	0.093	0.167	−0.004	0.212	−0.237
<i>BF</i>	0.021	−0.042	0.106	−0.008	0.233	−0.272
<i>Forest</i>	0.048	0.285	−0.170	−0.052	−0.097	−0.039
<i>Farm</i>	−0.056	0.291	0.152	0.048	0.094	0.031
<i>D</i>	0.292	0.296	0.536	0.083	0.134	0.182

Notes: ^a RCs means Regression Coefficients; *W** means component.

4. Discussion

4.1. Control Factors for Runoff

The hydrological response of a watershed to a rainfall event is determined by several interacting factors that control runoff generation [30]. In this study, only sediment flux data from the outlet of the watershed were available, and thus, the within-watershed pattern of erosion and deposition remains uncertain. Soil moisture content is reflected in the antecedent conditions with *BF*, *P1D*, *P5D* and *P10D*.

The results indicated that *TQ*, *Q*_{max}, and *TL* are important for runoff in Rainfall Regimes I and II. The rainfall characteristics (*P* and *I*_m) are more important for runoff in Rainfall Regime I, whereas soil moisture content is more important for Rainfall Regime II. The former pattern is in agreement with the results of Oeurng *et al.* [29]. In Regime I, *TQ* and *Q*_{max} dominate runoff, and base flow has only a minor effect due to its small magnitude. The effect of soil moisture is greater in Rainfall Regime II than in Rainfall Regime I. The sensitivity of the runoff response to soil moisture depends on the predominant runoff mechanisms [31]. Surface runoff generated by saturation-excess flow is driven from spatially and temporally dynamic variable source areas and requires lower rainfall intensities for its initiation. Both infiltration excess and saturation excess runoff processes may occur during heavy rainfall [32]. Discharge is affected by drainage network density, slope, channel roughness, and soil infiltration characteristics [33], and peak flow rates are typically affected by within-storm rainfall characteristics. Many studies have indicated that storm runoff from steep well-vegetated headwater catchments in humid areas is produced by saturation-excess mechanisms [34,35]. Central to the saturation-excess mechanism is the concept of runoff contributing zones, which expand and contract seasonally and during storms depending on antecedent wetness and storm magnitude [36]. The initial soil water content is more important in catchments with good vegetation. The runoff response is more uniform and is not dependent on initial soil moisture when the infiltration excess overland flow is predominant because of

high rain depth or less permeable soils [30]. Runoff from lower-intensity storms in soils of higher permeability is controlled by the soil water content of the surface soil layers and is more dependent on the initial conditions [30]. We attributed the different control factors of the two Rainfall Regimes to rainfall characteristics and consequent runoff mechanisms.

4.2. Control Factors of Sediment Yield

The sediment response of catchments is controlled by a complex function of ecological, climatic, and geomorphic processes [37]. All of the rainfall-related variables have important effects on TL for Rainfall Regime I. Raindrop splash detachment and surface overland flow are the two basic drivers of soil peel-back, flood generation, water loss, and sediment mobilization [38]. However, these two drivers are both rooted in rainfall energy; vegetation reduces this energy, and thus, the Q_{max} observed at the outlet of the watershed may reflect the rainfall energy. The sediment yield of a catchment represents only a part of the total erosion or sediment production within the catchment, as a considerable portion of the sediment is often deposited before reaching the outlet [39]. In the Wangjiaqiao watershed, 76% of the area has slope gradients exceeding of 30%. Cultivated sloping lands are major contributors to sediment yield, and a considerable portion of the sediment may move by gravity rather than by shear forces alone [15]. Lu and Higgitt [14] conclude that 60% of sediment is from arable land in 32 catchments in the TGA. The soil parent material of the watershed is predominantly purple sandy shale, and bedrock is typically exposed in the channel bed; thus, channel erosion is rare [18]. Sediments stored in the channel and distributed within tributaries are transported after flood events with sufficient transport capacity. The PLSR model of TL for Rainfall Regime II illustrates that events with low rainfall depth and short duration typically cause very limited hydrological responses and limited sediment transport, in agreement with previous study [40].

Land use change is important for sediment yield [41]. Soil erosion is largely determined by the absence of protective land cover, whereas sediment export to rivers is determined by onsite sediment production and the connectivity of sediment sources and the river [42,43]. Few studies have focused on how changes in land use can influence inter-event hydrologic process and sediment yield at a small watershed scale. Our results indicate that the percentage of forest and farmland in the study area has an important influence on sediment caused by Rainfall Regime I, *i.e.*, heavy rainfall events. The rainfall energy of an event is typically characterized by I_m and I_{30} , which had important effects on TL in Rainfall Regime I. Thus, heavy rainfall events, which cause a large sediment yield, are influenced by more factors than lower-intensity rainfall events. A similar conclusion has been reported by Ni *et al.* [44], who demonstrates that soil erosion has many forms for heavy rainfall events, including sheet erosion, rill erosion, gully erosion, landslip, and even collapses. Thus, the sediment processes are complex, heterogeneous, and controlled by multiple factors.

5. Conclusions

In this paper, K-means clustering was used to classify 29 rainfall events into two rainfall regimes. Four separate PLSR models were constructed to identify the main variables that control runoff and sediment yield in a small agriculture watershed in the TGA. The results confirmed the complex and heterogeneous nature of the hydrology and sediment response in the watershed.

- (1) Rainfall Regime I, which is characterized by a high mean rainfall depth and long duration, produced 2.6-fold greater mean runoff and 4.9-fold greater mean sediment yield than Rainfall Regime II.
- (2) The initial soil water content was more important for controlling runoff in Rainfall Regime II than in Rainfall Regime I, whereas rainfall characteristics played a greater role in controlling runoff in Rainfall Regime I.
- (3) Although land use changed significantly during the study years, these changes were not reflected in the control factors for runoff for Rainfall Regimes I or II or the control factors of sediment yield for Rainfall Regime II. At the inter-event scale, the percentages of forest and farmland only had a significant effect on sediment yield in Rainfall Regime I.

The PLSR methodology presented in this paper partially eliminates the co-dependency of the variables and facilitates a more unbiased view of the control factors for runoff and sediment yield. The variables used for clustering approach and PLSR are easy to obtain. Thus, this practicable and simple approach could be applied to a variety of other watersheds and enable better management of agricultural watersheds.

Acknowledgments

Financial support for this research was provided by the National Natural Science Foundation of China (41301294), the West Light Foundation of the Chinese Academy of Sciences, and the Fundamental Research Funds for the Central Universities (2014YB053).

Author Contributions

Both authors contributed equally to this work, including the related research and writing.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Zhu, A.X.; Wang, P.; Zhu, T.X.; Chen, L.J.; Cai, Q.G.; Liu, H.P. Modeling runoff and soil erosion in the Three-Gorge Reservoir drainage area of China using limited plot data. *J. Hydrol.* **2013**, *492*, 163–175.
2. Estrany, J.; Garcia, C.; Batalla, R.J. Hydrological response of a small mediterranean agricultural catchment. *J. Hydrol.* **2010**, *380*, 180–190.
3. Zabaleta, A.; Martínez, M.; Uriarte, J.A.; Antigüedad, I. Factors controlling suspended sediment yield during runoff events in small headwater catchments of the Basque Country. *Catena* **2007**, *71*, 179–190.
4. Oeurng, C.; Sauvage, S.; Sánchez-Pérez, J.M. Dynamics of suspended sediment transport and yield in a large agricultural catchment, southwest France. *Earth Surf. Process. Landf.* **2010**, *35*, 1289–1301.

5. López-Tarazón, J.; Batalla, R.J.; Vericat, D.; Balasch, J. Rainfall, runoff and sediment transport relations in a mesoscale mountainous catchment: The River Isábena (Ebro basin). *Catena* **2010**, *82*, 23–34.
6. Wine, M.L.; Zou, C.B.; Bradford, J.A.; Gunter, S.A. Runoff and sediment responses to grazing native and introduced species on highly erodible Southern Great Plains soil. *J. Hydrol.* **2012**, *450*, 336–341.
7. Mayor, Á.G.; Bautista, S.; Bellot, J. Factors and interactions controlling infiltration, runoff, and soil loss at the microscale in a patchy Mediterranean semiarid landscape. *Earth Surf. Process. Landf.* **2009**, *34*, 1702–1711.
8. Jarvis, D.; Stoeckl, N.; Chaiechi, T. Applying econometric techniques to hydrological problems in a large basin: Quantifying the rainfall-discharge relationship in the Burdekin, Queensland, Australia. *J. Hydrol.* **2013**, *496*, 107–121.
9. Taguas, E.; Ayuso, J.; Pérez, R.; Giráldez, J.; Gómez, J. Intra and inter-annual variability of runoff and sediment yield of an olive micro-catchment with soil protection by natural ground cover in Southern Spain. *Geoderma* **2013**, *206*, 49–62.
10. Onderka, M.; Wrede, S.; Rodný, M.; Pfister, L.; Hoffmann, L.; Krein, A. Hydrogeologic and landscape controls of dissolved inorganic nitrogen (DIN) and dissolved silica (DSi) fluxes in heterogeneous catchments. *J. Hydrol.* **2012**, *450*, 36–47.
11. Carrascal, L.M.; Galván, I.; Gordo, O. Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos* **2009**, *118*, 681–690.
12. Shi, Z.; Ai, L.; Li, X.; Huang, X.; Wu, G.; Liao, W. Partial least squares regression for linking land-cover patterns to soil erosion and sediment yield in watersheds. *J. Hydrol.* **2013**, *498*, 165–176.
13. Rommens, T.; Verstraeten, G.; Bogman, P.; Peeters, I.; Poesen, J.; Govers, G.; Van Rompaey, A.; Lang, A. Holocene alluvial sediment storage in a small river catchment in the loess area of central Belgium. *Geomorphology* **2006**, *77*, 187–201.
14. Lu, X.; Higgitt, D. Sediment delivery to the Three Gorges: 2: Local response. *Geomorphology* **2001**, *41*, 157–169.
15. Shi, Z.H.; Cai, C.F.; Ding, S.W.; Wang, T.W.; Chow, T. Soil conservation planning at the small watershed level using RUSLE with GIS: A case study in the Three Gorge Area of China. *Catena* **2004**, *55*, 33–48.
16. Liao, C.Y. Soil and water conservation in Yangtze River Basin during past 60 years: Review and perspective. *Yangtze River* **2010**, *41*, 2–6. (In Chinese)
17. Fang, N.F.; Shi, Z.H.; Li, L.; Guo, Z.L.; Liu, Q.J.; Ai, L. The effects of rainfall regimes and land use changes on runoff and soil loss in a small mountainous watershed. *Catena* **2012**, *99*, 1–8.
18. Fang, N.F.; Shi, Z.H.; Li, L.; Jiang, C. Rainfall, runoff, and suspended sediment delivery relationships in a small agricultural watershed of the Three Gorges area, China. *Geomorphology* **2011**, *135*, 158–166.
19. Bidin, K.; Greer, T. A spreadsheet-based technique (Lotus 1-2-3) for separating tropical forest storm hydrographs using Hewlett and Hibbert's slope. *Earth Surf. Process. Landf.* **1997**, *22*, 1231–1237.

20. Latron, J.; Soler, M.; Llorens, P.; Gallart, F. Spatial and temporal variability of the hydrological response in a small Mediterranean research catchment (Vallcebre, Eastern Pyrenees). *Hydrol. Process.* **2008**, *22*, 775–787.
21. Lana-Renault, N.; Regues, D.; Martí-Bono, C.; Beguería, S.; Latron, J.; Nadal, E.; Serrano, P.; Garcia-Ruiz, J. Temporal variability in the relationships between precipitation, discharge and suspended sediment concentration in a small Mediterranean mountain catchment. *Nord. Hydrol.* **2007**, *38*, 139–150.
22. Wei, W.; Chen, L.; Fu, B.; Huang, Z.; Wu, D.; Gui, L. The effect of land uses and rainfall regimes on runoff and soil erosion in the semi-arid loess hilly area, China. *J. Hydrol.* **2007**, *335*, 247–258.
23. Cerdan, O.; Govers, G.; Le Bissonnais, Y.; Van Oost, K.; Poesen, J.; Saby, N.; Gobin, A.; Vacca, A.; Quinton, J.; Auerswald, K. Rates and spatial variations of soil erosion in Europe: A study based on erosion plot data. *Geomorphology* **2010**, *122*, 167–177.
24. Piccarreta, M.; Capolongo, D.; Boenzi, F.; Bentivenga, M. Implications of decadal changes in precipitation and land use policy to soil erosion in Basilicata, Italy. *Catena* **2006**, *65*, 138–151.
25. Nyssen, J.; Poesen, J.; Moeyersons, J.; Haile, M.; Deckers, J. Dynamics of soil erosion rates and controlling factors in the Northern Ethiopian Highlands—Towards a sediment budget. *Earth Surf. Process. Landf.* **2008**, *33*, 695–711.
26. Martens, H.; Martens, M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual. Prefer.* **2000**, *11*, 5–16.
27. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.
28. Abdi, H. Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 97–106.
29. Ouyang, W.; Skidmore, A.K.; Hao, F.; Wang, T. Soil erosion dynamics response to landscape pattern. *Sci. Total. Environ.* **2010**, *408*, 1358–1366.
30. Castillo, V.; Gomez-Plaza, A.; Martínez-Mena, M. The role of antecedent soil water content in the runoff response of semiarid catchments: A simulation approach. *J. Hydrol.* **2003**, *284*, 114–130.
31. McDonnell, J.J. Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response. *Hydrol. Process.* **2003**, *17*, 1869–1875.
32. Srinivasan, M.; Gburek, W.; Hamlett, J. Dynamics of stormflow generation—A hillslope-scale field study in east-central Pennsylvania, USA. *Hydrol. Process.* **2002**, *16*, 649–665.
33. Fiedler, F.R.; Ramirez, J.A. A numerical method for simulating discontinuous shallow flow over an infiltrating surface. *Int. J. Numer. Methods Fluids* **2000**, *32*, 219–239.
34. Latron, J.; Gallart, F. Seasonal dynamics of runoff-contributing areas in a small Mediterranean research catchment (Vallcebre, Eastern Pyrenees). *J. Hydrol.* **2007**, *335*, 194–206.
35. Easton, Z.M.; Fuka, D.R.; Walter, M.T.; Cowan, D.M.; Schneiderman, E.M.; Steenhuis, T.S. Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. *J. Hydrol.* **2008**, *348*, 279–291.
36. Pearce, A.; Stewart, M.; Sklash, M. Storm runoff generation in humid headwater catchments: 1. Where does the water come from? *Water Resour. Res.* **1986**, *22*, 1263–1272.

37. Krishnaswamy, J.; Halpin, P.N.; Richter, D.D. Dynamics of sediment discharge in relation to land-use and hydro-climatology in a humid tropical watershed in Costa Rica. *J. Hydrol.* **2001**, *253*, 91–109.
38. Sanchis, M.; Torri, D.; Borselli, L.; Poesen, J. Climate effects on soil erodibility. *Earth Surf. Process. Landf.* **2008**, *33*, 1082–1097.
39. Walling, D. Tracing suspended sediment sources in catchments and river systems. *Sci. Total. Environ.* **2005**, *344*, 159–184.
40. Shi, Z.H.; Chen, L.D.; Fang, N.F.; Qin, D.F.; Cai, C.F. Research on the SCS-CN initial abstraction ratio using rainfall-runoff event analysis in the Three Gorges Area, China. *Catena* **2009**, *77*, 1–7.
41. Cantón, Y.; Solé-Benet, A.; De Vente, J.; Boix-Fayos, C.; Calvo-Cases, A.; Asensio, C.; Puigdefábregas, J. A review of runoff generation and soil erosion across scales in semiarid south-eastern Spain. *J. Arid. Environ.* **2011**, *75*, 1254–1261.
42. Van Rompaey, A.J.; Govers, G.; Puttemans, C. Modelling land use changes and their impact on soil erosion and sediment supply to rivers. *Earth Surf. Process. Landf.* **2002**, *27*, 481–494.
43. Bakker, M.M.; Govers, G.; van Doorn, A.; Quetier, F.; Chouvardas, D.; Rounsevell, M. The response of soil erosion and sediment export to land-use change in four areas of Europe: The importance of landscape pattern. *Geomorphology* **2008**, *98*, 213–226.
44. Ni, J.P.; Wei, C.F.; Xie, D.T.; Gao, M.; He, B.H. Influence of slope gradient on runoff erosion of purple soil slope in the Three Gorges Reservoir area. *J. Sediment Res.* **2009**, *2*, 29–33. (In Chinese)

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).