MODELS TO PREDICT SLOPE LENGTH FROM OTHER WATERSHED ATTRIBUTES

K. M. Mohammed[1*] T. H. Karim[1]
Assist. Lecturer Professor
[1]Dept. Soil and Water Coll. Agric. Engin. Sci. Salahaddin University-Erbil- Iraq.
kamyar.mohammed@su.edu.krd solavtariq@yahoo.com

ABSTRACT

Soil erosion by water is an extensive and increasing problem worldwide. Albeit, this problem has been recognized as a significant hazard in Iraq, yet the number of studies on this topic is very limited. Most of the models used for estimating soil erosion contain parameters for slope length factor (LS). A major constraint is the difficulty in extracting the LS factor. Accordingly, the current study was initiated with the main objective of deriving models to predict the slope length from relatively easy to measure basin characteristics with a reasonable accuracy. To achieve the above objective, standard methodologies were employed to describe 30 main basins with the upper part of Iraq in terms linear, areal and relief morphometric parameters. The majority of the delineated watersheds were characterized by having high slope lengths indicating lower drainage density and higher erosion rate. Linear and non-linear least squares techniques were applied to predict the slope length from other basin characteristics. Different indicators were used to test the performance of the proposed models and the approach was validated using K-fold procedure at independent basins. The results indicated that the 4-parameter regression model outperformed the remaining models of watershed slope length. The regressors of this model are bifurcation ratio, perimeter, and basin length and slope gradient.

Keywords: Slope length, watershed attributes, morphometric characteristics, modeling

مجلة العلوم الزراعية العراقية -2020: 51 (4):1025-1037 محمد وكريم

أشتقاق نماذج للتنبؤ بطول منحدرات أحواض الأنهر بالأعتماد على خصائصها

كاميار مطلب محمد طارق حمه كريم
مدرس مساعد استاذ
قسم علوم التربة و المياه– كلية علوم الهندسة الزراعية – جامعة صلاح الدين – اربيل

المستخلص

تعد التعرية المائية أحدى المشاكل واسعة النطاق ومتزايدة الحدة مع الزمن على المستوى العالم. وبالرغم من خطورة هذه العملية فأنه لاتزال الدراسات محدودة في هذا المجال وخصوصا في العراق. وتتضمن معظم النماذج المستخدمة لتقدير مفقودات التربة عاملى الطول ودرجة الانحدار. و ما يعيق هذا التقدير هو عدم توفر البيانات اللازمة والمتعلقة بهذه المعايير. وعلية نفذت هذه الدراسة بهدف التنبؤ بطول المنحدر على مستوى حوض النهر بالأعتماد على الخصائص المورفوميترية لثلاثين حوض ضمن المنطقة الجبلية في العراق وأظهرت الدراسة تميز معظم الاحواض بمنحدر طويل مما يستدل على انخفاض كثافة الصرف و و ارتفاع معدلات التعرية. كما أستخدمت تقنيات المربعات الصخرى الخطية وغير الخطية لبناء مجموعة من النماذج للتنبؤ بطول المنحدر. كذللك أستخدمت مجموعة من المؤشرات الأحصائية لأختباركفاءة النماذج وأظهرت النتائج بأن أحسن نموذج لتقدير هذا العامل هو نموذج ذو أربع متغيرات شملت على نسبة التشعب و محيط الحوض وطول الحوض و درجة الأنحدار.

لكلمات الدالة: طول المنحدر، خصائص حوض النهر، خصائص مورفوميترية، نمذجة

*Received:12/9/2019, Accepted:19/12/2019

INTRODUCTION

Soil erosion by water is an extensive and increasing problem worldwide. This problem has reached a stage of irreversibility in the Mediterranean region including the region under study. The reasons behind this problem are long dry periods followed by heavy bursts of erosive rainfall, falling on steep slopes with fragile soils. Therefore, quantification of soil loss becomes a significant issue for soil and water conservation practitioners and policy makers (20). Albeit, soil erosion has been recognized as a significant hazard in Iraq, yet the number of studies on this topic is very limited (12).Conservation of soil and water is in need of both knowledge of the factors affecting these natural resources and approaches for controlling these factors to preserve those resources (23). Both the universal soil loss equation (USLE) and its revised version (RUSLE) are often used to estimate soil erosion at regional landscape scales. These models contain parameters for slope length factor (L) and slope steepness factor (S). A major constraint is the difficulty in extracting the LS factor (37). The slope length, which affects the velocity of the overland flow and sediment transport processes, is generally acknowledged as the most important topographical factor in estimating, simulating and modeling soil erosion (38). Roose (24) indicated that the longer the slope, the more runoff will accumulate, gathering speed and gaining its own energy, causing rill erosion and then more serious gulling. Haan et al. (8) also demonstrated that increase in slope length and slope steepness can yield higher overland flow velocities and correspondingly higher erosion. The slope length can be defined as the horizontal distance between the origin of overland flow and the point where either the slope gradient decreases to such a degree that deposition begins or runoff becomes concentrated in a defined channel (6). The increasing availability of DEMs has promoted the use of image-processing techniques and software for deriving terrain properties (17); (2); (30) such as the LS-factor. Panagos et al. (19) revealed that the main drawback of this technique is the existence of landscape features such as roads, paths, fences, contours, stone walls and grass margins that may interrupt the water runoff and reduce the slope length. These features are not identified in the DEM. Wilson (33) stated that estimating LS values on a watershed basis is not an easy task because the field surveys are in need of the large number of profile measurement that are time consuming and costly, especially in large watershed. Hickey (9) pointed out that, traditionally, the best estimates for λ are obtained from field measurements, but these are rarely available or practical. Hickey et al. (10) clarified that the estimation of slope length factor from direct measurement requires high financial and human resources which might not always be available. Reimers (22) pointed out that it is possible to estimate hydrological parameters for a given catchment from other basin characteristics by applying a linear equation. Slope lengths estimated from contour maps are usually too long because most maps lack the detail to indicate all concentrated flow areas that ends RUSLE slope lengths (35). Horton (11) pointed out that the average length of overland flow is, in most cases, approximately half the average distance between the stream channels and is therefore approximately equal to half the reciprocal of drainage density. On the other hand, it was reported it is hard to obtain a realistic value for slope length and will vary for different users because in involves substantial judgment (34) An accurate and speedy method is required for slope length factor to improve the application of soil erosion models on watershed scales. The current study was conducted with the main objective of deriving models to predict the slope length from relatively easy to measured basin characteristics with a reasonable accuracy.

MATERIALS AND METHODS

Study area descriptions

The study area is located in the realm of the mountainous area of Erbil province, Iraq, between 35^o 30′ to 37^o 15′ 00"north latitudes and 43^o 30′ to 45^o 15′00" East longitudes. The Greater Zab and the Lesser Zab are the major upland tributaries of the Tigris River covering an area of 47285 km^2

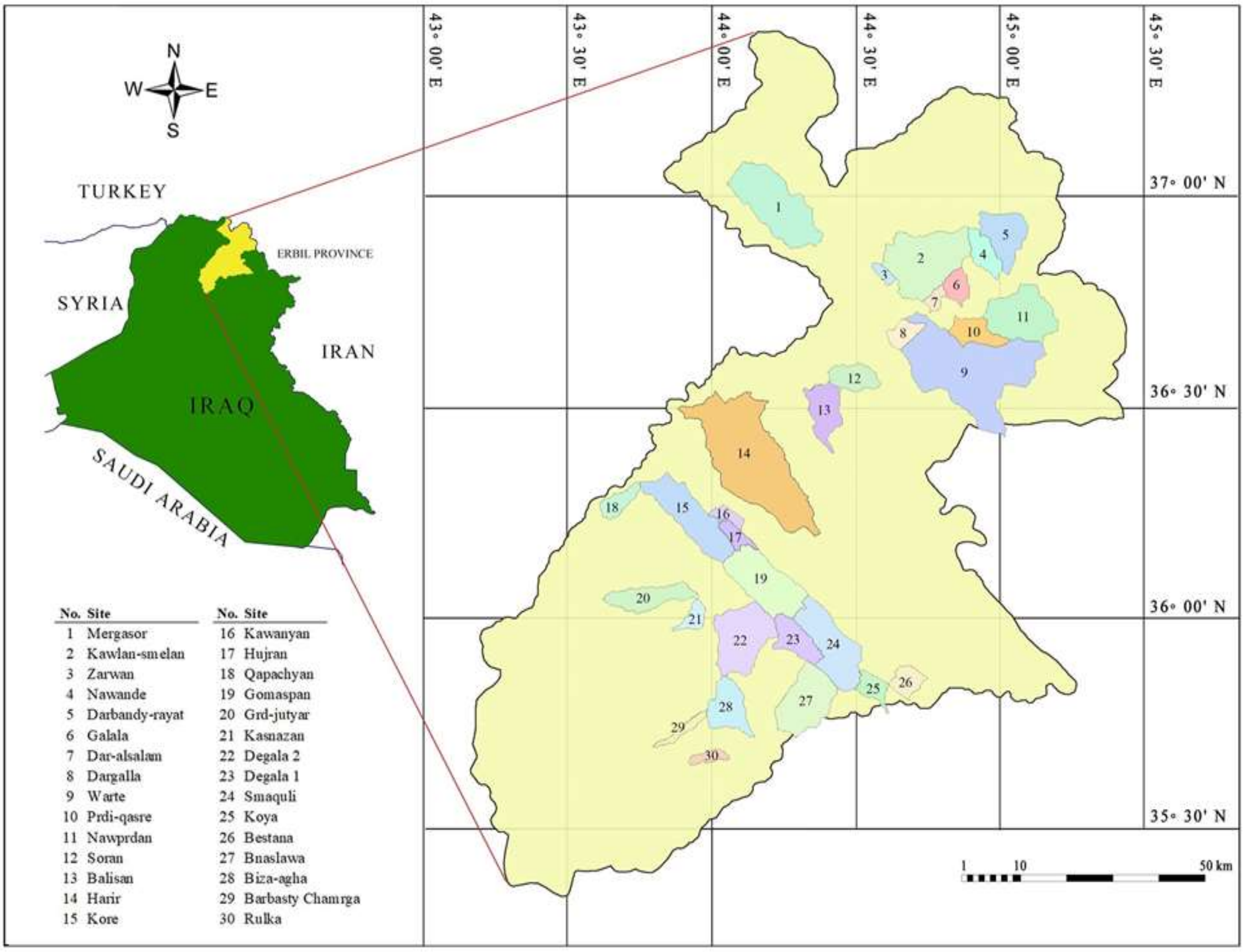

Figure1. Location map showing the delineated watersheds under study

The majority of the study areas are within this mountainous region. This region is considered as a part of the Zagros Mountains separated by broad valleys. Generally, the Iraqi Kurdistan Region is a mountainous area with relief difference ranging from few hundred meters up to 3000 m, and locally more. Almost all of the mountains form anticlines that have NW–SE trend changing westwards of longitude to E–W (25). The mountains are composed mainly of different kinds of limestone's except the peaks of the mountains on the Iranian borders that mainly consists of metamorphic and igneous rocks (3). A few basins at the lower part of the study area fall in the semiarid class (0.2 < AI < 0.5), while the remaining basins fall in the dry subhumid class (class (0.5 < AI < 0.65) according to the aridity index (AI) proposed (28). Additionally, based on the annual and monthly averages of temperature and precipitation for the study area, with no exception, the study basins fall in Csa climatic class according to the scheme proposed by Koppen. The study region is characterized by having a wide range of annual precipitation with a unimodal distribution. It ranges from about 400 mm at its lower part to more than 1000 mm at the borders. There is water surplus during the months of December to March. On the other hand, there is water deficit over the remaining months. The middle and upper parts of the study area can be generally described as rough broken and stony lands. These soils are either truncated or completely removed so that the diagnostic horizons of all orders other than Entisols are absent in most cases. The existing soils are variables due to variation in exposure, runoff, relief, parent materials, soil depth and maturity (6). The most common great groups on the sloppy lands are Rendolls and Xerorthents. On the other hand, Calcixerolls and Chromoxererts are the most abundant great groups over the plains intermountain valley. Considerable area were occupied by forest lands in the past, but at the present, the forest density ranges from treeless lands near the urban areas to very dense forestlands at remote places. The dominant forest tree species is oak trees. Dry farming is practiced on a large scale over the region. Wheat, barley, lentil, chick pea and faba been are the principal winter crops. Further, fallowing is also a common practice. Additionally many forestlands were converted to vine yards on steep slopes. Summer cropping can be observed over narrow strips of light to medium textured soils along the main rivers in the study area**.**

Measurement of watershed's attributes

The study area was partitioned into 30 main catchment delineations and each of these catchments were further divided into a group of sub-catchments depending on the nature of each main catchments using ArcMap ver. 10.0 Further, the same software was employed for determining different catchment attributes, including, area, perimeter, basin length total length of stream segments, slope. These databases were used for characterizing the basins in term of morphometric parameters, which fell in three categories, namely, linear, areal and relief using standard methodologies. The slope length was determined after subdividing each sub-catchment into component slopes, which simplified specification of the slope length perpendicular to the contours. The prerequisites for this task were identification of drainage networks, delimiting non-slope areas and classifying the sloping area into valley head, spur-end and valley side slopes (33); (36). The weighted average slope lengths were obtained for each basin according to the areas they represent.

Data processing

Before conducting regression analysis, the overall accuracy of the measured slope length by this method was assessed by ground truthing or spot checking the slope length at a number of sites or using measuring tape, range poles and theodolite. Linear and non-linear least squares techniques have been employed to determine the parameters of the proposed models in the present study for predicting slope length using Microsoft Excel Spread sheet and IBM SPSS software ver. 22. Following models derivation, the Shapiro-Wilk test was used to examine the normality of the residuals. Statgraphics software Release plus 4 was employed to detect multicolinearity among the regressors of proposed models. Microsoft Excel Spread sheet was also used to test the performance of the proposed model after calculating different model performance indicators. The sixfold cross-validation in which the data are divided into six groups, and one-sixth of the data were withheld at a time, was also used as the basis for model selection by using the same software. Additionally, this approach was validated at independent basins.

RESULTS AND DISCUSSION

General aspects about the delineated watersheds

The whole catchment was subdivided into 30 sub-basins (watersheds) for analysis with codes SW1 up to SW 30 on the basis of ridge lines, water divide, contours, and topographical variables. The drainage network map with 30 delineated subwatersheds is illustrated in (Figure1). Based on land use, most of the delineated watersheds fell in agricultural, natural forest and mountainous of high altitude classes where agricultural, forest cover and tourism dominate other uses. As a whole they are watersheds having 4th order streams covering areas varying from as slow as 1061.2 ha to as high as 34901.9 ha. Based on the classification scheme reported by Suresh (27), 70% of them fall in the Milliwatershed class (1000 – 10000 ha) (Table 1). Based on the classification scheme proposed by Walsh and Lawler (29) most of the stations at the lower part of the study area like Ankawa, Qushtapa and Khabat showed markedly seasonal with a long drier season (SI = 0.80 - 0.99). By contrast the stations within the mountainous area (Soran and Pirmam) showed rather seasonal (SI = 0.40 - 0.59) to seasonal (SI = 0.6 - 0.79) rainfall distribution.

Table 1. Some selected attributes of the watersheds within the study area

Watershed	CW	λ (km)	Rb (-)	A (km^2)	Dd (km^{-1})	Lb (km)	S (%)	Seg (km)	Re (-)	Rc (-)	Rf (-)
Balisan	SW1	0.368	4.094	64.050	1.476	10.609	28.778	94.545	0.851	0.416	0.122
Barbasty Chamrga	SW2	0.274	3.500	12.872	1.876	10.923	7.658	24.146	0.371	0.206	0.036
Bestana	SW3	0.329	3.574	30.892	1.578	6.730	16.573	48.759	0.932	0.568	0.074
Biza-agah	SW4	0.371	4.391	66.490	1.248	11.569	14.170	83.007	0.795	0.189	0.036
Bnaslawa	SW5	0.337	8.545	59.307	1.578	13.009	12.107	93.594	0.668	0.329	0.042
Darbandy-rayat	SW6	0.405	3.113	72.809	1.322	12.291	48.896	96.288	0.783	0.542	0.194
Dargalla	SW7	0.412	3.278	26.454	1.307	8.217	54.167	34.572	0.706	0.592	0.226
Dar-alsalam	SW8	0.399	3.500	12.926	1.319	6.259	43.694	17.052	0.648	0.496	0.225
Degala 1	SW9	0.363	4.311	64.444	1.252	7.571	24.598	80.669	1.196	0.563	0.084
Degala 2	SW10	0.360	5.315	127.658	2.006	13.951	14.009	256.091	0.914	0.574	0.055
Galala	SW11	0.337	3.367	27.536	1.160	7.580	47.578	31.938	0.781	0.670	0.239
Gomaspan	SW12	0.352	2.365	131.057	1.490	9.746	23.325	195.304	1.325	0.539	0.119
Grd-jutyar	SW13	0.294	4.045	55.254	2.311	13.739	7.599	127.689	0.610	0.312	0.049
Harir	SW14	0.382	4.486	349.019	1.624	25.403	18.243	566.926	0.830	0.366	0.057
Hujran	SW15	0.364	5.000	20.582	1.375	6.677	31.581	28.297	0.767	0.426	0.166
Kasnazan	SW16	0.334	2.833	20.428	1.436	7.000	17.398	29.338	0.728	0.471	0.061
Kawanyan	SW17	0.334	4.400	15.569	1.333	3.677	17.833	20.750	1.211	0.460	0.155
Kawlan-smelan	SW18	0.378	3.920	172.358	1.289	15.603	36.350	222.241	0.949	0.568	0.298
Kore	SW19	0.359	5.390	125.864	1.465	21.613	19.522	183.207	0.586	0.364	0.043
Koya	SW20	0.305	3.078	26.958	1.545	10.137	21.968	41.659	0.578	0.391	0.071
Mergasor	SW21	0.380	6.096	174.189	1.524	22.134	29.968	265.412	0.673	0.557	0.067
Nawande	SW22	0.377	3.778	28.842	1.319	11.190	45.986	51.234	0.541	0.393	0.217
Nawprdan	SW23	0.393	3.426	120.359	1.434	15.263	47.224	172.621	0.811	0.631	0.165
Prdi-qasre	SW24	0.355	3.894	43.969	1.330	8.616	39.413	58.476	0.868	0.438	0.296
Qapachyan	SW25	0.303	3.583	28.308	1.435	10.110	10.188	40.614	0.594	0.446	0.032
Rulka	SW26	0.339	3.833	27.072	0.763	6.919	13.534	20.662	0.848	0.732	0.048
Smaquli	SW27	0.323	3.974	126.040	1.554	11.645	23.292	195.831	1.088	0.423	0.065
Soran	SW28	0.355	6.500	31.248	1.424	8.462	30.567	44.559	0.745	0.599	0.032
Warte	SW29	0.437	3.444	333.459	1.226	29.928	46.709	408.832	0.688	0.333	0.095
Zarwan	SW30	0.349	3.625	10.612	1.080	8.462	55.031	11.457	0.434	0.603	0.134

CW = Code of watershed, λ = Slope length, Rb = Mean bifurcation ratio, A = Area, Dd = Drainage density, Lb = Basin length, S = Average slope, Seg = Total length of stream segments, Re = Elongation ratio, Rc = Circularity ratio and Rf = Relief ratio

The measured land slope on watershed scale ranged from a minimum of 0.274 km for WS2 to a maximum of 0.437 km for WS29 and those of the remaining watersheds fell between these two extremes. With a few exceptions the watershed slope length are (> 0.30 km). A high value of slope length means gentle slopes and long flow paths, more infiltration, and reduced runoff (21); (4).The slope length is synonymous with the length of sheet flow to a large degree or overland flow. The latter is related inversely to the average slope of the channel. Slope length being greater than 0.328 km (1000 ft) is an indication an undeveloped flow-path. According to Coates (5) other factors being constant, areas more advanced into maturity appear to contain smaller overland flow lengths than youthful areas because a drainage basin on an average develops maximum stream segments in its late youth and early mature stages and thus minimum (shorter) length of overland flow is found. About 57 % of the delineated watersheds were characterizing by having relief ratio of less than 0.1. Relatively low values of the watersheds (< 0.1) are suggesting gentle slope. The results also indicated that, 80% of the watersheds are characterized by having bifurcation ratio (Rb) of less than 5.0 it can be concluded that less than five may be classified into low, and more than five into high (4). Low class means the drainage pattern is not affected by the geologic structures whereas the high class means the drainage pattern is controlled by the geologic structures. As can be seen from Table 1, most of the watershed falls in the coarse drainage density class (Dd < 2.0). Low class of Dd shows a poorly drained basin with a slow hydrologic response. Surface runoff is not rapidly removed from the watershed making it highly susceptible to gully erosion (21). Judging from elongation ratio (Re), it was observed that the study area encompasses a wide range of shapes. Ranging from no-elongated (Re < 0.5) such as WS2 and WS30 to circular shape (Re >0.90) like SW3, SW8, SW9, SW11, SW15, SW16 and SW25. However, it was noticed that 70% of the values of this parameter are within the range of 0.5 - 0.9. This implies that most of them fall in the classes of less elongated (0.7- 0.8) and elongated (0.5-0.7). The Re values indicate elongated basin shape with high relief and gentle to steep slope (26). Judging from circularity ratio (Rc) and based on the scheme proposed by Miller (18), about 56% of the delineated watershed falls in the medium class (0.4 - 0.6). Only WS10, WS21, WS24 and WS30 have circular shape class (Rc > 0.6) and the rest fell in the non-circular shaped class. From the present study, it is clear that the area is susceptible to flooding to some extent (7). It is worth mentioning that the above mentioned characteristics can be considered as erosion risk assessment parameters and useful for watershed prioritization (14).

Sensitivity analysis

Prior to model building, Pearson's correlation and simple linear regression analyses were conducted and used as guides or simple sensitivity analysis to identify the influential factors affecting the overall slope length of the study watersheds. Table 2 displays the correlation matrix for the study variables. It is obvious from Table 2 that among the variables, only slope gradient and watershed area are positively and high significantly ($P \leq 0.01$) correlated with slope length, while each of perimeter, basin length and total segments length are positively and significantly ($P \leq 0.05$) correlated with slope length. On the other hand, the rest of the input variables offered insignificant correlation with slope length. It is also apparent from Table 2 that the slope gradient offered the strongest correlation. Therefore, it can be considered as the primary sensitivity parameter.

Table 2. Correlation matrix shown interrelationships between the study watershed attributes

Variable	λ (km)	Rb (-)	A (km^2)	P (km)	Dd (km^{-1})	L (km)	Seg (km)	S (%)	Rc (-)
Slope Length, λ (km)	1.000	-0.180	0.508	0.444	-0.396	0.407	0.433	0.675	0.210
Bifurcation ratio, Rb (-)		1.000	0.078	0.147	0.186	0.203	0.123	-0.283	-0.146
Area, A (km^2)			1.000	0.951	0.111	0.894	0.980	0.039	-0.152
Perimeter, P (km)				1.000	0.194	0.902	0.933	-0.088	-0.377
Drainage density, Dd (km^{-1})					1.000	0.208	0.247	-0.494	-0.443
Basin length, L (-)						1.000	0.872	0.032	-0.306
Total length of stream segments, Seg (km)							1.000	-0.045	-0.164
Average watershed slope, S (%)								1.000	0.416
Circularity ratio, Rc (-)									1.000

It is commendable to mention that in spite of poor correlation between slope length and bifurcation ratio, the latter was considered in building model 4 because it interacted differently with the other independent variables. This supports the findings of Willmott (32), who reported that some commonly used correlation measures such as Pearson's correlation coefficient and its square, (R^2) and test of statistical significance are often misleading. Furthermore, the area variable was treated with caution during the current study to avoid the detrimental effect of multicollinearity. Based on conducting sensitivity analysis, 4 variables were eliminated from 8 independent variables. The selection was based on the variable sensitivity and lack of multicollinearity problems. The variance inflation factor (VIF) for each of the selected variables of the proposed model (model 4) was less than 10.

Model calibration

The all possible cases algorithm was followed to specify which predictor variables were to be included in the regression equations. It was discerned that among the one variable model, model 1 offered the best performance in term of R^2. It was also noticed that among the two variable models, model 2 exhibits more acceptable results. Model 2 was constructed after inserting watershed perimeter as an additional variable, and the accuracy of prediction was considerably improved. The results also revealed that among three variable models, model 3 was the best, but the accuracy of prediction was slightly improved. The coefficient of determination increased from 0.712 to 0.742. To further improve the accuracy of prediction, four variable models were also tested to predict slope length. The results indicated that Model 4 exhibited the highest performance (Table 3). The findings also revealed that a step by step insertion of all the study variables did not give rise to a considerable improvement in the accuracy of prediction. The most complex model is not necessarily the most appropriate model overfitting occurs when too many variables are included in the model (39). It is also apparent from the results displayed in Table 3; the 4-variables regression model outperformed the remaining models of watershed slope length

Table 3. The models which exihibited the highest performance for predicting length of overland flow from watershed characteristics based on coefficient of determination

ModelID	MC	Regressors	PNC	Int	Slope								R^2
					Rb	A	P	Dd	Lb	Seg	S	Rc	
1	One - Variable	S	8	0.311							0.002		0.456
2	Two - Variable	P, S	28	0.276			0.001				0.002		0.712
3	Three-Variable	P, L, S	56	0.278			0.001		-0.003		0.002		0.742
4	Four-Variable	Rb, P, L, S	70	0.255	0.005		0.001		-0.003		0.002		0.77
5	Five-Variable	Rb, P, L, Seg, S	56	0.256	0.005		0.001		-0.003	1.35E-05	0.002		0.77
6	Six-Variable	Rb, A, P, Dd, L, S	56	0.288	0.005	5.70E-05	0.001	-0.017	-0.003		0.002		0.785
7	Seven-Variable	Rb, P, Dd, L,Seg, S, Rc	8	0.259	0.005		0.002	-0.013	-0.003	5.9 E-5	0.002	-0.001	0.791
8	Eight-Variable	Rb, A, P, Dd, L,Seg, S, Rc	1	0.287	0.005	0.00027	0.001	-0.016	-0.003	-1.9E-05	0.002	0.034	0.794

MC = Model classification, PNC = Possible number of cases, Int = Intercept, Rb = Bifurcation ratio, A= Area, P = perimeter, Dd = Drainage density, Lb = Baisn length, Seg = Total length of stream segments, S= Average watershed slope and Rc = Circularity ratio. This implies that the watershed slope length can largely be explained by four major factors: bifurcation ratio, watershed perimeter, basin length and average slope gradient. According to our findings, the slope gradient has emerged to be the most effective watershed characteristic on the overall slope length. The strong relationships in terms of R^2 values signify the importance and reliability of the obtained results through regression. The plot of the percent of bias from Models 1 through 8 versus the estimated values of slope length revealed that the residuals had no systematic distribution (Figure 2). This implies that these models are appropriate for estimating slope length. Additionally, the statistic for residuals from Shapiro-Wilk test (0.259) and from Kolmogorov-Smirnov test (0.117) proved that the residuals yielded by the indicated models are normally distributed.

Comparison between the measured and the predicted slope length Figure 3 shows a comparison between the measured slope length and the predicted values by the proposed models. The solid line is 1:1 line denoting the location where the observed and predicted values are the same. Overall, the observed-predicted plot Figure 3 shows a very limited scatter over the entire range of slope length. As it can be seen in this Figure, the predicted values were less dispersed from the observed data over the intermediate range of the data as compared to the lower and upper range of the predicted values.

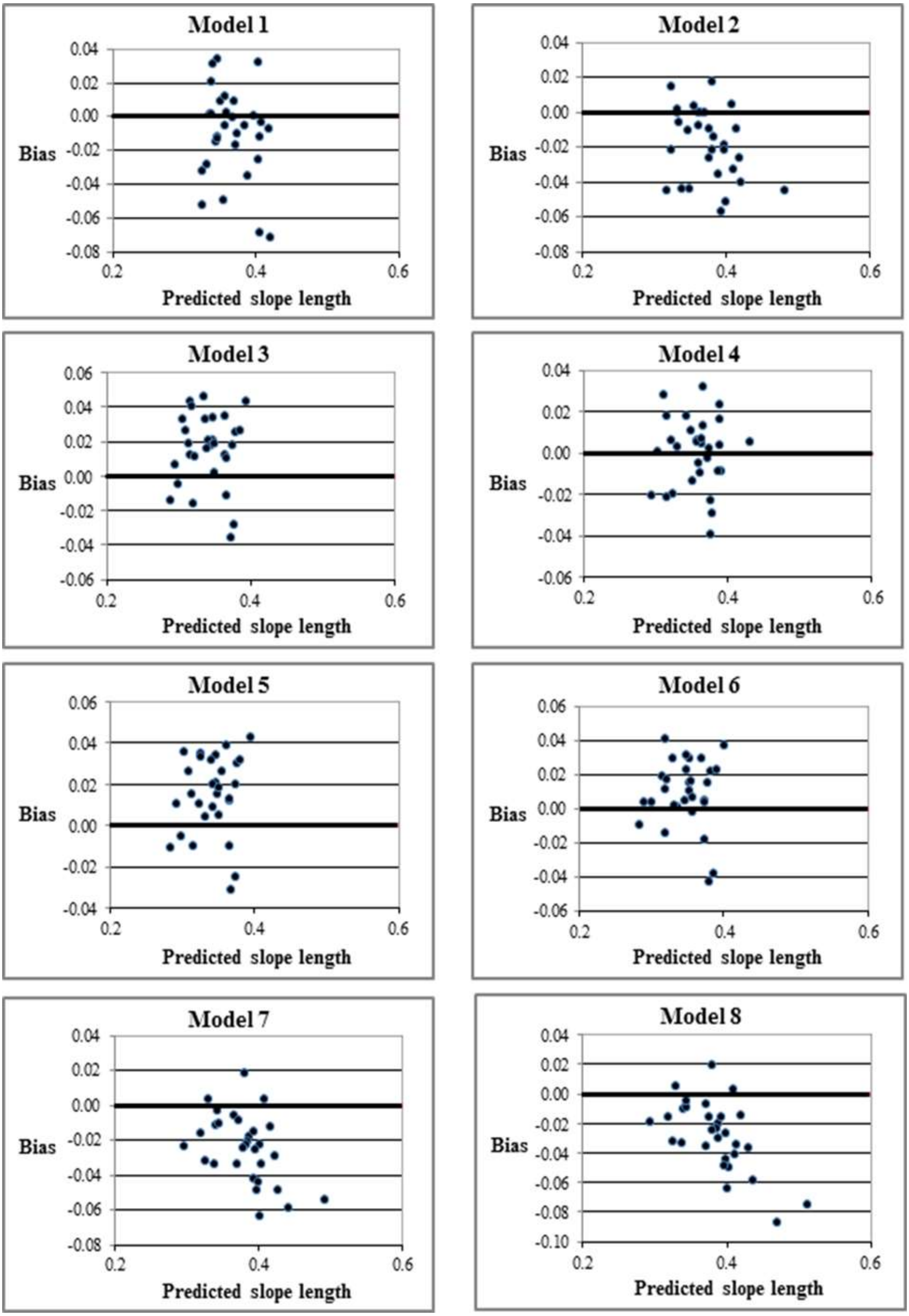

Figure 2. Plot of bias versus predicted slope length

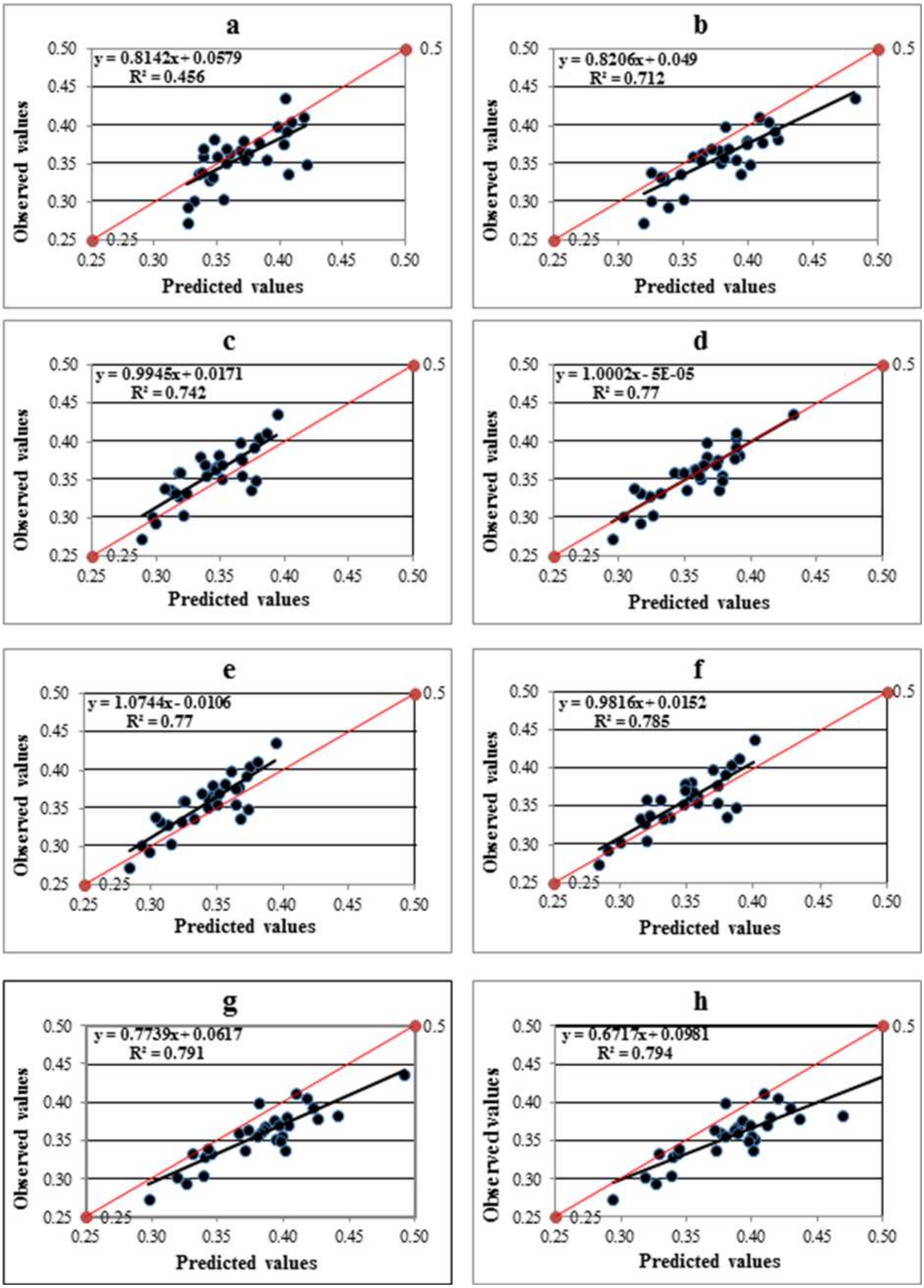

Figure 3. Plot of predicted slope length values verses observed values in relations to line 1:1 a=model 1, b=model 2, c=model 3, d=model 4, e=model 5, f=model 6, g=model 7, h=model 8

The proximity of the intercept of the relation from zero and closeness of their slopes from unity are indication of accurate prediction of overall watershed slope length. The fall of the majority of the data over the 1:1 line is an indication of the fact that the proposed model underestimated the slope length.

Evaluation of the models performance

To further confirm the results, different performance indicators of eight different modeling approaches were calculated and depicted in (Table 4). A steady increase in R^2 value can be detected with an increase in number of input variables. It is also evident from Table 4 that some indicators like the adjusted R^2 and d increased with an increase in number of input variable to a point (4 variables) beyond which it starts to decline with an increase in number of input variables. Contrarily, the reverse of this is true for the rest of the indicators. Along with d, the adjusted R^2 suggest that model 4 is calibrated well enough to simulate the slope length.

Table 4. The models which exhibited the highest performance for predicting slope length from watershed characteristics based on coefficient of determination

Model	Model code	R^2	R^2 adj	MBE	MAE	MAPE	CRM	RMSE	CV	d	AIC
One-Variable	1	0.456	0.44	-0.01	0.021	6.11	-0.034	0.028	7.95	0.78	-35.96
Two -Variable	2	0.712	0.69	-0.018	0.021	6.08	-0.053	0.027	7.6	0.84	-35.67
Three-Variable	3	0.742	0.71	-0.015	0.023	6.25	0.04	0.026	7.19	0.83	-35.63
Four-Variable	4	0.77	0.73	0.00	0.014	3.88	0.002	0.017	4.17	0.93	-45.12
Five-Variable	5	0.77	0.72	0.015	0.021	5.74	0.039	0.024	6.65	0.85	-34.63
Six-Variable	6	0.785	0.73	0.009	0.017	4.77	0.022	0.021	5.97	0.88	-35.84
Seven-Variable	7	0.791	0.72	-0.025	0.026	7.4	-0.069	0.031	8.73	0.82	-24.36
Eight-Variable	8	0.794	0.71	-0.028	0.03	8.44	-0.079	0.037	10.3	0.78	-18.45

MBE = mean biased error, MAE = mean absolute error, MAPE = mean absolute percentage error, CAM= coefficient of residual mass, RMSE= root mean square error, CV= coefficient of variability, d= index of agreement, AIC= akaike information criteria

Judging from the R^2 values, the 8-variable model the offered the highest performance. When a model contains an excessive number of independent variables and polynomial terms, it becomes overly customized to fit the peculiarities and random noise in samples rather than reflecting the entire population. Statisticians call this overfitting the model, and it produces deceptively high R-squared values and a decreased capability for precise predictions. Conversely, Model 4 scored best in term of all the performance indicators along with R^2. It is worth mentioning that some commonly used correlation measures such as Pearson's correlation coefficient and its square, (R^2) are often misleading when used to compare the predicted and observed values. Various measures seem to contain appropriate and insightful information (32). Both adjusted R^2 and the index of agreement suggest that model 4 is calibrated well enough to simulate the overall slope length. Judging from the Akaike information criteria (AIC) values, the four-variable model (model 4) offered the highest performance. The best model is one which minimizes the AIC. This index combines the sum squares of error and the number of model variables, implying that the best model combines the lowest sum squares of error (SSE) and the lowest number of model variables (1). Smaller MAE, MAPE and RMSE, values from a given approach indicate the closeness of the modeled values to the observed ones. The mean absolute percentage error (MAPE) is one of the most widely used measures of forecast accuracy, due to its advantages of scale-independency and interpretability (15). With no exception, all the models enlisted in Table 4 fell within the forecast potentially very good "based on (MAPE < 20%) (16). MBE describes the direction of the error bias. Its value, however, is related to magnitude of values under investigation. A negative MBE occurs when predictions are grater in value than observations, indicating overestimation Based on the classification scheme proposed by Wilding (31) the coefficient of variability of the predicted and observed land slope for all the candidate models are low (CV < 15%). Model 4 exhibited the lowest value for CV (4.17%) followed by model 6 and 5. The higher the CV, the greater the dispersion in the variable. The lower the CV, the smaller the residuals relative to the predicted values and is suggestive of a good model fit. Close inspection of Table 4 and judging from coefficient of residual mass (CRM) indicated that Model 4 slightly underestimated the overall slope length, while some models like model 1, 2, 7 and 8 overestimated the overall slope length for the study watersheds.

Model validation: As can be observed in Table 6 there is a slight fluctuation in coefficient of the variables and in the coefficient of determination and standard error of estimation. By judging from the values of standard error of estimates, it can be noticed the correspondence between the measured and predicted values still remain very good in spite of holding out. The maximum value of the standards of estimates is below 20 m. Additionally, the mean absolute percent of error for the testing data (not shown here) was blow 10%.

Table 5. Validation of the proposed model using K-fold method

Model	Skipped Fold	Intercept	Slope				Performance Indicator	
			Rb	P	L	S	R^2	Standard error of estimation
4	Fold 1	0.252	0.007	0.001	-0.003	0.002	0.724	0.019
	Fold 2	0.255	0.005	0.001	-0.003	0.002	0.844	0.0157
	Fold 3	0.25	0.005	0.002	-0.003	0.002	0.778	0.0197
	Fold 4	0.265	0.004	0.002	-0.004	0.002	0.772	0.0193
	Fold 5	0.245	0.006	0.002	-0.004	0.002	0.782	0.0187
	Fold 6	0.257	0.006	0.001	-0.002	0.002	0.745	0.0183
Average value for performance indicators							0.7742	0.0185

Close examination of Table 6 disclosed that the proposed models during the current study are powerful enough to capture the salient pattern of both testing and training sets. In other words, the models did not cause neither under fitting nor overfitting. The mean absolute percentage error (MAPE) was about 16%.

Table 6. Validation of some proposed models for predicting slope length by using a set of test data out of the training set

Watershed	Input variables				Slope length (km)		MAPE (%)
	Rb (-)	P (km)	L (km)	S (%)	Observed values	Estimated values	
Bira-jnah	3.50	25.75	7.17	11.92	0.386	0.301	22.13
Bnaslawa 2	3.33	25.47	8.50	6.36	0.34	0.284	16.37
Gomagru	3.48	31.15	9.16	10.45	0.368	0.297	19.30
Kawarta	3.50	11.34	1.81	7.49	0.308	0.293	4.74
Average Value							15.63

Regression analysis using non-classical techniques

It is evident from the displayed results of Table 7 that employing non-linear models did not enhance the predictive ability of the regression models. As a consequence, it is not recommended to apply nonlinear models for estimating slope length during the current study

Table 7. Some selected non-linear models for estimating slope length in the study region

Model	Input variables	Formula	R^2	Comments
1	Rb, P, L, S	$SL = 0.105+0.702\ Rb^{0.071}\ P^{0.188}\ L^{-0.068}\ S^{0.212}$	0.771	Very slight improvement
2	Rb, P, L, S	$SL= 0.149\ Rb^{0.047}\ P^{0.130}\ L^{-0.058}\ S^{0.147}$	0.770	No improvement

Multicollinearity analysis revealed that (no shown here) the values of the VIF for some of the explanatory variables for models 5 through 8 are > 10 and the tolerance < 0.1and these variables suffer from inflation in the variance of their parameters cause of the multicollinearity problem. The presence of multicollineearity leads to poor predictive power of the model and statistical inferences from such model might not to reliable (13). The results also indicated that the VIF values reduced from 11.01 to 0.741 for perimeter and from 8.11 to 0.739 for basin length through using the ridge regression. But, the accuracy of prediction was reduced in term of R^2 and MAPE (Table 8). It is also apparent from Table 9 a substantial reduction in the coefficients of the regressors.

Table 8. Test of performance of the proposed model for estimating the length of overland flow in the study region using ridge regression method

Model	Type of Analysis	Formula	R^2	Comments
4	Multiple linear regression	SL= 0.280 + 0.005 Rb - 0.001 P - 0.003 L+0.002 S	0.77	MAPE = 3.88
	Ridge Regression	SL= 0.280 + 0.00226 Rb - 0.00058 P +0.000075 L+0.00147 S	0.591	MAPE = 4.38; Ridge parameter= 0.20

An additional trial was also was made to moderate the problem of multicollinearity by reducing the number of the variables through principal component regression analysis. It was possible to shrink the number of the regressors from 8 to 2, but there is a decline the predictive ability of the model in term of R^2 and standard error of estimates (Table 9). Under this situation, no input variable was excluded. The two principal components explained 63% of variation in the overall watershed slope length

Table 9. Test of performance of the proposed model for estimating the length of overland flow in the study region using the study region Principal component regression method

Type of Analysis	Input variables	Formula	R^2	Std. error of estimates
PC regression	Rb, A, P, Dd, L, Seg, S, Rc	SL= 0.357+0.02 PC1 -0.021 PC2	0.632	0.022

REFERENCES

1. Akaike, H. 1974. A new look at the statistical model identification. In Selected Papers of Hirotugu Akaike. Springer, New York, NY. pp: (215-222).

2. Ali, H.Z. 2013. Using digital processing of satellite images for land-cover classification. Iraqi Journal of Agricultural Sciences, 43 (4), 119-128

3. Buringh, P. 1960. Soils and soil conditions in Iraq. Ministry of agriculture

4. Chandrashekar, H., K.V. Lokesh, M. Sameena and G. Ranganna. 2015. GIS–based morphometric analysis of two reservoir catchments of Arkavati River, Ramanagaram District, Karnataka. Aquatic Procedia, (4) 1345-1353

5. Coates, D.R. 1958. Quantitative geomorphology of small drainage basins of southern Indiana. Technical report (Columbia University. Department of Geology); no. 10

6. Foster, G.R. 2005. Revised Universal Soil Loss Equation, Version 2 (RUSLE2) Science Documentation (In Draft). USDA-ARS, Washington, DC

7. Gabale, S.M. and N.R. Pawar. 2015. Quantitative morphometric analysis of AmbilOdha (Rivulet) in Pune, Maharashtra, India. IOSR Journal of Environmental Science, Toxicology and Food Technology, 9 (7), 41-48

8. Haan, C.T., B.J. Barfield and J.C. Hayes. 1994. Design hydrology and sedimentology for small catchments. Elsevier. pp: 588

9. Hickey, R. 2000. Slope angle and slope length solutions for GIS. Cartography, 29 (1), 1-8

10. Hickey, R., A. Smith and P. Jankowski. 1994. Slope length calculations from a DEM within ARC/INFO GRID. Computers, environment and urban systems, 18(5), 365-380

11. Horton, R.E. 1945. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. Geological society of America bulletin, 56(3), 275-370

12. Hussein, K.S. 2016. Conservation planning for Bastora catchment based on detection of erosion risk prone areas. A thesis submitted to the college of Agriculture, University of Salahaddin as a partial fulfillment of the requirement of the degree of M.Sc in Soil and Water Science (Soil Conservation). Part four

13. Ibrahim, S. A. and W. B. Yahya. 2017. Effects of Outliers and Multicollinearity on Some Estimators of Linear Regression Model

14. Jawad, L. A. 2019. Utilizing integration of some remotely sensed morphometric aspects and hypsometric analyses to determine the geomorphological characteristics of Al-Abeadh Valley Watershed. Iraqi Journal of Agricultural Sciences, 50(1).

15. Kim, S. and H. Kim. 2016. A new metric of absolute percentage error for intermittent demand forecasts. International Journal of Forecasting, 32(3), 669-679

16. Lewis, C. 1997. Demand Forecasting and Inventory Control. The institute of Operation Researches. Woodhead Publishing Limited. pp: 391
17. Mashimbye, Z.E., W.P. de Clercq and A. Van Niekerk. 2014. An evaluation of digital elevation models (DEMs) for delineating land components. *Geoderma*, *213*, 312-319
18. Miller, A. 1953. Skin of the Earth, Methuen & Co. Ltd., London. pp: 54
19. Panagos, P., P. Borrelli and K. Meusburger. 2015. A new European slope length and steepness factor (LS-Factor) for modeling soil erosion by water. *Geosciences*, *5*(2), 117-126
20. Phuong, T.T., R.P. Shrestha and H.V. Chuong. 2017. Simulation of soil erosion risk in the upstream area of Bo River watershed. In Redefining Diversity & Dynamics of Natural Resources Management in Asia, Volume 3 (87-99). Elsevier
21. Rai, P.K., V.N. Mishra and K. Mohan. 2017. A study of morphometric evaluation of the Son basin, India using geospatial approach. Remote Sensing Applications: Society and Environment, 7, 9-20
22. Reimers, W. 1990. Estimating hydrological parameters from basin characteristics for large semiarid catchments. IAHS Publication. 191, 187-194
23. Renard, K.G., D.C. Yoder, D.T. Lightle and S.M. Dabney. 2011. Universal soil loss equation and revised universal soil loss equation. Handbook of erosion modelling, 8, pp: 135-167
24. Roose, Eric. 1996. "Land husbandry: components and strategy." FAO soils bulletin 70, pp: 313
25. Sissakian, V., A. Shihab and A. Othman. 2017. Factors controlling the development of straight valleys and streams in the Kurdistan Region, North and Northeast of Iraq. ARO-The Scientific Journal of Koya University, 5(2), pp: 32-48
26. Strahler, A. N. 1964. Quantitative geomorphology of drainage basins and channel networks. In V. T. Chow (Ed.) New York: McGraw Hill., Handbook of Applied Hydrology. 4, 39-4 and 76
27. Suresh, R. 2007. Soil and water conservation engineering. Standard Publishers and Distributors, Delhi, 799–812
28. UNESCO, 1979. Aridity definition (UN documents) United Nation Educational Scientific and Cultural Organization, New York. pp: 12
29. Walsh, R.P.D. and D.M. Lawler. 1981. Rainfall seasonality: description, spatial patterns and change through time. Weather, 36(7), 201-208
30. Wheib, K.A. 2013. Spectral reflectance properties of soil surface and land covers of AL-Salman depression in southern Iraq. Iraqi Journal of Agricultural Science, 43(4), 129-140
31. Wilding, L.P. 1985. Spatial variability: its documentation, accommodation and implication to soil surveys. In Soil spatial variability. Workshop (166-194).
32. Willmott, C.J. 1982. Some comments on the evaluation of model performance. Bulletin of the American Meteorological Society, 63(11), 1309-1313
33. Wilson, J.P. 1986. Estimating the topographic factor in the universal soil loss equation for watersheds. Journal of Soil and Water Conservation, 41(3), 179-184
34. Woreka, B.B. 2004. Evaluation of soil erosion in the Hararge region of Ethiopia using soil loss models, rainfall simulation and field trials. PhD Dissertation submitted to Faculty of Natural and Agriculture Sciences, University of Pretoria.
35. Yang, X. 2015. Digital mapping of RUSLE slope length and steepness factor across New South Wales, Australia. *Soil Research*, *53*(2), 216-225
36. Young, A. 2000. Land resources: now and for the future. Cambridge University Press
37. Zhang, H., J. Wei, Q. Yang, J.E. Baartman, L. Gai, X. Yang, S. Li, J. Yu, C.J. Ritsema and V. Geissen. 2017. An improved method for calculating slope length (λ) and the LS parameters of the Revised Universal Soil Loss Equation for large watersheds. *Geoderma*, *308*, 36-45
38. Zhang, R. J. 1989. "Sediment dynamics in rivers", Water Resources Press
39. Zhang, Z. 2014. Too much covariates in a multivariable model may cause the problem of overfitting. Journal of thoracic disease, 6(9), 196.