REVIEW OF THREE DATA-DRIVEN MODELLING TECHNIQUES FOR HYDROLOGICAL MODELLING AND FORECASTING

Oluwaseun Oyebode*, Fred Otieno and Josiah Adeyemo

Department of Civil Engineering and Surveying, Faculty of Engineering and the
Built Environment, Durban University of Technology, PO Box 1334, Durban, 4000, South Africa.

ABSTRACT

Various modelling techniques have been proposed and applied for modelling and forecasting of hydrological systems in different studies. These modelling techniques are majorly categorized into two namely, process-based and data-driven modelling techniques. While the process-based techniques provides detailed description of hydrological processes, data-driven techniques however describe the behaviour of hydrological processes by taking into account only limited assumptions about the underlying physics of the system being modelled. Although, process-based techniques have been widely applied in numerous hydrological modelling studies, the application of data-driven modelling techniques on the other hand has not been fully embraced in the hydrological domain. This paper provides a comprehensive review of several studies relating to three data-driven modelling techniques namely, K-Nearest Neighbours (K-NN), Model Trees (MTs) and Fuzzy Rule-Based Systems (FRBS). Modern trends with respect to their applications in hydrological modelling and forecasting studies are also discussed. The structure of this review encapsulates an introduction to each of the modelling techniques, their applications in hydrological modelling and forecasting, identification of areas of concern in their use, performance improvement methods, as well as summary of their advantages and disadvantages. The review aims to make a case for the application of data-driven modelling techniques by discussing the benefits embedded in its integration into water resources applications.

KEYWORDS:
data-driven models, fuzzy rule-based systems, hydrological modelling and forecasting, k-nearest neighbours, model trees

1. INTRODUCTION

The need to manage water resources across all regions of the world has always been of high importance to water managers and decision-makers, most especially in this era of climate variability. The management of water resources in arid and semi-arid regions is of crucial importance as the demand for water is these regions are far above the quantity available for supply, which consequently leads to water stress.

Water researchers, the government and other related stakeholders have been making concerted efforts towards developing various approaches to managing the amount of water available in their regions. Various strategies and policies are being formulated towards ensuring continuous availability of water. However, increase in population, economic growth, improved standard of living, higher agricultural water demand among other factors continue to place freshwater resources under agglomerative pressure. Consequently, constant availability of water for domestic, industrial, agricultural, energy, mining, ecological, transportation and recreational purposes can be considered to be under threat.

The pressure on freshwater resources is now being further aggravated by impacts of climate variability on the water cycle [1]. Important components of the water cycle such as precipitation, streamflow, evapotranspiration etc. are severely being impacted upon [2]. In hydrology, changes in water availability remain a major consequence of the complex, nonlinear and dynamic nature of hydrological and climatological processes within and around a catchment area [3]. Thus, hydrological forecasts both on short term and long term basis is critically important as it forms the basis upon which water managers, consumers, policy makers and other stakeholders put in place planning, allocation, control and adaptive strategies in order to ensure water availability and security.

Knowledge of the trends of hydrological processes also influence the making of financially-related decisions, especially when there is need to maximize returns on investments made on available water resources [4]. Thus, results from accurate and reliable hydrological modelling studies are of crucial importance to all stakeholders as it yields significant economic and social benefits.

1.1 Modelling techniques in hydrological studies

Considering the importance of hydrological modelling, a significant number of modelling techniques have

* Corresponding author

been developed and adopted for the purpose of modelling and forecasting water resource components. Based on the internal and spatial representations of hydrological processes within a catchment, modelling techniques can be categorized into two. The two categories are namely, process-based models and data-driven models (DDMs) [5].

The process-based models are generally referred to as "*knowledge-driven*" models based on their ability to provide detailed representation and interpretation of hydrological processes. This is achieved by incorporating laws based on physics of water movement in a catchment. The process-based models include the lumped conceptual and distributed physically-based models [6]. Examples of popular process-based models include the European Hydrological System (SHE) [7], the ACRU model – developed in South Africa [8], the HBV (*Hydrologiska Bryång Vattenbalansavdelning*) model [9,10], the Hydrologic Simulation Program – Fortran (HSPF) model [11], and the HYDROLOG model [12].

However, the complexities and extensive detailing involved in the representation of hydrological processes within a river catchment make the development of models from first principle extremely challenging, thereby inflicting certain drawbacks on the use of process-based models. The major drawbacks to the use of process-based models arise from mixed problems such as mis-calibration, over-parameterization, parameter instability, insensitivity or redundancy, high computational requirements and huge data demand [13-16]. These limitations tend to generate some uncertainty in the predictive capability of the process-based models, and thus affect the reliability of their results. Recently, attempts have been made to reduce the influences of these uncertainties through uncertainty evaluation techniques [17-19], but the processes involved are rather computationally expensive.

Data-driven models, on the other hand, define the relationships between system state variables (input, internal and output) variables while characterizing the behaviour of hydrological processes within a river catchment. DDMs achieve this by taking into account only few assumptions on the physics of the system being modelled. These models rely majorly upon the methods of computational intelligence and machine learning, and thus assume the presence of a considerable amount of data describing the physics of the modelled system [5]. Popular DDMs include artificial neural networks (ANNs) [20-22]; fuzzy rule-based systems [23-25]; tree-based methods [26,27]; evolutionary computational methods such as genetic programming (GP) [28-30]; and support vector machines (SVM) [31,32].

DDMs are relatively quicker to develop and easier to use when compared to the process-based models. In addition, due to the ability of DDMs to directly define input-output relationships, the large computational and data requirements often associated with process-based models are to some extent reduced in DDMs. The use of DDMs has also been seen as a promising technique to solving the sensitivity and uncertainty challenges inherent in the use of process-based models [33-35].

In the light of this, DDMs are now being considered as an alternative and promising approach that will complement or replace the knowledge-driven models [36,37]. This review provides an in-depth appraisal of the application of three DDMs (k-nearest neighbour, model trees and fuzzy rule-based systems) in hydrological modelling and forecasting studies, and suggests areas in which they can be integrated to achieve better results.

2. K-NEAREST NEIGHBOUR (K-NN) METHOD

The K-nearest neighbour method belongs to the class of methods based on the working principle of instance-based learning (IBL) algorithms. IBL algorithms are algorithms which initialize by storing information from the training samples using specific instances, and delays generalization effort until need arises for the prediction of a new query instance [38]. They apply the experiential knowledge gained from initial events to generate details about relatively new instances. They achieve this by retrieving salient information from a set of nearest Neighbours and building localized models based on them. The K-NN method is a typical representative of IBL, and thus operates by describing complex functions as a collection of less complex local approximations, with the letter, K, symbolizing the number of nearest Neighbours. K-NN combines the target variables of K selected Neighbours to determine the target outputs of a given test pattern. The pattern is represented by a limited number of illustrative observations referred to as "features", and consequently characterized by a vector known as "feature vector" [39]. This enables K-NN to recognize the feature vectors as a subgroup of the original pattern. Thus, K-NN is considered to be intuitive in nature, though it also exhibits highly influential statistical features [40].

In K-NN, the nearest Neighbours are interpreted as a function of a Euclidean distance which is a measure of the proximity or similarity of a feature vector of query distance and any feature vector of the training sample [41]. The Euclidean distance is often estimated as a weighted Euclidean norm. Furthermore, based on the Euclidean distance, each of the K Neighbours is assigned a weight factor, so as to reveal the relative impact on the prediction value. The weights are computed such that they generate the lowest mean square error of forecasting over the training samples. A number of kernels have been used for the implementation of K-NN. They include linear, inverse, square inverse, exponential and Gaussian kernels [38].

2.1 Application of K-NN in hydrological modelling

Karlsson and Yakowitz [40] was the first to subject the K-NN method to use in hydrologic studies. K-NN method was applied to solve a univariate rainfall-runoff forecasting problem. Results obtained showed a competitive performance between the K-NN, autoregressive moving average with auxiliary input (ARMAX) method and instantaneous unit hydrology (IUH) forecast methods. The satisfactory results obtained thereafter defined a robust theoretical foundation for subsequent use of the K-NN method. Galeati [41] investigated the potential of K-NN in predicting daily average discharge in a rocky basin in Italy. Results yielded comparable results between K-NN and an autoregressive model with exogenous input (ARX). Kember et al. [42] used the K-NN method to forecast daily river flow at a single site, and found it to provide improved forecasts. Shamseldin and O'Connor [39] applied K-NN to fine-tune parameters of a linear perturbation model in river flow forecasting study, obtaining an improved and more reliable forecast. Solomatine et al. [43] compared the performance of K-NN, ANN and M5 model trees for hourly and daily rainfall prediction. Results demonstrated that K-NN is comparable to other DDMs, especially when a Gaussian kernel is employed.

K-NN has also found application in climate change impact studies. Yates et al. [44] found that K-NN is capable of generating alternative climate information when conditioned upon hypothetical climate scenarios. Leander and Buishand [45] applied the nearest neighbour method to resample outputs from a regional climate model (RCM), and its performance was remarkably satisfactory. Bannayan and Hoogenboom [46] further testified to the reliable performance of the K-NN method when used for resampling daily temperature and precipitation events. Sharif and Burn [47] used an improved K-NN for perturbation of historical datasets in a climate change assessment study of the upper Thames River Basin, Canada. Results showed that the K-NN was effective in producing the desired precipitation amounts.

2.2 Areas of concern

Some issues relating to the efficacy of the K-NN method have been raised via its application in some assessment studies. Toth et al. [48] investigated the potential of the K-NN method comparatively with the ARMA and ANN methods for short-range rainfall prediction. Results demonstrated that the K-NN model failed to provide noteworthy predictive performance when compared with the ARMA and ANN methods. It was discovered that the improvement of the performance with an increasing number of nearest Neighbours was less noticeable, with no marginal improvement in overall performance when K is increased beyond a certain limit. An investigative study conducted by Scheuber [49] further corroborates Toth et al. [48]'s findings, as it was remarked that the selection of suitable parameters in the development of K-NN is quite challenging and could have negative impacts on algorithm performance. He further stressed that increasing K translates to the introduction of spatially more distant reference data, which consequently leads to higher degree of bias.

Kim and Tomppo [50] carried out a prediction error uncertainty assessment of K-NN method, and found out that K-NN lacks the logical approach to compute error estimates for domains of arbitrary size. This is in agreement with results from Maltamo et al. [51] and Gjertsen [52]'s study in which over- and under-estimations of the lowest and highest historic observations were evident. This further shows that limitations exist to the methodological and analytical characteristics of the K-NN method.

2.3 Performance improvement methods

Several methods have been devised by experts towards improving the performance of the K-NN algorithm. Akbari et al. [38] established that inconsistencies do occur in query instances in K-NN, which thereafter shows up in the data points of output values, thus leading to the deterioration forecasting results. A clustered K-NN (CKNN) was introduced to capture inconsistencies in data points, and was found to be robust against a set of noisy data. In addition, the CKNN demonstrated high level efficacy for daily inflow forecasting of a reservoir in Iran.

Magnussen et al. [53] tested a model-based calibration method to reduce the out-of-sample extrapolation bias often associated with K-NN. Results showed that the calibrated K-NN predictions were considerably closer to observed values than regular non-calibrated predictions. Prairie et al. [54] also developed a modified K-NN which involved the use of a probability metric to resample residuals from a traditional K-NN. The approach entails giving more weight to the nearest Neighbours and less to the farthest. The resultant model was applied to monthly streamflow in the Colorado River, United States, and was found to exhibit better performance in terms of capturing the patterns inherent in the datasets. Finally, in order to simplify the computational task of the K-NN algorithm, there is need for a reduction in the dimensionality of the K-NN feature space through its synergetic use with other transformation methods such as principal component analysis (PCA) [50].

2.4 Advantages and disadvantages

The K-NN is a learning algorithm which exhibits simplicity and robustness, and thus can tolerate noise and irrelevant attributes. K-NN also possesses the ability to give a good representation of probabilistic and overlapping concepts simultaneously, and as well capable of naturally exploiting inter-attribute relationships [55]. K-NN also gives room for the identification of past events (instances), which makes it more explicit in nature when compared with the ANN. This feature makes it suitable for use in hydrological studies, and hence its acceptability by forecasters.

On the other hand, K-NN does not have the ability to discover any input-output mapping functions, not even *a posteriori* like the ANN does [56]. Thus, it has no extrapolation ability when presented with unfamiliar input vectors, meaning, it has no ability whatsoever to predict values higher than those in the range of the historic observations [41,57]. This serves as a major disadvantage to the use of K-NN, as its credibility is severely limited when used for real forecasting.

3. MODEL TREES (MTS)

The model tree (MT) is a piecewise linear model unlike other nonlinear models such as the ANN and GP. MTs are extensions of classification and regression trees, in which computational process is represented by a hierarchical tree-like structure [58]. It comprises of a root node or decision point that subdivides into several other nodes and leaves. The process of developing the nodes and branches into a tree is based on the idea of splitting the input space into mutually exclusive domains, according to a predefined splitting criterion, progressively narrowing down the size of the domains [27]. When the number of instances in a domain becomes smaller than the predefined value, the splitting of that domain is finally brought into a halt with the creation of a leaf. Each time a new instance is fed into the tree, it follows after a specified path in accordance with the splitting rules defined in the tree-building procedure. A linear regression (LR) model is thereafter developed for each domain, and thus formulates a piecewise linear function for the estimation of nonlinear relationship between input-output variables as shown in Jung et al. [60].

The growing of model trees is carried out using algorithmic rules, which majorly is a function of the splitting criterion utilized. The M5 algorithm [59] is commonly used for inducing a MT and its working principle is as follows. Assuming there exist a set of training samples (initial instance), *P*, characterized by the values of a fixed set of (input) attributes and a corresponding target (output) value. The objective is to develop a model that relates a target value of the training samples to the values of their input attributes based on a divide-and-conquer method [36].

The overall quality of the model will be determined by the accuracy with which it predicts the target values of set of new unseen data (new instance). The set *P* is either associated with a leaf, or some test is chosen that splits *P* into subsets corresponding to the test outcomes and the same process is applied recursively to the subsets. The splitting criterion for the M5 algorithm is based on treating the standard deviation of the class values that reach the node, as a measure of the error at that node. Thus, the variable that maximizes this error reduction is chosen for splitting at that node. The mathematical expression for representing the standard deviation reduction (SDR) is given in equation (1):

$$SDR = sd(P) - \sum_i \frac{|P_i|}{|P|} sd(P_i) \qquad (1)$$

where SDR = standard deviation reduction; $sd(P)$ = standard deviation of all the training samples having a total number, P; P_i = i-th subset of P; and $sd(P_i)$ = standard deviation of the i-th subset.

Upon the exploration of all possible splits of input space, the M5 algorithm finally selects the one with the maximum value of SDR, which it then uses to develop linear regression models in the individual domains. Splitting terminates when the outputs of all the data that reach the node vary slightly or when only a few remain [34]. Pruning of the trees is carried out right away to prevent the problem of over-fitting which may occur as a result of monotonic increase in the training samples during tree growth [60]. Finally, a 'smoothing' operation is performed to compensate for sharp discontinuities that inevitably occur between adjacent linear models at the leaves of the pruned tree. The smoothing operation thus updates the predicted values from neigbhouring equations in order to achieve better agreement as a whole.

3.1 Application of MTs in hydrological modelling

The M5 MT technique can be considered to be relatively new in solving hydrological problems, with its first application in rainfall-runoff prediction reported by Kompare et al. [61]. Water-related areas in which MT have been successfully applied include rainfall-runoff and streamflow modelling [26, 62, 63]; sedimentation modelling [64]; modelling of water pollutants [65, 66]; and climate change impact modelling [67, 68].

Comparative studies have also been conducted with the aim of estimating the potential of M5 MTs against other DDMs. Solomatine and Dulal [26] applied M5 MTs for rainfall-runoff modelling at different hourly time-slices, and compared its performance to that of ANNs. The performance of both models were said to be comparable, although the ANNs performed slightly better at higher lead times. However, the M5 MTs produced more interpretable models and also allowed for the development of modular models of varying complexity and accuracy. Khan and See [62] employed one statistical model and three DDMs namely MLR, ANN, M5 MTs and evolutionary neural network (Evo-NN) in river level forecasting. The study was carried out in the Ouse River catchment located in Northern England. Results showed that the M5 and Evo-NN models provided the best performance based on global performance measures. Furthermore, it was observed that the M5 model demonstrated its ability to make explicit its internal structures, unlike other black-box models such as the ANNs.

In a quest for more accurate predictive models in hydrology, M5 MTs have gained some recognition in the development of modular and hybrid models, due to its transparent and intelligible nature. Solomatine and Xue [69] built a modular model comprising of the M5 MT and ANN which was applied to flood forecasting in the upper Huai River, China. Flood samples with different hydrological features also split into groups using separate M5 and ANN models. Improved accuracy in predicting high floods was generated by the modular model when compared with both individual models. The authors also attested to the fact that the incorporation of M5 MTs ensured transparency of the internal features inherent in the hydrological processes. Likewise, Bhattacharya and Solomatine [70] constructed a model with an ANN model and M5 MT for river stage-discharge modelling. The model was found to be superior in accuracy to a conventional stage-discharge rating curve, most especially during periods of high flow. It was finally submitted that the M5 algorithm did not only allow for higher accuracy, but was also being transparent, simple, verifiable and easily demonstrable. Results from Ajmera and Goyal [71]'s stage-discharge modelling study also supported this claim, as the M5 MTs outperformed three different ANN algorithms as well as the conventional stage-discharge method.

3.2 Areas of concern

Some issues relating to the application of M5 model trees in hydrological studies have been identified. One of such issues which have been reported in literature is the partitioning or splitting problem. Solomatine and Dulal [26] observed from their study that the results from the M5 model at higher lead times and peak flows produced sub-standard predictions compared to the ANN model. This was ascribed to the splitting criteria used to build linear models at the leaves. They noticed that model trees do not use all available attributes to make linear models at any leaf. Only attributes which fulfill the condition of certain criteria (such as SDR) go under one sub-tree, terminating to a leaf. This therefore may have resulted into the non-inclusion of influencing attributes. In addition, the resultant model tree was deemed to be so large, and an attempt made to prune it to a smaller size led to deterioration of accuracy. Thus, it can be inferred that unsupervised pruning operation of model trees could result into poor predictive performance.

Londhe and Charhate [36] also reported cases of over-estimations of peak values in their river flow forecasts, which were attributed to the absence of influencing attributes in the development of the linear models. Bhattacharya and Solomatine [64] additionally suggested that further improvement in terms of performance could be achieved by including additional information about physical processes, using larger datasets or by exploring other machine learning methods.

In furtherance to the abovementioned, the ability of MTs to rapidly increase computational requirements when confronted with high dimensionality have also attracted a bit of concern. Although, such ability enables MTs to learn efficiently and undertake tasks with high dimensionality [26]; it could however lead to extremely high computational demands [72].

3.3 Performance improvement methods

Some efforts have been directed towards addressing the aforementioned areas of concern. An M5' algorithm, a modification to the M5 algorithm, was presented by Wang and Witten [73]. M5' allows for the pruning of the tree size with minimal penalty in prediction performance via incorporation of a pruning factor, which can be specified by the user, thus it seemingly outperformed the original M5 algorithm. Samadi et al. [74] tested the M5' MT for prediction of scour depth below free overall spillways, and evaluated its performance against classification and regression trees (CART). The results indicated that the M5' MT produced better predictions than the CART method. Mafi et al. [75] also applied the M5' for modelling long-shore sediment transport rate (LSTR), and its performance compared to that of existing empirical equations initially proposed for such purpose. Results found that equations derived from the M5' model tree predicted the LSTR more accurately than the existing formulas, producing lesser error estimates. This result is also agreement with that obtained from Nasseri et al. [68]'s climate change modelling study. They employed a combination of M5' algorithm and three other nonlinear data mining methods in developing a nonlinear data mining downscaling model (NDMDM). The performance of the NDMDM model was evaluated comparatively with a popular statistical downscaling model (SDSM), using daily precipitation events. Results indicate better performance of the NDMDM model when a combination of M5' and multivariate adaptive regression splines (MARS) methods is employed.

Asides the applications of the M5' algorithm, some experts have also proposed other methods aimed at solving the partitioning problem in M5 model trees. Rather than use piecewise linear regression models at the leaf nodes of the model tree, Jung et al. [60] employed partial least square regression (PLSR), which is an extension of multivariate linear regression. The M5-PLSR MTs was then compared to the M5' MTs, MLF- and RBF-ANN and K-NN for algal growth prediction in a reservoir. Results show improved prediction by both M5' and M5-PLSR MTs using partitioned datasets, with the M5-PLSR MTs performing better than other algorithms via the use of more closely correlated multivariate datasets.

Hong and Chen [76] proposed an extension of the sample efficient regression tree (SERT) approach, which entails the fusion of the forward selection of regression analysis and the regression tree methodologies. This was done with the aim of maximizing the degree of freedom of the datasets and also to obtain unbiased model estimates. The outcome of the study was the realization of an unbiased MT. Recently, an innovative method termed turning point regression tree induction (TPRTI) was developed by Amalaman et al. [77] to determine optimal split points in MTs. The workings of the TPRTI method include: division a set of data into subsets using a sliding window, computation of a centroid for each subset, use of the centroid for identifying turning points which indicates where general trend changes in the input space. The novel approach was compared to the M5 algorithm using synthetic and real-life datasets for experimentation. Results from the TPRTI showed higher predictive accuracy, improvement in scalability and low model complexity when compared with the M5 algorithm.

However, irrespective of the recent developmental concepts which have resulted into improved accuracy of MTs, the need to reduce its high computational demands is still of concern to modellers [72]. Galelli and Castelletti

[27] recently advised on the need for the development of techniques that will strike a balance between accuracy and computational requirements.

3.4 Advantages and disadvantages

The application of MTs in hydrological studies has been found to have certain advantages. Such advantages include its ability to produce simple, accurate, transparent, provable and understandable models [60, 70]. M5-derived models have also been found to consistently showcase high rate of convergence, especially under fewer events [71, 78]. The pruning operation performed newly built model trees also help to counter overfitting problems [60], provided the operation is well supervised. Furthermore, the partitioning of the input space of model trees allows for the combination of several local linear models, and consequently results in improved model accuracy [79].

However, certain drawbacks to the use of model trees have also been identified. They include: (i) partitioning problems which normally occurs when the ratio of instances (observations) is smaller than the number of attributes (variables) [60]; (ii) need for high level of expertise for model implementation, especially as it relates to the achievement of optimal partitioning and pruning [80]; (iii) high computational demand and fallible performance when used to model data with high dimensionality and nonlinear features [81,82]; and (iv) generation of equations that are easily verifiable but not realistic in terms of physical interpretation [36].

4. FUZZY RULE-BASED SYSTEMS (FRBS)

Fuzzy rule-based systems (FRBS) are based on the method of fuzzy logic, established by Zadeh [83]. FRBS are built by simulating the reasoning process of humans in order to achieve transparency in modelling processes [84]. In FRBS, knowledge is represented using IF-THEN rules [85]. A typical rule is represented as IF-(antecedent part)-THEN (consequence part) [86]. The initialization procedure of an FRBS model entails choosing input and output variables which it uses in defining fuzzy sets. The FRBS model thereafter uses the fuzzy sets to construct membership functions on the input domain of the model. This is achieved by partitioning the domain into a number of overlapping regions. The membership functions are represented using quantitative and linguistic terms. Several types of membership functions that can be used for the fuzzy set in the antecedent of the rules include triangular and Gaussian functions.

Relationship between membership functions of the inputs and that of the outputs are expressed by using linguistic logical statements based on the subjective knowledge of the modeller. For example, river flow can be categorized as “low”, “intermediate” and “high”. Thus, the matching of the inputs and outputs with fuzzy rules can be expressed as:

“IF *Input* 1 is LOW and *Input* 2 is HIGH THEN *Output* is INTERMEDIATE”

The structure of the rule could either be a fuzzy set such as the Mamdani model [87], or a function, often linear referred to as a Takagi-Sugeno-Kang (TSK) model [88, 89]. Upon the formulation of a rule, an iterative and tuning process is introduced to relate observations to the rules. Finally, FRBS combines the fuzzy rules through an inference engine, and “defuzzification” is used to collapse the fuzzy or estimated model output into a single crisp value [84]. Conventional defuzzification methods include center of area (or gravity) method, bisector methods, and some other methods which focuses on the maximum membership value attained by the set. Following the working principle of the FRBS, hydrological systems can be modelled through the processing of historical observations and mapping of input-output variables, thus forming a DDM.

4.1 Applications of FRBS in hydrological modelling

FRBS has found application in several water-related studies, such as rainfall-runoff modelling, management of reservoir operations, streamflow and water-level forecasting, ecological and sediment yield modelling [85, 86, 90-94].

FRBS has also been found to be effective for river flow prediction purposes. Valença and Ludermir [95] used a neuro-fuzzy network model for monthly streamflow forecasting for a hydropower plant in Brazil. Results showed that the neuro-fuzzy model provided better predictions than models based on the traditional Box-Jenkins method. Aqil et al. [96] evaluated the potential of a neuro-fuzzy model for the purpose of predicting flow from local source in the Citarum River in Indonesia. The performance of the neuro-fuzzy model was compared to a MLR. It was reported that the neuro-fuzzy model gave better performance for low and medium flows, but underestimated the magnitude of high flows. Katambara and Ndiritu [86] applied FRBS for simulating daily streamflows at three reaches of the Letaba River in South Africa. Satisfactory results were obtained from the FRBS model despite the irregular and intermittent water abstractions that characterize the river.

Recently, the use of an adaptive neural fuzzy inference systems (ANFIS), a hybrid of FRBS and ANN approaches have been widely reported in river flow prediction studies [97-99], with results showing that the ANFIS performed slightly better when compared to ARMA and/ or ANN.

FRBS have also found application in the development of rainfall-runoff models. See and Openshaw [90] integrated a hybrid neural network, an ARMA model and a fuzzy rule-based model for rainfall-runoff forecasting. Hundecha et al. [100] formulated a set of fuzzy rule-based routines to independently simulate components of snowmelt, evaporation, runoff and basin response for a physically-based (HBV) model. With the aim of automatically generating IF-THEN rules from historical observations, the first-order Takagi-Sugeno models were used to simulate rainfall-runoff transformation [91, 101, 102]. Luchetta and Manetti [103] employed fuzzy logic to predict rainfall-runoff dynamics, and comparison with ANN approach indicated a better performance by the fuzzy method. Nayak et al. [104] developed a rainfall-runoff model by combining fuzzy logic and ANN. Results from the hybrid models were found to be comparable with an ANFIS model, with a suggestion that the hybrid model could be used as a promising alternative to ANFIS for rainfall-runoff modelling purposes.

FRBS has also been used conjunctively with other models or optimization algorithms. Chang and Chen [105] used a counter-propagation fuzzy-neural network (CFNN) rainfall-runoff model for hourly streamflow forecasting, and was found to be better in terms of prediction accuracy when compared to the ARMAX model. Maskey et al. [23] used a combination of FBRS and genetic algorithm (GA) for treating precipitation uncertainty in a rainfall-runoff model. Shiri and Kisi [24] applied a hybrid wavelet neuro-fuzzy model to investigate daily, monthly and annual streamflows in the Filyos River in Turkey. Bagis and Karaboga [106] compared FRBS and neural-fuzzy network system, and results indicated that the former was effective for regular reservoir operations, while the latter was effective primarily for flood control. Katambara and Ndiritu [107] developed a calibrated hybrid conceptual-fuzzy-logic model and investigated its potential in simulating daily streamflow in the Letaba River, South Africa. The model performance was satisfactory regardless of the intricacy of the system and insufficiency of relevant data. Results from the study suggest that hybrid conceptual-fuzzy logic-modelling could be adopted for more detailed and reliable planning analysis than single FRBS modelling, especially when used to model complex data-scarce river systems.

4.2 Areas of concern

Despite the successful application of FRBS in modelling studies, some areas of concern have been identified. An area that seems to be of major challenge in the use of FRBS is the identification of the optimal number of rules to achieve the best performance. Abebe et al. [108] noted that improper selection of fuzzy-rules could lead to extreme generalization and overfitting, especially when faced with data insufficiency. It was however suggested to carry out test runs in order to determine the optimal number of rules. Aqil et al. [109] also attributed the failure of their neuro-fuzzy model in underestimations of high flows to the FRBS, which was unable to make proper fuzzy rules corresponding to a high-range from the training datasets. They however encouraged the inclusion of another input variable for the refinement of the fuzzy rule. Katambara and Ndiritu [107] corroborates this claim by stating that their stand-alone fuzzy-systems performed poorly in simulating baseline recession due to the lack of subjective transformation rainfall and evaporation, and therefore failed to obtain realistic modelling. Thus, the representation of input-output relationships by FRBS may be hampered if optimal fuzzy rules are not employed.

In addition, since the determination of optimal rules in FRBS entails choosing appropriate input variables and membership functions, the choice of input variables and membership functions is therefore important. It has been established that FRBS rules increase exponentially with increase in input variables or membership functions, imposing a curse of dimensionality of the system, which translates into poor model performance [110]. Thus improper selection of input variables and membership functions may result into the addition of unnecessary variables that will create a more complex model than required, and further complicate the rule selection process.

Another issue that is of concern in the application of FRBS to hydrological modellers is the partitioning of the input domain. In FRBS, the number of partition represents the number of fuzzy sets, and the corresponding membership function defined in that order [111]. As a result, the partitioning of the input domain is vital to achieving optimal configuration of the model. The partitioning of input domain is done using different methods - grid partitioning and fuzzy clustering. However, the challenge remains as to which method to employ, as both methods come with their strength and limitations. The major limitation of grid partitioning are that the number of rules increase exponentially [111], and that the membership functions of the variables are independent of each other, leading to the disregard of the relationship between the variables [101]. Moreover, attempt to optimize the antecedent parameters leads to the more complexity. In fuzzy clustering, the antecedent parameters are obtained from fuzzy clusters. The drawback to the use of fuzzy clustering is that, the falling off of any data point from the cluster center or outside the cluster affects the model performance.

4.3 Performance improvement methods

Following the above-mentioned areas of concern, some techniques have been employed with respect to improving the performance of FRBS. Chang and Chang [112] introduced GA for the purpose of synergizing it with the ANFIS model. GA was used to achieve several objectives such as searching for the optimal reservoir operation histogram, construction of suitable model structure and parameters for the ANFIS model and estimation of optimal water release from the reservoir. Abolpour et al. [113] developed a new approach called “adaptive neural fuzzy reinforcement learning” (ANFRL) which was derived by the combination of ANFIS and fuzzy reinforcement learning. The ANFRL method was used for optimizing allocation of water resources and to obtain optimal values of decision parameters. Results also showed that ANFRL brought about an increase of 16% in water allocation; a value considered not attainable by the regular ANFIS model. However, two limitations to the use of ANFRL were identified. The first is its requirement for long series of hydrological data in deriving a robust set of rules, and the second - the performance of the ANFRL-derived models is a function of the ability of the ANFIS model to handle data variability. Thus, if the ANFIS model does not yield suitable estimation of water resources variability, then the ANFRL results will not be accurate.

Recently, Akrami et al. [114] applied a new method earlier proposed by Jovanovic et al. [115] to model the dynamic nonlinear behaviour of rainfall. The new method, referred to as modified ANFIS (MANFIS) entails the modification to the structure of the conventional ANFIS model, with the aim of improving its performance. MANFIS was used to find the optimal number of rules, discover the appropriate membership functions and learning algorithm. A hybrid learning algorithm which combines the least square and back-propagation gradient descent methods was used to train membership function parameters. Results showed that the MANFIS method when compared to the conventional ANFIS method produced faster convergence and low computational while maintaining outstanding performance. However, improved performance was not noticeable when using additional input member functions.

Generally, majority of the studies have demonstrated that FRBS has good predictive abilities. However, some studies have shown that the predictions produced by FRBS are quite inferior to that of other simple conventional models, such as ARMA [90] or MLR [116]. These results therefore substantiate the need to subject the quality of a model to test for each given situation, as there is no such flawless model that will perform well, at all times, in solving all modelling problems [117].

4.4 Advantages and disadvantages

Going by the applications of FRBS discussed in the previous sections, the advantages that can be derived from its use as stated by Jacquin and Shamseldin [102] include: (i) ability to infer the nature of complex systems purely from data, while also providing insight about their internal operations; (ii) ability to present knowledge that can easily be interpretable by humans; (iii) subjective knowledge by expert can be incorporated into the model in a natural and transparent way; (iv) flexibility of use, as their architecture and inference mechanisms can be adapted to a given modelling problem. Additional advantages of FRBS systems according to Aqil et al. [109] include: (v) ability to handle large amount of noisy data from dynamic and nonlinear systems, especially where the underlying physics of the system is not fully understood; (vi) ability to improve the performance of other models, and (vii) fast model development with less computation time, provided the input vector space is well-dimensioned.

The disadvantages of FRBS however include: (i) inability to provide adequate representation of input-output relationships in cases where too many variables are required or when used for modelling highly complex systems [110]; (ii) suffers from "the curse of dimensionality", where the number of fuzzy rules increases exponentially with little increment of inputs, leading additional complexity and higher computational time [114]; (iii) attempts to reduce the number of rules generally decreases model generalization ability [110]; and (iv) lacks an appropriate set of guidelines for calibrating model parameters in a way that will maximize model interpretability [102].

5. CONCLUSIONS

Following this extensive review of the application of three popular DDMs in the hydrological domain, one cannot boldly say one modelling technique is superior to the other, as all DDMs have their strengths and drawbacks. With a wide-range of modelling options to choose from, the challenge remains as to which particular technique will generate the best results for a given task, as one cannot continue to subject each DDM to test one after the other. Thus, it is of crucial importance for modellers and other decision-makers to subject DDMs to test comparatively, in order to determine the approach that best suites the given problem. Furthermore, another promising approach that has been producing improved results is the development of modular and hybrid models, which allows for complementary modelling. The K-nearest neighbours, model trees and fuzzy rule-based techniques have showcased great potential in this respect, as they can be easily be integrated into the process-based models and other DDMs. More importantly, adoption of these techniques can help reduce uncertainty inherent in the use of process-based models. Thus, this review showcases the possibility of achieving better effective management of water resources via the integration of DDMs to produce more accurate and reliable hydrological forecasts.

The authors have declared no conflict of interest.

REFERENCES

[1] Graham, L., Andersson, L., Horan, M., Kunz, R., Lumsden, T., et al. (2011) Using multiple climate projections for assessing hydrological response to climate change in the Thukela River Basin, South Africa. Physics and Chemistry of the Earth, Parts A/B/C 36: 727-735.

[2] Makkeasorn, A., Chang, N.B. and Zhou, X. (2008) Short-term streamflow forecasting with global climate change implications – A comparative study between genetic programming and neural network models. Journal of Hydrology 352: 336-354.

[3] Maity, R. and Kashid, S. (2011) Importance analysis of local and global climate inputs for basin-scale streamflow prediction. Water Resources Research 47: W11504.

[4] Robertson, D. and Wang, Q. Selecting predictors for seasonal streamflow predictions using a Bayesian joint probability (BJP) modelling approach; 2009 13-17 July 2009; Cairns. pp. 376-382.

[5] Solomatine, D. and Ostfeld, A. (2008) Data-driven modelling: some past experiences and new approaches. Journal of hydroinformatics 10: 3-22.

[6] Babovic, V. and Keijzer, M. (2002) Rainfall runoff modelling based on genetic programming. Nordic hydrology 33: 331-346.

[7] Abbot, M., Bathurst, J., Cunge, J., O'Connell, P. and Rasmussen, J. (1986) An introduction to the European Hydrological System-Systeme Hydrologique Europeen," SHE", 1: History and philosophy of a physically-based, distributed modelling system. Journal of Hydrology 87: 45-59.

[8] Schulze, R.E. (1995) Hydrology and agrohydrology: A text to accompany the ACRU 3.00 agrohydrological modelling system. Pretoria: Water Research Commission. WRC Report TT 69/9/95 WRC Report TT 69/9/95.

[9] Bergstrom, S. (1995) The HBV Model. In: Singh VP, editor. Computer models of watershed hydrology. Littleton: Water Resources Publications. pp. 443-476.

[10] Lindström, G., Johansson, B., Persson, M., Gardelin, M. and Bergström, S. (1997) Development and test of the distributed HBV-96 hydrological model. Journal of hydrology 201: 272-288.

[11] Johanson, R.C., Imhoff, J.C., Kittle, J.L. and Donigan, A. (1984) Hydrological Simulation Program--FORTRAN(HSPF): Users Manual for Release 8. 0. Athens, GA: Environmental Protection Agency.

[12] Porter, J. and McMahon, T. (1971) A model for the simulation of streamflow data from climatic records. Journal of Hydrology 13: 297-324.

[13] Poulin, A., Brissette, F., Leconte, R., Arsenault, R. and Malo, J.-S. (2011) Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. Journal of Hydrology 409: 626-636.

[14] Il-Won, J., Moradkhani, H. and Chang, H. (2012) Uncertainty assessment of climate change impacts for hydrologically distinct river basins. Journal of Hydrology: 73-87.

[15] Liu, Y. and Gupta, H.V. (2007) Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research 43.

[16] Rossa, A., Liechti, K., Zappa, M., Bruen, M., Germann, U., et al. (2011) The COST 731 Action: A review on uncertainty propagation in advanced hydro-meteorological forecast systems. Atmospheric Research 100: 150-167.

[17] Bastola, S., Murphy, C. and Sweeney, J. (2011) The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments. Advances in Water Resources 34: 562-576.

[18] Brigode, P., Oudin, L. and Perrin, C. (2013) Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change? Journal of Hydrology 476: 410-425.

[19] Montanari, A. and Di Baldassarre, G. (2013) Data errors and hydrological modelling: The role of model structure to propagate observation uncertainty. Advances in Water Resources: 498-504.

[20] Coulibaly, P. and Evora, N. (2007) Comparison of neural network methods for infilling missing daily weather records. Journal of Hydrology 341: 27-41.

[21] Besaw, L.E., Rizzo, D.M., Bierman, P.R. and Hackett, W.R. (2010) Advances in ungauged streamflow prediction using artificial neural networks. Journal of Hydrology 386: 27-37.

[22] Heng, S. and Suetsugi, T. (2013) Using Artificial Neural Network to Estimate Sediment Load in Ungauged Catchments of the Tonle Sap River Basin, Cambodia. Journal of Water Resource and Protection 5: 111-123.

[23] Maskey, S., Guinot, V. and Price, R.K. (2004) Treatment of precipitation uncertainty in rainfall-runoff modelling: a fuzzy set approach. Advances in water resources 27: 889-898.

[24] Shiri, J. and Kisi, O. (2010) Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. Journal of Hydrology 394: 486-493.

[25] Kisi, O. and Shiri, J. (2011) Precipitation Forecasting Using Wavelet-Genetic Programming and Wavelet-Neuro-Fuzzy Conjunction Models. Water Resources Management 25: 3135-3152.

[26] Solomatine, D. and Dulal, K.N. (2003) Model trees as an alternative to neural networks in rainfall—Runoff modelling. Hydrological Sciences Journal 48: 399-411.

[27] Galelli, S. and Castelletti, A. (2013) Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling. Hydrol Earth Syst Sci Discuss 10: 1617-1655.

[28] Guven, A. (2009) Linear genetic programming for time-series modelling of daily flow rate. Journal of earth system science 118: 137-146.

[29] Kashid, S.S., Ghosh, S. and Maity, R. (2010) Streamflow prediction using multi-site rainfall obtained from hydroclimatic teleconnection. Journal of Hydrology 395: 23-38.

[30] Ni, Q., Wang, L., Ye, R., Yang, F. and Sivakumar, M. (2010) Evolutionary modeling for streamflow forecasting with minimal datasets: a case study in the West Malian River, China. Environmental Engineering Science 27: 377-385.

[31] Auria, L. and Moro, R.A. (2008) Support Vector Machines (SVM) as a technique for solvency analysis. DIW Berlin. German Institute for Economic Research Mohrenstr. Berlin.

[32] Shabri, A. and Suhartono (2012) Streamflow forecasting using least-squares support vector machines. Hydrological Sciences Journal 57: 1275-1293.

[33] Tripathi, S., Srinivas, V. and Nanjundiah, R.S. (2006) Downscaling of precipitation for climate change scenarios: a support vector machine approach. Journal of Hydrology 330: 621-640.

[34] Shrestha, D.L. and Solomatine, D.P. (2008) Data‐driven approaches for estimating uncertainty in rainfall‐runoff modelling. International Journal of River Basin Management 6: 109-122.

[35] Selle, B. and Muttil, N. (2011) Testing the structure of a hydrological model using Genetic Programming. Journal of Hydrology 397: 1-9.

[36] Londhe, S. and Charhate, S. (2010) Comparison of data-driven modelling techniques for river flow forecasting. Hydrological Sciences Journal 55: 1163-1174.

[37] Mohanty, S., Jha, M.K., Kumar, A. and Panda, D. (2013) Comparative Evaluation of Numerical Model and Artificial Neural Network for Simulating Groundwater Flow in Kathajodi-Surua Inter-Basin of Odisha, India. Journal of Hydrology 495: 38-51.

[38] Akbari, M., van Overloop, P.J. and Afshar, A. (2011) Clustered K nearest neighbor algorithm for daily inflow forecasting. Water resources management 25: 1341-1357.

[39] Shamseldin, A.Y. and O'Connor, K.M. (1996) A nearest neighbour linear perturbation model for river flow forecasting. Journal of Hydrology 179: 353-375.

[40] Karlsson, M. and Yakowitz, S. (1987) Nearest‐neighbor methods for nonparametric rainfall‐runoff forecasting. Water Resources Research 23: 1300-1308.

[41] Galeati, G. (1990) A comparison of parametric and non-parametric methods for runoff forecasting. Hydrological sciences journal 35: 79-94.

[42] Kember, G., Flower, A. and Holubeshen, J. (1993) Forecasting river flow using nonlinear dynamics. Stochastic Hydrology and Hydraulics 7: 205-212.

[43] Solomatine, D., Maskey, M. and Shrestha, D.L. (2008) Instance‐based learning compared to other data‐driven methods in hydrological forecasting. Hydrological Processes 22: 275-287.

[44] Yates, D., Gangopadhyay, S., Rajagopalan, B. and Strzepek, K. (2003) A technique for generating regional climate scenarios using a nearest‐neighbor algorithm. Water Resources Research 39.

[45] Leander, R. and Buishand, T.A. (2007) Resampling of regional climate model output for the simulation of extreme river flows. Journal of Hydrology 332: 487-496.

[46] Bannayan, M. and Hoogenboom, G. (2008) Predicting realizations of daily weather data for climate forecasts using the non‐parametric nearest‐neighbour re‐sampling technique. International Journal of Climatology 28: 1357-1368.

[47] Sharif, M. and Burn, D.H. (2006) Simulating climate change scenarios using an improved K-nearest neighbor model. Journal of Hydrology 325: 179-196.

[48] Toth, E., Brath, A. and Montanari, A. (2000) Comparison of short-term rainfall prediction models for real-time flood forecasting. Journal of Hydrology 239: 132-147.

[49] Scheuber, M. (2010) Potentials and limits of the k-nearest-neighbour method for regionalising sample-based data in forestry. European Journal of Forest Research 129: 825-832.

[50] Kim, H.-J. and Tomppo, E. (2006) Model-based prediction error uncertainty estimation for< i> k</i>-nn method. Remote Sensing of Environment 104: 257-263.

[51] Maltamo, M., Malinen, J., Kangas, A., Härkönen, S. and Pasanen, A.M. (2003) Most similar neighbour‐based stand variable estimation for use in inventory by compartments in Finland. Forestry 76: 449-464.

[52] Gjertsen, A.K. (2007) Accuracy of forest mapping based on Landsat TM data and a kNN-based method. Remote Sensing of Environment 110: 420-430.

[53] Magnussen, S., Tomppo, E. and McRoberts, R.E. (2010) A model-assisted k-nearest neighbour approach to remove extrapolation bias. Scandinavian Journal of Forest Research 25: 174-184.

[54] Prairie, J.R., Rajagopalan, B., Fulp, T.J. and Zagona, E.A. (2006) Modified K-NN model for stochastic streamflow simulation. Journal of Hydrologic Engineering 11: 371-378.

[55] Aha, D.W., Kibler, D. and Albert, M.K. (1991) Instance-based learning algorithms. Machine learning 6: 37-66.

[56] Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T.B. and Bobée, B. (2007) A review of statistical water temperature models. Canadian Water Resources Journal 32: 179-192.

[57] Hessenmoller, D. and Elsenhans, A. (2002) Comparison between parametric and non-parametric methods to predict the annual increment of beech. Allgemeine Forst Und Jagdzeitung 173: 216-223.

[58] Deo, M.C. (2009) Recent Data Driven Methods and Applications in Coastal and Hydrologic Data Analysis. ISH Journal of Hydraulic Engineering 15: 310-327.

[59] Quinlan, J.R. Learning with continuous classes; 1992 16-18 November 1992; Hobart. World Scientific. pp. 343-348.

[60] Jung, N., Popescu, I., Kelderman, P., Solomatine, D. and Price, R. (2010) Application of model trees and other machine learning techniques for algal growth prediction in Yongdam reservoir, Republic of Korea. Journal of Hydroinformatics 12: 262-274.

[61] Kompare, B., Steinman, F., Cerar, U. and Dzeroski, S. (1997) Prediction of rainfall runoff from catchment by inteligent data analysis with machine learning tools within the artificial intelligence tools. Acta hydrotechnica 16: 79-94.

[62] Khan, S.A. and See, L. Rainfall-runoff modelling using data driven and statistical methods; 2006 2-3 September 2006; Islamabad. IEEE. pp. 16-20.

[63] Jagtap, S.A. (2013) Rainfall Runoff Modeling Using M5 Model Tree Techniques. International Journal of Global Technology Initiatives 2: A79-A83.

[64] Bhattacharya, B. and Solomatine, D.P. (2006) Machine learning in sedimentation modelling. Neural Networks 19: 208-214.

[65] Preis, A., Tubaltzev, A. and Ostfeld, A. (2006) Kinneret watershed analysis tool: a cell-based decision tree model for watershed flow and pollutants predictions. Water science and technology 53: 29-36.

[66] Ostfeld, A. and Salomons, S. (2005) A hybrid genetic—instance based learning algorithm for CE-QUAL-W2 calibration. Journal of Hydrology 310: 122-142.

[67] Goyal, M.K. and Ojha, C. (2013) Evaluation of Rule and Decision Tree Induction Algorithms for Generating Climate Change Scenarios for Temperature and Pan Evaporation on a Lake Basin. Journal of Hydrologic Engineering.

[68] Nasseri, M., Tavakol-Davani, H. and Zahraie, B. (2013) Performance assessment of different data mining methods in statistical downscaling of daily precipitation. Journal of Hydrology 492: 1-14.

[69] Solomatine, D.P. and Xue, Y. (2004) M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. Journal of Hydrologic Engineering 9: 491-501.

[70] Bhattacharya, B. and Solomatine, D.P. (2005) Neural networks and M5 model trees in modelling water level–discharge relationship. Neurocomputing 63: 381-396.

[71] Ajmera, T.K. and Goyal, M.K. (2012) Development of stage–discharge rating curve using model tree and neural networks: An application to Peachtree Creek in Atlanta. Expert Systems with Applications 39: 5702-5710.

[72] Ganesan, S. (2011) Regression Leaf Forest: A Fast and Accurate Learning Method for Large & High Dimensional Data Sets: University of Georgia.

[73] Wang, Y. and Witten, I.H. (1996) Induction of model trees for predicting continuous classes. New Zealand: Department of Computer Science, University of Waikato. Working Paper 96/23: 1170-487X Working Paper 96/23: 1170-487X.

[74] Samadi, M., Jabbari, E. and Azamathulla, H.M. (2012) Assessment of M5′ model tree and classification and regression trees for prediction of scour depth below free overfall spillways. Neural Computing and Applications: 1-10.

[75] Mafi, S., Yeganeh-Bakhtiary, A. and Kazeminezhad, M.H. (2013) Prediction formula for longshore sediment transport rate with M5'algorithm. Journal of Coastal Research: 2149-2154.

[76] Hong, A. and Chen, A. (2012) Piecewise regression model construction with sample efficient regression tree (SERT) and applications to semiconductor yield analysis. Journal of Process Control 22: 1307-1317.

[77] Amalaman, P.K., Eick, C.F. and Rizk, N. (2013) Using Turning Point Detection to Obtain Better Regression Trees. Machine Learning and Data Mining in Pattern Recognition. Berlin Heidelberg: Springer. pp. 325-339.

[78] Singh, K.K., Pal, M. and Singh, V. (2010) Estimation of mean annual flood in indian catchments using backpropagation neural network and M5 model tree. Water resources management 24: 2007-2019.

[79] Sattari, M.T., Pal, M., Apaydin, H. and Ozturk, F. (2013) M5 model tree application in daily river flow forecasting in Sohu Stream, Turkey. Water Resources 40: 233-242.

[80] Solomatine, D.P. and Siek, M.B. (2006) Modular learning models in forecasting natural phenomena. Neural networks 19: 215-224.

[81] Pal, M., Singh, N. and Tiwari, N. (2012) M5 model tree for pier scour prediction using field dataset. KSCE Journal of Civil Engineering 16: 1079-1084.

[82] Garg, V. and Jothiprakash, V. (2013) Evaluation of reservoir sedimentation using data driven techniques. Applied Soft Computing 13: 3567-3581.

[83] Zadeh, L.A. (1965) Fuzzy sets. Information and control 8: 338-353.

[84] Bourdin, D.R., Fleming, S.W. and Stull, R.B. (2012) Streamflow Modelling: A Primer on Applications, Approaches and Challenges. Atmosphere-Ocean 50: 507-536.

[85] Adriaenssens, V., Baets, B.D., Goethals, P.L. and Pauw, N.D. (2004) Fuzzy rule-based models for decision support in ecosystem management. Science of the Total Environment 319: 1-12.

[86] Katambara, Z. and Ndiritu, J. (2009) A fuzzy inference system for modelling streamflow: Case of Letaba River, South Africa. Physics and Chemistry of the Earth, Parts A/B/C 34: 688-700.

[87] Mamdani, E.H. (1974) Application of fuzzy algorithms for control of simple dynamic plant. Electrical Engineers, Proceedings of the Institution of 121: 1585-1588.

[88] Sugeno, M. and Kang, G. (1988) Structure identification of fuzzy model. Fuzzy sets and systems 28: 15-33.

[89] Takagi, T. and Sugeno, M. (1985) Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man and Cybernetics: 116-132.

[90] See, L. and Openshaw, S. (2000) A hybrid multi-model approach to river level forecasting. Hydrological sciences journal 45: 523-536.

[91] Xiong, L., Shamseldin, A.Y. and O'connor, K.M. (2001) A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi–Sugeno fuzzy system. Journal of hydrology 245: 196-217.

[92] Chang, F.-J. and Chang, Y.-T. (2006) Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Advances in Water Resources 29: 1-10.

[93] Goyal, M.K., Ojha, C., Singh, R., Swamee, P. and Nema, R. (2013) Application of ANN, fuzzy logic and decision tree algorithms for the development of reservoir operating rules. Water Resources Management 27: 911-925.

[94] Kisi, O., Shiri, J. and Nikoofar, B. (2012) Forecasting daily lake levels using artificial intelligence approaches. Computers & Geosciences 41: 169-180.

[95] Valença, M. and Ludermir, T. Monthly stream flow forecasting using an neural fuzzy network model; 2000 22-25 November; Rio de Janeiro. IEEE. pp. 117-119.

[96] Aqil, M., Kita, I., Yano, A. and Nishiyama, S. (2007) A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff. Journal of hydrology 337: 22-34.

[97] Chau, K., Wu, C. and Li, Y. (2005) Comparison of several flood forecasting models in Yangtze River. Journal of Hydrologic Engineering 10: 485-491.

[98] Keskin, M.E., Taylan, D. and Terzi, O. (2006) Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series. Hydrological sciences journal 51: 588-598.

[99] Wang, W.-C., Chau, K.-W., Cheng, C.-T. and Qiu, L. (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. Journal of hydrology 374: 294-306.

[100] Hundecha, Y., Bardossy, A. and WERNER, H.-W. (2001) Development of a fuzzy logic-based rainfall-runoff model. Hydrological Sciences Journal 46: 363-376.

[101] Vernieuwe, H., Georgieva, O., De Baets, B., Pauwels, V., Verhoest, N.E., et al. (2005) Comparison of data-driven Takagi–Sugeno models of rainfall–discharge dynamics. Journal of Hydrology 302: 173-186.

[102] Jacquin, A.P. and Shamseldin, A.Y. (2006) Development of rainfall–runoff models using Takagi–Sugeno fuzzy inference systems. Journal of Hydrology 329: 154-173.

[103] Luchetta, A. and Manetti, S. (2003) A real time hydrological forecasting system using a fuzzy clustering approach. Computers & geosciences 29: 1111-1117.

[104] Nayak, P., Sudheer, K. and Jain, S. (2007) Rainfall ‐ runoff modeling through hybrid intelligent system. Water Resources Research 43.

[105] Chang, F.-J. and Chen, Y.-C. (2001) A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. Journal of hydrology 245: 153-164.

[106] Bagis, A. and Karaboga, D. (2004) Artificial neural networks and fuzzy logic based control of spillway gates of dams. Hydrological processes 18: 2485-2501.

[107] Katambara, Z. and Ndiritu, J.G. (2010) A hybrid conceptual–fuzzy inference streamflow modelling for the Letaba River system in South Africa. Physics and Chemistry of the Earth, Parts A/B/C 35: 582-595.

[108] Abebe, A., Solomatine, D. and Venneker, R. (2000) Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. Hydrological Sciences Journal 45: 425-436.

[109] Aqil, M., Kita, I., Yano, A. and Nishiyama, S. (2007) Analysis and prediction of flow from local source in a river basin using a Neuro-fuzzy modeling tool. Journal of environmental management 85: 215-223.

[110] Li, T.-S., Tong, S.-C. and Feng, G. (2010) A novel robust adaptive-fuzzy-tracking control for a class of nonlinear-multi-input/multi-output systems. IEEE Transactions on Fuzzy Systems 18: 150-160.

[111] Nayak, P., Sudheer, K. and Ramasastri, K. (2005) Fuzzy computing based rainfall–runoff model for real time flood forecasting. Hydrological Processes 19: 955-968.

[112] Chang, L.C. and Chang, F.J. (2001) Intelligent control for modelling of real ‐ time reservoir operation. Hydrological processes 15: 1621-1634.

[113] Abolpour, B., Javan, M. and Karamouz, M. (2007) Water allocation improvement in river basin using adaptive neural fuzzy reinforcement learning approach. Applied soft computing 7: 265-285.

[114] Akrami, S.A., El-Shafie, A. and Jaafar, O. (2013) Improving Rainfall Forecasting Efficiency Using Modified Adaptive Neuro-Fuzzy Inference System (MANFIS). Water Resources Management: 1-17.

[115] Jovanovic, B.B., Reljin, I.S. and Reljin, B.D. Modified ANFIS architecture-improving efficiency of ANFIS technique; 2004 23-25 September 2004; Belgrade. IEEE. pp. 215-220.

[116] Jacquin, A., Shamseldin, A.Y., Webb, B., Arnell, N., Onof, C., et al. Development of a river flow forecasting model for the Blue Nile River using Takagi-Sugeno-Kang fuzzy systems; 2004 12-16 July 2004; Imperial College, London. British Hydrological Society. pp. 291-296.

[117] Maier, H.R. and Dandy, G.C. (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental modelling & software 15: 101-124.

Received: November 19, 2013
Accepted: February 03, 2014

CORRESPONDING AUTHOR

Oluwaseun Oyebode
Department of Civil Engineering and Surveying
Faculty of Engineering and the Built Environment
Durban University of Technology
PO Box 1334
Durban 4000
SOUTH AFRICA

Phone: +27 84 807 3576
E-mail: oluwaseun.oyebode@gmail.com
21243519@dut4life.ac.za