Open Access
Volume 5 Issue 2
Research Article

Journal of Water Engineering and Management
ISSN: 2582-6298

Modeling Stage–Discharge and Runoff–Sediment Dynamics using Soft Computing and Statistical Approaches in the Nagavali River Basin, Andhra Pradesh

Anita[1], Pravendra Kumar[1]
[1]Department of Soil and Water Conservation Engineering, G. B. P. U. A & T., Pantnagar, India

Article Info

Article History:
Received on: March 12, 2024
Revised on: July 28, 2024
Accepted on: September 25, 2024
Published on: September 30, 2024

Published by Academic Hope

***Corresponding author:** Pravendra Kumar
Email: drpravendra123@gmail.com

How to Cite:
Anita and Kumar, P. 2024. Modeling Stage–Discharge and Runoff–Sediment Dynamics using Soft Computing and Statistical Approaches in the Nagavali River Basin, Andhra Pradesh, Journal of Water Engineering and Management 5(2):19-30. DOI: https://doi.org/10.47884/jweam.v5i2pp19-30

Abstract

Soil and water are critical natural resources underpinning agricultural productivity and ecological sustainability, particularly in monsoon-dominated regions such as the Nagavali River basin in eastern India. Effective estimation of runoff additionally and sediment yield hand in hand is essential for workable watershed planning, hydraulic structure design, and sediment management, yet remains evergreen challenging due to the stochastic and nonlinear nature of hydrological processes. The Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS), two data-driven soft computing techniques, are assessed in this study in conjunction with Multiple Linear Regression (MLR) and Sediment Rating Curve (SRC) methods for modeling daily stage–discharge and runoff–sediment relationships. The Water Resources Information System (WRIS) provided monsoon-season stage, discharge, and suspended sediment content data for a total of twelve years (2001–2012), with 2001–2009 used for training and 2010–2012 for testing. While ANFIS models used Gaussian and triangular membership functions with hybrid learning, ANN models used feed-forward back-propagation with Levenberg–Marquardt optimization. RMSE, correlation coefficient (r), coefficient of efficiency (CE), and pooled average relative error (PARE) were applyed to evaluate the model's performance. Results indicate that soft-computing models outperform traditional approaches for both runoff and sediment prediction. ANFIS with triangular membership functions demonstrated the highest accuracy, followed by double-hidden-layer ANN. MLR provided acceptable results, whereas SRC exhibited limited capability due to pronounced nonlinearity in sediment–runoff dynamics. Antecedent flow and sediment conditions (up to three- models day lag) significantly influenced current-day discharge and sediment yield. Overall, ANFIS and ANN proved to be robust, efficient, and reliable tools for hydrological forecasting in the Nagavali basin, highlighting their potential for adoption in similar tropical catchments to enhance sediment–runoff prediction and support sustainable watershed management.

Keywords: Artificial neural network: Adaptive neuro-fuzzy inference system, Runoff–sediment modelling, Stage–discharge relationship, Nagavali river basin.

Introduction

Soil and water are two essentially important natural resources that form the foundation of agricultural productivity and ecosystem stability. With the growing population and increasing pressure on water resources, the judicious and efficient utilization of land and water has become indispensable for sustainable development. Runoff serves as the primary driving force behind the detachment and transportation of soil particles, thereby accelerating the processes of soil erosion and sediment generation. The eroded and dissolved soil material transported through surface flow ultimately enters the stream network and is collectively referred to as sediment yield from the catchment. Sedimentation, as a cumulative and aggravated process, can cause severe and often irreversible damages to hydraulic structures. It leads to the accumulation of sediments behind reservoirs and barriers, a consequent reduction in storage capacity, increased maintenance costs for irrigation canals, and even damage to ports, coasts, and other water infrastructure. Therefore, accurate and precise estimation of runoff and eventual sediment yield is crucial for effective watershed management and design of water resource structures.

Several empirical and conclusive data-driven pattern have developed several models to estimate runoff rates and sediment transport processes. Among them, the stage–discharge relationship provides a simple, cost-effective, and reliable technique for discharge estimation when properly calibrated. Such rating curves and modeling approaches are essential tools in hydrological analysis, sediment management, and catchment-scale planning. Over the decades, a reckless wide range of models have been evolved and studied, broadly categorized as conceptual and concise models, which is physically based models, and data-driven models. Physically based and conceptual models simulate watershed processes by representing physical laws and empirical relationships, whereas data-driven models focus on learning statistical or computational relationships directly from observed input–output datasets, without explicitly describing the underlying physical processes. Such models are particularly advantageous when detailed process information is limited or when data are complex and nonlinear.

The ANN, ANFIS and FL models are examples of data-driven methods that have shown great promise for simulating dynamic and nonlinear hydrological processes. These models are capable of extracting hidden patterns from hydrological time series data and establishing robust predictive relationships between rainfall, runoff, and sediment yield. Their key advantage and significance lies in their ability to self-organize and adaptively learn the underlying system behavior without requiring prior knowledge of explicit functional forms. Earlier studies, such as that of Sudheer et al. (2002), reported that ANN models provided superior performance compared to traditional empirical equations in estimating sediment concentration. The ANFIS combines and reconciles the advantages of both methods into a single framework by fusing the reasoning and interpretability of fuzzy logic with the learning powers of meshy neural networks. ANFIS, which is based on the TSK and fuzzy inference system (Loukas, 2001), uses a hybrid and advanced learning approach that combines or recapitulates back-propagation and even are least-mean-square optimization techniques with language rules from fuzzy logic. Because of this, it can effectively record intricate nonlinear interactions. Simultaneously, a traditional yet powerful statistical method for modelling and portraying the combined effects of several explanatory variables and a specific dependent variable is MLR.

MLR has been extensively used in hydrological applications for predicting flow, sediment, and water quality parameters based on available datasets. The uses of ANN, ANFIS, and MLR provides a comprehensive comparative framework for modeling and forecasting hydrological responses, particularly for runoff and sediment yield prediction in diverse watershed conditions. Several researchers have modelled runoff using ANN model (Bhattacharya et al., 2005; Raghuwanshi et al., 2006; Agarwal et al., 2009; Nourani, 2009; Bisht et al., 2010; Kisi et al., 2004; Bharti et al., 2017). Neuro-fuzzy models have also been used to forecast river flow, sediment rating curve and inflow (Shafie et al., 2007; Cobaner et al., 2009; Firat et al., 2010; Chang and Tsai, 2016). Eisazadeh et al. (2013) analysed the application of multiple linear regression (MLR) and neural network (NN) methods for calculating sediment yield. Patil and Valunjkar (2014) used a multilayer perceptron (MLP) for the Gunjwani watershed in the Bhima sub-basin of Maharashtra, India, to forecast next-day runoff.

In order to calculate the daily (SSC) at the Tekra gauging site on the Pranhita River, a significant tributary (branch) of the Godavari basin in Andhra Pradesh, India, Malik et al. (2017) created a set of computational models (approaches). MLR, MNLR, SRC, MLPNN and CANFIS were all built (developed) and tested in this analysis. The study used historical data from the Nagavali River to examine the effectiveness of various data-driven (machine-based) and statistical methodologies (techniques) for calculating discharge and SSC. ANN, ANFIS, MLR and SRC were among the models that were examined, offering useful benchmarks (comparisons) for sediment prediction.

To appraise the performance of these models, statistical metrics, including (RMSE), (r), coefficient of efficiency CE and PARE, are deployed. The overarching objective is to generate stage–discharge models utilizing ANN, ANFIS, and MLR for the Nagavali River basin, alongside the development of runoff–sediment models involving ANN, ANFIS, MLR, and SRC methods. Following the formulation, the models undergo validation against observed hydrological data, with their performance assessed through comparative analysis to determine the most effective and precise approach for discharge and sediment estimation in the study region.

Materials and Methods

Description of the Study Area

The Nagavali river basin is a medium-sized, that flows eastward situated in peninsular India, with significant ecological and geographical relevance. It spans an area of 4,462 km² in Odisha and 5,048 km² in Andhra Pradesh, showcasing its interstate importance. The river has its source in the Eastern Ghats, specifically close to Lakhbahal in the district Odisha of Kalahandi, at a height of approximately 1,300 meters. Flowing towards the east, it gets in Andhra Pradesh near the Parvathipuram mandal in the Srikakulam district, and concludes its journey by discharge into the Bay of Bengal, which is located near Mahfuzbandar in the same district. In terms of its geographical positioning, the Nagavali river basin is surrounded by other significant water bodies, including the Godavari River to the west and the Bay of Bengal to the east, indicating its integration into the larger hydrological framework of the region. The Champavathi and Peddagedda rivers define the southern boundary, while the Vamsadhara and Mahanadi rivers border the north. The Nagavali river basin is a crucial part of the local ecology and hydrology because of its complex hydrographic setting.

The research area drains parts of Andhra Pradesh's Srikakulam, Vijayanagaram, and Visakhapatnam districts as well as Orissa's Kalahandi, Rayagada, and Koraput districts. The examined stretch is roughly 256 kilometers long, and the catchment area is 9510 square kilometers in total. Orissa and Andhra Pradesh, two Indian states, make up the majority of the watershed covered by the geographic area described in the paper. It states that the watershed's first 161 kilometers are in Orissa, and the others are in Andhra Pradesh. Barha, Baldiya, Satklnala, Sitagurha, Srikona, Janjhvati, Gumidigedda, Vottigedda, Suvarnamukhi, Vonigedda, Vagavathi, and Relligedda are among the notable streams that contribute to this watershed. The hydrological characteristics and significance of these tributaries in the entire watershed system are highlighted in this data. Geographically, the region is located between latitudes 18° 10' and 19° 44' N and longitudes 82° 52' and 84° 05' E. Figure 1 shows the precise location.

With three different seasons—summer (March to May), monsoon (June to September), post-monsoon (October to November), and winter (December to February)—the Nagavali river basin has a tropical wet climate. The typical yearly rainfall in this basin is 1131 mm. The South-West monsoon has a major effect on this region, bringing heavy rainfall that enhances the region's general climate. Summertime temperatures can be as high as 37 °C and as low as 25 °C. On the other hand, temperatures drop throughout the winter, with highs of 20 °C and lows of 14 °C. Around 96% of the total soil coverage in the research region is made up of a variety of soil types, such as sandy loams, loamy soils, and red soils. The region has a variety of classes of wastelands in addition to these soil types, including gullied lands, salt-affected zones, degraded pastures, shrub-dominated lands, waterlogged and marshy lands, barren rocky regions and mining and industrial wastelands. Additionally, 68,641 hectares, or 12% of the basin's total area, are covered by forests.

Data Acquisition and Analysis

The Water Resources Information System of India provided Hydrological information on stage, discharge, and suspended sediment every day content throughout the monsoon season, which ran from June 1 to September 30, 2001 to 2012. Stage graphs produced by an automatic water stage recorder were used to calculate the stage, which was measured in meters, and the discharge, which was recorded in cubic meters per second (m^3/s). A 1-liter water sampler was used to measure the sediment load, which was represented in g/l as the total suspended sediment content. For analysis, the complete dataset covering the 12-year monsoon season has been split into two separate sets.

In this investigation, sediment, discharge, and stage data from 2001 to 2009 were employed to calibrate (train) the models. Data from 2010 to 2012 were used to verify (test) the developed models. In particular, 25% of the data from 2001 to 2009 was set aside for validation throughout the testing phase, and the remaining 75% was used for model calibration.

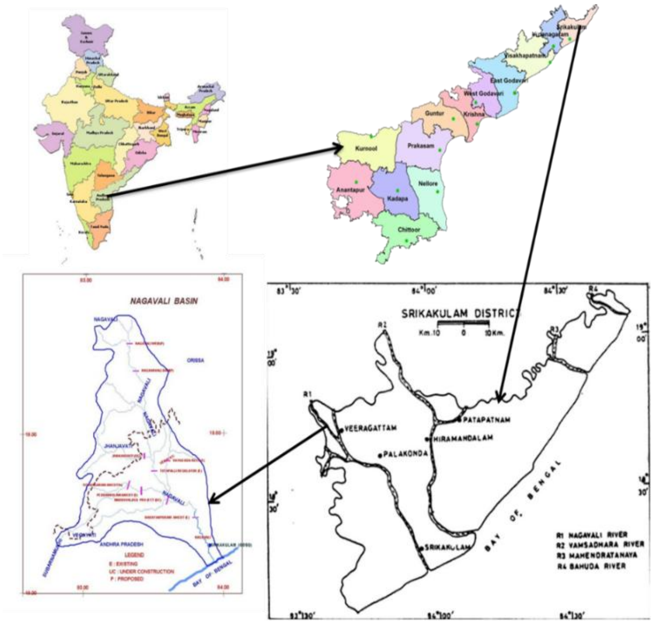

Fig. 1 The study area's location

Artificial Neural Networks (ANNs)

A computational model created to mimic parts of the human brain, especially its learning processes, is called an ANNs. These networks are made up of linked artificial neurons that exchange messages with one another. The neural network can adapt to various inputs and learn from them over time because the connections between these neurons are linked to numerical weights that may be changed by experience. A crucial feature that enables artificial neural networks to enhance their performance based on prior knowledge and data is this flexibility. All things considered, they play a crucial role in machine learning by simulating how the human brain functions to promote learning.

A "learning rule" that modifies the connection weights in response to the input patterns it receives is how the ANN functions. The main purpose of ANNs, which are is formed by several tightly connected, nonlinear operational units called neurons, is to map input sets to output sets. Incoming signals are amplified by predetermined weights throughout this phase before being sent to the neurons. These signals are combined (summed) at the neuron level to provide the net input, which is further processed by an activation function to generate the final output. This methodical technique demonstrates the fundamental ideas of artificial intelligence and machine learning by enabling the ANN to handle information efficiently and change over time.

The neuron then passes the net input through an activation function $f()$, which computes the final output y of the node as,

$$y = f(net) \quad \dots (1)$$

A neuron's activation level can be converted into an output using the activation function, also known as the transfer function. Every neuron has an activation function that controls how it responds to different stimuli. The most common kind used in neural networks is the log sigmoid function. This function effectively achieves a balance between linear and nonlinear behavior and is described as strictly rising. It is especially well-suited for multilayer networks that use the back-propagation technique for training because of its differentiable character, which enables efficient learning and optimization within these models.

The log sigmoid function is defined as,

$$f(net) = \frac{1}{1+e^{-net}} \quad \dots (2)$$

The step activation function and the log sigmoid activation function are similar, but the log sigmoid activation function adds another area of uncertainty. one or more hidden layers an input layer, an output layer and positioned in between are the several layers of neurons that make up a multilayer feed-forward network. Through connections that are determined by particular weights, each of these layers is tightly connected to the layer before it. This structure is demonstrated in Fig. 2. which shows a network with a single hidden layer and highlights the fundamental architecture and connectivity present in such neural network designs.

Among the most crucial stages in creating an (ANN) model is training its weight matrix. Supervised learning and unsupervised learning are the two main categories of learning algorithms that can be used to complete this training process. Supervised learning enables immediate feedback and correction throughout the learning process by training the network using a predetermined set of input patterns and their corresponding known output patterns. Unsupervised learning, on the other hand, doesn't require an outside teacher. In this paradigm, the ANN relies on its capacity to find correlations without explicit feedback or instruction in order to automatically recognize patterns and regularities within the input data. The system can develop its own understanding based on the intrinsic structure of the input space thanks to this learning autonomy.

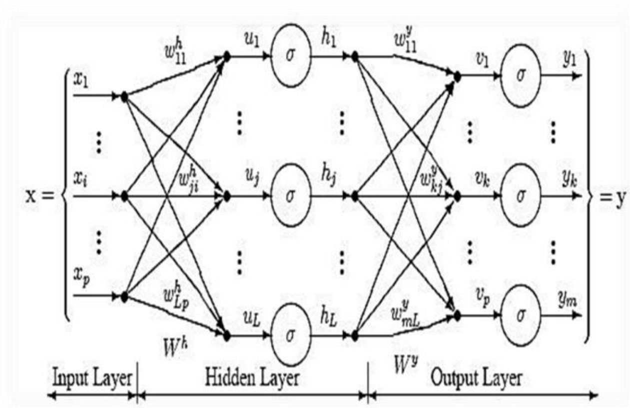

Fig. 2 Multilayer artificial neural network

Neural networks are distinguished by their learning algorithms, the most common of which is supervised learning. Reinforcement learning works best in pattern classification tasks, in contrast to unsupervised learning, which doesn't possess the paired input-output data required to train the networks efficiently. Supervised learning techniques are very good at solving time series forecasting problems since they require both inputs and appropriate outputs for network training. The Levenberg-Marquardt learning algorithm, that is described in the following sections, is especially used in this study.

Levenberg–Marquardt

In 1994, Hagan and Menhaj presented the Levenberg–Marquardt algorithm (LM), a sophisticated higher-order adaptive technique for neural network training that aims to minimize the mean square error (MSE). This algorithm belongs to the group of learning strategies known as pseudo second-order approaches.The Levenberg-Marquardt approach improves weight adjustment by combining gradient and curvature information of the error surface, in contrast to typical gradient descent algorithms that only use a local essessment of the performance surface's slope. This dual strategy makes it easier to find the best weight combinations, which eventually improves neural network training performance. Because of this dual consideration, the LM method can traverse the weight space more effectively, leading to faster convergence and better neural network training results.

The Levenberg-Marquardt (LM) approach's ability to transition to gradient search techniques when the performance surface's local curvature deviates from a parabolic shape—a scenario that frequently arises in neural computing tasks—is one of its main advantages. According to Ham and Kostanic (2001), the training of the Multilayer Perceptron (MLP) must be framed as a nonlinear optimization problem in order to execute the LM algorithm. However, a major drawback of the LM approach is the high computing load needed for matrix inversion, particularly when working with a large number of variables—often thousands. When trying to train neural networks effectively, this intricacy can provide difficulties.

$$W_{k+1} = W_{k+1} W_k - \left(J_k^T J_k + \mu I\right)^{-1} J_k^T \quad \ldots (3)$$

In this case, the algorithm's behavior during training is largely dependent on the parameter μ. The algorithm functions according to the LM (Levenberg-Marquardt) approach, which is used to effectively solve non-linear least squares problems, when μ is set to 0. On the other hand, the algorithm changes to function more like the steepest descent method when the value of μ climbs considerably. This change implies that the algorithm modifies its strategy according to the particular value of μ, affecting its efficiency and computing approach in optimization jobs.

Adaptive Neuro-Fuzzy Inference System (ANFIS)

An (ANFIS) is a sophisticated artificial intelligence model that synthesizes elements from both (ANN) and(FL). Introduced by Jang in 1993, ANFIS has significantly impacted various applications, particularly in modeling hydrological processes. One of the primary advantages of (FL) within this system is its capability to transform qualitative knowledge and subjective human observations into quantitatively precise analyses. Despite these strengths, ANFIS encounters challenges that stem largely from fuzzy logic itself. A critical limitation is the lack of a clear methodology to effectively convert human thought processes into a structured rule-based fuzzy inference system (FIS). Additionally, the adjustment of membership functions (MFs)—integral components of fuzzy logic—can be a labor-intensive and time-consuming process, complicating the practical implementation of the system.

The Takagi-Sugeno type fuzzy (FIS), on which ANFIS is based, expresses each rule's output as a the input variables are combined linearly with a constant term. The total output of the system is derived from the average weighted of the outputs from each rule, allowing for greater versatility in modeling complex relationships. To illustrate, consider a simplified scenario with two inputs, denoted as x and y, and a single output, represented as f. A typical first-order Takagi-Sugeno fuzzy model might include two basic fuzzy if-then rules, serving to simplify the modeling process while retaining the essential characteristics of both ANN and FL. Overall, ANFIS stands at the intersection of these two domains, offering promising enhancements in the modeling and analysis of intricate systems.

Rule 1: if x is A_1 and y is B_1, then $f_1 = a_1x + b_1y + c_1$ … (4)

Rule 2: if x is A_2 and y is B_2, then $f_2 = a_2x + b_2y + c_2$ … (5)

where the output functions' parameters are a1, b1, c1, and a2, b2, c2, and the membership functions for inputs x and y are A1, A2, B1, and B2. The ANFIS architecture is divided into five layers. A fixed node is present in the other layers, but an adaptive node is present in the first and fourth layers. Here is a quick summary of each layer:

Multiple Linear Regression (MLR)

A well-known technique for hydrological prediction, the regression model can be applied to two or more variables that are systematically related by a linear connection. In this work, hydrological parameters like runoff and silt concentration were estimated using (MLR) analysis on the same data set. The following is the expression for the regression equation that was used:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \quad \text{… (6)}$$

in which β_0 is an intercept, β_1, β_2,, β_n are constants and x_1, x_2 and x_n are independent variables. It follows that each independent variable is assumed to have an additive effect on Y and to be linearly related to Y.

Sediment Rating Curve (SRC)

When sampling procedures are insufficient to adequately capture the continuous sediment concentration record, (SRC) are used to estimate suspended sediment loads. these curves are typically defined using a power function that links water and sediment discharge over long stretches of time, usually more than ten years. The crucial connection between sediment concentration or load and discharge can be represented mathematically, underscoring the importance of long-term data in evaluating sediment transport dynamics.

$$S_t = a\, Q_t^{\,b} \quad \text{… (7)}$$

Where S_t is the suspended sediment load at time t, Qt is the discharge at time t, and a and b are regression constant

Development of Models

In this investigation, the writers aim to forecast runoff and suspended sediment concentration by developing and implementing various predictive models, including (ANN), (ANFIS) and (MLR). The dataset utilized comprises daily stage, discharge, and suspended sediment concentration measurements collected over a twelve-year duration, specifically during the monsoon season from June 1, 2001, to September 30, 2012. The analysis involved the use of MATLAB (R2015a) for constructing the ANN and ANFIS models, while Microsoft Excel 2010 was employed for performing the regression analysis. The primary focus was on creating effective daily runoff prediction models based on daily stage (H) and discharge (Q) data collected during the monsoon periods within the specified timeframe. To develop these runoff prediction models, various input combinations were explored, including the current day's stage, stages from the previous one to three days, and runoff values from the previous one to two days, with the current day's runoff serving as the output variable. The resulting model can be expressed in terms of these inputs and outputs, allowing for accurate predictions of runoff based on historical data.

$Q_t = f\,(H_t, H_{t-1}, H_{t-2}, H_{t-3}, Q_{t-1}, Q_{t-2},)$ … (8)

The variables H and Q, which stand for the stage at times t, t-1, t-2, and t-3 and runoff at times t, t-1, and t-2, are defined in the given text. The stage and runoff training data's patterns serve as the foundation for the function "f," which links these variables. The minimum and maximum values of the model's inputs, which are restricted to a specific range, fluctuate greatly depending on the variables. The input data were standardized to a common scale to make sure that no single variable unduly affects the model's output. In particular, a particular normalization algorithm was used to standardize the stage and runoff data between 0 and 1.

$$y_{norm} = \frac{y_i - y_{min}}{y_{max} - y_{min}} \quad \dots (9)$$

In sediment concentration prediction, the challenge lies in the complexity of the factors influencing the suspended sediment concentration on a given day. Key variables include Rainfall patterns, runoff, soil properties, and vegetation, all of which necessitate careful consideration of time lags for accurate modeling. The methodology for developing predictive models involves utilizing various combinations of runoff data from both the present and the past several days—specifically runoff over the past one, two, and three days—as well as sediment concentration values from the previous one, two, and three days. This comprehensive approach aims to accurately output the current day's sediment concentration, highlighting the interconnected nature of hydrological processes in sediment transport. The formulation incorporates the input data's upper and lower bounds to normalize the variables, enhancing the model's predictive capability.

$$S_t = f(Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, S_{t-1}, S_{t-2}, S_{t-3}) \quad \dots (10)$$

where S_t is the concentration of sediment at time t, t-1, t-2, and t-3, and Qt is the runoff.

Table 1 Specifics of the MLR model for SSC and runoff forecasts

Model	**Output-Input variables**
MLR-I (Runoff)	$Q_t = a_1 + b_1H_t + c_1H_{t-1} + d_1H_{t-2} + e_1H_{t-3} + f_1Q_{t-1} + g_1Q_{t-2}$
MLR-II (SSC)	$S_t = a_2 + b_2Q_t + c_2Q_{t-1} + d_2Q_{t-2} + e_2Q_{t-3} + f_2S_{t-1} + g_2S_{t-2} + h_2S_{t-3}$

In this document, the concept of (MLR) is presented through two models: MLR-I and MLR-II, which analyze the relationship between discharge and sediment levels over specified time periods. MLR-I utilizes input variables from current and previous days' stages, as well as discharges from prior days, to predict the current day's discharge. Specifically, the input variables for MLR-I include the discharge measurements from today, yesterday, two days ago, and three days ago, in addition to discharge values from the preceding one and two days, with the current day's discharge serving as the output. Conversely, MLR-II focuses on sediment levels, defining its output variable as the current day's sediment level. The input variables for this model encompass the current day's discharge, sediment levels from the current and previous days, and sediment values from the prior three days. This model similarly correlates the sediment levels to discharges observed over comparable time frames.

To support the training and evaluating MLR, ANN, and ANFIS, a daily stage time series dataset is utilized. This dataset is structured with parameters Hij indexed by year (i = 1 to M) and day (j = 1 to N), where M symbolizes the total number of years and N the total number of days during the monsoon season. The detailed temporal arrangement of the dataset is critical for effectively capturing the underlying patterns in the observed discharge and sediment data. Likewise, the necessary daily runoff time series Qij, where i = 1 to M and j = 1 to N, were likewise accessible. N = 122 days (June 1st to September 30th) in a year and M = 12 years (2001-2012) were discovered for the Srikakulam location's Nagavali river basin. The necessary time series data for suspended sediment content and daily discharge were accessible. For the Nagavali river basin, it was found that M = 12 years (2001-2012) and N = 122 days (June 1–September 30) every year. Neural networks with single and double hidden layers, which are ideally suited for nonlinear regression, were trained for the minimal error function using the back-propagation technique. The quantity of neurons in every hidden layer varies depending on the ANN's structure and how well they converge to the intended result. Table 2 enumerates the training variables for the ANN model along with their assigned values.

Table 2 Training variables and the values assigned to them for ANN

Variables for training	**Assigned values**
Neural network type	Feed forward back-propagation
Quantity of input	6
Quantity of output	1
Quantity of hidden layers	1, 2
Quantity of neurons in hidden layer	1 to 55
Total Quantity of layers	3, 4
Training function	Levenberg-Marquardt
Transfer function in hidden layer	Sigmoid
Transfer function in output layer	Pureline
Maximum epoch	1000

A variety of membership functions, such as triangular, trapezoidal, generalized bell, and Gaussian types, are used by the (ANFIS). There are two to three different inputs for each function. A fuzzy model of the Takagi-Sugeno-Kang type is used in the study, and training was done across 30 epochs the algorithm for back-propagation learning is utilized to evaluate the effectiveness of model training. Using both training and testing datasets, statistical indicators were used to forecast runoff and suspended sediment concentration in the research area in order to assess model performance. Table 3 provides specific training variables for the ANFIS architecture along with their assigned values.

Table 3 Training variables and the ANFIS models' assigned values

Training variables	**Assigned values**
Input membership function type	'trimf', 'trapmf', 'gbellmf' and 'gussmf'
Membership function based on input	2 to 3
Fuzzy model type	Takagi Sugeno
Epochs	30
Learning algorithm	Back-propagation

Results and Discussion

ANNs Based Runoff and (SSC) Prediction Models

The Gamma test is utilized to establish the input parameters necessary for modeling runoff-sediment and stage-discharge, while eliminating models that increase complexity without significant changes to results. By employing the Gamma test, one can predict the minimum modeling error in advance. A data-splitting technique facilitated the evaluation of the test's attributes, allowing for optimal input vector selection based on minimum gamma values, standard error, and V-ratio. Additionally, the research delved into optimal model selection and analyzed various input combinations to assess their effects on simulating runoff and (SSC).

A systematic evaluation of 2n-1 meaningful combinations of inputs was conducted to identify the best prediction models for runoff and SSC, through analysis based on gamma, standard error, and V-ratio. In total, seven input parameters were included in the runoff prediction model: today's stage (Ht), along with the stages and runoff values from the previous three days (Ht-1, Ht-2, Ht-3, Qt-1, Qt-2, Qt-3). Similarly, the SSC prediction model considered runoff values from today and the preceding three days (Qt, Qt-1, Qt-2, Qt-3), in conjunction with sediment values from the prior three days (St-1, St-2, St-3). The study ultimately examined 127 combinations for both models, finding that the optimal parameters for runoff prediction involved today's stage and the last two days of runoff, while the best parameters for SSC prediction incorporated today's runoff and sediment values from the previous two days. The models were implemented using (ANNs) designed in MATLAB (R2015a) to facilitate both runoff and SSC predictions.

The runoff prediction model was structured to include varied input variables such as today's runoff measurements and those from the previous three days. The Levenberg-Marquardt learning algorithm was applied to both single and double hidden layers, pureline activation was used in the output layer, and log-sigmoid activation functions were used in the hidden layers. The training process was configured for up to 1000 epochs applying the feed-forward back-propagation method. Model performance was evaluated through several statistical and hydrological metrics, including (RMSE), (r), (CE) and (PARE). Findings indicated that the architectures of (6-28-1) and (6-52-52-1) produced outstanding results, exhibiting superior correlation (r) and (CE) with a correspondingly lower RMSE compared to the other models assessed.

ANN Based SSC Prediction Models

The present study focuses on developing prediction models for (SSC) and runoff using (ANN) and (ANFIS) methodologies. For ANN models, inputs consisted of current and previous three days' runoff and sediment concentrations. The models employed a log-sigmoid activation function in hidden layers and a pureline activation function in the output layer, using the Levenberg-Marquardt learning algorithm up to 1000 maximum epochs. The evaluation of the ANN performance relied on hydrological and statistical indices, namely RMSE, correlation coefficient (r), coefficient of efficiency (CE), and predictive accuracy (PARE), identifying networks such as (7-55-1) and (7-30-30-1) as superior based on higher CE and r values along with lower RMSE.

Further, ANFIS models were also constructed for the same outputs, utilizing present and prior day stage, discharge, and sediment data from the Nagavali river basin. The Takagi-Sugeno-Kang fuzzy inference system was deployed within ANFIS, with an epoch limit of 30

for training optimal network configurations. Inputs for the ANFIS-based daily runoff prediction included current and previous three days' stage and runoff data, calibrated using a variety of membership functions (triangular, trapezoidal, Gaussian, and generalized bell), with the model structure confirmed through grid partitioning. The hybrid learning algorithm was utilized with an error tolerance of 0.001.

Ultimately, for SSC prediction, similar input parameters were used, where the best inputs were identified through a gamma (Γ) test, and the model structure was further refined using grid partitioning. During both methodologies, a parallel emphasis was placed on statistical performance metrics, leading to the conclusion that the triangular membership function with three input partitions exhibited the best performance for runoff prediction while also optimizing SSC predictions effectively. The effectiveness of the generated models for predicting suspended sediment concentration (SSC) was evaluated using a variety of statistical and hydrological indices. (RMSE), (r), and (CE) were important measurements. The most successful model for SSC prediction among the models examined was Triangular, 3, which showed the highest values for CE and r along with the lowest RMSE.

MLR Based Runoff and SSC Prediction Models

The study utilized the Multiple Linear Regression (MLR) technique to establish models for predicting daily runoff. The models incorporated several variables, including the current day's stage of water (Ht), the stage from the previous day (Ht-1), and the stages from the two and three days prior (Ht-2, Ht-3). Additionally, previous runoff values were considered: the runoff from the previous day (Qt-1) and the runoff from two days prior (Qt-2). Through this MLR approach, the researchers determined the coefficients of these independent variables and the necessary intercepts, thereby formulating a mathematical representation that aids in predicting daily runoff based on historical water stages and runoff levels. This method underscores the relevance of historical data in improving predictions for runoff, which is critical for water resource management and planning.

Qt = a1 + b1Ht + c1Ht-1 + d1Ht-2 + e1Ht-3 + f1Qt-1 + g1Qt-2 … (11)

Table 4 displays the runoff prediction model's regression equation during the training phase.

Table 4 MLR model for predicting runoff

Models	Model equations
MLR-I (training)	Qt = - 118.021 + 370.8057*Ht - 269.41*Ht-1 - 14.2196*Ht-2 - 11.8504*Ht-3 + 0.6872*Qt-1 + 0.0729*Qt- 2

The creation of an MLR-based SSC model that combines runoff from the current day with sediment from the previous one, two, and three days is expressed as follows:

St = a2 + b2Qt + c2Qt-1 + d2Qt-2 + e2Qt-3+ f2St-1 + g2St-2 + h2St-3 … (12)

Regression equation for MLR based SSC model is depicted in Table 5 during training period.

Table 5 MLR model for predicting SSC

Models	Model equation
MLR-II (training)	St = - 276.547 + 384.1661*Qt - 242.319*Qt-1 + 18.2608*Qt-2 + 13.9898*Qt-3 + 0.6463*St-1 - 0.01939*St-2 - 0.0828*St-3

The performance evaluation metrics for the runoff and SSC models that were developed using MLR displays Table 6.

Table 6 Performance metrics for SSC and MLR-based runoff models

Model	RMSE	r	CE
MLR-I (runoff)	36.895 m3/sec	0.967	0.931
MLR-II (sediment)	0.502 g/l	0.835	0.696

SRC-based Prediction Model for SSC

The connection between sediment and river discharge is discussed in the text, with a particular emphasis on the impact of river discharge (represented by Qt) on suspended sediment concentration (SSC). This relationship was established by a regression study, which produced a power equation that is expressed as St = 0.2952 Qt^0.07810. This formula provides a mathematical representation of how changes in daily river discharge affect the amount of suspended silt. Furthermore, Table 7 displays the performance indices of the SRC model used to predict SSC, demonstrating the model's efficacy in predicting suspended sediment levels based on the established relationship

Table 7 Performance metrics for the SSC model based on SRC

Model	RMSE	r	CE
SRC	0.733 g/l	0.731	0.620

Performance Evaluation of Developed Models

It has been demonstrated that the ANFIS model performs better at predicting runoff and SSC than both ANN and MLR models. Visual observations of the model's outputs lead to this conclusion. Several statistical and hydrological metrics, such as the (r), (RMSE), coefficient of efficiency (CE), and pooled average relative error (PARE), were used to evaluate the ANFIS model against its competitors. These metrics offer a thorough assessment of the models' efficacy and forecast accuracy in simulating intricate hydrological processes. For the ANN models predicting runoff, correlation coefficients during the training period (2001-2009) ranged from 0.995 for the (6-28-1) network, with a testing score of 0.979 repeating for the (6-52-52-1) setup. In terms of SSC predictions, the calibration and validation phases offered values of r at 0.904 and 0.765 for the (7-55-1) network, while the (7-30-30-1) network yielded 0.911 and 0.767, respectively.

The study analyzes the performance of various predictive models, notably the ANFIS, MLR, ANN, and SRC, in forecasting runoff and SSC. The ANFIS (Triangular, 3) model demonstrated exceptional correlation coefficients of 0.997 and 0.988 during calibration and validation for runoff predictions, with corresponding r values of 0.941 and 0.779 during SSC training and testing. Conversely, the MLR-I model produced lower correlation coefficients of 0.967 for runoff during training and 0.964 during testing, while the MLR II model for SSC showed values of 0.835 and 0.832.

In terms of RMSE, the ANFIS model excelled with values of 17.130 m^3/s during calibration and 30.644 m^3/s in validation for runoff predictions, alongside RMSEs of 0.307 m^3/s and 0.113 m^3/s for SSC training and testing periods, respectively. MLR-I's RMSE values for runoff were higher, at 36.895 m^3/s during training and 53.063 m^3/s during testing, while for MLR-II SSC predictions, RMSE values were 0.502 m^3/s and 0.121 m^3/s. The SRC-based SSC model yielded even lower correlation values, with RMSEs of 0.733 m^3/s during training and 0.235 m^3/s for testing, indicating poorer predictive performance.

A critical metric for evaluating these models is the CE. For the ANN runoff prediction model with network configuration (6-28-1), CE values were recorded at 0.989 and 0.935 for calibration and validation. In contrast, the ANFIS runoff model reached CE values of 0.994 and 0.966, while SSC predictions resulting CE values were lower compared to runoff but still indicative of good predictive performance at 0.886 and 0.775. In comparison, MLR-I and II showed CE values of 0.931 and 0.824 for runoff and significantly lower values for SSC. Another important performance metric utilized was the PARE. The ANFIS runoff predictions demonstrated PARE values of 0.001% and 0.018%, indicating slight over-prediction during training and testing phases. MLR-I and II models also showed minor over-prediction tendencies, whereas SRC consistently yielded negative PARE values, indicating under-prediction in both capacities. Thus, the results indicate that the ANFIS model generally outperforms its counterparts in terms of predictive accuracy and reliability for both runoff and SSC, evidenced by higher correlation coefficients, lower RMSE, superior CE values, and more favorable PARE statistics.

Table 8 Performance assessments of created ANN, ANFIS, and MLR models for runoff forecasting throughout training and testing phases

Model	Training				Testing			
	RMSE (m3/s)	r	CE	PARE (%)	RMSE (m3/s)	r	CE	PARE (%)
ANN (6-28-1)	22.973	0.995	0.989	0.141	39.840	0.979	0.935	-0.994
ANN (6-52-52-1)	22.565	0.995	0.989	0.113	42.152	0.979	0.941	0.807
ANFIS (Triangular, 3)	17.130	0.997	0.994	0.001	30.644	0.988	0.966	0.018
MLR-I	36.895	0.967	0.931	0.006	53.063	0.964	0.823	0.117

The models' performance evaluations showed that they could estimate silt and runoff with sufficient precision. The ANFIS model performed better in runoff prediction than the and MLR models, according to the qualitative evaluation generated from the graphs comparing observed against anticipated values of daily runoff. The ANFIS model outperforms ANN, MLR, and SRC (Support Vector Regression) models for SSC prediction, according to the examination of daily SSC using corresponding scatter plots.

Table 9 Performance assessments of created models for SSC prediction using ANN, ANFIS, MLR, and SRC during training and testing

Model	Training				Testing			
	RMSE (g/l)	R	CE	PARE (%)	RMSE (g/l)	r	CE	PARE (%)
ANN (7-55-1)	0.390	0.904	0.816	-0.003	0.171	0.865	0.736	-0.130
ANN (7-30-30-1)	0.376	0.911	0.829	-0.001	0.146	0.868	0.748	-0.026
ANFIS (Triangular, 3)	0.307	0.941	0.886	0.002	0.113	0.930	0.775	0.028
MLR-II	0.501	0.835	0.696	0.003	0.121	0.832	0.691	0.027
SRC	0.733	0.731	0.620	-0.023	0.235	0.729	0.608	-0.147

Conclusions

Soil is a vital natural resource for sustaining land productivity, particularly in regions dominated by subsistence agriculture such as the Nagavali River basin. Understanding sediment transport and runoff dynamics is crucial for effective watershed management. Hydrological prediction, however, remains challenging due to the stochastic nature of rainfall–runoff and sediment processes. In recent years, soft-computing techniques have emerged as robust alternatives to conventional methods, requiring fewer input data and capable of capturing complex nonlinear hydrological behaviour. In this study, the researchers aimed to model daily stage–discharge and runoff–sediment relationships using various advanced methodologies, specifically ANN, ANFIS and MLR, in conjunction with a SRC approach. The research utilized data over a twelve-year period from 2001 to 2012, focused on the monsoon season, which included metrics such as stage, discharge, and suspended sediment concentration. This data was sourced from the WRIS and was strategically divided into training (2001–2009)and testing (2010–2012) datasets.

The geographical focus of the study was the Nagavali basin, which spans 9,510 square kilometers across the states of Odisha (4,462 km²) and Andhra Pradesh (5,048 km²). This basin experiences a mean annual rainfall of approximately 1131 mm, situated between the latitudes of 18°10'–19°44' N and longitudes of 82°52'–84°05'E. To develop the ANN models, a feed-forward back-propagation method was employed, utilizing Levenberg–Marquardt training for optimization. In the case of the ANFIS models, grid partitioning methods were used along with triangular and Gaussian membership functions, complemented by a hybrid training approach. The MLR models were implemented using Microsoft Excel, while the SRC method adopted a power-law relationship to correlate runoff and suspended sediment concentration. The performance of these models was rigorously evaluated through several statistical measures, including RMSE, r, CE and PARE, ensuring a comprehensive assessment of model accuracy and reliability in predicting sediment dynamics within the basin.

Key findings

- Soft-computing models produced accurate stage–discharge and runoff–sediment predictions.
- Double-hidden-layer ANN models performed better than single-layer networks for both runoff and SSC.
- ANFIS (Triangular, 3 MFs) achieved the best performance among all models for both runoff and SSC estimation.
- MLR models showed acceptable performance, whereas the SRC method performed poorly for this basin, reflecting nonlinearity in sediment–runoff dynamics.
- Current-day runoff and SSC were strongly dependent on antecedent flow and sediment conditions (up to three-day lag).

ANFIS and ANN models demonstrated superior predictive capability over traditional regression and rating-curve methods, underscoring the effectiveness of intelligent, data-driven approaches in monsoon-dominated hydrological basins. These techniques offer powerful tools for sediment–runoff forecasting and sustainable watershed management in similar tropical catchments.

Data Availability Statement

The datasets generated during the current study are available from the corresponding author on reasonable request.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research received no external funding.

References

Agarwal, A., Rai, R. K. and Upadhyay, A. 2009. Forecasting of runoff and sediment yield using artificial neural networks. Journal of Water Resource and Protection, 1(5): 368–375.

Bharti, B., Pandey, A., Tripathi, S. K. and Kumar, D. 2017. Modelling of runoff and sediment yield using ANN, LS-SVR, REPTree and M5 models. Hydrology Research. https://doi.org/10.2166/nh-153-2017

Bhattacharya, B., Price, R. K. and Solomatine, D. P. 2005. Data-driven modelling in the context of sediment transport. Physics and Chemistry of the Earth, 30(4): 297–302.

Bisht, D. C. S., Raju, M. M. and Joshi, M. C. 2010. ANN-based river stage-discharge modelling for Godavari River, India. Computer Modelling and New Technologies, 14(3): 48–62.

Chang, F.-J. and Tsai, M.-J. 2016. A nonlinear spatio-temporal lumping of radar rainfall for modelling multi-step-ahead inflow forecasts by data-driven techniques. Journal of Hydrology, 535: 256–269.

Cobaner, M., Unal, B. and Kisi, O. 2009. Suspended sediment concentration estimation by adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data. Journal of Hydrology, 367 (1): 52–61.

Eisazadeh, L. L., Sokouti, R., Homaee, M. and Pazira, E. 2013. Modelling sediment yield using artificial neural network and multiple linear regression methods. International Journal of Biosciences, 3 (9): 116–122.

Firat, M. and Gungor, M. 2007. River flow estimation using adaptive neuro-fuzzy inference system. Mathematics and Computers in Simulation, 75 (3): 87–96.

Ham, F. M. and Kostanic, I. 2001. Principles of neurocomputing for science and engineering, pp. 24–91. Arnold Publishers.

Kisi, O. 2004. River flow modelling using artificial neural networks. Journal of Hydrologic Engineering, 9 (1): 60–63.

Loukas, Y. L. 2001. Adaptive neuro-fuzzy inference system: An instant and architecture-free predictor for improved QSAR studies. Journal of Medicinal Chemistry, 44 (17): 2772–2783.

Nourani, V. 2009. Using artificial neural networks for sediment load forecasting of Talkherood River mouth. Journal of Urban and Environmental Engineering, 3 (1): 1–6.

Nourani, V., Baghanam, A. H., Adamowski, J. and Kisi, O. 2014. Applications of hybrid wavelet–artificial intelligence models in hydrology: A review. Journal of Hydrology, 514:358–377.

Patil, S. and Valunjkar, S. 2014. Forecasting of daily runoff using artificial neural networks. International Journal of Civil Engineering and Technology, 5 (1):13–20.

Raghuwanshi, N. S., Singh, R. and Reddy, L. S. 2006. Runoff and sediment yield modelling using artificial neural networks: Upper Siwane River, India. Journal of Hydrologic Engineering, 11 (1):71–79.

Shafie, A. E., Taha, M. R. and Noureldin, A. 2007. A neuro-fuzzy model for inflow forecasting of the Nile River at Aswan High Dam. Water Resources Management, 21:533–556.

Sudheer, K. P., Gosain, A. K. and Ramasastri, K. S. 2002. A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrological Processes, 16 (6):1325–1330.