

Article

# Energy Loss in Skimming Flow over Cascade Spillways: Comparison of Artificial Intelligence-Based and Regression Methods

Meysam Nouri <sup>1</sup> , Parveen Sihag <sup>2</sup>, Farzin Salmasi <sup>3</sup> and Ozgur Kisi <sup>4,5,\*</sup> 

<sup>1</sup> Department of Water Engineering, Faculty of Agriculture, University of Urmia, Urmia 57561–51818, Iran; m.nourighanli@urmia.ac.ir

<sup>2</sup> Department of Civil Engineering, Shoolini University, Himachal Pradesh, Solan 173229, India; parveen12sihag@gmail.com

<sup>3</sup> Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz 51666–16471, Iran; salmasi@tabrizu.ac.ir

<sup>4</sup> Ozgur Kisi, Department of Civil Engineering, School of Technology, Ilia State University, 0162 Tbilisi, Georgia

<sup>5</sup> Ozgur Kisi, Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

\* Correspondence: ozgur.kisi@iliauni.edu.ge or ozgurkisi@duytan.edu.vn

Received: 7 August 2020; Accepted: 27 September 2020; Published: 1 October 2020



**Abstract:** In this study, the energy dissipation of cascade spillways was studied by conducting a series of laboratory experiments. Five spillways slope angles ( $\alpha$ ) ( $10^\circ$ ,  $20^\circ$ ,  $30^\circ$ ,  $40^\circ$ , and  $50^\circ$ ), various step numbers ( $N$ ) ranging from 4 to 75, and a wide range of discharges ( $Q$ ), were considered. Some data-based models were developed to explain the relationships between hydraulic parameters. Multiple linear and nonlinear regression-based equations were developed based on dimensional analysis theory to compute energy dissipation over cascade spillways. For testing the robustness of developed data-based models, M5P, stochastic M5P, and random forest (RF) were used as new artificial intelligence (AI)-based techniques. To relate the input and output variables of energy dissipation, AI-based and regression approaches were developed. It was found that the formulation based on the stochastic M5P approach in solving energy dissipation problems over cascade spillways is more successful than the other regression and AI-based methods. Sensitivity analysis suggests that spillway slope in degrees ( $\alpha$ ) is the most influential input variable in predicting the relative energy dissipation (%) of the spillway in comparison to other input variables.

**Keywords:** cascade; energy dissipation; spillway; M5P; stochastic M5P; random forest

## 1. Introduction

These days, hydraulic structures have great importance in power generation and hydropower energy. The Ertan projects dam in China with a height of 240 m, design discharge of  $20,000 \text{ m}^3/\text{s}$ , and flood power of 39,000 MW is a good example of great hydropower projects [1]. In the similar hydraulic structures, because of the high reservoir elevation, there is additional kinetic energy in comparison with other structures and the safe passage should be fulfilled for these overtopping flows [2–4]. Under uncontrolled kinetic energy condition flood power can easily cause downstream erosion, structure collapse, etc. [5].

Stepped and cascade spillways are commonly used as energy dissipaters in hydraulic and hydropower projects all over the world. Usage of the stepped channels and cascade/stepped spillways goes back to at least three centuries ago [6,7]. They are also used in other projects, such as aeration projects and waterscape projects [8–11]. In the literature, three classical flow regimes on the stepped spillways have been reported, as follows: nappe flow regime, skimming flow regime, and transition

flow regime. When a small discharge occurs, nappe flow regime is shown over the stepped structures. Additionally, skimming and transition flow regimes belong to large discharge and intermediate flow rates, respectively [12–17]. Additionally, Ren et al. [8], Bai et al. [18], Bung [19], Chen et al. [20], Chinnarasri and Wongwises [21], Sorensen [22] have hydraulically studied stepped spillways.

In the stepped channels and cascade spillways, considerable values of kinetic energy and flow power are dissipated by steps. Due to this energy dissipation along the structure, their stilling basin is made with short length [23]. Due to significant amount of kinetic energy and velocity reduction and also high value of air entrainment, the risk of cavitation will be reduced in these structures [24,25]. The process of energy dissipation over the cascade spillways has been studied experimentally and numerically. Takahashi et al. [26] studied the effect of steps with end sills at 30° on the energy dissipation and large value of energy loss in comparison with horizontal steps was reported. Zare and Doering [27] used baffles and sills on the stepped spillways and reported that the energy dissipation due to the baffle edged spillways is more than the sill edged spillways. To determine the optimum values of dissipated energy, Chatila and Jurdi [28] studied stepped spillways with different step sizes. They suggested that stepped spillways with the maximum possible number of steps could have better performance in comparison to traditional smooth back spillways. Chanson [29] conducted an experimental study to analyze energy dissipation in stepped chutes. He reported that skimming flow in long chutes and nappe flow in short chutes enable higher energy dissipation, respectively. For analysis of the effect of step pool porosity on energy dissipation, a pooled stepped spillway was investigated [30]. They reported that the flat steps have large energy dissipation in comparison to the porous pooled steps. Relvas and Pinheiro [31] used stepped chutes lined wedge shaped blocks and analyzed the velocity distribution and the energy dissipation in this structure. Results showed that by increasing the steps slope, the energy dissipation reduces in comparison to common stepped chutes.

Multivariate data analysis techniques such as artificial intelligence (AI)-based models and soft computing techniques were recently employed as decision-support tenets in water science [32–37]. The operational costs and the expended time of experimental measurements are reduced by development of predictive models which are integrated into decision-support systems [38,39]. Instead of solving the governing equations, these techniques use effective parameters and their background to estimate the problems [40]. In other words, because of higher uncertainty and complexity of the solutions, a large number of effective parameters existing, and their interactions, these techniques can be used as a direct method to solve such problems [32]. Jiang et al. [41] used genetic algorithm (GA) and support vector regression (SVR) for prediction of stepped spillways energy dissipation and their results showed that the GA-SVR model can be successfully used for prediction of energy dissipation. The multiple regression equations, artificial neural networks (ANN), and gene expression programming (GEP) were used for modeling energy dissipation over stepped gabion weirs and the ANN results were recommended for reliable estimation [42].

It is evident from the previous studies that evaluation of energy dissipation using new stochastic methods such as M5P, stochastic M5P, and random forest has not been done yet. Evaluation of new models can introduce the optimum model for prediction of energy dissipation. The primary focus of this research is to investigate the accuracy of three new artificial intelligence-based methods (M5P, stochastic M5P, and random forest) for estimation of skimming flow regime energy dissipation along the cascade spillways. In addition, several regression equation based models are also tested for energy dissipation estimation. For evaluation of used models, experimental data are used. The experimental variables are spillway steps ( $N$ ), spillway slope ( $\alpha$ ), spillway height ( $H_{dam}$ ), and discharge over spillway ( $Q$ ).

## 2. Materials and Methods

### 2.1. Experimental Setup

In the present study, using effective parameters changes, experiments were done on stepped spillways. Study experiments were done at the two hydraulic laboratories: (i) University of Tabriz,

Iran and (ii) Shahid Chamran University (SCU), Iran. The fabricated physical models were installed in a flume with the following dimensions: 0.5 and 0.8 m width, 10 m length, and 1.0 and 2.0 m height. Table 1 shows the geometrical characteristic of physical models. Variable parameters include spillway steps ( $N$ ), spillway slope ( $\alpha$ ), spillway height ( $H_{dam}$ ), and discharge over spillway ( $Q$ ). With these variable parameters, several runs were performed on physical models in the laboratory. Using a gate in the downstream of the experimental flume, the water level was controlled in the form of hydraulic jump and a sharp triangle weir with angle of  $53^\circ$  was used for flow measurements. Additionally, a point gauge with the accuracy of  $\pm 0.1$  mm was used upstream water level measurement instrumentation.

Table 1. Physical models’ parameters and their values.

Number of Steps ( $N$ )	Spillway Height over Critical Depth $H_{dam}/d_c$	Spillway Slope $\alpha$ (Degrees)	Spillway Height $H_{dam}$ (m)
4–75	1.9–37.3	10–50	0.6–1.7

Upstream water level,  $y_0$ , was measured at a distance of 0.60 m from the spillway and  $y_2$  is the secondary depth for the hydraulic jump. In studied cascade spillways and similar structures, due to air entrainment and thin flow, measuring the flow depth ( $y_1$ ) is difficult. Pegram et al. [43] used the conjugate water depth of hydraulic jump ( $y_2$ ) for calculation of energy dissipation.

Using different values for the effective parameters including spillway slope ( $\alpha$ ), step number ( $N$ ), discharge rates ( $Q$ ), and structure height ( $H_{dam}$ ), 3772 experimental data were collected and used for evaluation of M5P, stochastic M5P, random forest, multiple linear regression (MLR), multiple nonlinear regression (MNLR), and logarithmic regression. In total, 70% of the data was used for training and the remaining 30% was used for testing the implemented models.

### 2.2. Dimensional Analysis

Calculation of upstream energy head ( $E_{max}$ ) was done using Equation (1). Additionally, Equations (2) and (3) were used for calculation of downstream energy head ( $E_1$ ) and relative energy dissipation ( $\Delta E/E_{max}$ ), respectively. In these equations,  $g$  is the acceleration due to gravity,  $H_{dam}$  is the height of spillway,  $y_0$  is the upstream water depth,  $V_0$  is the approach velocity, and  $q$  is the discharge rate per unit width. The conjugate depth ( $y_2$ ) was used to calculate flow depth ( $y_1$ ) as given by Equation (4).

$$E_{max} = H_{dam} + y_0 + \frac{V_0^2}{2g} = H_{dam} + y_0 + \frac{q^2}{2g(H_{dam} + y_0)^2} \tag{1}$$

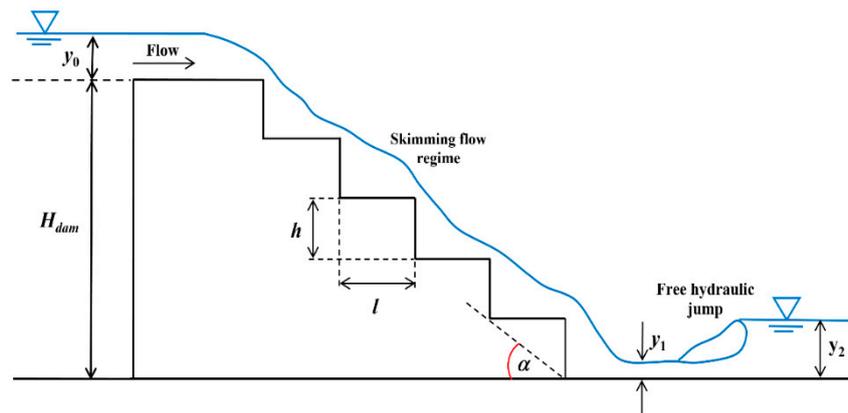
$$E_1 = y_1 + \frac{V_1^2}{2g} = y_1 + \frac{q^2}{2gy_1^2} \tag{2}$$

$$\frac{\Delta E}{E_{max}} = \frac{E_{max} - E_1}{E_{max}} = 1 - \frac{E_1}{E_{max}} \tag{3}$$

$$y_1 = \frac{y_2}{2} (\sqrt{1 + 8Fr_2^2} - 1) \tag{4}$$

where  $Fr_2$  is Froude number ( $Fr_2 = V_2/\sqrt{gy_2}$ ),  $y_2$  and  $V_2$  are water depth and velocity after hydraulic jump, respectively. Generally, effective parameters in the energy dissipation process can be written as Equation (5). In this equation,  $h$ ,  $l$ , and  $N$  are step height, step length, and step number, respectively. Both of  $y_1$  and  $y_2$  were calculated in corresponding locations that are shown in Figure 1, but using Equations (1)–(4), only  $y_2$  is used in calculations of energy dissipation.

$$\frac{\Delta E}{E_{max}} = f(q, l, h, g, N, H_{dam}) \tag{5}$$



**Figure 1.** Location of measurement for  $y_1$  and  $y_2$  depths at the downstream of cascade spillway.

In summary, the most effective parameters in hydraulic of stepped spillways are divided in two groups: geometrical parameters and hydraulic parameters. Step length ( $l$ ), the spillway height ( $H_{dam}$ ), spillway slope ( $\alpha$ ), step height ( $h$ ), and steps number ( $N$ ) are the geometrical effective parameters. Additionally, upstream energy head ( $E_{max}$ ), downstream energy head ( $E_1$ ), and discharge per unit width of canal ( $q$ ) can be classified as hydraulic effective parameters. The Buckingham  $\pi$  theorem was used to present relative energy dissipation as provided in Equation (6):

$$\Delta E/E_{max} = f(H_{dam}/d_c, l/h, N) \tag{6}$$

Equation (6) can be rewritten as follows:

$$\Delta E/E_{max} = f(H_{dam}/d_c, \alpha, N) \tag{7}$$

where,  $d_c$  is critical depth ( $d_c = (q^2/g)^{1/3}$ ) and  $\alpha = \tan^{-1}(h/l)$  is spillway slope in degrees.

### 2.3. Artificial Intelligence and Regression-Based Models

#### 2.3.1. Multiple Linear Regression (MLR)

MLR is a linear method to develop the relationship between dependent and independent variables. The general equation of the MLR is as follows:

$$E = c_0 + x_1^{c_1} + x_2^{c_2} + x_3^{c_3} + x_4^{c_4} + \dots + x_n^{c_n} \tag{8}$$

#### 2.3.2. Multi Nonlinear Regression (MNLRL)

Multiple regression (MNLRL) can also be applied to more than one predictor's parameters. The general equation of the MNLRL model is given by Equation (9), where, E is the dependent variable and independent variables are  $x_1, x_2, x_3, \dots, x_n$ .

$$E = c_0 x_1^{c_1} x_2^{c_2} x_3^{c_3} x_4^{c_4} \dots x_n^{c_n} \tag{9}$$

#### 2.3.3. Random Forest

The random forest algorithm was initially proposed by Breiman [44]. A highly flexible random forest (RF) algorithm was successfully used in the solution of several engineering-related complex problems. A large number of trees are developed in this algorithm with the root node attaining a dissimilar bootstrap (bagging) sample of the original data set. At every node, division is executed using arbitrarily chosen subset of the estimator parameters. The random forest algorithm is simple, less sensitive in training, and has higher precision in prediction [45]. Only two user defined parameters

are required in this algorithm. The number of trees grown (k) and number of input parameters (m) are the user defined parameters. Trial and error process is used for the model development. In this study, WEKA 3.9 software was used for the analysis.

### 2.3.4. M5P Model

M5P tree is a binary regression tree model which can produce continuous numerical attributes by using linear regression function [46]. M5P uses a divergence metric to produce a decision tree by constructing the tree model. M5P tree algorithm uses the standard deviation of a particular class value at the terminal node by measuring the error value at that node by evaluating the expected outcome. M5P generates the compact and relative comprehensible model. M5P combines a conventional decision tree with the possibility of linear regression functions at the nodes.

### 2.3.5. Stochastic M5P Model

Gradient boosting (stochastic) based assembly method was initially introduced by Friedman [47] for the nonlinear problems. Stochastic develops an additive regression model by in succession applying a traditional technique to current by least squares method at every iteration. Stochastic simplifies them through optimizing the random differentiable loss function, concerning the based model. It was found that both the estimation correctness and speed of model development of gradient boosting can be significantly raised by integrating randomization into the method. Particularly, at every iteration, a subsample of the calibration data is picked randomly (without substitution) from the total data set of calibration. This arbitrarily chosen subsample is then applied in the place of the total data set to fit the base model update for the present iteration. The size of the subsample selected for each iteration is a primary parameter and is considered as a small part of the total calibrating dataset. This arbitrary selection method also raised the robustness beside overcapacity of the base model. In this study, M5P is used as the base model.

## 2.4. Performance Assessment Parameters

For the model performance assessment, four statistical indexes, coefficient of correlation (CC), root mean square error (RMSE), mean absolute error (MAE) were used. The above-mentioned assessment indexes can be calculated by the equations as follows:

$$CC = \frac{a \sum_{i=1}^a OP - (\sum_{i=1}^a O)(\sum_{i=1}^a P)}{\sqrt{a(\sum_{i=1}^a O)^2 - (\sum_{i=1}^a P)^2} \sqrt{a(\sum_{i=1}^a O - \sum_{i=1}^a P)^2}} \quad (10)$$

$$RMSE = \sqrt{\frac{1}{a} \left( \sum_{i=1}^a (O - P)^2 \right)} \quad (11)$$

$$MAE = \frac{1}{a} \sum_{i=1}^a |O - P| \quad (12)$$

$$SI = \frac{RMSE}{\bar{O}} \quad (13)$$

where  $O$  = observed values,  $P$  = predicted values,  $\bar{O}$  = mean of observed values,  $a$  = number of observations.

## 2.5. Dataset

A total of 3772 observations of relative energy dissipation in spillway were used for model development and validation. The total data set was split into two different groups. The process of splitting was arbitrary. A larger group (70% observations) was considered as training data set

for model development and the remaining group was considered as testing data set for the model validation. Three independent variables, the number of steps ( $N$ ), the ratio of spillway height and critical depth ( $H_{dam}/d_c$ ), and spillway slope in degrees ( $\alpha$ ) were considered as inputs whereas relative energy dissipation in % was considered as a target for model development and validation. A correlation matrix plot of the total data set is shown in Figure 2. Characteristics of the data used for model development and validation are illustrated in Table 2.

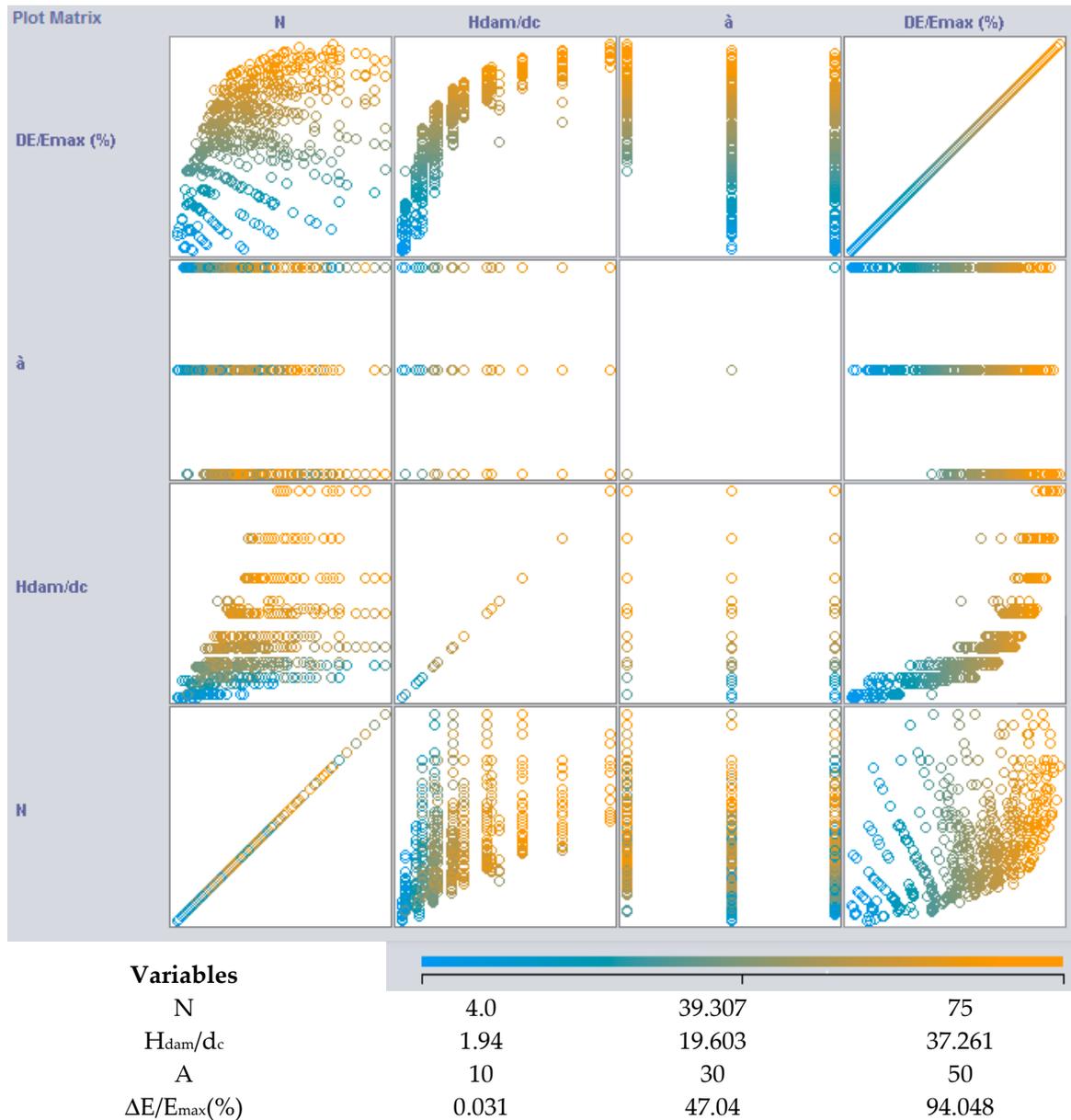


Figure 2. Correlation matrix plot of the total data set.

**Table 2.** Used data and their characteristics.

Range	N	H <sub>dam</sub> /d <sub>c</sub>	α	ΔE/E <sub>max</sub> (%)	N	H <sub>dam</sub> /d <sub>c</sub>	α	ΔE/E <sub>max</sub> (%)
Minimum	4.0	1.945	10.0	0.1239	4.0	1.945	10.0	0.0308
Maximum	75.0	37.26	50.0	92.63	75.0	37.26	50.0	92.63
Mean	27.0	11.57	28.49	57.68	27.0	11.61	29.04	57.34
Standard Deviation	15.3217	7.043	14.078	22.17	15.14	7.063	14.10	22.28
Kurtosis	0.2455	2.081	-1.269	-0.3790	0.3720	2.059	-1.291	-0.3704
Skewness	0.7987	1.344	0.1373	-0.6655	0.8477	1.348	0.1052	-0.6635
Confidence Level (95.0%)	0.5964	0.2741	0.5480	0.8628	0.8604	0.4014	0.8015	1.266

### 3. Results and Discussion

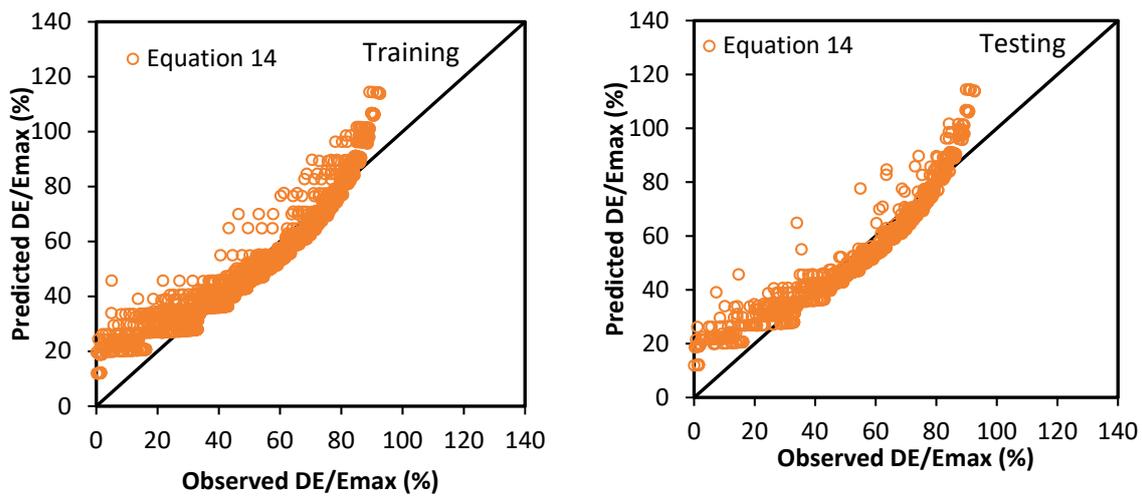
#### 3.1. Results of Linear and Nonlinear Regression Based Models

Different linear and nonlinear regression-based models were also applied in the current study. These equations which were developed by least square technique to drive regression coefficients using the training data set are listed in Table 3. XLSTAT software was used for the development of these equations. The performance evaluation parameter values are listed in Table 3 for all developed equations.

**Table 3.** Regression equations for the relative energy dissipation in spillway.

Equation No.	Linear and Nonlinear Regression Based Equations	CC	RMSE	MAE	SI	CC	RMSE	MAE	SI
		Training Data Set				Testing Data Set			
(14)	$DE/E_{max} (\%) = -2.2166 \text{Log}_{10}(N) + 74.6 \text{Log}_{10}(H_{dam}/d_c) - 52.6495 \text{Log}_{10}(\alpha) + 60.376$	0.9492	6.9755	5.4797	0.1209	0.9495	6.9864	5.4785	0.1218
(15)	$\frac{DE}{E_{max}} (\%) = 51.0958 + 0.1722 N + 2.2017 \left(\frac{H_{dam}}{d_c}\right) - 0.82750 \alpha$	0.8572	11.4138	9.0656	0.1979	8524	11.6493	9.2226	0.2032
(16)	$DE/E_{max} (\%) = 49.7274N^{0.0223} \left(\frac{H_{dam}}{d_c}\right)^{0.4519} \alpha^{-0.3078}$	0.8916	10.1579	7.9174	0.1761	0914	10.2312	7.9501	0.1784
(17)	$\frac{DE}{E_{max}} (\%) = -1399.148N^{-49.658} + 2191.05 \left(\frac{H_{dam}}{d_c}\right)^{0.0139} - 210.779\alpha^{0.0842} - 1927.49$	0.9491	5.4838	6.9833	0.1211	0494	5.4919	6.9953	0.1220
(18)	$\frac{DE}{E_{max}} (\%) = 328.003N^{-0.0054} \left(\frac{H_{dam}}{d_c}\right)^{0.098} \alpha^{-0.0719} - 262.94$	0.9380	6.0021	7.6858	0.1332	0376	6.0485	7.7486	0.1351
(19)	$\frac{DE}{E_{max}} (\%) = \left[ 1.899N^{-0.9316} - 0.6805 \right] / \left[ (0.0899N^{-1.404}) - 0.0412 \left(\frac{H_{dam}}{d_c}\right)^{-0.9962} - 0.00209\alpha^{0.355} \right]$	0.8784	8.1376	10.7333	0.1861	0.8860	7.9582	10.5422	0.1839
(20)	$\frac{DE}{E_{max}} (\%) = -0.8925 / \left[ (0.02935N^{0.0083}) - 0.0876 \left(\frac{H_{dam}}{d_c}\right)^{-1.1634} + 0.03405\alpha^{-0.1494} \right]$	0.8691	8.6033	11.0739	0.1920	0.8759	8.4493	10.9294	0.1906

Four statistical performance assessment statistics, CC, RMSE, MAE, and SI, were used to compare the performance of the MLR and MNLR based models (Table 3). According to the Table 3, the Equation (14) based model with higher value of CC (0.9492) and lower value of SI, MAE, and RMSE (0.1218, 5.4797%, and 6.9755%, respectively) has better results in comparison to other MLR and MNLR based models. Figure 3 provides agreement plots among observed and predicted relative energy dissipation (%) by Equation (14) (logarithmic regression) based models for training and testing stages, respectively. As observed from the graph that Equation (14) based model considerably overestimates the low (<40) and high (>90) DE/E<sub>max</sub> values.

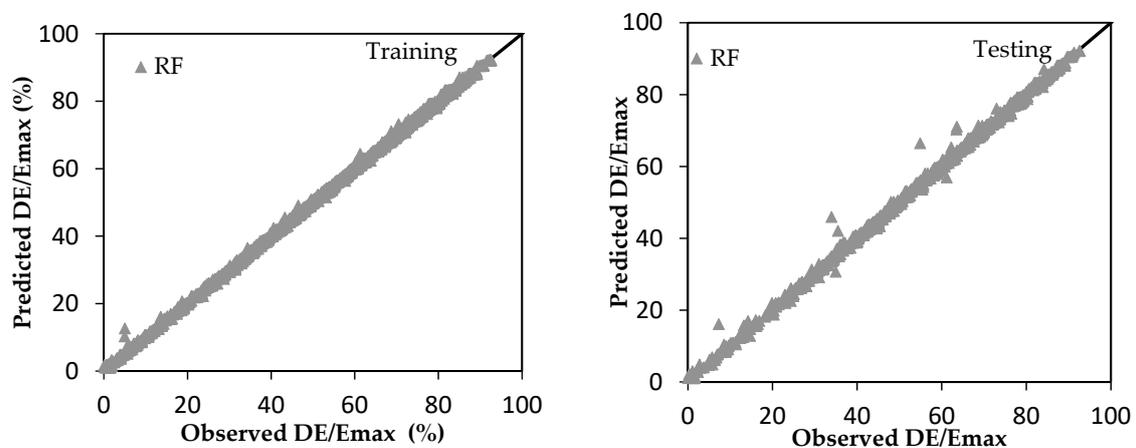


**Figure 3.** Agreement plot of observed and predicted relative energy dissipation by the Equation (14) based model using training and testing data sets, respectively.

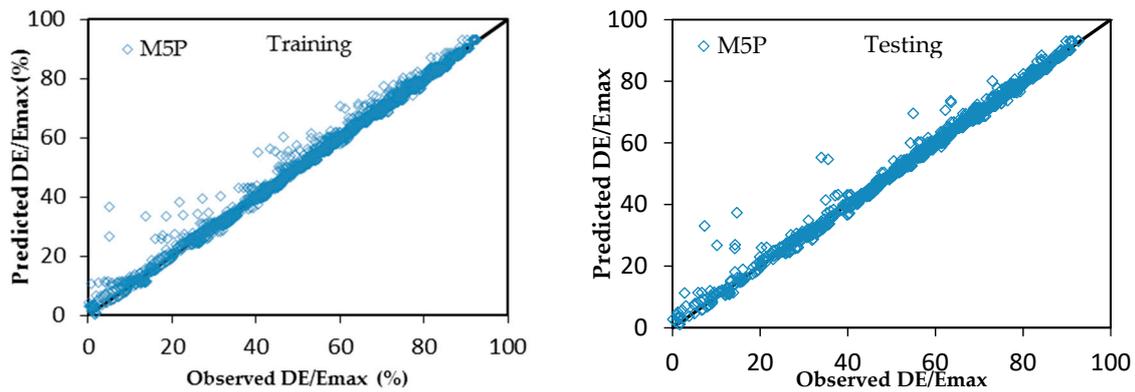
Further, single-factor ANOVA showed that the  $F$ -value (0.005627) was less than  $f$ -critical (3.8454) and  $p$ -value (0.9402) was greater than 0.05, which suggested that the difference in predicted values by the Equation (14) based model and actual values is insignificant.

### 3.2. Results of Soft Computing Based Models

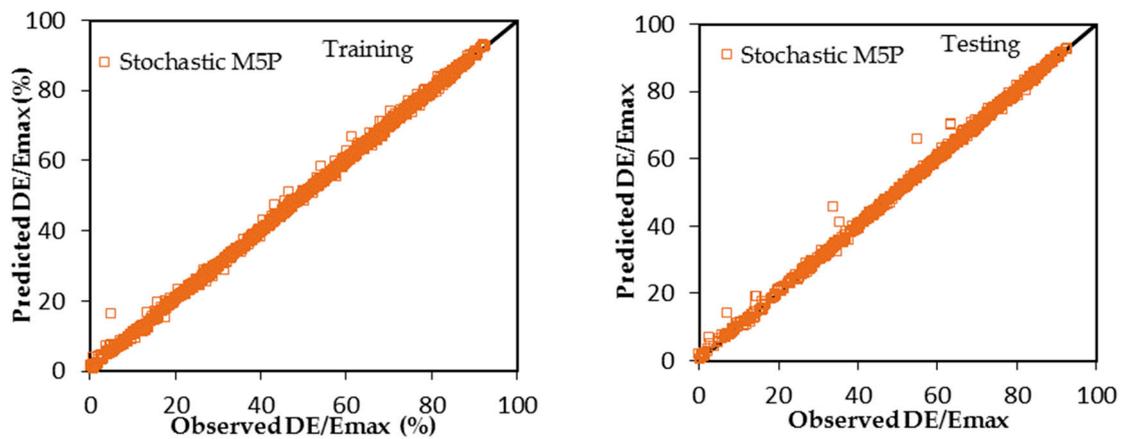
Figures 4–6 provide agreement plots between observed and predicted relative energy dissipation (%) of spillway by RF, M5P, and stochastic M5P based models for training and testing stages, respectively. Predicted values from RF, M5P, and stochastic M5P based models closely follow the observed values. It is clear that the predictions of RF are less scattered than those of the M5P. However, the stochastic M5P provided better predictions compared to RF. Results of test data set from Table 4 indicates that the performance of the stochastic M5P approach is the best among the implemented approaches. The stochastic M5P based model with correlation coefficient of 0.9993 and RMSE of 0.8246% presents reliable results in comparison to the RF-based model with a correlation coefficient of 0.9992 and RMSE of 0.9093%. Additionally, the M5P based model with correlation coefficient of 0.9958 and RMSE of 2.054% has poor results in comparison to stochastic M5P and RF-based models. Comparison of M5P and stochastic M5P based models reveals that the latter model considerably improved the first model (traditional M5P); an increase in CC by 0.35% and decrease in RMSE by 60% (Table 4).



**Figure 4.** Agreement plot of observed and predicted relative energy dissipation by the random forest (RF) model using training and testing data sets, respectively.



**Figure 5.** Agreement plot of observed and predicted relative energy dissipation by the M5P model using training and testing data sets, respectively.



**Figure 6.** Agreement plot of observed and predicted relative energy dissipation by the stochastic M5P model using training and testing data sets, respectively.

**Table 4.** Performance assessment measures for soft computing-based models using training and testing data sets.

Approaches	Training Data Set				Testing Data Set			
	CC	RMSE	MAE	SI	CC	RMSE	MAE	SI
M5P	0.9965	1.853	1.048	0.0321	0.9958	2.054	1.071	0.0358
Stochastic M5P	0.9996	0.6142	0.3803	0.0106	0.9993	0.8246	0.4278	0.0144
RF	0.9997	0.5019	0.3673	0.0087	0.9992	0.9093	0.5054	0.0159

The brief statistical properties of the observed and predicted values using regression and soft computing-based models are summarized in Table 5. The statistics of predictions by the RF and stochastic M5P are very close to those of the observations. Using the results of single-factor ANOVA presented in Table 6, it can be seen that used regression and soft computing-based models results have an insignificant difference in comparison to actual values. For the comparison of regression and soft computing-based models, an agreement, performance, and relative error plot are shown in Figure 7 using the testing data set. It can be inferred from this figure that the predicted values provided by the stochastic M5P based model lies very closer to the line of perfect agreement and predicted relative energy dissipation (%) of the spillway was found to follow the same patterns of actual relative energy dissipation (%) of the spillway with minimum relative error.

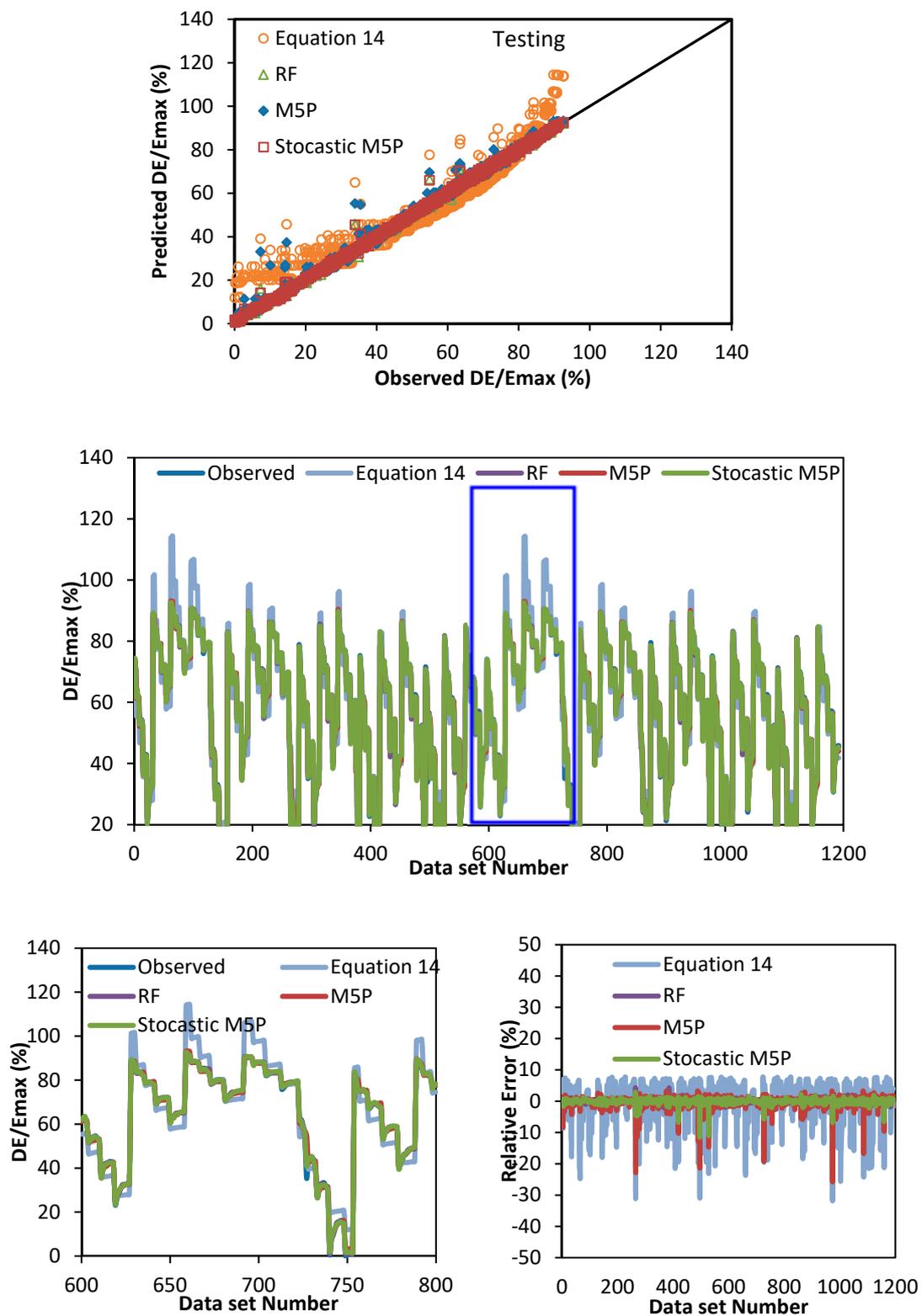


Figure 7. Performance of various regression and soft computing-based models during the testing stage.

**Table 5.** Features of the observed and predicted values using regression and soft computing-based models for both training and testing stages.

Models	Maximum	Minimum	Mean	Standard Deviation	Kurtosis	Skewness	Confidence Level (95%)
Training data set							
<i>Observation</i>	92.63	0.1240	57.68	22.17	−0.3790	−0.6655	0.863
<b>Equation (14)</b>	114.43	11.91	57.68	21.04	−0.5525	0.2231	0.819
<i>RF</i>	92.12	1.051	57.71	22.11	−0.3899	−0.6629	0.861
<i>M5P</i>	93.14	−0.1610	57.72	21.91	−0.4433	−0.6331	0.853
<i>Stochastic M5P</i>	92.98	0.6010	57.68	22.11	−0.4127	−0.6531	0.861
Testing data set							
<i>Observation</i>	92.63	0.0310	57.34	22.28	−0.3704	−0.6635	1.266
<b>Equation (14)</b>	114.41	11.89	57.27	21.10	−0.5077	0.2328	1.199
<i>RF</i>	92.12	1.051	57.41	22.19	−0.3806	−0.6627	1.261
<i>M5P</i>	93.15	−0.7090	57.38	22.04	−0.4447	−0.6250	1.253
<i>Stochastic M5P</i>	92.63	0.4820	57.37	22.22	−0.4052	−0.6499	1.263

**Table 6.** Single-factor ANOVA results.

Method	SS	df	MS	F	p-Value	F Crit	Variation Among Groups
<b>Equation (14)</b>	2.649	1	2.649	0.0056	0.9402	3.8454	Insignificant
<b>RF</b>	3.534	1	3.534	0.0071	0.9326	3.8454	Insignificant
<b>M5P</b>	1.0522	1	1.052	0.0021	0.9631	3.8454	Insignificant
<b>Stochastic M5P</b>	0.7684	1	0.7684	0.0016	0.9686	3.8454	Insignificant

For the assessment of the obtained outcomes, box plots (Figure 8) were drawn, in which overall error distribution was presented. Accordingly, the negative and positive error values correspond to the over and under estimation behavior of the models, respectively. The statistics of the overall error distribution is listed in Table 7. The values of minimum error, first quartile, median, third quartile, and maximum error are listed and presented in Table 7 and Figure 8 for all applied models. The lower quartile value in the stochastic M5P is −0.2508 which is less than all other discussed models. On the other hand, in upper quartile, the stochastic M5P with Q75 = 0.2928 performed better than the other applied models. As can be seen in Figure 8 and Table 7, the maximum and minimum errors in stochastic M5P model are −11.53 and 2.675, respectively, which verify the capability of this model to predict the relative energy dissipation (%) of the spillway. The lower box width in the stochastic M5P model confirms the maximum errors around the zero.

Taylor diagrams are plotted in Figure 9 for all applied models. A Taylor diagram was used to illustrate schematically the performance of the applied models [48]. Three statistic parameters including standard deviation, correlation, and root mean square error, evaluated the accuracy of the applied models. Figure 9 indicates that the stochastic M5P model achieves higher correlation with minimum RMSE value. The Taylor diagram also confirms that the stochastic M5P model is performing better than the other applied models.

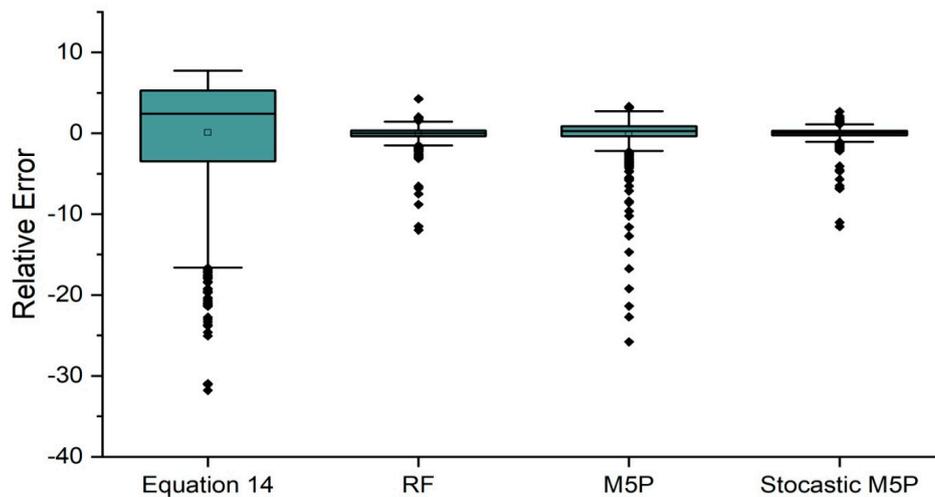


Figure 8. Box plots of error distribution for each model using the test data set.

Table 7. The statistics of the applied models’ error distribution using the test data set.

Statistic	Equation (14)	RF	M5P	Stochastic M5P
Minimum	−31.79	−11.97	−25.78	−11.53
1st Quartile	−3.468	−0.3842	−0.3555	−0.2508
Median	2.428	0.0230	0.2870	0.0245
3rd Quartile	5.291	0.3545	0.8805	0.2928
Maximum	7.758	4.253	3.294	2.675

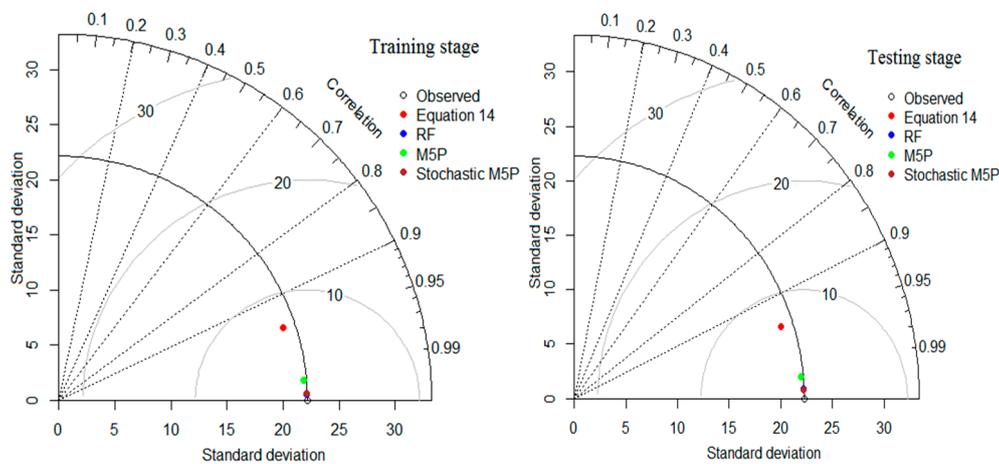


Figure 9. Taylor diagram for the applied models for training and testing stages.

Figure 10 presents contours of energy loss in stepped spillways by means of  $H_{dam}/d_c$  vs.  $N$ . With a constant value for  $N$ , an increase in  $H_{dam}/d_c$  values, shows an increase in energy loss up to 97.5%. This means that any decline in discharge over the spillway or an increase in spillway height, results in increase of energy loss. In addition, for a constant value of  $H_{dam}/d_c$ , an increase in  $N$  results in increase of energy loss especially for  $N < 90$ . Higher number of steps ( $N > 90$ ) has no significant effect on energy loss. Transformation of contours of energy loss to vertical type demonstrates this phenomenon. Figure 10 is for spillways with 30 degrees in slope and the other contours of energy loss can be provided for different spillway slopes. For brevity, the other contour lines are not presented here.

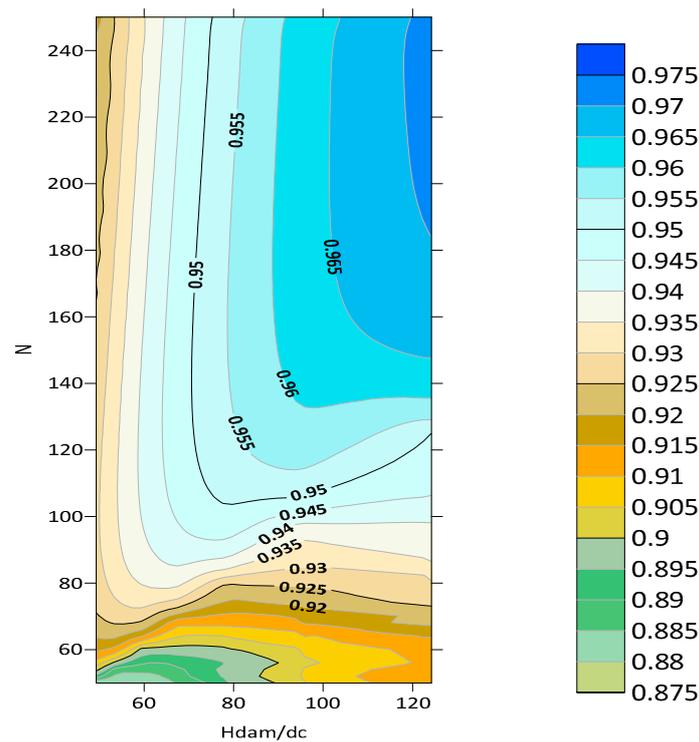


Figure 10. Contours of energy loss for stepped spillways with 30 degrees in slope.

### 3.3. Sensitivity Investigation

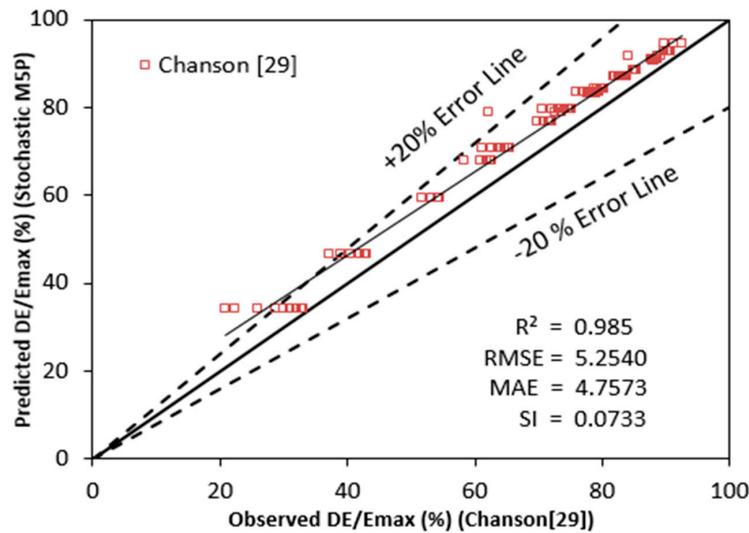
Sensitivity investigation was performed to find the most substantial input variable in predicting relative energy dissipation (%) of the spillway (Table 8). For this investigation, the stochastic M5P based model was used because it produced the best results. It is apparent from Table 8 that four input combinations were considered and in three of them, one input variable was omitted. Results from the table conclude that spillway slope in degrees ( $\alpha$ ) is the most influential input variable in predicting the relative energy dissipation (%) of the spillway in comparison to other input variables used in this study. The second essential parameter is  $H_{dam}/d_c$  in the prediction of the relative energy dissipation with respect to CC and RMSE statistics.

Table 8. Sensitivity investigation based on the stochastic M5P model.

Input Combination	Input Parameter Removed	Stochastic M5P	
		CC	RMSE (%)
$N, H_{dam}/d_c, \alpha$		0.9993	0.8246
$H_{dam}/d_c, \alpha$	N	0.9862	3.691
$N, \alpha$	$H_{dam}/d_c$	0.7953	13.515
$N, H_{dam}/d_c$	$\alpha$	0.7818	13.90

For validation of the best model, the stochastic M5P model, it was also used for prediction based on the data set that resulted from the original equation by Chanson [29]. The developed empirical formula, by Chanson [29], for calculating the energy dissipation of stepped spillways in the skimming flow is listed as Equation (21), where  $f$  is the friction factor and  $\alpha$  is the spillway slope. Figure 11 shows the scatter plots of the results obtained from the stochastic M5P model and resulting data from Equation (21). Figure 11 illustrates the agreement between the predicted values and the resulting data from Equation (21) and it is clear that the stochastic M5P model has powerful generalization ability for the prediction of stepped spillways energy dissipation in skimming flow.

$$\frac{\Delta E}{E_{max}} = 1 - \frac{\left(\frac{f}{8 \sin \alpha}\right)^{1/3} * \cos \alpha + \frac{1}{2} * \left(\frac{f}{8 \sin \alpha}\right)^{-2/3}}{\frac{2}{3} + \frac{H_{dam}}{d_c}} \quad (21)$$



**Figure 11.** Scatter plot of predicted relative energy dissipation using the stochastic M5P versus results of Equation (21) by Chanson [29].

#### 4. Conclusions

This paper investigates the potential of various regression-based equations, RF, M5P, and stochastic M5P based models in predicting the relative energy dissipation (%) of the cascade spillway. Three input parameters and one output parameter were spillway slope ( $\alpha$ ), step number ( $N$ ), rate of spillway height to critical depth ( $H_{dam}/d_c$ ), and relative energy dissipation of stepped spillway ( $\Delta E/E_{max}$ ), respectively. Experimental data were used to evaluate the models' exactness. From the comparison of performance evaluation parameters, it was found that the stochastic M5P based model works better ( $CC = 0.9993$ ,  $RMSE = 0.8246\%$ ) in comparison to other aforementioned regression and soft computing-based models. Another major conclusion was that the RF-based model works better than the M5P and various equation-based models. The stochastic approach significantly improved the performance of the traditional M5P algorithm, with an increase in RMSE by 60%. Sensitivity investigation concludes that spillway slope in degrees ( $\alpha$ ) is the most influential input variable in predicting the relative energy dissipation (%) of the spillway in comparison to other input variables used in this study.

The study revealed that a more reliable energy dissipation predictor model can be built based on the stochastic M5P model and experimental data. In other words, using such studies, the design efficiency can be improved due to time saving and low operational costs. The main limitations of the proposed model are the fact that it has a black box structure and is highly based on used data. In case of lacking 3D modeling methods which are able to perform as well as or better than the black box, the M5P can be recommended. In the present study, the data driven methods were only compared with linear and nonlinear regression methods. In future works, the proposed methods may be compared with 3D modeling methods and their accuracies can be assessed by considering their advantages and disadvantages.

**Author Contributions:** M.N.; analyzed experimental results and wrote the paper. P.S.; used the predictor models and analyzed their results. F.S.; carried out experiments and contributed in the results analysis. O.K.; reviewed and revised the paper and also contributed to the models' results analysis. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Qian, S.-T.; Wu, J.-H.; Ma, F. Hydraulic Performance of Ski-Jump-Step Energy Dissipater. *J. Hydraul. Eng.* **2016**, *142*, 05016004. [[CrossRef](#)]
2. Wu, J.; Qian, S.; Wang, Y.; Zhou, Y. Residual Energy on Ski-Jump-Step and Stepped Spillways with Various Step Configurations. *J. Hydraul. Eng.* **2020**, *146*, 06020002. [[CrossRef](#)]
3. Gou, W.; Li, H.; Du, Y.; Yin, H.; Liu, F.; Lian, J. Effect of Sediment Concentration on Hydraulic Characteristics of Energy Dissipation in a Falling Turbulent Jet. *Appl. Sci.* **2018**, *8*, 1672. [[CrossRef](#)]
4. Hennig, T.; Wang, W.; Feng, Y.; Ou, X.; He, D. Review of Yunnan's hydropower development. Comparing small and large hydropower projects regarding their environmental implications and socio-economic consequences. *Renew. Sustain. Energy Rev.* **2013**, *27*, 585–595. [[CrossRef](#)]
5. James, C. *Hydraulic Structures*; Springer: Cham, Switzerland, 2020. [[CrossRef](#)]
6. Chanson, H. History of stepped channels and spillways: A rediscovery of the “wheel”. *Can. J. Civ. Eng.* **1995**, *22*, 247–259. [[CrossRef](#)]
7. Felder, S.; Chanson, H. Energy dissipation, flow resistance and gas-liquid interfacial area in skimming flows on moderate-slope stepped spillways. *Environ. Fluid Mech.* **2011**, *9*, 427–441. [[CrossRef](#)]
8. Ren, W.; Wu, J.; Ma, F.; Zhou, Y. Prediction of Flow-Regime Transition over a Stepped Channel. *J. Irrig. Drain. Eng.* **2019**, *145*, 06019004. [[CrossRef](#)]
9. Khdir, H.; Potier, O.; Leclerc, J.-P. Aeration efficiency over stepped cascades: Better predictions from flow regimes. *Water Res.* **2014**, *55*, 194–202. [[CrossRef](#)]
10. Boes, R.M.; Hager, W.H. Hydraulic Design of Stepped Spillways. *J. Hydraul. Eng.* **2003**, *129*, 671–679. [[CrossRef](#)]
11. Tian, J. Hydraulic Characteristics of Stepped Release Structure. Ph.D. Thesis, Institute of Water Resources and Hydroelectric Engineering, Xi'an University of Technology, Xi'an, China, 2005.
12. Toombes, L.; Wagner, C.; Chanson, H. Flow patterns in nappe flow regime down low gradient stepped chutes. *J. Hydraul. Res.* **2008**, *46*, 4–14. [[CrossRef](#)]
13. Chamani, M.R.; Rajaratnam, N. Characteristics of Skimming Flow over Stepped Spillways. *J. Hydraul. Eng.* **1999**, *125*, 361–368. [[CrossRef](#)]
14. Rajaratnam, N. Skimming Flow in Stepped Spillways. *J. Hydraul. Eng.* **1990**, *116*, 587–591. [[CrossRef](#)]
15. Chanson, H.; Toombes, L. Hydraulics of stepped chutes: The transition flow. *J. Hydraul. Res.* **2004**, *42*, 43–54. [[CrossRef](#)]
16. Peyras, L.; Royet, P.; Degoutte, G. Flow and Energy Dissipation over Stepped Gabion Weirs. *J. Hydraul. Eng.* **1992**, *118*, 707–717. [[CrossRef](#)]
17. Ohtsu, I.; Yasuda, Y.; Takahashi, M. Flow Characteristics of Skimming Flows in Stepped Channels. *J. Hydraul. Eng.* **2004**, *130*, 860–869. [[CrossRef](#)]
18. Bai, Z.L.; Peng, Y.; Zhang, J. Three-Dimensional Turbulence Simulation of Flow in a V-Shaped Stepped Spillway. *J. Hydraul. Eng.* **2017**, *143*, 06017011. [[CrossRef](#)]
19. Bung, D.B. Developing flow in skimming flow regime on embankment stepped spillways. *J. Hydraul. Res.* **2011**, *49*, 639–648. [[CrossRef](#)]
20. Chen, Q.; Dai, G.; Liu, H. Volume of Fluid Model for Turbulence Numerical Simulation of Stepped Spillway Overflow. *J. Hydraul. Eng.* **2002**, *128*, 683–688. [[CrossRef](#)]
21. Chinnarasri, C.; Wongwises, S. Flow Patterns and Energy Dissipation over Various Stepped Chutes. *J. Irrig. Drain. Eng.* **2006**, *132*, 70–76. [[CrossRef](#)]
22. Sorensen, R.M. Stepped Spillway Hydraulic Model Investigation. *J. Hydraul. Eng.* **1985**, *111*, 1461–1472. [[CrossRef](#)]
23. Chanson, H. *The Hydraulics of Stepped Chutes and Spillways*; A.A. Balkema: Rotterdam, The Netherlands, 2002; p. 384. [[CrossRef](#)]
24. Pfister, M.; Hager, W.H. Self-entrainment of air on stepped spillways. *Int. J. Multiph. Flow* **2011**, *37*, 99–107. [[CrossRef](#)]
25. Zhang, G.; Chanson, H. Effects of Step and Cavity Shapes on Aeration and Energy Dissipation Performances of Stepped Chutes. *J. Hydraul. Eng.* **2018**, *144*, 04018060. [[CrossRef](#)]
26. Takahashi, M.; Yasuda, Y.; Ohtsu, I. Discussion of “Flow Patterns and Energy Dissipation over Various Stepped Chutes” by Chaiyuth Chinnarasri and Somchai Wongwises. *J. Irrig. Drain. Eng.* **2008**, *134*, 114–116. [[CrossRef](#)]
27. Zare, H.K.; Doering, J.C. Energy dissipation and flow characteristics of baffles and sills on stepped spillways. *J. Hydraul. Res.* **2012**, *50*, 192–199. [[CrossRef](#)]

28. Chatila, J.G.; Jurdi, B.R. Stepped Spillway as an Energy Dissipater. *Can. Water Resour. J./Rev. Can. Ressour. Hydr.* **2013**, *29*, 147–158. [[CrossRef](#)]
29. Chanson, H. Comparison of energy dissipation between nappe and skimming flow regimes on stepped chutes. *J. Hydraul. Res.* **1994**, *32*, 213–218. [[CrossRef](#)]
30. Felder, S.; Chanson, H. Effects of Step Pool Porosity upon Flow Aeration and Energy Dissipation on Pooled Stepped Spillways. *J. Hydraul. Eng.* **2014**, *140*, 04014002. [[CrossRef](#)]
31. Relvas, A.T.; Pinheiro, A.N. Velocity Distribution and Energy Dissipation along Stepped Chutes Lined with Wedge-Shaped Concrete Blocks. *J. Hydraul. Eng.* **2011**, *137*, 423–431. [[CrossRef](#)]
32. Seyedzadeh, A.; Maroufpoor, S.; Maroufpoor, E.; Shiri, J.; Bozorg-Haddad, O.; Gavazi, F. Artificial intelligence approach to estimate discharge of drip tape irrigation based on temperature and pressure. *Agric. Water Manag.* **2020**, *228*, 105905. [[CrossRef](#)]
33. Akbari, M.; Salmasi, F.; Arvanaghi, H.; Karbasi, M.; Farsadzadeh, D. Application of Gaussian Process Regression Model to Predict Discharge Coefficient of Gated Piano Key Weir. *Water Resour. Manag.* **2019**, *33*, 3929–3947. [[CrossRef](#)]
34. Roushangar, K.; Akhgar, S.; Salmasi, F. Estimating discharge coefficient of stepped spillways under nappe and skimming flow regime using data driven approaches. *Flow Meas. Instrum.* **2017**, *59*, 79–87. [[CrossRef](#)]
35. Bahrami, B.; Mohsenpour, S.; Noghabi, H.R.S.; Hemmati, N.; Tabzar, A. Estimation of flow rates of individual phases in an oil-gas-water multiphase flow system using neural network approach and pressure signal analysis. *Flow Meas. Instrum.* **2019**, *66*, 28–36. [[CrossRef](#)]
36. Jahanpanah, E.; Khosravinia, P.; Sanikhani, H.; Kisi, O. Estimation of discharge with free overfall in rectangular channel using artificial intelligence models. *Flow Meas. Instrum.* **2019**, *67*, 118–130. [[CrossRef](#)]
37. Al-Khatib, I.A.; Gogus, M. Prediction models for discharge estimation in rectangular compound broad-crested weirs. *Flow Meas. Instrum.* **2014**, *36*, 1–8. [[CrossRef](#)]
38. Brandhorst, N.; Erdal, D.; Neuweiler, I. Soil moisture prediction with the ensemble Kalman filter: Handling uncertainty of soil hydraulic parameters. *Adv. Water Resour.* **2017**, *110*, 360–370. [[CrossRef](#)]
39. Maroufpoor, S.; Maroufpoor, E.; Bozorg-Haddad, O.; Shiri, J.; Yaseen, Z. Soil moisture simulation using hybrid artificial intelligent model: Hybridization of adaptive neuro fuzzy inference system with grey wolf optimizer algorithm. *J. Hydrol.* **2019**, *575*, 544–556. [[CrossRef](#)]
40. Salmasi, F.; Nouri, M.; Abraham, J. Laboratory Study of the Effect of Sills on Radial Gate Discharge Coefficient. *KSCE J. Civ. Eng.* **2019**, *23*, 2117–2125. [[CrossRef](#)]
41. Jiang, L.; Diao, M.; Xue, H.; Sun, H. Energy Dissipation Prediction for Stepped Spillway Based on Genetic Algorithm–Support Vector Regression. *J. Irrig. Drain. Eng.* **2018**, *144*, 04018003. [[CrossRef](#)]
42. Khatibi, R.; Salmasi, F.; Ghorbani, M.A.; Asadi, H. Modelling Energy Dissipation Over Stepped-gabion Weirs by Artificial Intelligence. *Water Resour. Manag.* **2014**, *28*, 1807–1821. [[CrossRef](#)]
43. Pegram, G.G.S.; Officer, A.K.; Mottram, S.R. Hydraulics of Skimming Flow on Modeled Stepped Spillways. *J. Hydraul. Eng.* **1999**, *125*, 500–510. [[CrossRef](#)]
44. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
45. Breiman, L.; Last, M.; Rice, J. Random Forests: Finding Quasars. *Stat. Chall. Astron.* **2006**, *45*, 5–32. [[CrossRef](#)]
46. Quinlan, J.R. Learning with continuous classes. *Aust. Jt. Conf. Artif. Intell.* **1992**, *92*, 343–348.
47. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [[CrossRef](#)]
48. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Space Phys.* **2001**, *106*, 7183–7192. [[CrossRef](#)]

