

Research Article

Evolutionary Prediction of Soil Loss from Observed Rainstorm Parameters in an Erosion Watershed Using Genetic Programming

Kennedy C. Onyelowe ¹, Ahmed M. Ebid ², and Light Nwobia ³

¹Department of Civil and Mechanical Engineering, Kampala International University, Kampala, Uganda

²Department of Structural Engineering, Faculty of Engineering and Technology, Future University, New Cairo, Egypt

³Department of Civil Engineering, Michael Okpara University of Agriculture, Umudike, Nigeria

Correspondence should be addressed to Kennedy C. Onyelowe; kennedychibuzor@kiu.ac.ug

Received 11 May 2021; Revised 4 September 2021; Accepted 6 September 2021; Published 21 September 2021

Academic Editor: Durgesh Jaiswal

Copyright © 2021 Kennedy C. Onyelowe et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Various environmental problems such as soil degradation and landform evolutions are initiated by a natural process known as soil erosion. Aggregated soil surfaces are dispersed through the impact of raindrop and its associated parameters, which were considered in this present work as function of soil loss. In an attempt to monitor environmental degradation due to the impact of raindrop and its associated factors, this work has employed the learning abilities of genetic programming (GP) to predict soil loss deploying rainfall amount, kinetic energy, rainfall intensity, gully head advance, soil detachment, factored soil detachment, runoff, and runoff rate database collected over a three-year period as predictors. Three evolutionary trials were executed, and three models were presented considering different permutations of the predictors. The performance evaluation of the three models showed that trial 3 with the highest parametric permutation, i.e., that included the influence of all the studied parameters showed the least error of 0.1 and the maximum coefficient of determination (R^2) of 0.97 and as such is the most efficient, robust, and applicable GP model to predict the soil loss value.

1. Introduction

1.1. Soil Loss and Influencing Rainstorm Parameters. As part of the earth with the utmost organic matter content, topsoil plays a significant role in farming and other fertile activities. Soil erosion, a naturally occurring process where topsoil is being removed by the actions of rain or wind, is a serious environmental hazard that reduces soil fertility as well as the productivity of agricultural land [1–5]. Frankly, a momentous percentage of global land is permanently being lost to water erosion through runoff and to wind. In fact, there is a strong unresolved interconnectedness among precipitation, vegetation, and erosion [6]. Soil loss has also been reported to contribute to environmental degradation via air and water pollution, nutrient loss, and loss of soil organic matter or biota [1, 7]. Among the factors that constitute soil formation such as topography, original materials, organisms, and time, which also determine the rate at which soil loss occurs, climatic or hydrometeorological factors such as rainstorm

parameters play the most critical role [8]. Some authors have described soil temperature and soil loss as functions of meteorological variables which are also positively affected by hydrological factors such as rainfall parameters like rainfall intensity, kinetic energy, runoff, and many more [9, 10]. For instance, Kinnell [10] had earlier reported that the difference between rainfall intensity and the average infiltration rate could be used to estimate variations in the efficiency of the use of rainfall energy in causing sheet erosion. While investigating the connection of amount and kinetic energy of rainstorm with that of soil influence, Kinnell [10] developed an equation that modelled rainfall kinetic energy as a function of rainfall intensity, in which many researchers have always attributed to be critical factors of soil loss actions [11].

Moreover, another significant environmental threat many locations face is the continuous increase in the gully head or elevation as more and more dissection of the landscape is triggered by incessant soil erosion. Due to its

intricacies and criticality in exacerbating soil loss, gully head advancement has been modelled using various field studies, aerial photography, GIS analysis, and multiple regression, and the results all confirmed the spatial and temporal variation of gully longitudinal development which reveals the effects of the runoff and waterfall process on soil erosion [12, 13]. Another common occurrence within the environment is when coarse and medium sand size material get reduced to either larger or smaller particle sizes [14]. During this phenomenon of particle detachment either in the factored or unfactored form, high hydration energies from raindrops induce collision of particles, followed by a splash of drops of the fluidized soil, and then a disintegration of the fluidized soil by an overland flow [14, 15]. All these parameters affecting soil erosion have not been perfectly measured, and hence, the use of forecasting or predictions techniques has proven to be useful.

At present, various modelling techniques have emerged such as field studies, aerial photography, GIS analysis, remote sensing, and multiple regressions, which help to estimate soil loss under a wide range of climatic and land use conditions [12, 16–19]. There has also been a massive decrease in emphasis on erosion research dwelling strictly on empirically based models, such as the Universal Soil Loss Equation, due to the emergence of various evolutionary computation techniques, otherwise, known as machine learning methods that facilitate easy, accurate, and robust computations and predictions. Not only does the machine learning approach predicts outputs in a more precise form but also have the capability of inferring a decision boundary that separates the input data space into two distinctive regions of erosion and nonerosion segments, thereby, making soil loss prediction worthwhile [20]. As systems imitate the human brain, machine learning techniques have been applied in various areas of engineering and even beyond and are useful in making predictions, performing clustering, extracting association rules, or making decisions from a given dataset [21–23].

1.2. Genetic Programming (GP) and Potential for Soil Loss Prediction. Genetic programming (GP) is one of the data-driven evolutionary computation methods that explores a program space rather than just searching it in a process setting as shown in Figure 1, which is a unique advantage it claims over the genetic algorithm [24, 25]. GP performs this exploration effectively by selecting datasets accordingly, fitting, and introducing genetic variation via some genetic operators [26, 27]. GP has the potential of changing their sizes, shapes, and composition, much like a living organism, to learn and report findings made from inputted datasets [21, 28, 29].

As a model that deals with complex adaptive systems via its tree-structure modelling mechanism (Figure 2), gene expression programming is a new, popular evolutionary technique that produces modelling equations in addition to its robust prediction configuration [31–33]. GP has been applied to many studies in environmental, water resources, structural engineering, geotechnical engineering, and beyond [26, 27, 31, 33–35].

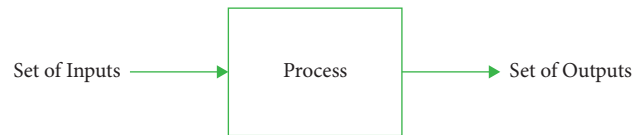


FIGURE 1: Genetic programming simplified activity flow for inputs, process, and outputs.

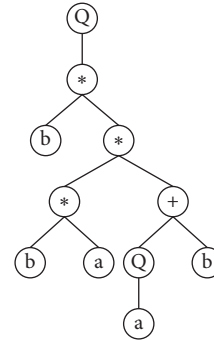


FIGURE 2: Typical tree structure of GP with varied string codes to allow for unaltered adaptation and evolution [30].

As earlier inferred, extensive care must be taken to ensure that the various constraints affecting soil erosion are all considered when predicting soil loss from the daily temporal and spatial variations of the concerned rainstorm parameters occasioned by different land use and climatic factors. Having being offered the amazing advantage of the GP technique as a data-driven tool that does not only successfully study and analyze all parameters of interest but also generates prediction equations and models based on lessons learnt from the input datasets. Use of GP in predicting soil loss from rainstorm parameters promises to aid in developing new algorithms or evaluate the existing ones for sustainable adjustment of soil parameters against continuous soil erosion actions [36, 37]. The present study aims at utilizing the GP approach in predicting soil loss from various rainstorm parameters, particularly rainfall amount (mm), rainfall intensity (mm/h), kinetic energy ($J/m^2/mm$), gully head advance (m), unfactored and factored soil detachments (g/m^2), runoff rate (m^3/s), runoff ($cm^3 \times 10^3$), and soil loss (kg/m^2), all having spatial and temporal variability modelled for the day-by-day period of 1993, 1995, and 1996.

2. Methodology

2.1. Data Collection and Tabulation. The data collection method adopted in this work was a literature collation of observed rainfall data from previous research works spanning for a period of three years in a watershed measuring about $35 m^2$ as presented in Table 1 [38] and graphically in Figures 3–11. The graphs represent the observed and estimated values of rainfall amount, kinetic energy, rainfall intensity, gully head advance, soil detachment, factored soil detachment, runoff, runoff rate, and soil loss, respectively, collated from the erosion watershed under the study within the three-year period.

TABLE 1: Measurement and computation of rainstorm parameter day by day for 1993, 1995, and 1996 seasons over bare soil [38].

Date	Rainfall amount (P) (mm)	Kinetic energy (E_K) ($J/m^2/mm$)	Rainfall intensity (I) (mm/hr)	Gully head advance (E) (m)	Soil detachment (D) (g/m^2)	Factored soil detachment (D_f) (g/m^2)	Runoff (cm^3) $\times 10^3$ (R)	Runoff rate (R_r) (m^3/s)	Soil loss (L) (kg/m^2)
10/7/93	13.5	30.36	25.23	0.0453382	9.46	9.56	32.75	0.0019	0.036
19/7/93	15.6	32.49	61.2	0.0466684	17.16	17.35	33	0.0019	0.013
20/7/93	32.1	37.66	43.6	0.0539136	18.27	18.47	39.5	0.0023	0.056
27/7/93	8.5	23.71	7.11	0.0413315	5.90	5.96	16.5	0.0009	0.048
3/8/93	22.3	35.39	26.38	0.0501255	13.64	13.79	43.5	0.0026	0.057
8/8/93	17.7	33.32	34.9	0.0478622	11.29	11.41	41.85	0.0025	0.045
12/8/93	12.3	29.19	48.4	0.0445019	19.22	19.43	31.95	0.0018	0.05
18/8/93	37.7	38.81	43.6	0.0556757	13.15	13.29	48.5	0.0028	0.104
21/8/93	19.7	34.33	48.26	0.0488980	18.56	18.76	42.15	0.0025	0.015
1/9/93	14.9	31.51	49.6	0.0462418	16.82	17.00	42.2	0.0025	0.029
6/9/93	12.7	29.60	49.6	0.0447877	5.91	5.97	32.95	0.0019	0.05
8/9/93	1.9	-11.91	1.91	0.0306301	3.11	3.14	—	—	—
10/9/93	16.8	32.60	66.8	0.0473652	6.64	6.71	32.25	0.0019	0.0085
17/9/93	6.9	19.98	6.7	0.0396431	7.92	8.01	21.2	0.0012	0.02
20/9/93	8.4	23.51	13.89	0.0412338	10.25	10.36	31.6	0.0019	0.042
23/9/93	2.3	-6.52	1.6	0.0318231	5.61	5.67	21.4	0.0013	0.01
5/7/95	13.2	30.0	9.0	0.0451349	8.22	8.38	53	0.0037	0.056
17/7/95	30.5	37.9	30.48	0.0533651	15.93	16.10	64.2	0.0045	0.05
19/7/95	1.7	-14.81	2.54	0.0299562	7.91	8.06	1	0.0007	0.0022
20/7/95	16.9	33.14	30.5	0.0474215	9.67	9.85	51	0.0036	0.001
23/7/95	6.8	19.94	16.9	0.0395275	16.9	17.22	15.5	0.0011	0.11
27/7/95	14.1	32.0	61.0	0.0457342	5.4	5.5	53	0.0037	0.05
30/7/95	13.0	30.17	55.9	0.0449973	7.3	7.44	55	0.0039	0.013
3/8/95	24.4	36.39	45.7	0.0510359	20.3	20.69	58	0.0041	0.06
14/8/95	16.6	32.96	53.4	0.0472519	10.51	10.71	53	0.0037	0.07
10/8/95	15.7	32.38	16.5	0.0467280	12.52	12.76	61	0.0043	0.04
18/8/95	22.0	35.57	30.5	0.0499899	13.93	14.19	63	0.0044	0.025
24/8/95	32.6	38.32	30.5	0.0540806	13.9	14.16	65	0.0046	0.096

TABLE 1: Continued.

Date	Rainfall amount (P) (mm)	Kinetic energy (E_K) ($J/m^2/mm$)	Rainfall intensity (I) (mm/hr)	Gully head advance (E) (m)	Soil detachment (D) (g/m^2)	Factored soil detachment (D_f) (g/m^2)	Runoff (cm^3) $\times 10^3$ (R)	Runoff rate (R_r) (m^3/s)	Soil loss (L) (kg/m^2)
26/8/95	13.1	30.28	9.3	0.0449973	18.6	18.95	51	0.0036	0.052
28/8/95	19.5	34.52	12.2	0.0487983	8.6	8.76	55	0.0039	0.015
7/9/95	2.6	-2.95	1.44	0.0326131	15.9	16.20	—	—	—
12/9/95	3.3	3.43	1.4	0.0342058	20.92	21.32	22	0.0015	0.0002
15/9/95	5.1	14.0	33.0	0.0373174	4.61	4.70	14	0.00098	0.027
16/9/95	18.8	34.20	58.40	0.0484428	9.1	9.30	55	0.0039	0.052
23/9/95	8.8	24.56	1.99	0.0416192	6.22	6.40	35	0.0025	0.045
24/9/95	6.6	19.37	30.5	0.0392922	6.93	7.06	42	0.0029	0.03
28/9/95	15.2	32.01	33.02	0.0464266	5.64	5.75	43	0.003	0.02
5/10/95	13.2	30.35	60.96	0.0451349	7.02	7.15	35	0.0025	0.04
8/10/95	14.7	31.64	61.00	0.0461170	3.67	3.74	38	0.0027	0.028
4/7/96	46.0	39.74	23.0	0.0579361	13.05	13.3	68.5	0.0022	0.05
7/7/96	15.0	31.58	18.0	0.0463037	5.32	5.42	43.0	0.003	0.04
14/7/96	16.0	32.29	24.0	0.0469053	7.61	7.76	29.0	0.002	0.02
16/7/96	12.0	28.86	16.0	0.0442827	32.14	32.75	62.5	0.0044	0.71
19/7/96	15.0	31.58	45.0	0.0463037	3.49	3.56	20.5	0.0014	0.023
20/7/96	40.0	39.11	48.98	0.0563391	11.08	11.29	52.0	0.0036	0.17
24/7/96	40.0	39.11	40.0	0.0563391	9.38	9.56	60.0	0.0042	0.073
25/7/96	35.0	38.42	48.84	0.0548544	8.51	8.67	60.0	0.0042	0.092
1/8/96	4.0	8.09	6.86	0.0355475	3.78	3.85	—	—	—
3/8/96	20.0	34.46	40.0	0.0490460	10.78	10.99	69.5	0.0049	0.1
9/8/96	23.0	35.63	24.64	0.0504363	19.16	19.53	39.5	0.0028	0.04
14/8/96	14.0	30.79	10.5	0.0456692	2.20	2.24	27.0	0.0019	0.02
16/8/96	43.0	39.45	41.62	0.0571599	8.79	8.96	64.0	0.0045	0.043
17/8/96	39.0	38.98	45.0	0.0560545	9.72	9.91	40.0	0.0028	0.054
20/8/96	16.0	32.29	16.27	0.0469053	8.22	8.38	59.5	0.0042	0.065
24/8/96	15.0	31.58	19.56	0.0463037	9.78	9.97	46.5	0.0033	0.056
26/8/96	7.0	20.25	6.46	0.0397573	1.96	2.00	12.5	0.0088	—
30/8/96	8.0	22.69	13.71	0.0408334	4.42	4.50	60.0	0.0042	0.01
31/8/96	10.0	26.29	8.57	0.0426970	3.13	3.19	3.20	0.0022	0.02

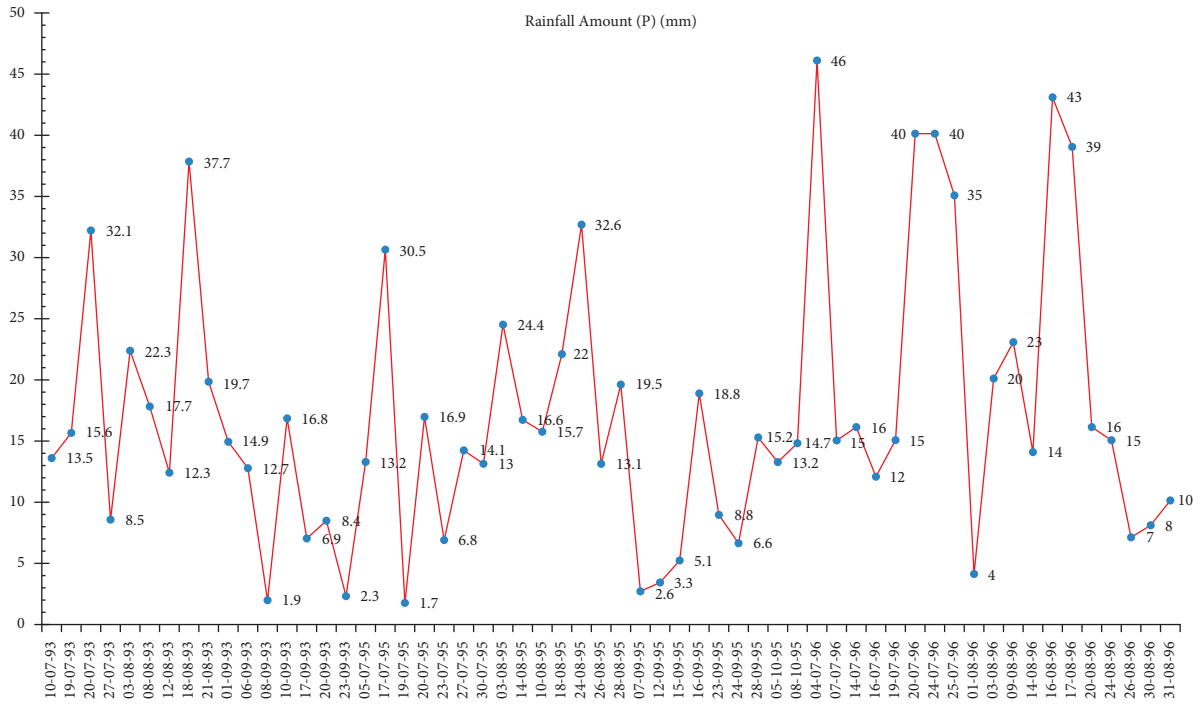


FIGURE 3: Rainfall amount with the selected dates of observation.

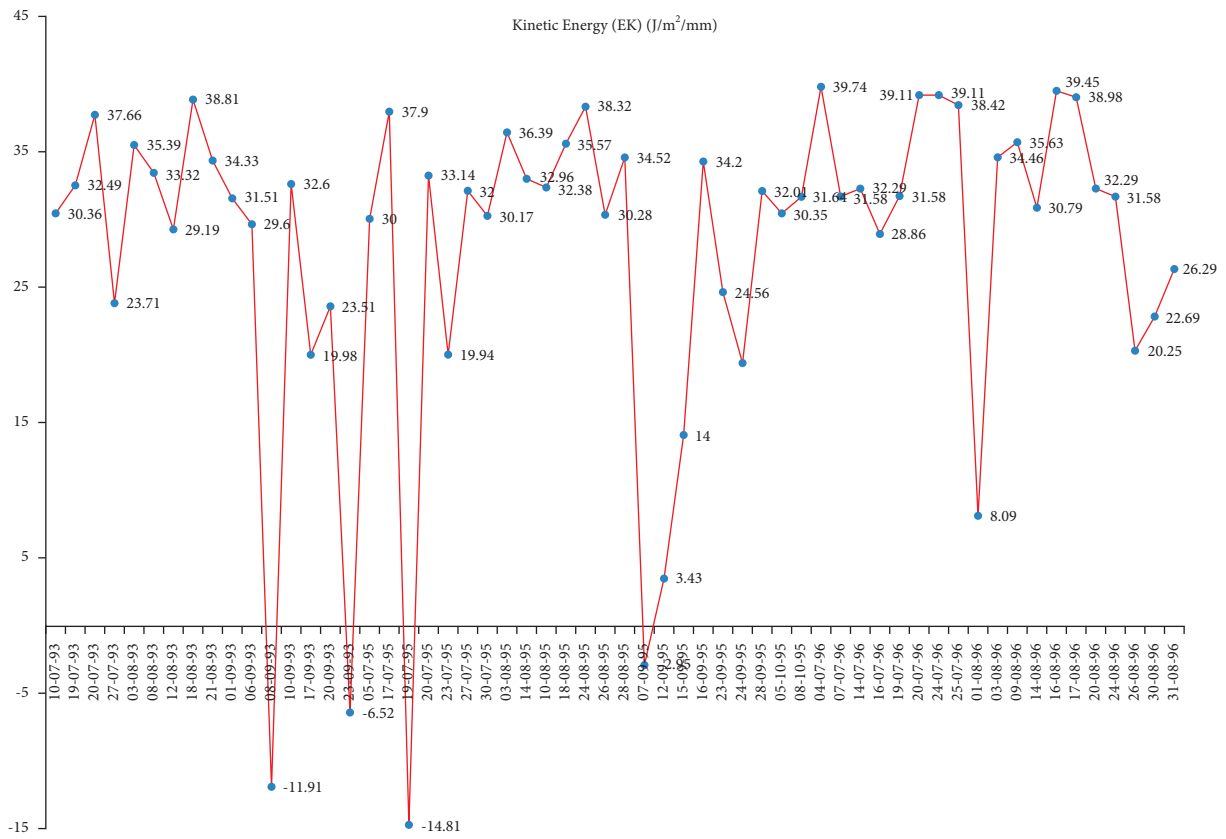


FIGURE 4: Average raindrop kinetic energy with the selected dates of observation.

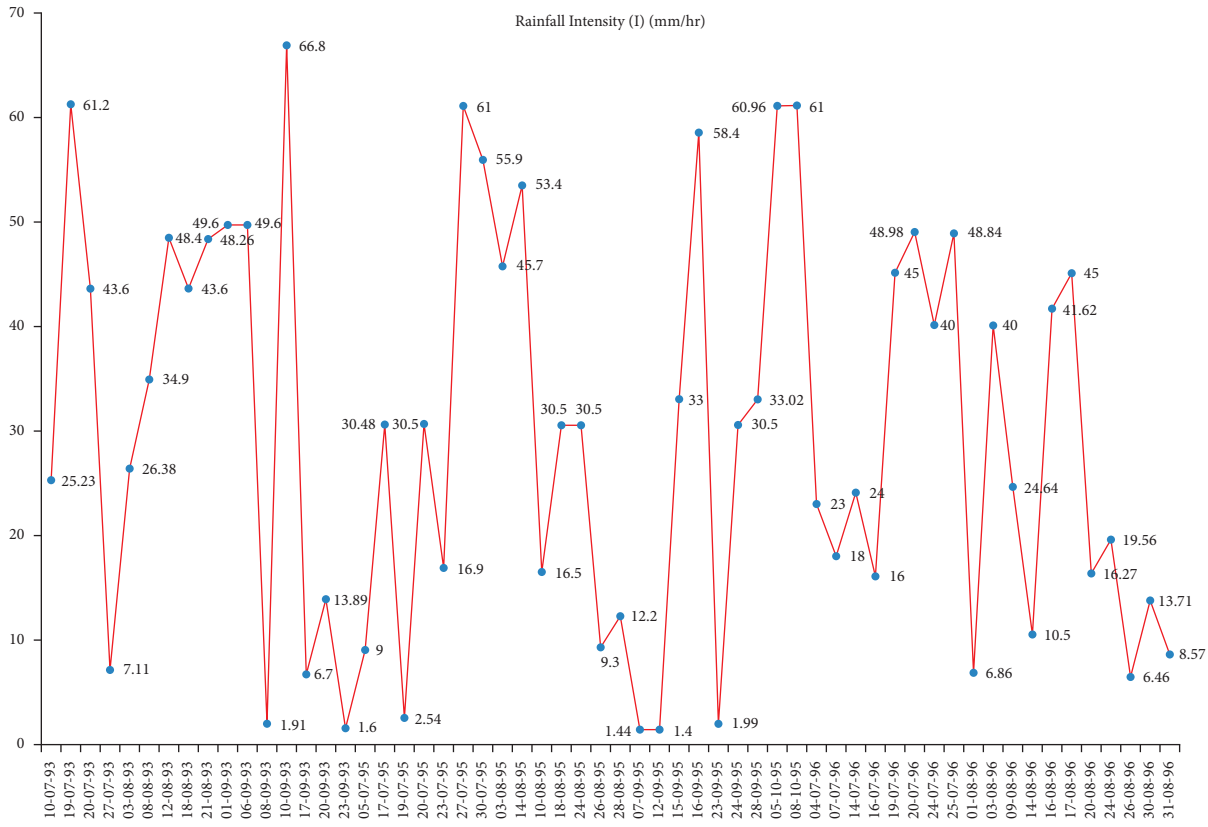


FIGURE 5: Observed rainfall intensity with the selected dates of observation.

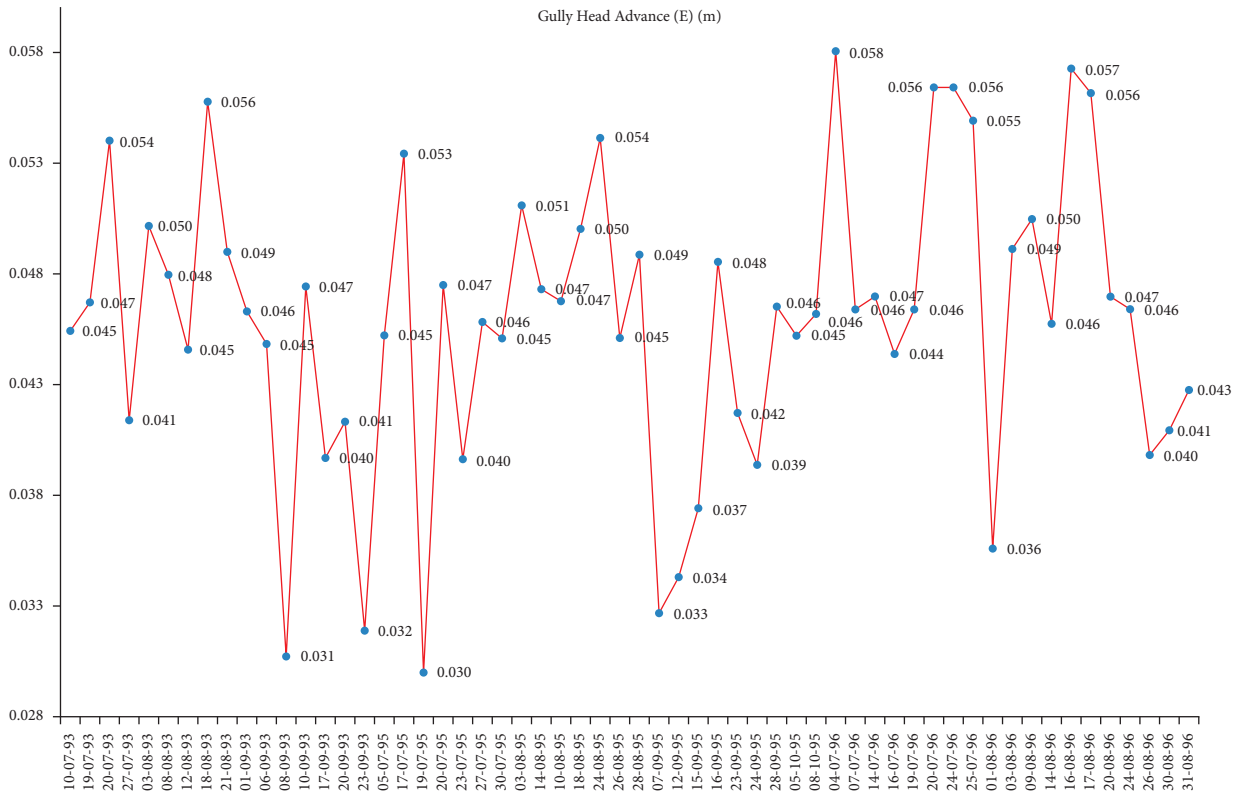


FIGURE 6: Estimated gully head advance with the selected dates of rainfall observation.

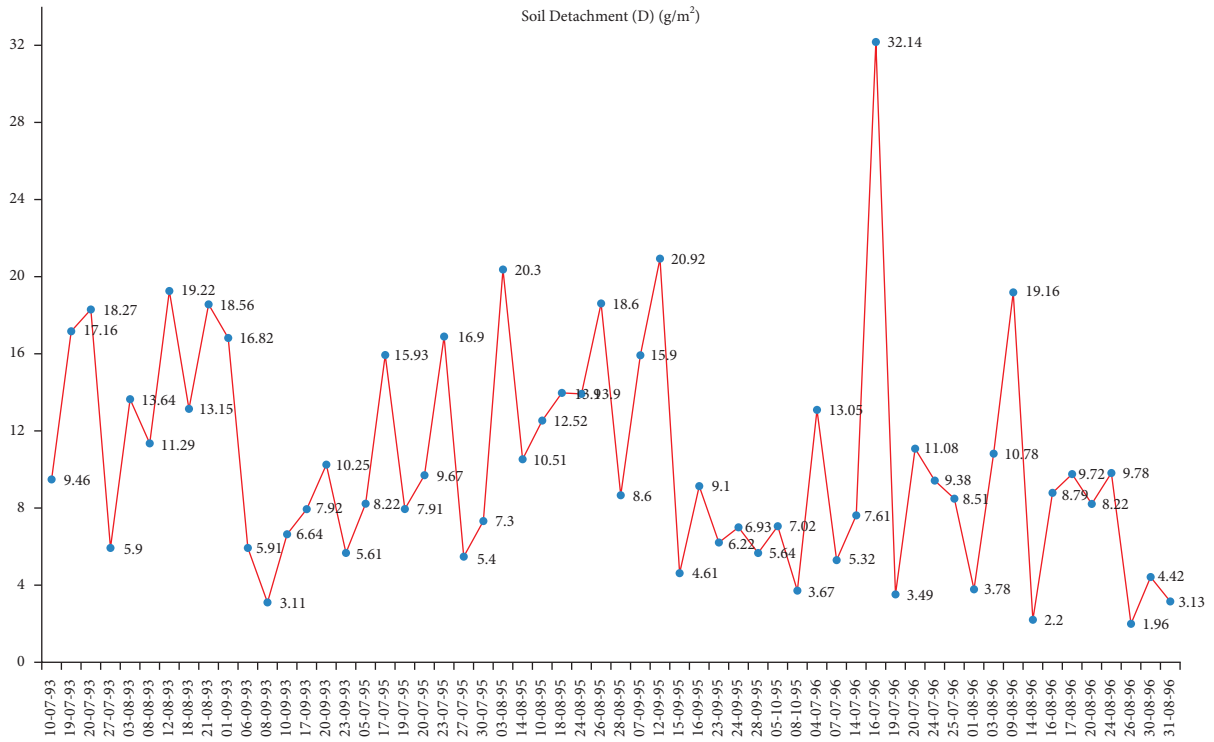


FIGURE 7: Soil detachment with the selected dates of observation.

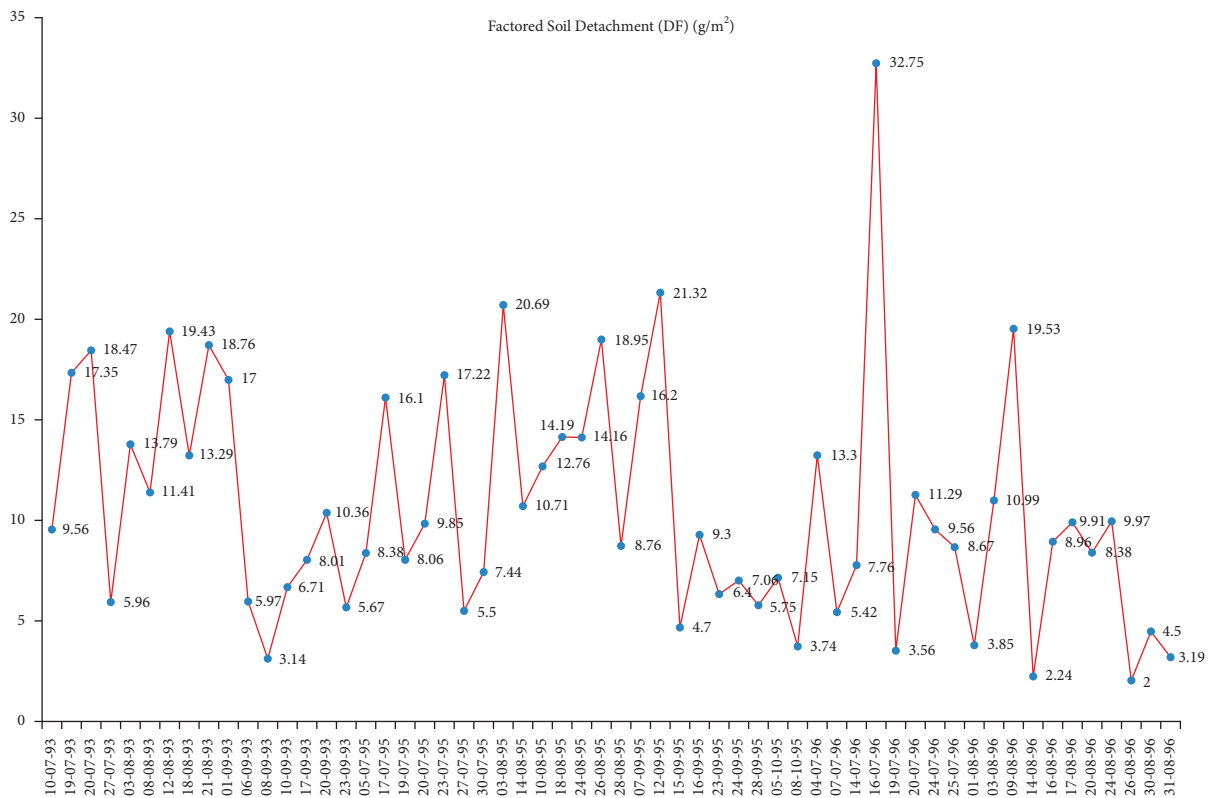


FIGURE 8: Factored soil detachment with the selected dates of observation.

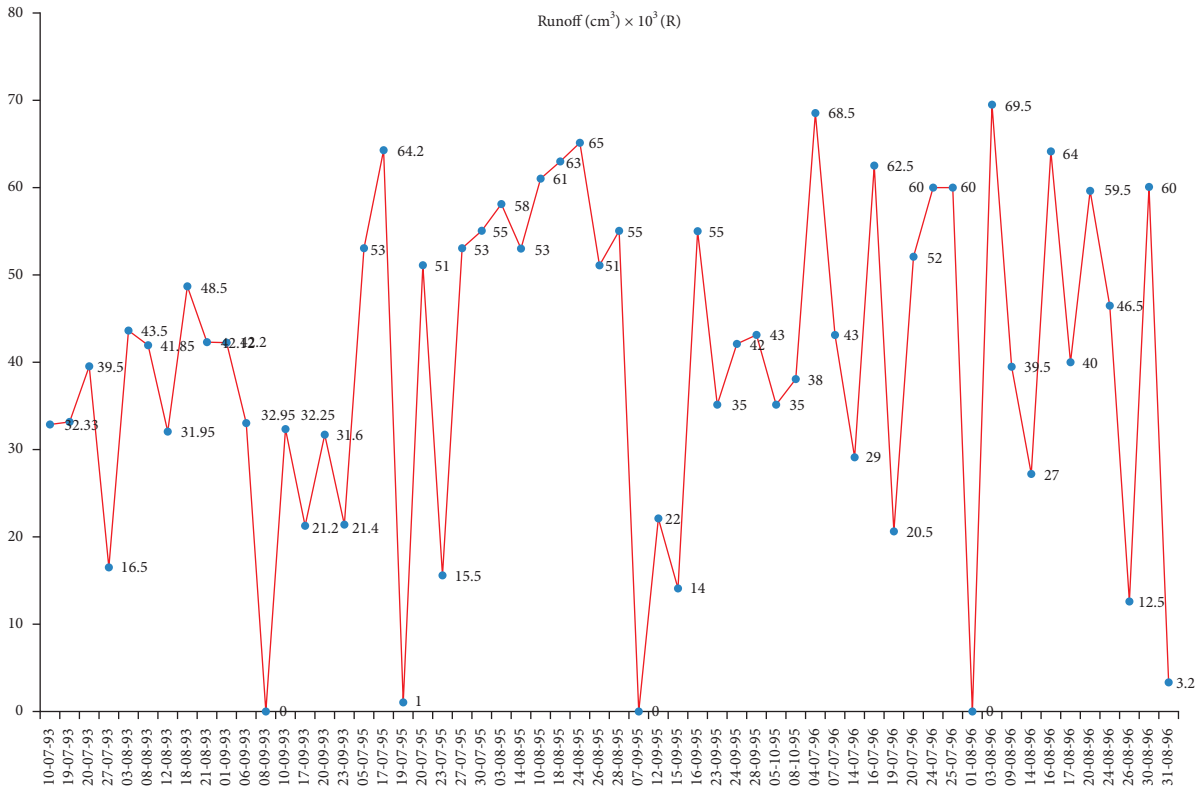


FIGURE 9: Observed runoff with the selected dates of observation.

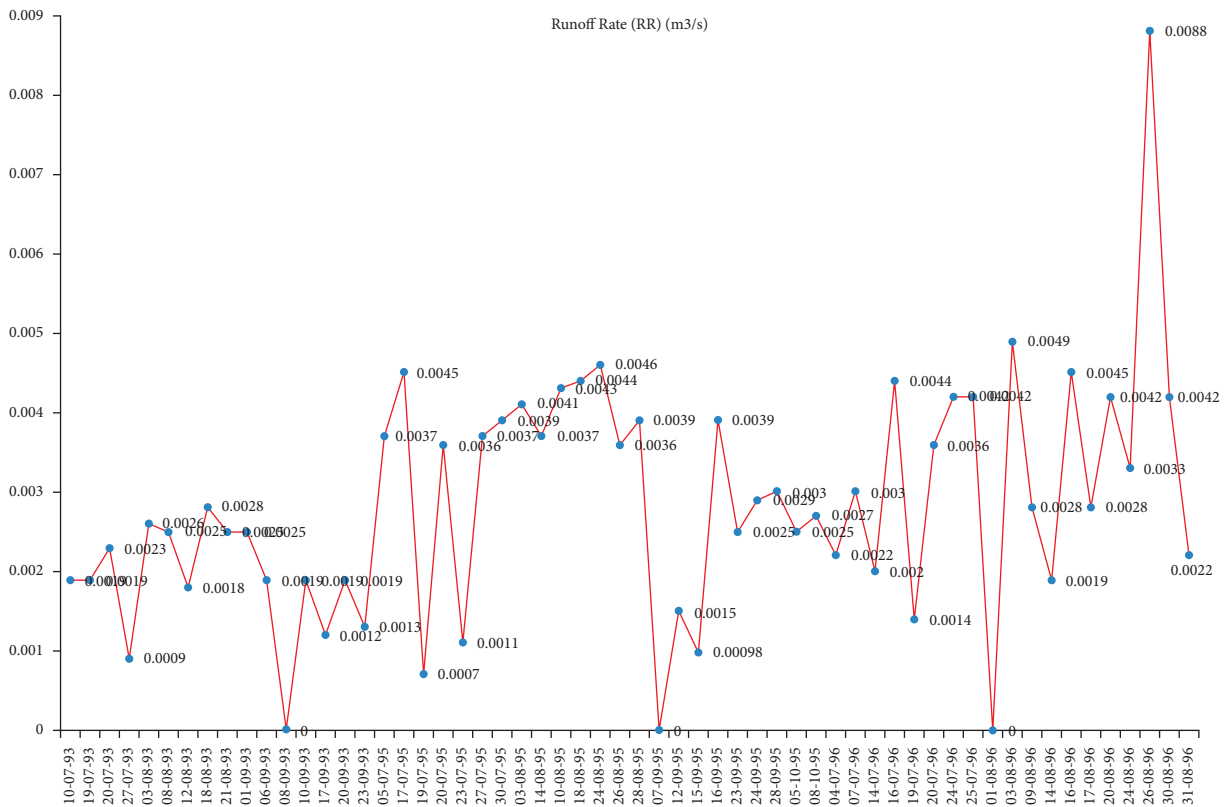


FIGURE 10: Observed runoff rate with the selected dates of observation.

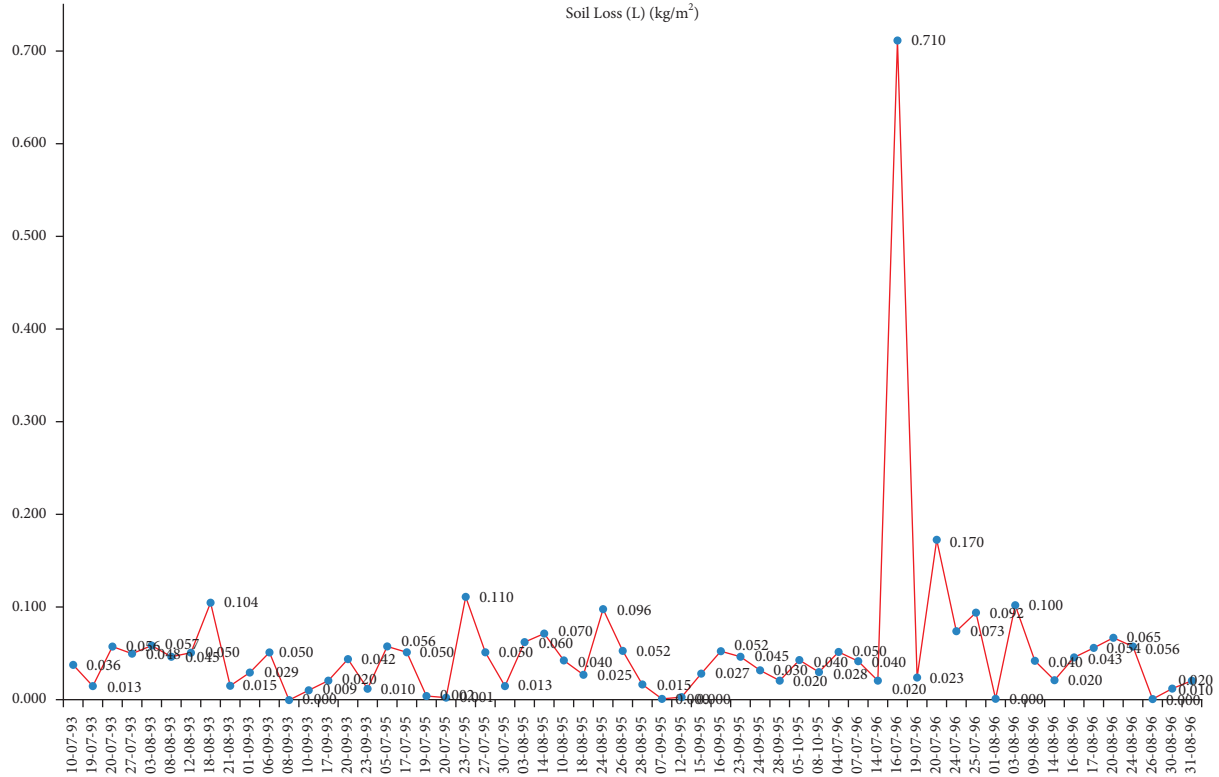


FIGURE 11: Observed soil loss with the selected dates of observation.

Furthermore, the gully head advance parameter was estimated by using the United States Soil Conservation Service simple parametric model equation estimating average gully head advance for erosion watersheds as presented in the following equation [39–41]:

$$E = 5.25 \times 10^{-3} A^{0.46} P^{0.20}, \quad (1)$$

where A is the drainage area above gully head in m^2 ($3.5 \times 10 = 35 m^2$) and $P=24$ hours rainfall in mm. The values of the literature data collation and parametric estimation of erosion parameter were tabulated and deployed to a prediction exercise using the gene expression programming learning technique.

The model functional relationship is presented in the following equation:

$$L = f(P, E_K, I, E, D, D_f, R, R_r). \quad (2)$$

From equation (2), it can be aptly deduced that the soil loss (L) is the dependent or target variable, while rainfall amount, kinetic energy, rainfall intensity, gully head advance, soil detachment, factored soil detachment, runoff, and runoff rate are the independent or predictor variables, which are functions of the function L .

Forty eight (48) records were collected to carry out this study. Each record contains the following data:

- (i) Rainfall amount (P) (mm)
- (ii) Kinetic energy (E_K) ($J/m^2/mm$)
- (iii) Rainfall intensity (I) (mm/h)
- (iv) Gully head advance (E) (m)

- (v) Soil detachment (D) (g/m^2)
- (vi) Factored soil detachment (D_f) (g/m^2)
- (vii) Runoff (litre) (R)
- (viii) Runoff rate (R_r) (m^3/s)
- (ix) Soil loss (L) (kg/m^2)

2.2. Statistical Analysis of Database. The collected records were divided into two sets. The first one is a training set that contains (32) records, while the second one is the validation set that contains (16) records. The complete dataset is given in Table 2. Statistical analysis for the utilized database is summarized in Table 3 to show the significant variability in the soil properties. It can be observed that the soil loss has the highest degree of deviation and variance compared to the other observed parameters with gully head advance showing the least degree of deviation and variance. However, the variation and variance of the parameters showed values less than 1.

3. Results and Discussion

3.1. Research Program. Three trials were carried out to correlate the soil loss value (L) to the corresponding field measurements. Each trial uses certain complexity level starting from 4 levels expression and up to 6 levels expression. Iterations are performed until achieving the minimum (SSR) of the set which reflects the most accurate expression at the considered level of complexity. Characteristics of each trial are as summarized in Table 4.

TABLE 2: Collected dataset.

Rainfall amount (P) mm	Kinetic energy (E_K) $J/m^2/mm$	Rainfall intensity (I) mm/hr	Gully head advance (E) m	Soil detachment (D) g/m^2	Factored soil detachment (D_f) g/m^2	Runoff (R) Litre	Runoff rate (R_r) m^3/s	Soil loss (L) kg/m^2
Training set								
13.5	30.4	25.2	45.3	9.5	9.6	32.8	6.8	36.0
15.6	32.5	61.2	46.7	17.2	17.4	33.0	6.8	13.0
32.1	37.7	43.6	53.9	18.3	18.5	39.5	8.3	56.0
8.5	23.7	7.1	41.3	5.9	6.0	16.5	3.2	48.0
22.3	35.4	26.4	50.1	13.6	13.8	43.5	9.4	57.0
17.7	33.3	34.9	47.9	11.3	11.4	41.9	9.0	45.0
12.3	29.2	48.4	44.5	19.2	19.4	32.0	6.5	50.0
37.7	38.8	43.6	55.7	13.2	13.3	48.5	10.1	104.0
19.7	34.3	48.3	48.9	18.6	18.8	42.2	9.0	15.0
14.9	31.5	49.6	46.3	16.8	17.0	42.2	9.0	29.0
12.7	29.6	49.6	44.8	5.9	6.0	33.0	6.8	50.0
16.8	32.6	66.8	47.4	6.6	6.7	32.3	6.8	8.5
6.9	20.0	6.7	39.7	7.9	8.0	21.2	4.3	20.0
8.4	23.5	13.9	41.2	10.3	10.4	31.6	6.8	42.0
13.2	30.0	9.0	45.1	8.2	8.4	53.0	13.3	56.0
30.5	37.9	30.5	53.4	15.9	16.1	64.2	16.2	50.0
6.8	19.9	16.9	39.5	16.9	17.2	15.5	4.0	110.0
14.1	32.0	61.0	45.7	5.4	5.5	53.0	13.3	50.0
13.0	30.2	55.9	45.0	7.3	7.4	55.0	14.0	13.0
24.4	36.4	45.7	51.0	20.3	20.7	58.0	14.8	60.0
16.6	33.0	53.4	47.3	10.5	10.7	53.0	13.3	70.0
15.7	32.4	16.5	46.7	12.5	12.8	61.0	15.5	40.0
22.0	35.6	30.5	50.0	13.9	14.2	63.0	15.8	25.0
32.6	38.3	30.5	54.1	13.9	14.2	65.0	16.6	96.0
13.1	30.3	9.3	45.0	18.6	19.0	51.0	13.0	52.0
19.5	34.5	12.2	48.8	8.6	8.8	55.0	14.0	15.0
5.1	14.0	33.0	37.3	4.6	4.7	14.0	3.5	27.0
18.8	34.2	58.4	48.5	9.1	9.3	55.0	14.0	52.0
8.8	24.6	2.0	41.6	6.2	6.4	35.0	9.0	45.0
6.6	19.4	30.5	39.3	6.9	7.1	42.0	10.4	30.0
15.2	32.0	33.0	46.4	5.6	5.8	43.0	10.8	20.0
13.2	30.4	61.0	45.1	7.0	7.2	35.0	9.0	40.0
Validation set								
14.7	31.6	61.0	46.1	3.7	3.7	38.0	9.7	28.0
46.0	39.7	23.0	57.9	13.1	13.3	68.5	7.9	50.0
15.0	31.6	18.0	46.3	5.3	5.4	43.0	10.8	40.0
16.0	32.3	24.0	46.9	7.6	7.8	29.0	7.2	20.0
15.0	31.6	45.0	46.3	3.5	3.6	20.5	5.0	23.0
40.0	39.1	40.0	56.3	9.4	9.6	60.0	15.1	73.0
35.0	38.4	48.8	54.9	8.5	8.7	60.0	15.1	92.0
20.0	34.5	40.0	49.1	10.8	11.0	69.5	17.6	100.0
23.0	35.6	24.6	50.4	19.2	19.5	39.5	10.1	40.0
14.0	30.8	10.5	45.7	2.2	2.2	27.0	6.8	20.0
43.0	39.5	41.6	57.2	8.8	9.0	64.0	16.2	43.0
39.0	39.0	45.0	56.1	9.7	9.9	40.0	10.1	54.0
16.0	32.3	16.3	46.9	8.2	8.4	59.5	15.1	65.0
15.0	31.6	19.6	46.3	9.8	10.0	46.5	11.9	56.0
8.0	22.7	13.7	40.8	4.4	4.5	60.0	15.1	10.0
10.0	26.3	8.6	42.7	3.1	3.2	3.2	7.9	20.0

The following paragraphs present and discuss the results of each trial. Results of all trials are summarized in Table 5. For all performed trials, the main target was to make a comparison between the predicted and the measured soil loss values and hence evaluate the accuracy of the developed expressions based on the statistical analysis.

3.2. Prediction of Soil Loss Value Using Genetic Programming (GP). As usual in the (GP) technique, the trials start with the lowest level of complexity and increased to improve the prediction accuracy. This is because the simple formulas allow for a limited number of variables; hence, the appeared variables are the most effective ones, while the complicated

TABLE 3: Statistical analysis of collected database.

	Rainfall amount (P) mm	Kinetic energy (E_K) J/m ² /mm	Rainfall intensity (I) mm/hr	Gully head advance (E) m	Soil detachment (D) g/m ²	Factored soil detachment (D_f) g/m ²	Runoff (R) Litre	Runoff rate (R_r) m ³ /s	Soil loss (L) kg/m ²
Training set									
Max.	5.1	14.0	2.0	37.3	4.6	4.7	14.0	3.2	8.5
Min.	37.7	38.8	66.8	55.7	20.3	20.7	65.0	16.6	110.0
Avg.	16.5	30.5	34.8	46.4	11.4	11.6	42.5	10.1	44.5
SD	7.9	6.0	18.8	4.5	4.8	4.9	14.0	3.9	24.7
Var.	0.5	0.2	0.5	0.1	0.4	0.4	0.3	0.4	0.6
Validation set									
Max.	8.0	22.7	8.6	40.8	2.2	2.2	3.2	5.0	10.0
Min.	46.0	39.7	61.0	57.9	19.2	19.5	69.5	17.6	100.0
Avg.	23.1	33.5	30.0	49.4	8.0	8.1	45.5	11.4	45.9
SD	12.4	4.7	15.3	5.2	4.2	4.3	18.4	3.8	25.7
Var.	0.5	0.1	0.5	0.1	0.5	0.5	0.4	0.3	0.6

TABLE 4: Characteristics of each (GEP) trial.

Trial no.	No. of levels	Used variables	Population size	Survivors size	No. of generations	Mutation present (%)
1	4		50,000	15000	500	
2	5	$P, E_K, I, E, D, D_f, R, R_r, 1, 3, 5, 7, 11, 13, 17, 19$	75,000	25000	750	5
3	6		100,000	30,000	1000	

TABLE 5: Summary for the performance accuracy of formulas generated using (GP).

Trial no.	No. of levels	Generated formula	Error %			R^2		
			Training	Validation	Total	Training	Validation	Total
1	4	Equation (1)	30	26	29	0.61	0.65	0.62
2	5	Equation (2)	25	21	24	0.76	0.82	0.79
3	6	Equation (3)	11	9	10	0.96	0.98	0.97

formulas allow for more less effective variables. And hence, the considered variables could be ranked according to their impact on the output.

3.2.1. GP Model Trial No. (1). Start with the simplest expression with four levels of complexity (chromosome length = 32 genes); the generated formula in this trial is shown in equation (3) and in Figure 12(a). The achieved error percent for training, validation, and total sets is (30%), (26%), and (29%), respectively, while the corresponding coefficient of determination (R^2) values of (0.61), (0.65), and (0.62), respectively. The archived (R^2) values indicate a fair correlation between the predicted and measured soil loss values.

$$L = \frac{2.1P - Df}{1.1} + \frac{(A2 + A3)^2}{A2.A3} + \frac{P - Df}{A3} + \frac{I}{Df - (D - 11)^2} - \frac{A3.A1}{A2.(A1^2 - 28)} + 12, \quad (3)$$

where

$$A1 = (17 - P + Df),$$

$$A2 = \frac{7.7 Df}{2.1P - Df}, \quad (4)$$

$$A3 = 2 A1 - \frac{126}{A1}.$$

3.2.2. GP Model Trial No. (2). In this trial, the complexity level is expanded to five levels (chromosome length = 64 genes). Equation (5) appears a good correlated relationship, whereas Figure 12(b) gives the comparison between the predicted and experimental values of soil loss (L). The achieved error percent for training, validation, and total sets is (25%), (21%), and (24%), respectively, and their corresponding (R^2) values are (0.76), (0.82), and (0.79). Comparing between the (R^2) values of equations (1) and (2) illustrates that the accuracy is significantly improved.

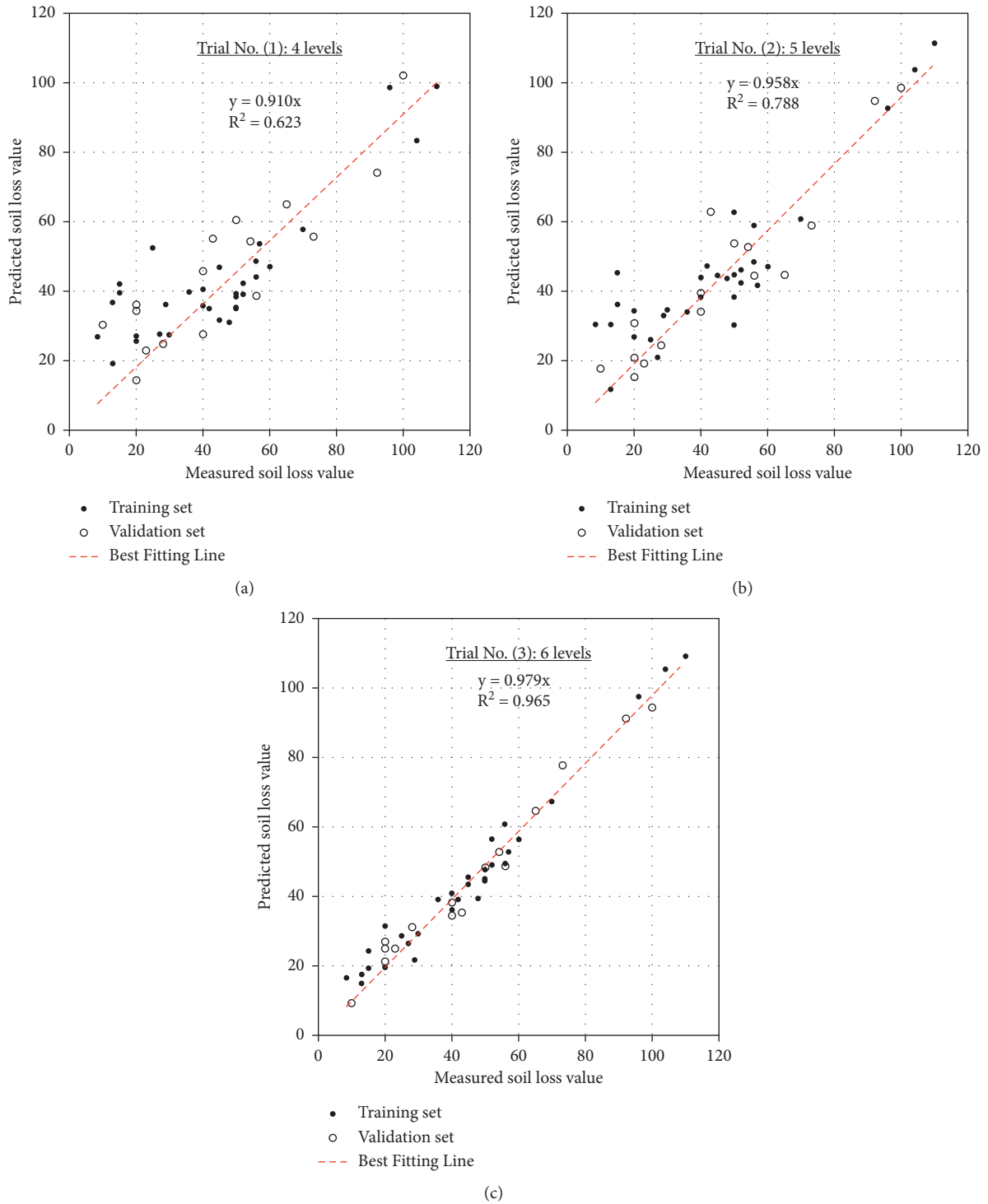


FIGURE 12: Relation between predicted and measured soil loss values using the different developed formulas.

$$L = \frac{A2 + A3 + Rr}{E(A1 + A3) - A4} + \frac{A1^2 - A2 + (Df/P)^4 + R + P - E - 11}{A4} + \left(\frac{Df}{P}\right)^4 + \frac{A5}{11 - D} + \frac{D.P}{Df} + A1^2 + A3 + A5 + Rr, \quad (5)$$

where

$$\begin{aligned}
A1 &= Ln(D) + \frac{1}{11 - D}, \\
A2 &= \frac{1}{(11 - D)(Df/P)^4}, \\
A3 &= \frac{11 - D}{1 - A2}, \\
A4 &= A2 + E + D - P, \\
A5 &= \frac{R - P}{P(A1 + A3)} - \frac{1}{A2 + 2} \\
&\quad + \frac{(Df + 1)(Df/P)^4 + A1^2 + R + D - 11}{A2 + 11}.
\end{aligned} \tag{6}$$

$$\begin{aligned}
L &= \frac{A7 - (02I/(A2 + 2.65))}{11.0A1 - A3} + \frac{P.A1 - ((A4 + A6)/(A2 - 2.0)) + A2 - Ek}{1 - Df + A2 + 2.65} + \frac{A5}{4.0 - Rr} - \frac{(Rr * (A4 + 51.0)/2.65A2) + A4 + A6}{A5(51.0 - 3.87Rr)} \\
&\quad + \frac{I}{19.0 - A2 * A3} + \frac{11.0A1 - A3}{A2 + 7.0} - \left(\frac{1}{A2}\right)^2 + \frac{A4}{P} + A7 + P + 22.0,
\end{aligned} \tag{7}$$

where

$$\begin{aligned}
A1 &= Ln\left(\frac{P}{7.0}\right), \\
A2 &= (D - Rr), \\
A3 &= \left(\frac{51.0}{D}\right), \\
A4 &= \frac{(I.P)}{(A2 * A3 + D - 51)}, \\
A5 &= (Ek - 4.25 Rr), \\
A6 &= (Rr - I), \\
A7 &= \left(\frac{13.25}{I}\right).
\end{aligned} \tag{8}$$

Studying the generated formulas equations (3)–(7) indicates that the most effective factors in predicting soil loss (L) values are (P), (D), and (D_f), while (E), (R), and (R_r) are the secondary factors. Finally, (I) and (E_K) have minor effects on the values of soil loss (L).

4. Conclusions

This study was concerned in predicting the soil loss value using the available database. The considered parameters were rainfall amount, kinetic energy, rainfall intensity, gully head advance, soil detachment, factored soil detachment, runoff, and runoff rate. Three GP models were developed with different complexity and accuracy levels, and comparing these models shows the following:

3.2.3. GP Model Trial No. (3). Six levels of complexity (chromosome length = 128 genes) were used in this trial to produce equation (7). The relation between measured and predicted values of soil loss is shown in Figure 12(c). The achieved error values are (11%), (9%), and (10%) for the training, validation, and total sets, and their (R^2) values are (0.96), (0.98), and (0.97), respectively. It could be noted that the equation (3) achieved an excellent accuracy.

- (i) The 1st model in the simplest one (with only 32 genes in the chromosome) depended mainly on rainfall amount (P), soil detachment (D), and factored soil detachment (D_f). A fair accuracy level was achieved.
- (ii) The 2nd model (with 64 genes in the chromosome) utilized the same parameters of the 1st model, besides gully head advance (E), runoff (R), and runoff rate (R_r) to enhance the accuracy level.
- (iii) The last and the most complicated model (with 128 genes in the chromosome) included all the eight considered parameters and showed an excellent level of accuracy

Based on the above brief, the following points could be concluded:

- (i) Genetic programming (GP) technique was successfully used to predict the soil loss value (L) with excellent accuracy.
- (ii) The accuracy of the predicted soil loss value increased with increasing the complexity of the used expression up to certain level (6 levels in this study), beyond that the extra accuracy is not worth the extra complexity.
- (iii) Rainfall amount (P), soil detachment (D), and factored soil detachment (D_f) are the main parameters controlling the soil loss (L)
- (iv) Gully head advance (E), runoff (R), and runoff rate (R_r) have secondary impact on soil loss (L)
- (v) Kinetic energy (E_K) and rainfall intensity (I) do not have a significant effect on the soil loss (L)

Data Availability

The data used to support the results of this research are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. Balasubramanian, "Soil erosion—causes and effects," *Technical Reports*, 2017.
- [2] K. C. Onyelowe, "The menace of the geo-environmental hazard caused by gully erosion in Abia State, Nigeria," *Environmental Technology & Innovation*, vol. 8, pp. 343–348, 2017.
- [3] J. K. Adewumi and R. P. C. Morgan, "Soil detachment by raindrop impact comparison of predictive models," in *Book on Proceeding of the First Regional Symposium on Hydrology of Tropical Watersheds*, J. C. Agunwamba, G. O. Chukwuma, and C. C. Mbajiorgu, Eds., DE- Adroit Innovation, Albuquerque, NM, USA, pp. 142–143, 1997.
- [4] K. C. Onyelowe, D. Bui Van, O. C. Ikpemo, O. A. Ubachukwu, and M. Van Nguyen, "Assessment of rainstorm induced sediment deposition, gully development at Ikot Ekpene, Nigeria and the devastating effect on the environment," *Environmental Technology & Innovation*, vol. 10, pp. 194–207, 2018.
- [5] M. I. Ume and G. U. Ojiako, "Determination of a dynamic model for rates of soil loss in gully erosions," in *Proceedings Presented at the 2nd UNIBEN Conference on Engineering and Technological Development*, Benin, Nigeria, 1989.
- [6] X. Zhang, G. Q. Yu, Z. B. Li, and P. Li, "Experimental study on slope runoff, erosion and sediment under different vegetation types," *Water Resources Management*, vol. 28, no. 9, pp. 2415–2433, 2014.
- [7] D. Pimentel, "Soil erosion: a food and environmental threat," *Environmental Development and Sustainability*, vol. 8, pp. 119–137, 2006, <https://doi.org/10.1007/s10668-005-1262-8>.
- [8] S. A. Bhat, M. U. D. Dar, and R. S. Meena, "Soil erosion and management strategies," in *Sustainable Management of Soil and Environment*, R. Meena, S. Kumar, J. Singh, and M. Jat, Eds., Springer, Berlin, Germany, pp. 73–122, 2019.
- [9] M. Zounemat-Kermani, "Hydrometeorological parameters in prediction of soil temperature by means of artificial neural network: case study in Wyoming," *Journal of Hydrologic Engineering*, vol. 18, no. 6, 2013.
- [10] P. Kinnell, "The effect of kinetic energy of excess rainfall on soil loss from non-vegetated plots," *Soil Research*, vol. 21, no. 4, pp. 445–453, 1983.
- [11] Q. Wang, R. Horton, and M. Shao, "Effective raindrop kinetic energy influence on soil potassium transport into runoff 1," *Soil Science*, vol. 167, no. 6, pp. 369–376, 2002.
- [12] A. Nazari Samani, H. Ahmadi, A. Mohammadi et al., "Factors controlling gully advancement and models evaluation (Hableh Rood Basin, Iran)," *Water Resources Management*, vol. 24, no. 8, pp. 1531–1549, 2010.
- [13] N. Kariminejad, M. Hosseinalizadeh, H. Pourghasemi, G. Campetella, A. Bernatek-Jakiel, and M. Ownegh, "Evaluation of factors affecting gully headcut location using summary statistics and the maximum entropy model: golestan Province, NE Iran," *The Science of the Total Environment*, vol. 677, pp. 281–298, 2019.
- [14] S. Yariv, "Comments on the mechanism of soil detachment by rainfall," *Geoderma*, vol. 15, no. 5, pp. 393–399, 1976.
- [15] B. Wang, G. Zhang, Y. Shi, and X. Zhang, "Soil detachment by overland flow under different vegetation restoration models in Loess Plateau of China," *Catena*, vol. 116, no. 1, pp. 51–59, 2014.
- [16] S. Halder, M. Roy, S. Bhattacharya, S. Mondal, and P. Roy, "Soil loss estimation for sustainable watershed conservation in semi-arid bengal basin," in *Advances in Water Resources Management for Sustainable Use*, M. Roy, P. Kumar, P. Roy, M. Biswas, and S. Pal, Eds., pp. 413–439, Singapore, 2021.
- [17] M. Nearing, E. Albert, J. Laflen, and L. Lane, "Prediction technology for soil erosion by water: status and research needs," *Soil Science Society of America Journal*, vol. 54, no. 6, 1990.
- [18] I. Gaubi, A. Chaabani, A. Mammou, and M. Hamza, "A GIS-based soil erosion prediction using the revised universal soil loss equation (RUSLE) (Lebna Watershed, Cap Bon, Tunisia)," *Natural Hazards*, vol. 86, no. 1, pp. 219–239, 2017.
- [19] S. VemuUdaya and U. Pinnamaneni, "Estimation of spatial patterns of soil erosion using remote sensing and GIS: a case study of Indravati catchment," *Natural Hazards*, vol. 59, no. 3, pp. 1299–1315, 2011.
- [20] D. Vu, X. Tran, M. Cao, T. Tran, and N. Hoang, "Machine learning based soil erosion susceptibility prediction using social spider algorithm optimized multivariate adaptive regression spline," *Measurement*, vol. 164, 2020, <https://doi.org/10.1016/j.measurement.2020.108066>.
- [21] K. Onyelowe, M. Iqbal, and F. Jalal, "Smart computing models of California bearing ratio, unconfined compressive strength, and resistance value of activated ash modified soft clay soil with adaptive neuro-fuzzy inference system and ensemble random forest regression techniques," *Multiscale and Multidisciplinary Modeling, Experiments and Design*, vol. 4, pp. 207–225, 2021a, <https://doi.org/10.1007/s41939-021-00092-8>.
- [22] K. Onyelowe, M. Iqbal, F. Jalal, M. Onyia, and I. Onuoha, "Application of 3 algorithms ANN programming to predict the strength performance of hydrated-lime activated rice husk ash treated soil," *Multiscale and Multidisciplinary Modeling, Experiments and Design*, <https://doi.org/10.1007/s41939-021-00093-7>, 2021b.
- [23] Y. Singh, P. Bhatia, and O. Sangwan, "A review of studies on machine learning techniques," *International Journal of Computer Science and Security*, vol. 1, 2007.
- [24] A. M. Ebid, "35 years of (AI) in geotechnical engineering: state of the art," *Geotechnical & Geological Engineering*, vol. 39, pp. 637–690, 2020.
- [25] K. C. Onyelowe, A. M. Ebid, L. Nwobia, and L. Dao-Phuc, "Prediction and performance analysis of compression index of multiple-binder-treated soil by genetic programming approach," *Nanotechnology for Environmental Engineering*, vol. 6, no. 2, p. 28, 2021.
- [26] D. Khash and D. Al-Mussawi, "Studying the effect of head difference on exit gradients and uplift pressures beneath hydraulic structures by gene expression programming," *IOP Conference Series: Materials Science and Engineering*, vol. 1094, no. 1, Article ID 012099, 2021.
- [27] M. Momeni, M. Hadianfard, G. Bedon, and A. Baghlani, "Damage evaluation of H-section steel columns under impulsive blast loads via gene expression programming," *Engineering Structures*, vol. 219, no. 9, pp. 1–19, 2020.

- [28] A. H. ELbosraty, A. M. Ebid, and A. L. Fayed, "Predicting (nk) factor of (CPT) test using (GP): comparative study of MEPX & GN7," *International Journal of Scientific Engineering and Research*, vol. 10, no. 3, pp. 613–620, 2019.
- [29] A. H. El-Bosraty, A. M. Ebid, and A. L. Fayed, "Estimation of the undrained shear strength of east port-said clay using the gene expression programming," *Ain Shams Engineering Journal*, vol. 11, no. 4, 2020.
- [30] K. C. Onyelowe, A. M. Ebid, and L. I. Nwobia, "Predictive models of volumetric stability (durability) and erodibility of lateritic soil treated with different nanotextured bio-ashes with application of loss of strength on immersion; GP, ANN and EPR performance study," *Cleaner Materials*, vol. 1, no. 1, pp. 1–10, 2021.
- [31] W. Al-Mussawi and H. Kais, "Prediction of the depth of local scouring at a bridge pier using a gene expression programming method," *SN Applied Sciences*, vol. 3, no. 2, 2021.
- [32] N. Ryan and D. Hibler, "Robust gene expression programming," *Procedia Computer Science*, vol. 6, no. 1, pp. 165–170, 2011.
- [33] G. Johari, G. Habibagahi, and A. Ghahramani, "Prediction of soil-water characteristic curve using gene expression programming," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 132, no. 5, p. 661, 2006.
- [34] W. Al-Mussawi, "Application of a genetic algorithm for the optimization of a location and inclination angle of a cut-off wall for anisotropic foundations under hydraulic structures," *Geotechnical & Geological Engineering*, vol. 37, no. 1, 2019.
- [35] A. Tenpe and A. Patel, "Utilization of support vector models and gene expression programming for soil strength modeling," *Arabian Journal for Science and Engineering*, vol. 45, pp. 4301–4319, 2020.
- [36] G. Zhang, Y. Liu, Y. Han, and X. Zhang, "Sediment transport and soil detachment on steep slopes: i. transport capacity estimation," *Soil Science Society of America Journal*, vol. 73, no. 4, pp. 1291–1297, 2009a.
- [37] G. Zhang, M. Tang, and X. Zhang, "Temporal variation in soil detachment under different land uses in the Loess Plateau of China," *Earth Surface Processes and Land Reforms*, vol. 34, no. 9, pp. 1302–1309, 2009b.
- [38] J. K. Adewumi, O. J. Mudiare, and J. J. Owonubi, "Soil loss estimation in samaru, zaria: using European soil erosion model," in *Book on Proceeding of the First Regional Symposium on Hydrology of Tropical Watersheds*, J. C. Agunwamba, G. O. Chukwuma, and C. C. Mbajorgu, Eds., pp. 133–145, DE- Adroit Innovation, Albuquerque, NM, USA, 2000.
- [39] United States Soil Conservation Service, "Procedures for Determining the Rates of Land Damage, Land Depreciation and Volume of Sediment Produced by Gully erosion," *FAO Conservation Guide*, pp. 99–114, FAO, Rome, Italy, 1977.
- [40] L. Li, D. Shuhan, L. Wu, and G. Liu, "An overview of soil loss tolerance," *Catena*, vol. 78, no. 2, pp. 93–99, 2009.
- [41] W. Wischmeier and D. Smith, "Rainfall Energy and its relationship to soil loss," *Advanced Earth and Space Science*, vol. 39, no. 2, pp. 285–291, 1958.