A COMPARATIVE STUDY OF USING ADAPTIVE NEURAL FUZZY INFERENCE SYSTEM (ANFIS), GAUSSIAN PROCESS REGRESSION (GPR), AND SMRGT MODELS IN FLOW COEFFICIENT ESTIMATION.

Ruya mehdi*

Gaziantep University, Civil Engineering Department, Yeditepe st., no 85088, sahinbey dist., Gaziantep, Turkey.

ruya.mehdi1991@gmail.com

Ayse Yeter GUNAL

2Gaziantep University, Civil Engineering Department, Osmangazi district, University Street, 27410 Sehitkamil / Gaziantep, Turkey .

agunal@gantep.edu.tr

Reception: 04/03/2023 **Acceptance**: 25/04/2023 **Publication**: 23/05/2023

Suggested citation:

Ruya M. And Ayse Yeter G. (2023). **A Comparative Study of Using Adaptive Neural Fuzzy Inference System (ANFIS), Gaussian Process Regression (GPR), and SMRGT Models in Flow Coefficient Estimation.** *3C Tecnología. Glosas de innovación aplicada a la pyme, 12(2)*, 125-146. https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

ABSTRACT

Estimating the flow coefficient is a crucial hydrologic process that plays a significant role in flood forecasting, water resource planning, and flood control. Accurate prediction of the flow coefficient is essential to prevent flood-related losses, manage flood warning systems, and control water flow. This study aimed to predict the flow coefficient for a period of 19 years (2000-2019) in the Aksu River Sub-Basin in Turkey, using historical climatic data, including precipitation, temperature, and humidity, provided by The Turkish State of Meteorological Service (TSMS). The study utilized three different approaches, namely, the Adaptive Neural Fuzzy Inference System (ANFIS), Simple Membership function and fuzzy Rules Generation Technique (SMRGT), and Gaussian Process Regression (GPR), to predict the flow coefficient. The models were evaluated using several statistical tests, such as Root Mean Square Error (RMSE), Coefficient of Determination (R2), Mean Absolute Error (MAE), and Mean Square Error (MSE), to determine their accuracy. Based on the evaluation criteria, it is concluded that the Simple Membership Functions and Fuzzy Rules Generation Technique (SMRGT) model has superior flow coefficient estimation performance than the other models.

KEYWORDS

ANFIS, SMRGT, Flow coefficient, Prediction, Gaussian process regression.

INDEX

1. INTRODUCTION

Hydrology is the study of the entire water cycle, with the most critical aspect being the section where rainfall causes water to flow. The flow is crucial in designing flood protection measures for urban and agricultural areas, as well as determining the amount of water that can be extracted from a river for irrigation or water supply. Turkey is located in a region that is prone to natural disasters, such as floods and earthquakes, and the amount of rainfall, particularly during the rainy season, is a significant factor in climate change. The rainy season is when floods and landslides are most likely to occur, and several factors, such as the condition of the catchment area [1], rain duration and intensity [2], land cover [3], topographic conditions [4], and drainage network capacity [5], can contribute to floods. However, climate change is the fundamental cause of these disasters. Urban floods, including flash floods, are considered the most distressing types of floods. Flood forecasts for Turkey indicate that 51% of flood events occur in late spring and early summer, with a significant portion observed during winter and a small portion in autumn. The Black Sea, Mediterranean, and Marmara Regions have the highest frequency of flood occurrence in that order.

In flood forecasting, the flow coefficient is the most important factor to consider. Flow coefficient is the ratio of the volume of water that drains superficially throughout rainfall to the total volume of precipitation over a specified period [6,7]. The flow coefficient, an essential tool in hydrologic processes of countless urban and rural engineering projects, etc. [8], can indicate the quantity of water flowing from specific precipitation and reflect the influence of natural geomorphological elements on the flow. Flow coefficients are also useful when contrasting watersheds to determine how various landscapes convert precipitation into rainfall events. [9,10] Precipitation is one of the most crucial variables when assessing and determining the flow coefficient [7, 11]. Precipitation may refer to a single rainfall event or an interval in which multiple rainfall events occur. Initial losses and infiltration capacity are attained when precipitation intensifies—consequently, flow increases, leading to a greater flow coefficient. In addition to precipitation properties such as intensity, duration, and distribution, specific physical aspects of watersheds, such as soil type, vegetation, slope, and climate, influence the occurrence and volume of the flow. The flow coefficient can be estimated by employing tables in which the flow is related to the surface type. According to [12], the effective study of the coefficient is a highly complex operation due to many influencing variables. This implies that the flow coefficients reported in the literature transmit less information than is required [9] and that their values, when tabulated as if they were constant, may not reflect reality.

Since the accurate estimation of the flow coefficient is crucial to our existence, improving models incorporating meteorological, hydrologic, and geological variables is necessary. Thus, effective water management and operation of water structures will be possible. Several models are used to model such a process. These models are separated into experimental models, conceptual black box models, or grey box models, and physically-based distributional models, or white box models.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

Experimental models (black box model) do not explicitly account for the physical laws of the processes and only connect the input and output via the conversion function. The second group consists of conceptual models, which are based on limited studies of the existing processes in the basin hydrology system, as opposed to the distributional physically-based models; their development has not been based on the total number of physical processes but rather on the designer's comprehension of the system's behavior. The third group consists of distributional physically-based models; these models attempt to account for all the processes within the desired hydrological system by applying physical definitions. In contrast, physically based models provide a more realistic approach by mathematically representing the real phenomenon. Even though physically-based models appear to be more suitable for modeling purposes, they lack acceptability because of their fundamental uncertainty and high computational cost.

Reports indicate that machine learning techniques such as ANN and FIS effectively model such complexities (flow coefficient). Their simplicity and capacity for dealing with nonlinearity without understanding the entire system distinguish them from others. Numerous examples in the literature demonstrate that fuzzy logic (FL)-based systems excelled at modeling different hydrological events such as precipitation, runoff, streamflow, etc. Due to the presence of uncertainty and vagueness in these domains, FL-based systems are well-suited for modeling.

This study proposes one of the pertinent machine learning algorithms, the Adaptive Neuro-Fuzzy Inference System (ANFIS), for estimating the flow coefficient. The ANFIS model employs Tagaki-Sugeno-Kang (TSK) first order [13,14]. As a flow coefficient prediction, the hybrid learning algorithm is selected from various algorithms for supervised learning. The widespread use of hybrid learning algorithms justifies their selection. An advantage of ANFIS is that it is a combination of ANN and fuzzy systems employing ANN learning capabilities to acquire fuzzy if-then rules with suitable membership functions, which can learn something from the inaccurate data that has been input and leads to the inference. Another benefit is that it can effectively utilize neural networks' self-learning and memory capabilities, resulting in a more sustainable training process [15].

These methods (ANFIS and other fuzzy systems) lack a definitive method for determining the number of fuzzy rules and membership functions (MF) required for each rule [13]. In addition, they have no learning algorithm for refining MF that can minimize output error. Therefore, Toprak in 2009 [16] proposed a new method known as the Simple Membership functions and fuzzy Rules Generation Technique (SMRGT). This new technique takes into account the physical cause-and-effect relationship and is designed to assist those who struggle to select the number, form, and logic of membership functions (MFs) and fuzzy rules (FRs) in any fuzzy set.

Gaussian Process Regression (GPR) is a statistical learning theory and Bayesian theory-based machine learning technique. It is well-suited for handling complicated regression tasks, such as high dimensions, a small number of samples, and

 https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

nonlinearity, and it has a substantial potential for generalization. Gaussian process regression has many favorable circumstances over neural networks, including simple implementation, self-adaptive acquisition of hyper-parameters, flexible inference of non-parameters, and probabilistic significance of its outcome. Results are less affected by bias and easier to read thanks to the GPR's seamless integration of hyperparameter estimates, model training, and security assessments. Processes with a Gaussian (GP) distribution take it for granted that the overall distribution of the model's probabilities is Gaussian.

The objectives of this study are to (1) compare the predictive power of the ANFIS, SMRGT, and GPR models and (2) select the model and algorithm with the highest degree of accuracy and the lowest error rate. This is the first attempt to compare the abovementioned models to determine the flow coefficient.

2. MATERIALS AND METHODS

2.1. AREA OF STUDY AND DATASET

The Aksu River basin is located in the Antalya Basin, southwest of Turkey. The total length of the Aksu River is approximately 145 km, with headwaters Akdag situated within Isparta Province and discharges to the Mediterranean from the Antalya-Aksu border. The southern part of the basin is narrower than the north. Two different climatic types, Mediterranean and continental climates, are observed in the Aksu River basin. The north part has low precipitation throughout the year, and the northwest and northeast mountain areas are the highest areas and have lower temperatures, intense precipitation, and snow, whereas the south plain areas are generally warmer with intense rainfall and evaporation. Several measurement data are collected to support the study. The primary data are obtained from TSMS (Turkish State of Meteorological Service). The data processed for this study are precipitation, temperature, and humidity.

2.2. CLIMATE PROPERTIES OF THE STUDY AREA

2.2.1. PRECIPITATION

The most severe effect of climate change is a rise in the frequency and intensity of extreme weather events in some parts of the world; the most obvious manifestation of this is the recent rise in the frequency and intensity of extreme precipitation in various parts of the world, which is causing infrastructure systems to become completely inadequate. Precipitation ranks among the most crucial elements of climatic parameters and atmospheric circulation, as well as the element that provides water to the land and is the primary flow source. In this work, the precipitation stations' data and locations are obtained from TSMS (Turkish State of Meteorological Service). 57

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

precipitation observation stations (POSs) with (1793) records of monthly precipitation data for 20 years are used. The precipitation increased in (Oct., Nov., Dec., Jan., and Feb.) and the minimum precipitation recordings showed in (June, July, August. and Sep.). The maximum monthly precipitation (907.2 mm) was recorded in (Nov. 2001). In comparison, the minimum record for most of the years was (0.1 mm), especially in August. The annual average rainfall was 963.60 mm based on 19 years of Aksu meteorological station measurements (see Fig.1). The maximum annual rainfall was 1891.8mm in 2001.

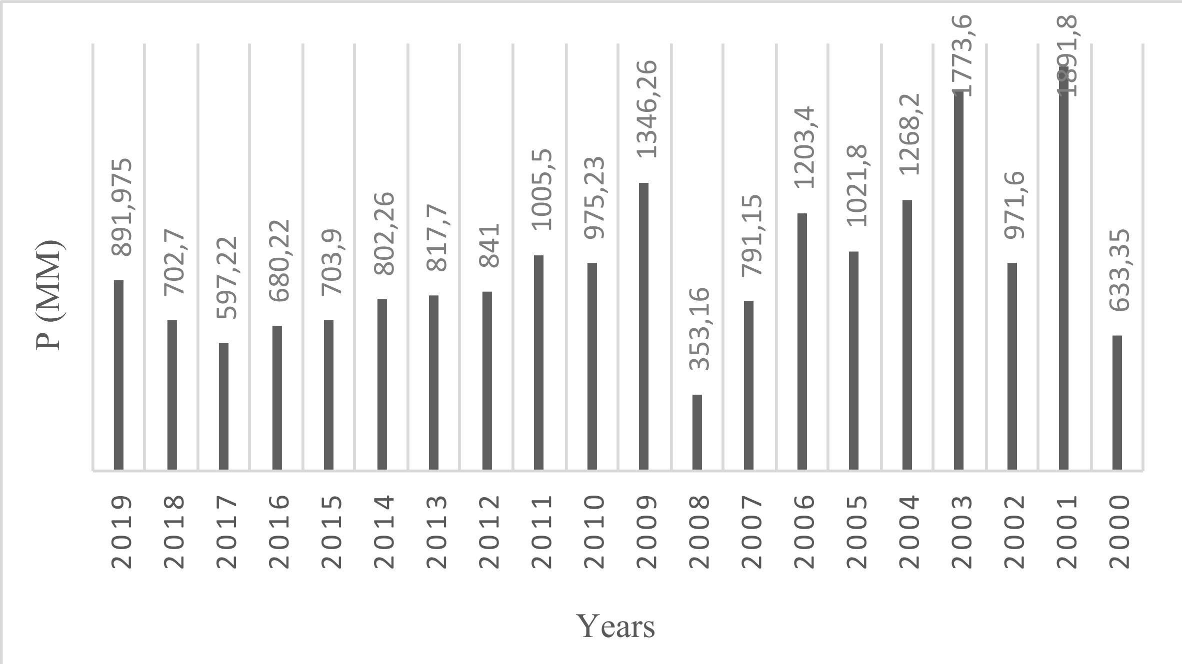

Figure 1. Average annual precipitation values for 19 years.

2.2.2. TEMPERATURE

The region is influenced by both moist tropical (MT) and warm and dry tropical air (CT) from the African and Arabian regions during the summer. (6996) Monthly temperature data have been studied in Aksu meteorological stations; the temperature showed an increase in (July, and Aug.), while the minimum temperature recordings showed in (Dec., and Jan.). The maximum monthly temperature was (31.4 °C) recorded in Aug. 2012, and the minimum record (- 5 °C) was shown in (Dec., and Jan.) 2016 and 2017. The annual average temperature was 16.03 °C based on 19 years of Aksu meteorological station measurements (see Fig.2). The maximum annual temperature was 16.92 °C in 2010.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

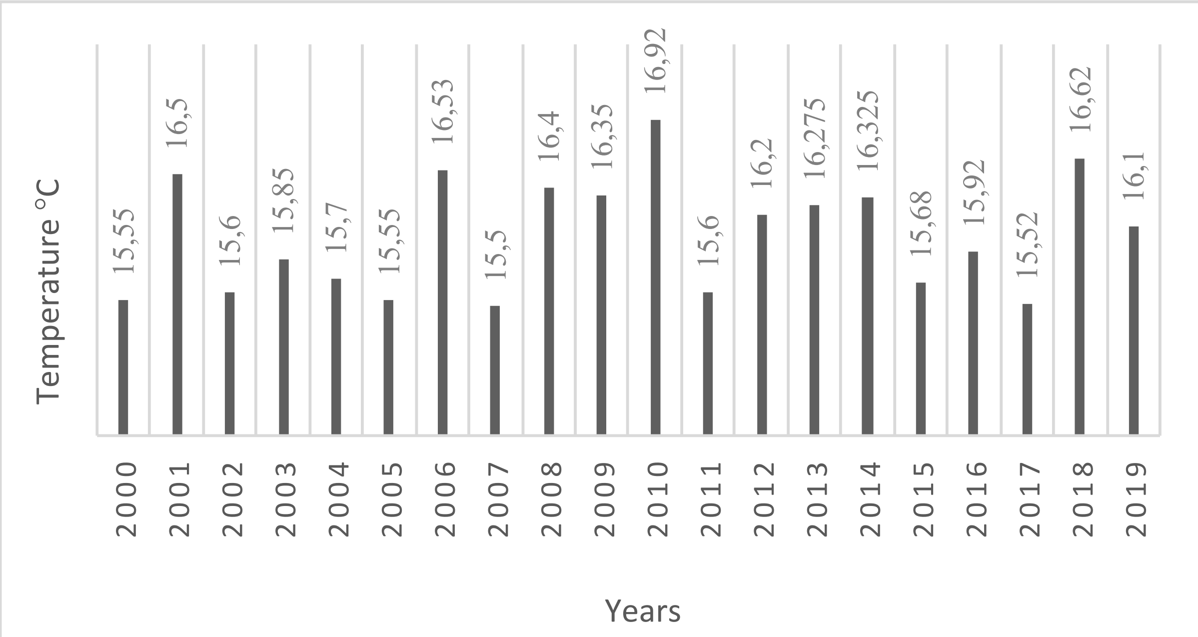

Figure 2. Average annual temperature values.

2.2.3. RELATIVE HUMIDITY

(6680) of monthly relative humidity data have been considered from the Aksu meteorological stations; it is recognized that the humidity increased in (Jan., and Dec.), while the minimum humidity recordings showed in (July and Aug.). The maximum monthly humidity was (97.7%) recorded in Jan. 2017, and the minimum record was (2.4%) in Dec. 2017. the annual average humidity was 63.3% (see Fig.3). The maximum annual humidity was 67.4 % in 2002, and the minimum was 58.85 % in 2013.

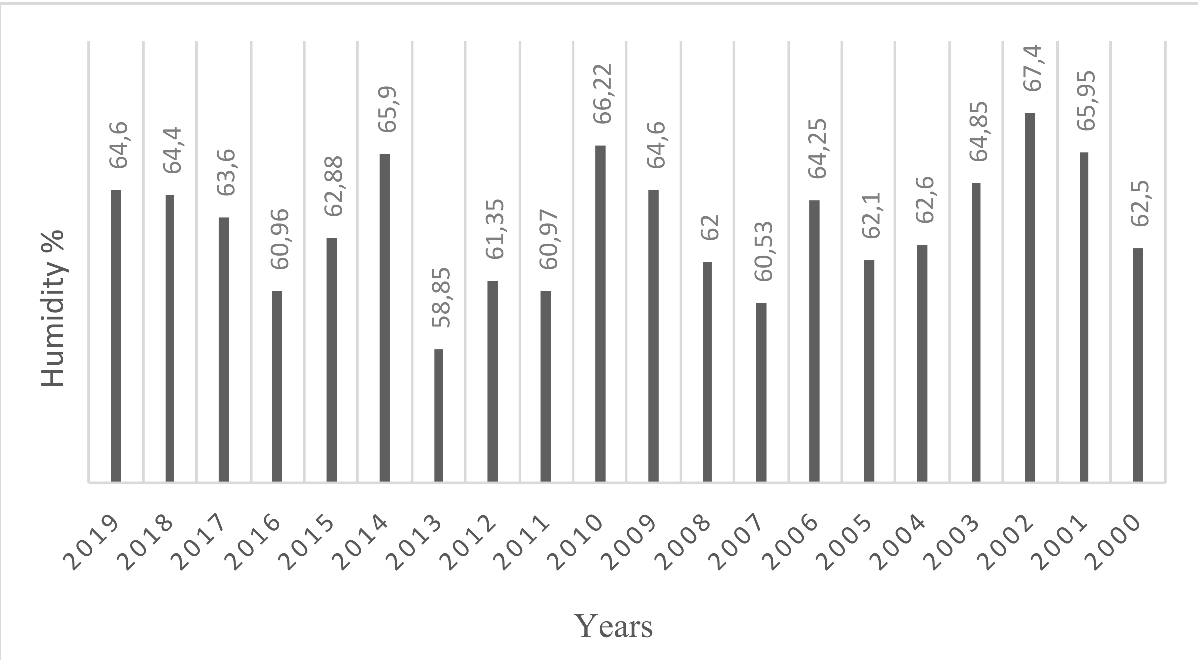

Figure 3. average annual relative humidity rates.

 https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

2.3. METHODS

2.3.1. ANFIS MODEL DEVELOPMENT

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a hybrid approach combining the advantages of two intelligent methods, neural networks, and fuzzy logic, to ensure qualitative and quantitative rationality. This new network can be effectively trained to interpret linguistic variables by utilizing neural networks and fuzzy logic. ANFIS implements a Sugeno-style first-order fuzzy system; it applies TSK Takagi Sugeno and Kang rules in its architecture [17] and effectively handles nonlinear real-time problems. ANFIS has been utilized extensively in disaster risk management, rock engineering [18,19], health services, finance, and other real-time areas [20–22]. It addresses regression and classification issues.

In first-order Sugeno's system, a typical set of IF/THEN rules for three inputs and one output can be expressed as follows:

Rule 1: If x is A1 and y is B1, then f1 = p1 x + q1 y + r1 (1)

Rule 2: If x is A2 and y is B2, then f2 = p2 x + q2 y + r2 (2)

Rule 3: If x is A3 and y is B3, then f3 = p3 x + q3 y + r3 (3)

Generally, ANFIS is composed of five layers:

Input Layer:

Nodes in the input layer stand in for the system's input variables. Each input node is associated with a membership function connecting an input value to a fuzzy set. If there are n input variables, the input layer can be denoted as:

$$y_1 = x_1, y_2 = x_2 \cdots y_n = x_n \tag{4}$$

Where Input variables are denoted by $x_1, x_2, \cdots, x_n$ and their corresponding nodes in the input layer are denoted by $y_1, y_2, \cdots, y_n$.

Fuzzification Layer:

The multiplicators and transmitters of this layer are their nodes. This product signifies the firing strength of a rule. Let $A_{ij}(x)$ be the membership function of node (i) for input (j) with parameters (p_{ij}). The output of the fuzzification layer can be denoted as:

$$u_{ij} = A_{ij}(x_j),\ for\ i = 1\ to\ m\ and\ j = 1\ to\ n \tag{5}$$

Where (m) is the number of membership functions per input variable and (u_{ij}) is the degree of membership of the j^{th} input variable in the i^{th} fuzzy set.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

Rule Layer:

The nodes in this layer calculate the i^{th} rule's firing strength relative to the total firing strength of all rules.

$$\bar{W} = \frac{W1 + W2 + W3}{W1} \tag{6}$$

Defuzzification Layer:

This layer's nodes are adaptive with node functions.

$$\bar{W}if = \bar{W} \cdot (pix + qiy + ri) \tag{7}$$

Where is the output of Layer 3 and $\{pi, qi, ri\}$ are the parameter set? Parameters of this layer are referred to as consequent parameters.

Output Layer:

All inputs are combined at a single fixed node to produce the final output. We can model the output layer as:

$$f = \sum_{i=1}^{n} \bar{W}ifi \tag{8}$$

In this model, 7-year data is used, where the training data were Precipitation, temperature, and humidity (input variables) data from January 2013 to December 2017 (5 years). On the other hand, the testing data from January 2018 to December 2019 (2 years), in this case, by trial and error, is 70%:30%. Training is conducted using a membership function such as the Gaussian membership function (gaussmf). In this step, a fuzzy inference system (FIS) is generated and evaluated, which can produce MSE and MARE.

To reduce computations that are too large in the pre-processing, the data is normalized into the range (0-1) using the following equation:

$$\bar{x} = \frac{x - m}{n - m} \tag{9}$$

Where $\bar{x}$ the normalized data x is the original data, and n, m are the maximum and minimum values of the original data, respectively.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

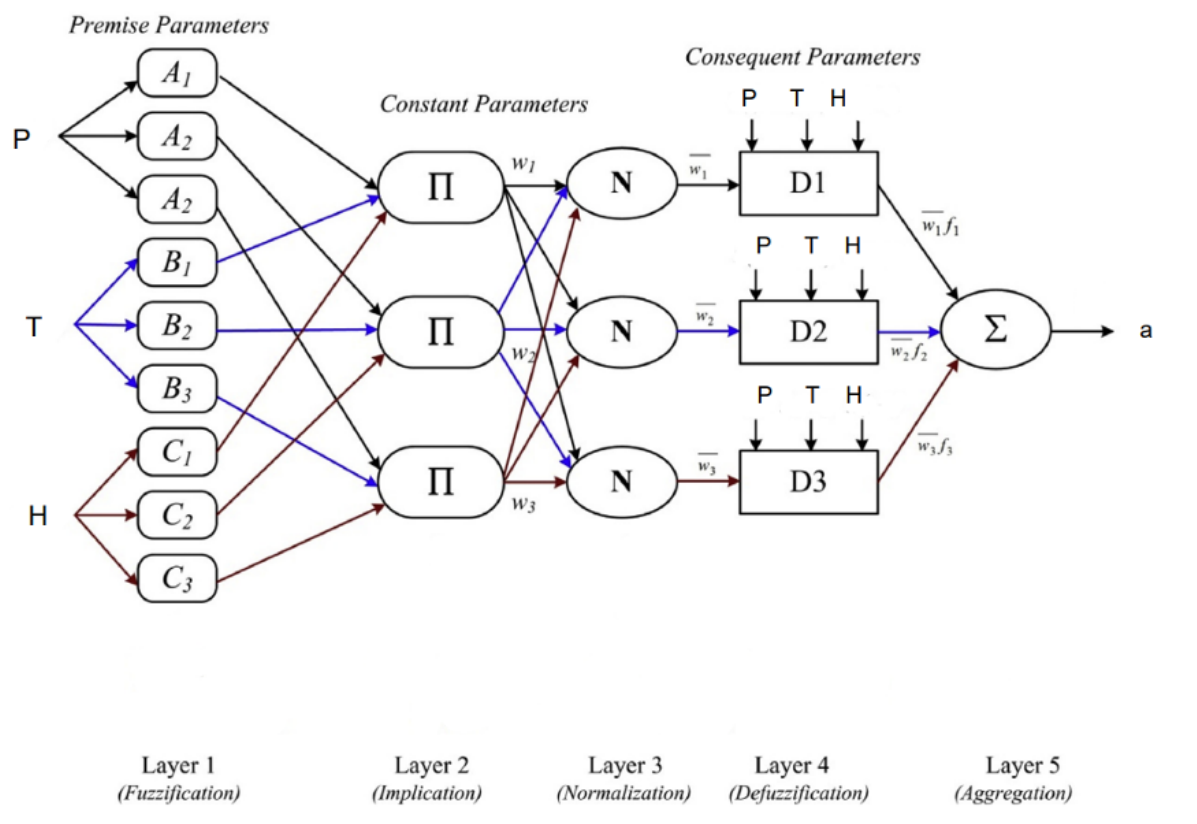

Figure 5. A framework of the ANFIS model.

2.3.2. SIMPLE MEMBERSHIP FUNCTIONS AND FUZZY RULES GENERATION TECHNIQUE (SMRGT)

The fuzzy-Mamdani method is used to construct the SMRGT model. A combination of expert judgment and data-driven experimentation determines both the fuzzy subset and the variable ranges in this model. This method streamlines incorporating the event's physics into a fuzzy model. The steps involved in the SMRGT procedure are as follows:

1. Define input and output variables: The first step is to define the input and output variables of the fuzzy logic system. This work used three inputs (Precipitation, temperature, and humidity) with one output (flow coefficient).

2. Determine membership functions: Membership functions (MFs) map input values to fuzzy sets. Five MFs were used and labeled as; Very low, Low, Medium, High, and Very high. Also, this step involves selecting the shape of the membership function, a triangular shape was selected.

3. Determine the key values: in this step, the unit width (UW), core value (C_i), the number of right-angled triangles (nu), the expanded base width (EUW), and key values (K_i) of the fuzzy sets were determined. Equations [10–18] were used to calculate the key values. Table 1 shows the obtained key values. These key values are the inputs of the model.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

$$Vr = (P,T,H)\max - (P,T,H)\min \tag{10}$$

$$Ci = K3 = \frac{Vr}{2} - (P,T,H)\min \tag{11}$$

$$UW = \frac{Vr}{nu} \tag{12}$$

$$O = \frac{UW}{2} \tag{13}$$

$$EUW = UW + 0 \tag{14}$$

$$K4 = Ki = C_i + 1 = \left(\frac{Ci - (P,T,H)\min}{2}\right) + (P,T,H)\min \tag{15}$$

$$K2 = C_i - 1 = (P,T,H)\max - \left((P,T,H)\max - \frac{Ki}{2}\right) \tag{16}$$

$$K1 = (P,T,H)\min + \left(\frac{EUW}{3}\right) \tag{17}$$

$$K_5 = (P,T,H)_{\max} - \frac{EUW}{3} \tag{18}$$

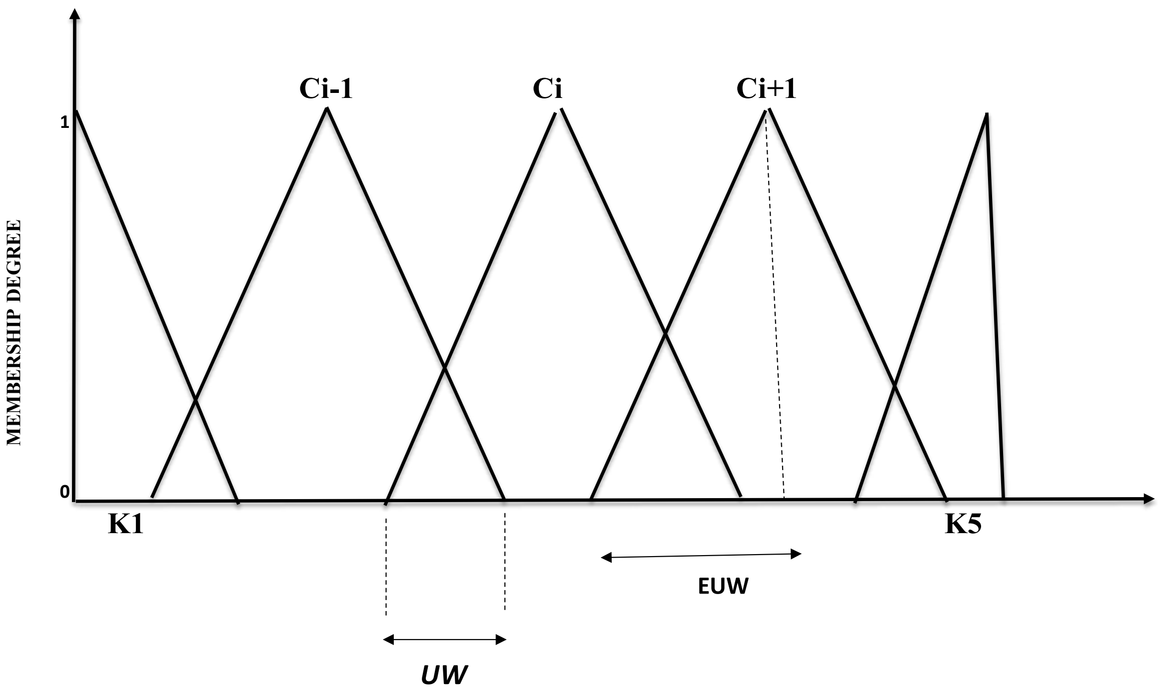

Figure 5. The parameters of the triangular MF.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

Table 1. key values of the SMRGT model.

	Ci-1 (K2)	Ci (K3)	Ci+1 (K4)	K1	K5
Precipitation	650	1100	1550	312.5	1887.5
Temperature	12.5	25	37.5	3.125	46.88
Humidity	25	50	75	6.25	93.75

4. Generate fuzzy rules: Fuzzy rules map input values to output values. Each rule consists of an antecedent (input) and a consequent (output). In this step, 125 rules were set in pertinent physical conditions such as "IF", "AND", and "THEN."

5. Run the model: MATLAB software was selected. As an operator, the Mamdani algorithm is implemented. The centroid method was selected for the defuzzification procedure. Input and output files prepared and added to the program with (.dat) extension. Then the program with the (.fis) extension is loaded. The (.m) extension file is prepared for running the prepared program. Model results can be obtained by running this file with the (.m) extension. Preparing the program with this procedure will reduce the trial and error process. Then the table of the fuzzy set was created.

2.3.3. GAUSSIAN PROCESS REGRESSION (GPR)

Numerous disciplines employ a potent instrument that can be considered a generalized regression model. This paper uses a Gaussian Process Regression (GPR) model to predict the flow coefficient values. Before explaining the Gaussian Process Regression, it is required to describe a regression model. In regression, i^{th} observation's (y_i) output is considered to be a function of the variables (x_i) input, plus some noise (ε_i).

$$y_i = f(x_i) + \varepsilon_i \tag{19}$$

The fundamental regression function is forecasted based on the input parameters and the given outputs. Once the regression model has been developed, a new value for the output variable can be determined for a given input variable. This is why regression models are widely used [23-25]. For GPR, it is assumed that the regression function (x) is derived from a Gaussian Process (GP) with a zero mean function and the covariance/kernel function (x, x').

$$f(x) \sim GP\left(0, k\left(x, x'\right)\right) \tag{20}$$

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

It is also assumed that the noise ε_i has a Gaussian distribution. The function (x, x') is known as a kernel function. This function represents the covariance between the x and x' values in a regression model given x and x' as inputs.

GPR offers numerous advantages over alternative regression models. For instance, it offers an indication of the uncertainty of the predictions, which is crucial for various practical applications. It can also model nonlinear relationships between input and output variables and accommodate missing data.

The procedure of the Gaussian Process Regression (GPR) in MATLAB can be summarised as follows:

i. Load the data of three inputs and one output (The training data was average monthly precipitation, temperature, and humidity records for 15 years) into MATLAB. The input data should be a size N x D matrix, where N is the number of data points and D is the number of input variables. The output data should be a column vector of size N x 1.

ii. Define the kernel function: the kernel function was defined using the 'make kernel' function.

iii. Specify the prior distribution: the prior distribution over the Gaussian process is specified using MATLAB's 'fitrgp' function. We can specify a mean function, a kernel function, and hyperparameters for the kernel function.

iv. Train the model: the GPR model is trained using the 'fitrgp' function. This function estimates the hyperparameters of the kernel function from the training data.

v. Make predictions: the trained GPR model uses the' predict' function to predict new input values. The 'predict' function returns a predicted mean and a variance for each input value.

vi. Load test data: 5 years of measurements of the abovementioned parameters were selected and loaded. Then the prediction was made. Cross-validation (v=5) was selected for GPR to protect the models against overfitting.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

```
% Load the input and output data
X = load('input_data.mat');
y = load('output_data.mat');

% Define the kernel function
kernel = fitrgp(X, y, 'KernelFunction', 'squaredexponential');

% Train the model
gprMdl = fitrgp(X, y, 'KernelFunction', 'squaredexponential');

% Make predictions
X_new = [1 2 3]; % example new input values
[y_pred, y_sd] = predict(gprMdl, X_new);

% Evaluate the model
mse = immse(y_pred, y_true);

% Load the trained GPR model
load('gpr_model.mat'); % the trained GPR model is stored in a .mat file

% Load the test data
X_test = load('test_input.mat'); % N x D matrix of new input data

% Make predictions
[y_pred, y_sd] = predict(gprMdl, X_test);

% Evaluate the predictions
y_true = load('test_output.mat'); % N x 1 vector of true output values
mse = immse(y_pred, y_true);
```

Figure 6. The generated code for the GPR model in MATLAB.

2.4. MODELS EVALUATION

Four parameters were used to evaluate the model's performance: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), the coefficient of determination (R^2), and Mean Square Error (MSE). They were given in Eq. (5-7). MAE, MSE, and RMSE are two measures of error. Thus ideal models would have MAE and RMSE values equal to zero. The coefficient of determination is the proportion of variability the

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

regression line indicates to the variability of data for linear regression. A regression line that is the mean value of data would have R^2=0, while an ideal model would have R^2=1.

$$MAE = \frac{1}{n}\sum_{1}^{n} |\, \mathrm{Ci},\ \text{measured} - \mathrm{Ci},\ \text{estimated}\ | \qquad (21)$$

$$MSE = \frac{1}{n}\sum_{1}^{n} (\mathrm{Ci},\ \text{measured} - \mathrm{Ci},\ \text{estimated}\ )^2 \qquad (22)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n} (Ci,\ \text{measured} - \text{Ci, estimated}\ )\text{^}2} \qquad (23)$$

3. RESULTS AND DISCUSSION

In this study, the flow coefficient in the Aksu river basin was estimated by using Adaptive Neural Fuzzy Inference System (ANFIS), Simple membership functions and fuzzy Rules Generation Technique (SMRGT), and Gaussian Process Regression (GPR) models. The results were compared with each other. The dataset belonging to the years 2000–2019 was used in modeling the SMRGT and GPR. For ANFIS, seven year's data from 2013-2019 were used; 70% was used for training and 30% for testing. Monthly Precipitation (P), Temperature (T), and Relative Humidity (H) were used as the input variables. To determine the success of the models used to estimate the flow coefficient value, RMSE (root mean square error), MAE (mean absolute error), MSE (mean square error), and R (correlation coefficient) were calculated, as explained in the previous section. The performance of the model results is shown in Table 2. When Table 2 was examined, all models gave similar results. According to the RMSE, MAE, MSE, and R criteria, the best results were obtained in the SMRGT, and the worst was in the GPR.

Table 2. The RMSE, MAE, MSE, and R^2 statistics of all models.

Models	Period	RMSE	MSE	MAE	R^2
ANFIS	Training	1.92	37	1.01	993
	Testing	15.67	2.45	12.15	561
	All data	8.53	728	4.19	863
SMRGT	All data	9.6	0.93	8.07	963
GPR	Training	26.9	7.24	20.26	0.61
	Testing	20.1	4.05	15.79	0.55

In ANFIS analysis, Gaussian parabolic 5 × 5 × 5 Membership Functions (MFs) and Grid partition section were analyzed with 100 iterations, assuming the output as linear.

 https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

Variation and scatter graphs for the ANFIS method are shown in Fig.7. The correlation coefficient of all data is seen as R: 0.863. As realized in the figure, ANFIS results were close to the observed values.

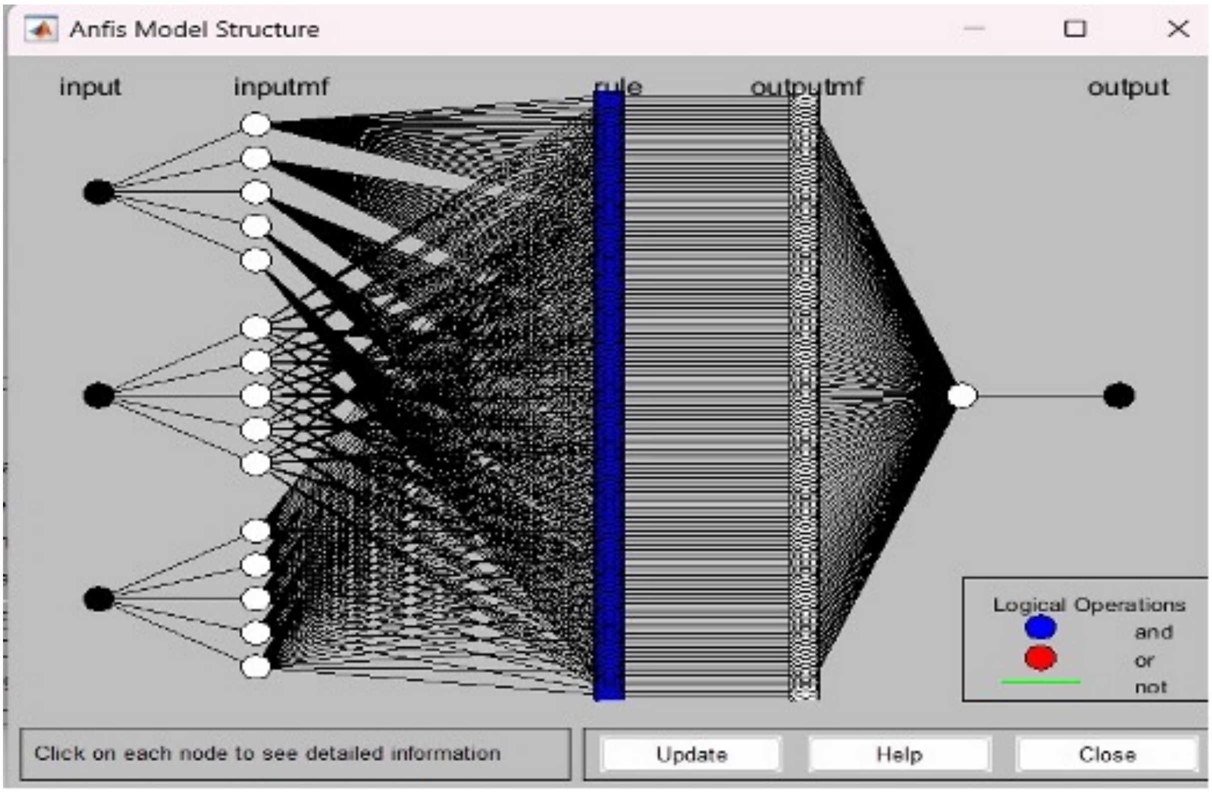

Figure 7. ANFIS structure uses three inputs and 5 MFs for each input with a type of Gaussmf.

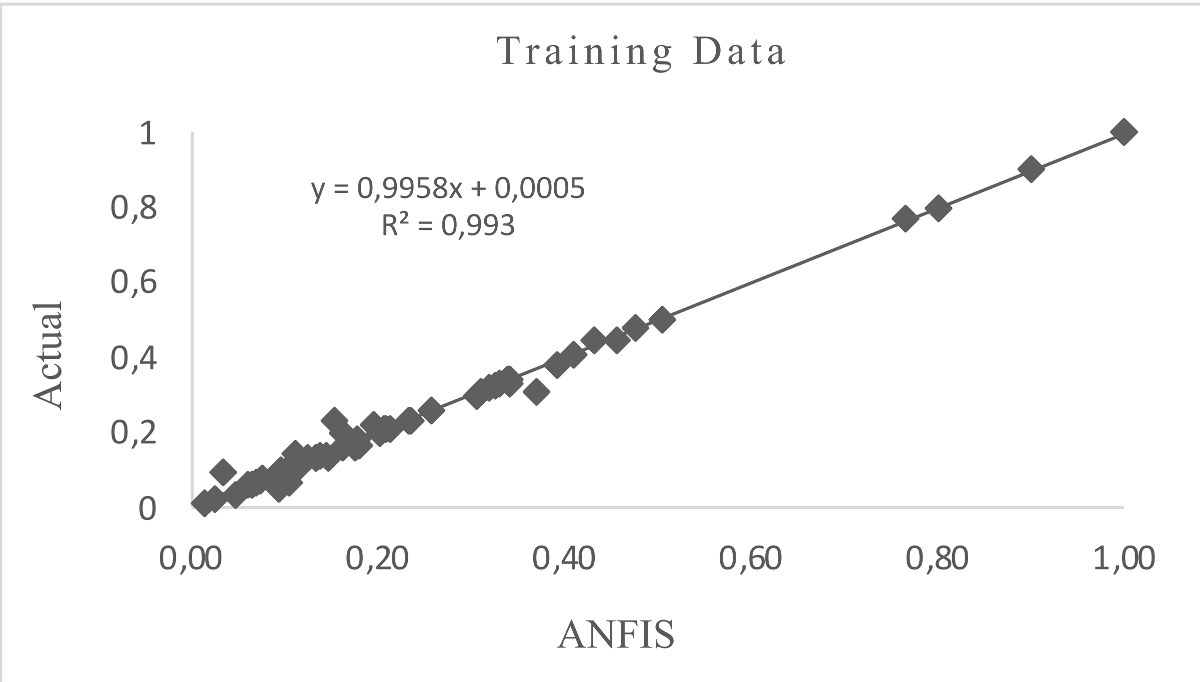

Figure 8. Scatter diagram of the trained data results.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

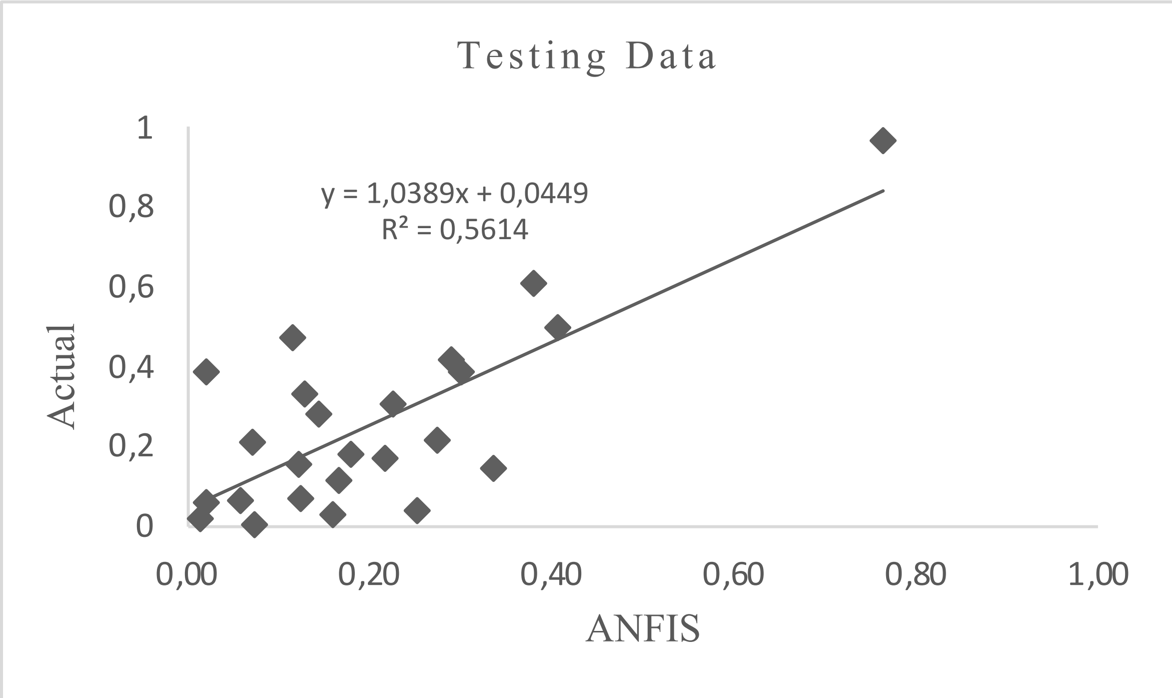

Figure 9. Scatter diagram of the tested data results.

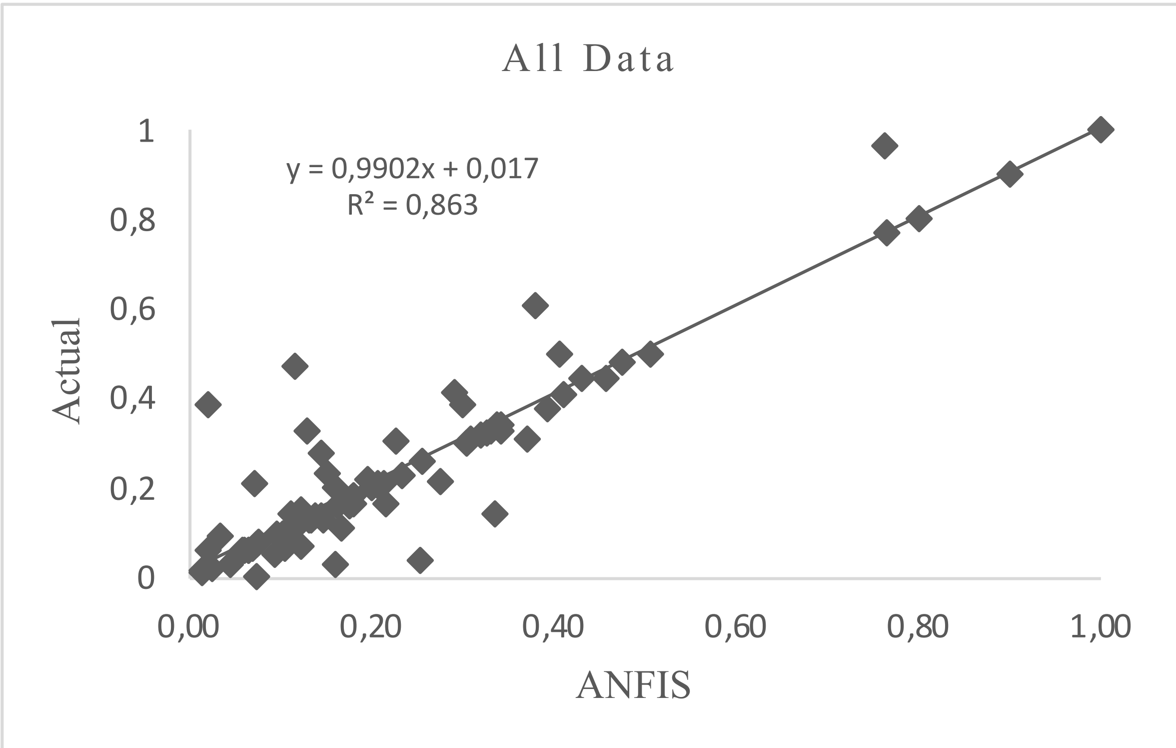

Figure 10. Scatter plot for all data results.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

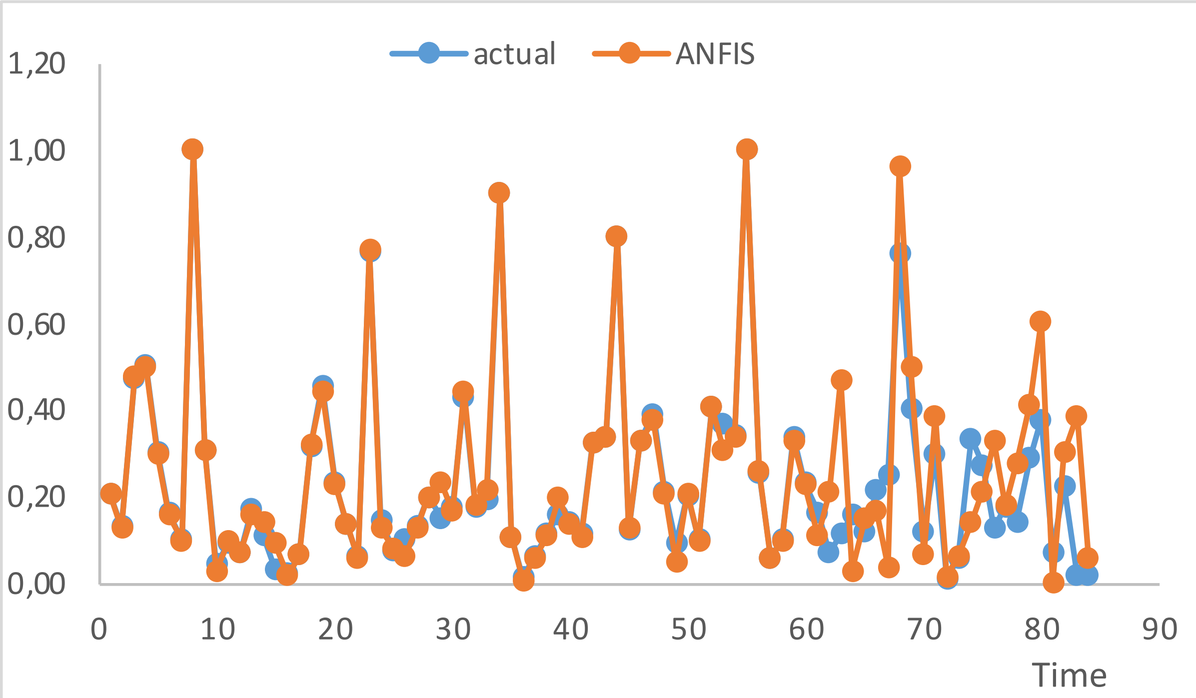

Figure 11. Variation graph of ANFIS model outcomes.

Simple Membership Functions and Fuzzy Rules Generation Technique (SMRGT) model results are given in Figure 12. When the scatter graph was examined, it was seen that the output values of the SMRGT model gave closer results to the actual values; also, the correlation coefficient was 0.96. In Table 2, it was found that the SMRGT model showed the best performance among all models. Compared to all models, it can be seen from Table 2 that SMRGT model results had the low error rates (RMSE: 9.6; MSE: 0.93; MAE: 8.07) and the highest correlation (R: 0.963). The figures show that the SMRGT approach shows almost the same trend as the actual values.

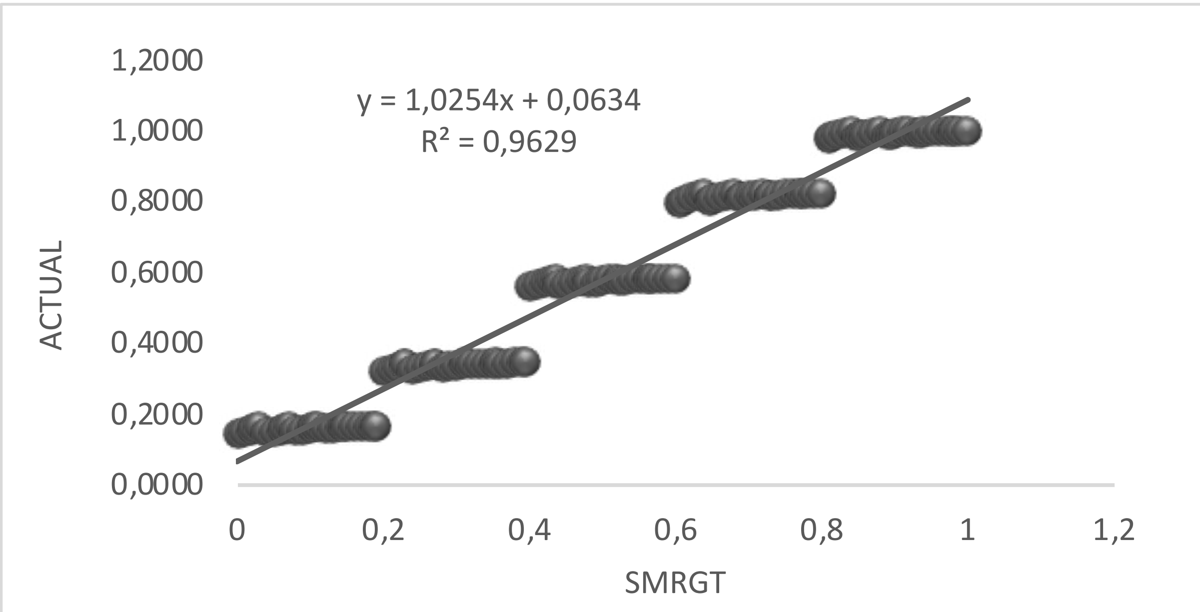

Figure 12. scatter diagram of SMRGT model.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

The forecasting result of the Gaussian Process Regression (GPR) model for training and testing data are given in Fig.13, 14. It can be seen clearly that some of the data fall along the regression line, while the rest were distributed far to the line. Moreover, the statistic error rate is higher than SMRGT and ANFIS models with a lower correlation coefficient (R^2:61 training; R^2:55 testing). In other words, the predicted data values are not highly fitted with the actual date values as in SMRGT and ANFIS.

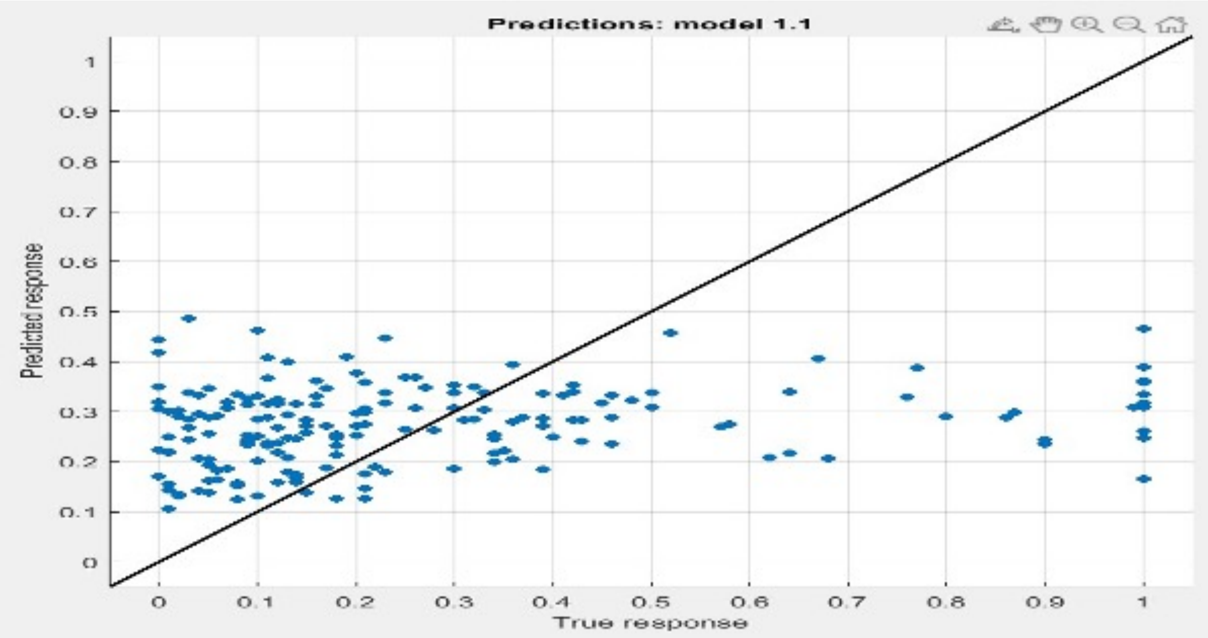

Figure 13. the prediction result of the training set for the GPR model.

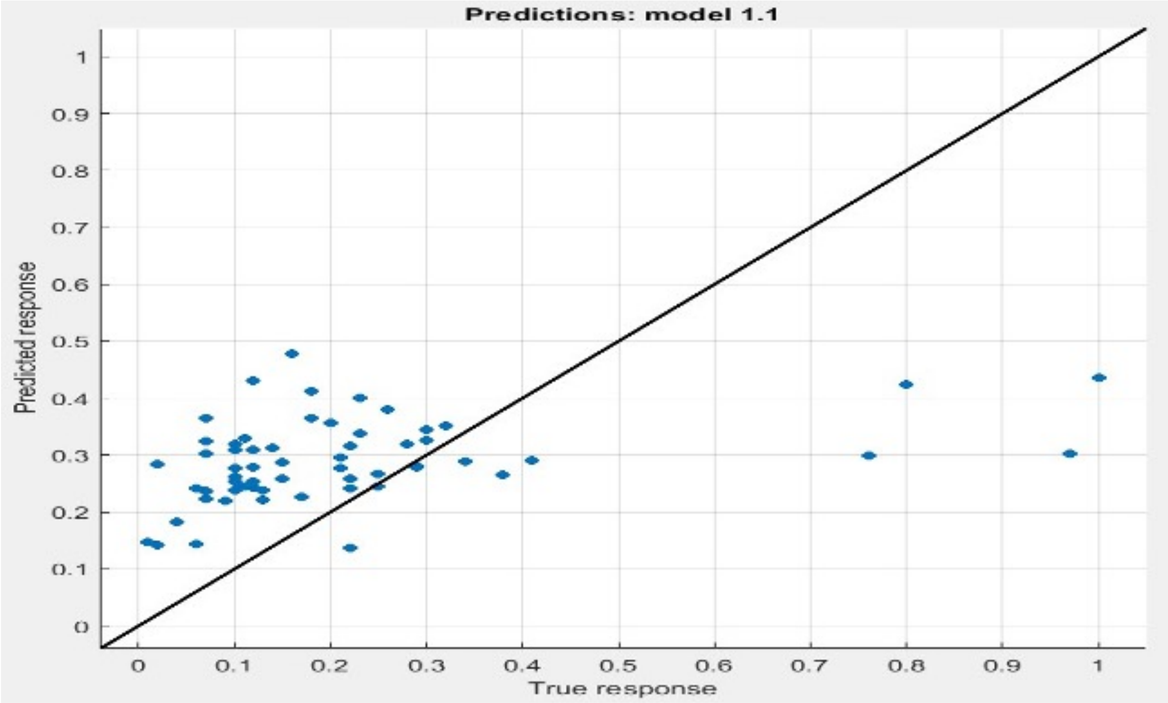

Figure 14. the prediction result of the training set for the GPR model.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

4. CONCLUSION

Determining river flows and variations is important to use water resources efficiently, construct water structures, and prevent flood disasters. However, accurate flow prediction is related to a good understanding of the hydrological and meteorological characteristics of the river basin. Artificial intelligence has taken a large portion of climate and water science research. The nonlinearity of meteorological variables and their dependency on many other properties and variables render machine-learning models beneficial and efficient in this field. This study used monthly average temperature, precipitation, and relative humidity values for flow coefficient prediction. The dataset belonging to the year range of 2000–2019 in the Aksu River Basin was examined. The flow coefficient was estimated by using Adaptive Neuro-Fuzzy Inference System (ANFIS), Simple Membership Functions and Fuzzy Rules Generation Technique (SMRGT), and Gaussian Process Regression (GPR) models.

The best models were found by applying statistical indicators such as RMSE, MAE, MSE, and R. The SMRGT model performed well with a low error rate and high correlation coefficient.

ANFIS model showed good performance with a lower error rate, but the correlation coefficient was lower than the SMRGT model.

The GPR model performed worse than other models; the error rate was higher, and the correlation coefficient was very low. The reason might be in using an inappropriate kernel function or overfitting or underfitting the data; when the model is too complex or has too many hyperparameters, it may fit the noise in the data rather than the true relationship between the input and output variables. It is important to examine the data and the statistical model carefully is used to identify the reasons for higher statistical errors and lower correlation coefficients. Appropriate statistical techniques and data-cleaning methods can address these issues and improve the accuracy of the results.

For future works, Scientists can improve the predictability of the flow coefficient by looking into the relationships between other variables and precipitation. These variables include wind speed, permeability, and land use information. Understanding what causes flash floods is essential in urban areas where rapid housing development or the conversion of marginal areas into housing is of interest. The overall study demonstrated the predictive ability of fuzzy logic models (SMRGT and ANFIS). Even though the available data size is relatively small, the prediction of the flow coefficient yields very good results and high performance. If more data becomes available, successful models can be used to estimate more accurately. The similarity between statistical parameters for the SMRGT model suggests that it can be relied upon to calculate the flow coefficient. The implementation of the algorithm demonstrates that model calibration does not require additional data. To begin using SMRGT, the modeler's knowledge is required.

https://doi.org/10.17993/3ctecno.2023.v12n2e44.125-146

REFERENCES

(1) Alexander, G. N. (1972). Effect of catchment area on flood magnitude. Journal of Hydrology, 16, 225-240.

(2) Ghazavi, R., Rabori, A. M., & Reveshty, M. A. (2016). Modelling and assessment of urban flood hazards based on rainfall intensity-duration-frequency curves reformation. Natural Hazards and Earth System Sciences Discussions. Advance online publication. https://doi.org/10.5194/nhess-2016-304

(3) Mousavi, S. M., Roostaei, S., & Rostamzadeh, H. (2019). Estimation of flood land use/land cover mapping by regional modelling of flood hazard at sub-basin level case study: Marand basin. Geomatics, Natural Hazards and Risk, 10, 1155-1175.

(4) Masoudian, M., & Theobald, S. (2011). Influence of land surface topography on flood hydrograph. Journal of American Science, 7, 354-361.

(5) Ngamalieu-Nengoue, U. A., Martinez-Solano, F. J., Iglesias-Rey, P. L., & Mora-Melia, D. (2019). Multi-objective optimisation for urban drainage or sewer networks rehabilitation through pipes substitution and storage tanks installation. Water, 11, 935. https://doi.org/10.3390/w11050935

(6) Bedient, P. B., Huber, W. C., & Vieux, B. E. (2008). Hydrology and Floodplain Analysis. Upper Saddle River, NJ: Prentice Hall.

(7) Junior, A. R. B. (2022). Elementos de Hidrologia aplicada. Editora Blucher.

(8) Sen, Z., & Altunkaynak, A. (2005). A comparative fuzzy logic approach to runoff coefficient and runoff estimation. Journal of Hydrology, 312, 1-14.

(9) Blume, T., Zehe, E., & Bronstert, A. (2007). Rainfall-runoff response, event-based runoff coefficients and hydrograph separation. Hydrological Sciences Journal, 52, 843-862. https://doi.org/10.1623/hysj.52.5.843

(10) Che, D., Liang, A., Li, X., & Ma, B. (2018). Remote sensing assessment of safety risk of iron tailings pond based on runoff coefficient. Sensors, 18, 4373. https://doi.org/10.3390/s18124373

(11) Campos, M. De R., & Machado, R. E. (n.d.). Time distribution of intense rainfalls at Campinas, Brazil. International Journal of Advanced Engineering Research and Science (IJAERS), 5(11).

(12) Lallam, F., Megnounif, A., & Ghenim, A. N. (2018). Estimating the runoff coefficient using the analytic hierarchy process. Journal of Water and Land Development, 38, 67-74. https://doi.org/10.2478/jwld-2018-0043

(13) Jang, J. S. R. (1993). ANFIS: adaptive network-based fuzzy inference systems. IEEE Transactions on Systems, Man, and Cybernetics, 23, 665-685.

(14) Jang, J. S., & Sun, C. T. (1995). Neuro-fuzzy modeling and control. Proceedings of the IEEE, 83(3), 378-406.

(15) Huang, M., Zhang, T., Ruan, J., & Chen, X. (2017). A new efficient hybrid intelligent model for biodegradation process of DMP with fuzzy wavelet neural networks. Scientific Reports, 7, 41239. https://doi.org/10.1038/srep41239

(16) Toprak, Z. F. (2009). Flow Discharge Modeling in Open Canals Using a New Fuzzy Modeling Technique (SMRGT). CLEN – Soil, Air, Water, September. https://doi.org/10.1002/clen.200900146

(17) Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics, 15(1), 116-132. https://doi.org/10.1109/TSMC.1985.6313399

(18) Jalalifar, H., et al. (2011). Application of the adaptive neuro-fuzzy inference system for prediction of a rock engineering classification system. Computers and Geotechnics, 38(7), 783-790.

(19) Rajesh, S., et al. (2012). Estimation of elastic constant of rocks using an ANFIS approach. Applied Soft Computing, 12(1), 40-45.

(20) Rai, A. A., et al. (2015). Prediction models for performance and emissions of a dual fuel CI engine using ANFIS. Sadhana, 40(2), 515-535.

(21) Sankar, S. G., et al. (2018). Prediction of PM 2.5 using an ensemble of artificial neural networks and regression models. Journal of Ambient Intelligence and Humanized Computing, 9, 1-11.

(22) Xianghong, W., & Baozhen, W. (2019). Research on prediction of environmental aerosol and PM2.5 based on artificial neural network. Neural Computing and Applications, 31(12), 8217-8227.

(23) Raposo, F., Borja, R., & Ibelli-Bianco, C. (2020). Predictive regression models for biochemical methane potential tests of biomass samples: Pitfalls and challenges of laboratory measurements. Renewable and Sustainable Energy Reviews, 127, 109890.

(24) Liu, K., Hu, X., Wei, Z., Li, Y., & Jiang, Y. (2019). Modified Gaussian process regression models for cyclic capacity prediction of lithium-ion batteries. IEEE Transactions on Transportation Electrification, 5(4), 1225-1236.

(25) Chen, X., Huang, J., & Yi, M. (2020). Cost estimation for general aviation aircrafts using regression models and variable importance in projection analysis. Journal of Cleaner Production, 256, 120648.