Predictive Modelling Techniques in Sediment Yield and Hydrological Modelling

Adesoji T. Jaiyeola, Josiah Adeyemo

Abstract—This paper presents an extensive review of literature relevant to the modelling techniques adopted in sediment yield and hydrological modelling. Several studies relating to sediment yield are discussed. Many research areas of sedimentation in rivers, runoff and reservoirs are presented. Different types of hydrological models, different methods employed in selecting appropriate models for different case studies are analysed. Applications of evolutionary algorithms and artificial intelligence techniques are discussed and compared especially in water resources management and modelling. This review concentrates on Genetic Programming (GP) and fully discusses its theories and applications. The successful applications of GP as a soft computing technique were reviewed in sediment modelling. Some fundamental issues such as benchmark, generalization ability, bloat, over-fitting and other open issues relating to the working principles of GP are highlighted. This paper concludes with the identification of some research gaps in hydrological modelling and sediment yield.

Keywords—Artificial intelligence, evolutionary algorithm, genetic programming, sediment yield.

I. INTRODUCTION

SEDIMENT yield can be considered to be the total sediment load that leaves a drainage basin. It could also be defined as the quantity of sediment per unit area removed from a watershed by flowing water during a specific period of time [1]. Sediment yield can be divided into three categories namely: (1) suspended sediment – these are the particles suspended due to turbulence in the upper portion of a river just below the water surface. They comprise of silt, clay, and sand size; (2) bed load – these are larger particles that move along the bottom of a river. These particles are large sand, gravel, rock and boulders; (3) saltation – these are particles that bounce up and down the top and bottom of a river. They are mostly sand size or gravel [2]. Fig. 1 shows an illustration of the various categories of particles found in a river.
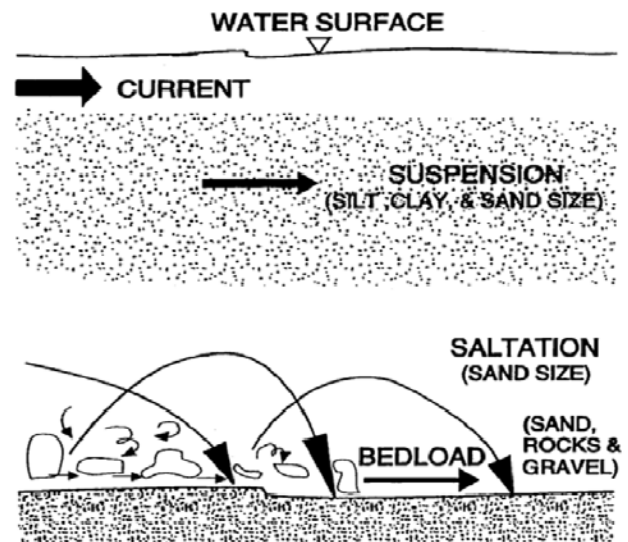

Fig. 1 Layers and movement of particles in a river [2]

The effect of rainfall splash detachment and entrainment through overland flow also generates sediments. Detachment takes place when locally induced shear stress exceeds the cohesive strength of the soil [3]. Geomorphic characteristics like vegetation cover, land use, precipitation, sediment storage, drainage density, topography, soil erodibility and sediment transport capacity affects the sediment yield of a river basin [2]. This is increased through soil disturbance during land use, unstable geological terrain and/or a high rainfall zone. All rivers contain sediments and when a river is stilled behind a dam, some of the sediments sink to the bottom of the dam. As sediments accumulate, the reservoir gradually loses its ability to store water for the purposes for which it was built. Every reservoir loses storage to sedimentation, although the rate at which this happens varies widely. Sedimentation is a major technical problem faced by marine industries. Apart from rapidly filling reservoirs, sediment-filled rivers also cause abrasion to turbines and other reservoir components [4]. The knowledge of the quantity of sediment present in a river at a particular time can lead to a better understanding of flood capacity in reservoirs and consequently help control over-bane flooding. The reduction of sediments in a reservoir also has the following advantages: (1) it improves water quality; (2) it makes the water more suitable for man and aquatic life and; (3) it allows the design storage to be maintained and it allows for better navigation [4]. The development of hydrological models to forecast the quantity of sediment that will be present in a river at a given time helps planners and managers of water resource systems to understand the system better in terms of

A.T. Jaiyeola is a Master's Student at the Department of Civil Engineering and Survey, Durban University of Technology, Durban, South Africa (Phone: +27730994561; e-mail: soj707@yahoo.com).

J. Adeyemo, Prof. is an Associate Professor at the Department of Civil Engineering and Survey, Durban University of Technology, Durban, South Africa (Phone: +27719087300; e-mail: josiaha@dut.ac.za).

World Academy of Science, Engineering and Technology
International Journal of Geological and Environmental Engineering
Vol:9, No:9, 2015

its problems and to find alternative ways to address them [5]. According to [6], many models are used for simulating and estimating sediment yield transport but they differ in the processes involved, the complexity and data required for calibration and usage. Furthermore, it was stated that the most appropriate model depends on the intended use and the features of the catchment in consideration. The following factors also affect the choice of a model: (1) the intended model capabilities and features; (2) the simplicity of the model and output scales and; (3) the assumptions made in using the model; (4) validity and accuracy of the model and (5) the data required by the model [6].

II. TYPES OF HYDROLOGICAL MODELS

Hydrological models can be categorized into three groups namely: empirical or statistical models, conceptual models and physic - based models.

A. Empirical or Statistical Models

Empirical models are based on the analysis of observations and the characterization of responses from the observed data [7]. These models require fewer amounts of data and lower computational requirements when compared to other types of models. They have a high level of spatial and temporal aggregation and also incorporate a small number of causal variables [8]. The parameter values are obtained from the calibration of experimental sites and are very effective in the identification of sediment sources and nutrient generation [6]. These models are criticized for using unrealistic assumptions concerning the physics of the catchment while ignoring the catchment heterogeneity inputs and characteristics such as soil types and rainfall [7]. They also do not respond to events thereby neglecting the impact of rainfall - runoff on the catchment being modelled [4]. Despite these shortcomings, the more complex and dynamic models in this regard cannot be considered as better when compared to other model groups. Empirical models can also be used as a first step when identifying sources of sediment and nutrient generation [6]. Fournier, Dendy and Bolton, and Revised Universal Soil Loss Equation (RUSLE) [9], [10] are examples of empirical models.

B. Conceptual Models

Conceptual models view flow path in catchment as a series of internal storages. They generally consider the description of catchment processes but neglect the specific interaction between the processes [11]. Therefore, both qualitative and quantitative effects of land use changes are indicated in these models. Parameter values in conceptual models are determined by the calibration against observed data, leading to problems associated with its identification [12], [13]. Spear [14], observed that as a result of the calibration techniques used for medium complexity models, many possible 'best' parameter sets can be made available. Thus calibration and identification of additional parameters using *a priori* knowledge of the system can limit the number of parameters to be estimated [15]. According to [16] conceptual models

can be used as an intermediary model between empirical models and physics-based models. They also reflect the principles that govern the system to be modelled. Examples of such models include Agricultural Nonpoint Source Pollution (AGNPS) and Morgan-Morgan-Finney (MMF).

C. Physics-Based Models

These are based on the use of basic physical equations which are solutions that describe discharge and sediment generation. Standard equations such as the equation of conservation of mass for sediment and the equations of conservation of mass and momentum for flow are used in physics-based models [17]. In theory, its parameters are measurable but conversely, this is not obtainable in practice due to their large numbers, hence they are calibrated against observed data [18]. The calibration of the model's parameter values when there are occurrences of missing values usually result in uncertainty in the model outcome. Furthermore, uncertainty can also occur when the parameter values cannot be measured. This is due to the likelihood of the occurrence of error during measurements [19]. Examples of widely-used physics-based models include Gridded Surface Subsurface Hydrologic Analysis (GSSHA) [20], [21], Hydrologic Simulation Program Fortran (HSPF) [22], Kinematic Runoff and Erosion Model (KINEROS2) [23], MIKE SHE [24] and Soil and Water Assessment Tool (SWAT) [25].

D. Selecting an Appropriate Model

The choice of a model depends mainly on its purpose, so a model may not necessarily be used for all modelling situations. From the literature [26]-[28], the choice of a model largely depends on where the emphasis will be laid, that is, either on the processes of the work or on the expected output that addresses the problem. For instance, in a study carried out by [28] both empirical and conceptual models were used to predict nitrate concentrations in groundwater aquifers in a catchment. The models' forecasting abilities were considered because of the semi-empirical nature of the process. Letcher et al. [26], in their technical report, argued that the combination of simple empirical and conceptual models can function more effectively when used within a developed framework. Also, [27] examined the relationship between the number of optimized parameters and model performance in 429 catchments. It was stated that over-parameterization of rainfall-runoff models can greatly affect the ability of the model to forecast stream flow [4].

III. SOFT COMPUTING AND ENGINEERING DESIGN

The desire to better understand life on earth has been in existence from the very beginning and this has been inspired by the efforts made by man to find ways to produce life-oriented behaviours [29]. The field of artificial intelligence has, year on year, produced a series of new methodologies incorporating the use of biologically-inspired computation methods to solve problems. These 'bio-inspired' models are related to and include the study of probabilistic reasoning, machine learning, emergence of novelty, complex adaptive

World Academy of Science, Engineering and Technology
International Journal of Geological and Environmental Engineering
Vol:9, No:9, 2015

systems, social behaviour, intelligence, sustainability and survival [30]. The past two decades have witnessed great improvement in the application of artificial intelligence techniques to solve problems where conventional approaches have failed or performed poorly [31]. A data-driven approach which utilizes soft computing or Artificial Intelligence (AI), namely, Artificial Neural Network (ANN), Fuzzy Logic (FL), and Genetic Algorithms (GA) has been used by researchers to address some of these issues [32]-[34]. AIs have been used successfully over the years in hydrology and water resources applications like sediment yield estimation models [35], rainfall-runoff modelling [36] and water quality prediction models [37]. In this study, emphasis is laid on past and recent applications of genetic programming (GP), a family of evolutionary algorithms (EAs) which adopts a biologically-inspired computation model to solve problems.

IV. ARTIFICIAL INTELLIGENCE IN HYDROLOGICAL STUDIES

Water is one of the most important natural resource. Even though its importance to our existence on earth cannot be over emphasized, it is also an important raw material for many industries [16]. Though this resource is abundant and almost 70% of the Earth's surface is covered with water, the amount of freshwater that can be effectively utilized is very limited [38]. There is an acute water shortage and people often have to face hardships in getting access to potable water due to depleting water tables, drying up of wells and rivers, and irregular rainfall. Due to these shortages, large swathes of land are being rendered barren. The excess water in the form of floods, which cause changes to river courses and rising sea levels, have resulted in large scale destruction of life and property [39]. The scarcity of water also has political implications in many countries around the globe. While some do not favour equitable sharing of river water, others oppose the construction of dams and the diversion of river waters.

From the foregoing, it is clear that proper management of this precious natural resource is very important for the survival of mankind and other forms of life [40]. This calls for research in finding and implementing new methods and techniques for the proper management of limited water resources to ensure a reliable supply of water for fulfilling the needs of the society [4]. Such new methods and techniques can be readily implemented if a good foundation for understanding water resources and the consumption patterns is laid. It is necessary to study the quality and quantity of water available over the years and then match this availability to the demands of various stakeholders. It is necessary to develop models that can predict the water inflow patterns and develop methods and techniques that can ensure proper utilization of water. The field of water resource modelling includes the accurate forecast of rainfall patterns and modelling of rainfall water flow [41]. It also involves the applications of various methods and techniques to purify collected water and wastewater for reuse. These processes are complex and thus the development of more accurate and reliable models for these processes is ongoing. There have been numerous attempts to develop these models, many of which have successfully served the purpose in specific situations, for example, an early attempt in forecasting rainfall patterns could be traced back to 1851 when Mulvany used self-registering rain and flood gauges to observe the relationship between flood discharges and rainfall [40].

On a broad scale, the different models used in water resource modelling are typically distinguished on the basis of the approach followed for describing the spatial extent of watershed and the hydrologic processes involved. The watershed models are typically classified as lumped or distributed models [42], [43] whereas the hydrological processes are classified as knowledge-driven models or data-driven models [44]. Another category of models called mechanistic models use differential equations to describe the processes at the surface and subsurface [45], [46]. A special class of mechanistic models which focus on storage elements are called conceptual models [47], [48]. Fewer numbers of parameters are required to describe the system in these models. There are also models that use information from hydro-meteorological data to map the relationship between rainfall and runoff, these models are known as precipitation-runoff data driven models.

V. COMPARISON OF ARTIFICIAL INTELLIGENCE (AI) MODELS

The important characteristics of all the models are briefly stated. K-nearest Neighbors algorithm (KNN) is one of the simplest techniques in AI. This KNN algorithm is very robust against noise and irrelevant attributes in the data. The disadvantage of KNN models is that they are unsuitable for real time forecasting and cannot be extrapolated [31]. Chaos Theory based models are accurate for long-term predictions but these models are applicable only if the data series is chaotic [34]. Similar to KNNs, the extrapolation abilities of these models are also poor. ANN based models have the advantage of being trained so fast and they map the complex relationships easily [49]. ANNs are also good for long term forecasting. These models are very robust against incomplete data, noise and outliers [41]. The disadvantages of ANN models are that they are not transparent, not easily interpretable and are difficult to generalize. Furthermore, these models cannot incorporate the knowledge about the systems within the model, hence it is called a black box model [50]. Fuzzy – Rule based systems (FRBS) can incorporate the structured knowledge within the model. These models are relatively transparent and provide some information about the rules used for mapping the inputs to outputs [33]. These are very robust to noise and very useful if the accuracy of sensors is low [51]. Limitations of these models include a slow convergence rate and an exponential increase in the number of rules with the increase in the number of input variables [52]. Support vector machines (SVMs) are easy to generalize and there is no increase in the number of parameters on giving multidimensional inputs. These models perform well even for small data sets [53]. The main disadvantages of SVMs is the requirement of a large computational capacity [53]. Furthermore, there is an exponential increase in training time when increasing the number of samples [54]. Genetic

World Academy of Science, Engineering and Technology
International Journal of Geological and Environmental Engineering
Vol:9, No:9, 2015

Programming (GPs) have an easily understandable model structure but they are also computationally intensive and hardly provide any physical insights about the underlying relationships [48], [49], [52], [55]-[60]. Each model reviewed in this study has its advantages and disadvantages. An understanding of the problems to be solved and the goals to be achieved determines which model is best suited for the task. Moreover, the models are complementary, so a hybrid approach is often better than using a single model. The next section reviews some recent studies that shed more light on the usefulness of these models in specific situations.

VI. RECENT APPLICATIONS OF ARTIFICIAL INTELLIGENCE IN WATER RESOURCES MODELLING

SVMs, nonparametric KNN, radial basis function NN were applied in the study of virtual water content of crops [61] and KNN was found to best serve the purpose. SVM performed better than the radial basis function Neutral Network (NN). In another study by [53], SVM was found to the best alternative out of a combination of SVM, probabilistic NN and KNN for the classification of water quality. KNN was the worst performer with the maximum number as well as value of errors [53]. GP and Adaptive Neuro-fuzzy Inference System (ANFIS) techniques were used for forecasting ground water levels by [51]. Based on different combinations of depth values of the water table from two stations, five GP and ANFIS models were developed, and forecasts were made for one, two and three-days ahead for the water table depths. Root mean square errors (*RMSE*), scatter index (*SI*), variance account for (*VAF*) and coefficient of determination (R^2) statistics were used for measuring the performance. Both models successfully forecasted water table depth fluctuations but GP performed better than the other model, giving explicit expressions for the problem [51]. In yet another study carried out by [62], hybrid wavelet-artificial neural network (WANN) and linear GP (LGP) techniques were proposed for forecasting monthly stream flow. RMSE and Nash-Sutcliffe efficiency (NSE) measures were used for comparing the performance. The ANN method was used as the primary reference model to model six different monthly stream flow scenarios based on the records of two successive gauging stations and the main time series of input(s) and output records were decomposed into sub-time series using the wavelet transform. These sub-time series of each model were imposed to ANN to develop WANN models as optimized versions of the reference ANN models. LGP performed better than WANN in all the reference models [62]. GP and ANN based methods were used as potential surrogate models for coastal aquifer management to determine the optimal rate of groundwater extraction. A 3-D simulation model for coupled flow and transport simulation together with an optimization algorithm in a linked simulation-optimization framework was used for the comparison [63]. GP again performed best against the other three models in estimating everyday suspended sediment load [63].

VII. OVERVIEW OF GENETIC PROGRAMMING (GP)

Genetic programming (GP), [64], which is a derivative of genetic algorithms (GA) is a systematic, domain-independent method that generates computer programs to solve problems automatically giving it a high level of what is expected from it [65]. GP involves a repeated random search for solutions from an existing pool of computer programs, which are potential solutions, by applying the principle of natural evolution such as crossover and mutation to form a new population [39]. This process continues until the best solution is obtained. These programs are expressed in the form of a syntax tree where the nodes represent the instructions called the functions, and the leaves, which are the terminals, represent the independent variables and random constants [65]. Five preliminary steps are necessary before the operation of GP. These include the determination of (i) the terminal set; (ii) the functional set; (iii) the fitness measure; (iv) the parameters for controlling the run; and (v) the termination criterion and method of designating the result of the run [66]. The steps involved in the implementation of GP are explained in details in the subsequent sections. According to [67] the major advantages of adopting GP over other soft computing techniques includes the following: (1) it is used when there is a large amount of data in computer readable form that needs to be examined, classified and integrated; (2) it is used in situations where small performance improvements are easily and routinely measured; (3) when the interrelationships between the variables are poorly understood; (4) when limited dataset is available; (5) when the ultimate solution to the problem is difficult to find; and (6) when conventional mathematical models cannot provide the required analytical solution [30], [39], [68], [69].

A. General Applications of Genetic Programming

GP is a robust and dynamic model. It has been widely applied to solve different kinds of difficult real world problems [41]. GP has the capability to select the best input from its variables, making it possible for the input and output variables to be expressed as a regression equation. GP has been applied to all aspects of life such as in the field of water resources engineering, photogrammetry, medicine, biology, electrical engineering, science, civil engineering, industrial engineering, electrical power and mechanical engineering [69]-[72]. The review of the applications of GP in all these fields is beyond the scope of this study but a brief overview of the application of GP in selected fields is provided, with emphasis on sediment modelling.

In the field of water resources management, [73] developed a rainfall-runoff model for predicting runoff using GP. A new formulation for bed concentration of suspended sediment was expressed by converting suspended sediment data into an equation for a better understanding of its generation process using GP techniques. In another study, the experimental flume data utilized by [74] was subjected to GP. This involves mining of data from sediment transportation near a riverbed. The results were compared with those from human experts and it was found to be very promising for mining of knowledge

World Academy of Science, Engineering and Technology
International Journal of Geological and Environmental Engineering
Vol:9, No:9, 2015

acquired data. Furthermore, GP was used to formulate sedimentary particle settling velocity equations by [75].

GP was applied by Liong, et al. [76] to predict rainfall-runoff in a catchment area in Singapore. The intensities and durations of six different storms were used to train and test the model. From the results obtained in the study, a consistent relationship between rainfall and runoff was identified. This implies that the application of the GP technique to forecast rainfall-runoff is a better alternative to other traditional models. The prediction of velocity of flow on wetlands and vegetated areas using the GP technique was explored by [77]. They discovered a symbolic expression from laboratory data that showed a better understanding of the effect of vegetation on velocity and discovery processes. GP and Artificial Neural Network (ANN) techniques were used to predict and model the rainfall-runoff relationship of a typical urban basin by [50]. Sivapragasam et al. [39] examined the relationship between storage and discharge in the Walla Walla river in the United State of America (USA). The researchers discovered that this relationship is insufficient for routing flood hydrographs on natural channels. Therefore, a GP model was developed for routing flood hydrographs. The developed model was very effective for routing complex flood hydrographs and it was able to express the route in a simple and mathematical expression.

Giustolisi [68] used GP to determine the coefficient of Chezy resistance in corrugated channels by using three corrugated plastic pipes to measure hydraulic parameters. His work produced two GP equations for Chezy resistance coefficients which represents the experiment data. Also, [78] used GP to predict soil characteristic curve by conducting pressure plate tests on silty clay, clay, loam and sandy loam using Soil Vision software. The test results were used for training and testing the GP model and the resultant model was compared with experimental results and other models, and GP model was found to be superior to them. Rabunal et al. [79] used GP to predict the unit hydrograph of a typical urban basin in conjunction with ANN. The two models were combined to establish an accurate relationship between rainfall and runoff in that basin.

GP was also used to predict short-term and long-term river flows and the result was found to be more accurate compared with that from ANFIS techniques [80]. In another study, [51] predicted groundwater table depth fluctuations using GEP and ANFIS. The results showed that both models can be used to successively predict the fluctuation but that GEP models were found to be more accurate. Shiri et al. [81] used limited climatic variables to model daily reference evapotranspiration, and also found that GEP performed better than ANFIS. In all the situations where the GP technique was applied, it proved itself to be accurate and superior to other techniques.

B. Sediment Modelling Using GP Approach

The GP approach has also been used successfully and intensively as a hydrological modelling tool especially for estimating sediment yield. Kizhisseri et al. [82] employed GP to develop a better sediment-temporal pattern of fluid field relationship, using numerical model results and Sandy Duck field data. Also, [35] developed an explicit relationship between daily suspended sediment and discharge using GP. Their results suggested that GP is a better technique than the sediment rating curve and multi-linear regression techniques and that GP is more practicable to use.

Garg [10] explored the ability of GP to estimate sediment yield in the Arno River basin in Italy which is susceptible to flooding. Five variables – river length, drainage density, yearly average rainfall, erodible area and watershed area, were used as input variables in this study and the results showed that GP is an efficient and reliable technique for estimating sediment yield even when the data set is limited. In a study carried out by [83], three soft computing techniques namely, Gene Expression Programming (GEP), Adaptive Neuro-Fuzzy Inference System (ANFIS) and Artificial Neural Networks (ANNs) were used to estimate daily suspended sediment load in the Eel River near Dos Rios, in California, USA. Suspended sediment load, stream flow and daily rainfall data were used as input for developing the models. The results, when compared, show that the GEP model is superior to the other developed models in predicting daily suspended load in the river.

LGP, GEP, and ANN techniques were used in the estimation of daily suspended sediment in the Tongue River in Montana, USA. Discharge and suspended sediment data from two stations on the river were used as inputs. GEP performed better than ANN but LGP models were found to be superior to the GEP models [84]. Also, in a study by [85], GP model trees, (MT) and ANN, which are data driven models, were used for the estimation of the quantity of sediment deposited in Gobindsagar reservoir [85]. It was found that both GP and ANN, which are nonlinear models, captured the trend of sediment deposition into the reservoir better than linear model trees (MT) [85].

GP models were also compared with SVM, ANFIS and ANN models by [86]. Daily discharge and sediment yield data from 1972 to 1989 obtained from two stations on the Cumberland River in the United States of America was used to test and train the models. The predicted outputs from the developed GP models were compared with those from the SVM, ANFIS and ANN models. The results showed that the GP models are superior to the other three models in predicting sediment yield. In another study, the data set of discharge and suspended sediment yield from Rio Valenciano and Quebrada Blanca Stations operated by the US Geological Survey (USGS) were also used as training and testing data by [87]. The performance of the developed LGP model, which is an extension of GP, and those of ANFIS and ANN models were compared using standard model evaluation criteria. Furthermore, it was discovered that the LGP model is superior and more accurate than both the ANFIS and ANN models.

In conclusion, GP models have been found to exhibit exceptional performance when used as regression models in the majority of the case studies mentioned, especially for pattern recognition and complex non-linear estimations. It was also found that GP is less prone to over-fitting during training with observation datasets [84]. In all, the application of GP

World Academy of Science, Engineering and Technology
International Journal of Geological and Environmental Engineering
Vol:9, No:9, 2015

may serve as a decisive factor in the planning, construction, operation, management and maintenance of water resources projects.

C. Evaluations of GP Theories and Principles

Irrespective of the wide acceptance and application of GP in solving complex problems and its application in diverse areas of life, there appears to be an eminent need to critically evaluate its theories and principles. According to [88], comparison between studies is very difficult because of the absence of standardization among the studies. It was claimed that the development of standard benchmarks is a very important step necessary for the maturation of the GP field. It should be noted that most of the studies reviewed applied GP to solve domain-specific and nontrivial problems; hence, simple benchmarks were used for analysis and comparison studies. This is detrimental to the advancement of the GP approach. O'Neill et al. [89] also stated that the development of a good benchmark suite is an issue to be considered in the next 10 years of GP. According to [88], a good benchmark should have the following qualities: relevance; speed; variance; accommodating to implementers; representation-independent; easy to interpret and compare; current; and precisely defined. Examples of current benchmarks include: predictive modelling; classification; binary functions; and symbolic regression [88]. It was further suggested by [88] that a candidate benchmark suite needs to be deliberated upon by the GP community. Another issue that needs serious attention is the generalization capability of GP solutions. GP should be able to produce the same generalization performance from training data set for unseen data. This ability is affected by bloating and over-fitting. Naik and Dabhi [90] surveyed and classified the various methods used in controlling bloating. Four bloat - control techniques were identified, namely: double tournament method; lexicographic parsimony pressure with ratio bucketing; lexicographic parsimony pressure with direct bucketing; and tarpeian method. These methods were applied to six different problems and the outcomes were analysed against each other. Based on this, tarpeian method and double tournament method were combined and used on the six problems. The study stated that the combination of these two methods performed better than the individual methods, except on a multi-valued regression problem without a constant [90].

It is tought that GP has a simple algorithm but the process of obtaining a sound theoretical model and precise mathematical results has been difficult to obtain spanning many years after the origin of the GP technique [91]. According to [92], the delay for this was as a result of the different versions of GP requiring different theoretical models. In addition, the different representations of GP such as tree-based, graph-based and linear, differ in dynamics and require different theoretical tools. Theoreticians are facing a lot of challenges due to the non-linearities, randomness, and numerous degrees of freedom present in a typical GP system as well as in Grammatical Evolution (GE) [93], Evolution Programming (EP) [94] and Cartesian GP (CGP) [95].

Poli et al. [92] also highlighted some fundamental questions that needs the attention of the GP community. Questions such as: What goes wrong when GP cannot solve a problem? Is the biasness of its genetic operators fully understood? Can the properties of both GP systems and GP problems be expressed in a common language? How should current GP theory be adapted to suit dynamic environments? These questions and many more need to be addressed and are meant to stimulate researchers into improving the theories behind GP.

Some very important open issues in GP were discussed in [89] which started at a panel discussion at the EuroGP series of conferences which began in 1998. Some of the issues include: representing GP appropriately; determining the level of difficulty of a problem for GP; comparing the performance of GP in both static and dynamic environments; determining the level of natural evolution. For complete and comprehensive details of these open issues readers are referred to [89]. These issues are meant to help researchers improve the techniques of GP and to stimulate future research to provide greater knowledge and strengthen the GP algorithm.

VIII. CONCLUSION

The applications of AI models and techniques to real life situations have brought greater understanding to evolutionary computing. GP approaches have been used successfully in providing solutions to many complex problems in the fields of hydrology, medicine, biology, photogrammetry, telecommunications and network, finance, gaming and engineering. In this paper, certain areas of the application of GP were also highlighted, with specific emphasis on the applications of GP to sediment yield modelling. The community of GP operators and researchers needs to be optimistic because as this approach matures, important advances in the theory of GP are being achieved.

Although progress has been made and opportunities have opened up in this field over the past decades, there are still some open issues that needs the attention of the GP communities. Although the review in this study may not be exhaustive, it is meant to give GP theoreticians and the general community of GP sufficient research direction and open issues that needs attention. These open issues and the successes made in the application of GP to solve complex and nonlinear problems should be discussed in the literature so that it can spread into the science and research communities. This will generate more intellectual and productive discussions that will eventually lead to the advancement of GP technique.

The following factors also affect the choice of a model: (1) the intended model capabilities and features; (2) the simplicity of the model and output scales and; (3) the assumptions made in using the model; (4) validity and accuracy of the model and (5) the data required by the model [6]. In addition, in selecting the appropriate model, the choice of a model depends mainly on its purpose, so a model may not necessarily be used for all modelling situations. The choice of a model largely depends on where the emphasis will be laid, that is, either on the processes of the work or on the expected output that addresses the problem.

World Academy of Science, Engineering and Technology
International Journal of Geological and Environmental Engineering
Vol:9, No:9, 2015

Recently a combination of different AI models has been used for finding optimal solutions. For example, [96] used a wavelet analysis and GP method to construct a hybrid model for optimizing rainfall-runoff time process modeling [96]. The hybrid model that linked wavelet analysis to GP used sensitivity analysis for identifying input variables of an ANN rainfall-runoff model. Furthermore, the time series of both the rainfall and runoff variables, were decomposed into many multi-frequency time series using the wavelet transform and these were imposed to the GP as input data to optimize the structure of ANN modeling. The results could be compared favorably to both GP and ANN models. The same methodology also worked in predicting the suspended sediment load in rivers [97]. The introduction of wavelet coefficient inputs from wavelet GP and wavelet neuro-fuzzy, for forecasting daily precipitation on the basis of previously recorded results, resulted in an improvement of the forecasting results. Further enhancement of the model by merging the inputs from both (best single model and hybrid models) and using these as the model inputs, enabled the new hybrid wavelet GP models to successfully predict the daily precipitation, although the neuro-fuzzy models still performed badly [98].

A combination of Extended Kalman Filter (EKF) with GP was used by [4] to forecast the water demand in urbanized areas. The latent variables were inferred from the EKF. Five models were presented where the first five-three lags of observed water requirement were used as the independent and most probable inputs. The effect of observation precision on water demand prediction was clearly evident in their findings.

REFERENCES

[1] P. G. Griffiths, R. Hereford, and R. H. Webb, "Sediment yield and runoff frequency of small drainage basins in the Mojave Desert, California and Nevada," ed, 2006.

[2] M. J. Bender, L. F. Sawatsky, D. Long, and P. Anderson, *A strategy for determining acceptable sediment yield for reclaimed mine lands.* Italy: UNESCO, 2005.

[3] R. Loch and D. Silburn, "Constraints to sustainability—soil erosion," *Sustainable Crop Production in the Sub-tropics: an Australian Perspective. QDPI,* 1996.

[4] M. Nasseri, A. Moeini, and M. Tabesh, "Forecasting monthly urban water demand using Extended Kalman Filter and Genetic Programming," *Expert Systems with Applications,* vol. 38, pp. 7387-7395, 6// 2011.

[5] D. P. Loucks and E. V. Beek, *An introduction to methods,models,and application.,* 2005.

[6] W. S. Merritt, R. A. Letcher, and A. J. Jakeman, "A review of erosion and sediment transport models," *Environmental Modelling & Software,* vol. 18, pp. 761-799, 2003.

[7] H. Wheater, A. Jakeman, and K. Beven, "Progress and directions in rainfall-runoff modelling," 1993.

[8] A. J. Jakeman, T. R. Green, S. G. Beavis, L. Zhang, C. R. Dietrich, and P. F. Crapper, " Modelling upland and in-stream erosion, sediment and phosphorus transport in a large catchment," *Hydrological Processes* vol. 13, pp. 745–752, 1999.

[9] M. Kouli, P. Soupios, and F. Vallianatos, "Soil erosion prediction using the revised universal soil loss equation (RUSLE) in a GIS framework, Chania, Northwestern Crete, Greece," *Environmental Geology,* vol. 57, pp. 483-497, 2009.

[10] V. Garg, "Modeling catchment sediment yield: a genetic programming approach," *Natural Hazards,* 2011.

[11] S. Sorooshian, "Parameter estimation, model identification, and model validation: conceptual-type models," in *Recent advances in the modeling of hydrologic systems*, ed: Springer, 1991, pp. 443-467.

[12] M. B. Abbott, J. C. Bathurst, J. A. Cunge, P. E. O'Connell, and J. Rasmussen, "An introduction to the European Hydrological System—Systeme Hydrologique Europeen, SHE. 1. History and philosophy of a physically-based, distributed modelling system," *Journal ofHydrology,* vol. 87, pp. 45–59., 1986.

[13] A. Jakeman and G. Hornberger, "How much complexity is warranted in a rainfall-runoff model?," *Water Resources Research,* vol. 29, pp. 2637-2649, 1993.

[14] R. C. Spear, "Large simulation models: calibration, uniqueness and goodness of fit," *Environmental Modelling & Software,* vol. 12, pp. 219-228, 1997.

[15] F. Kleissen, M. Beck, and H. Wheater, "The identifiability of conceptual hydrochemical models," *Water Resources Research,* vol. 26, pp. 2979-2992, 1990.

[16] M. B. Beck, "Water quality modeling: a review of the analysis of uncertainty," *Water Resources Research,* vol. 23, pp. 1393-1442, 1987.

[17] J. P. Bennett, "Concepts of mathematical modeling of sediment yield," *Water Resources Research,* vol. 10, pp. 485-492, 1974.

[18] M. B. Beck, A. J. Jakeman, and M. J. McAleer, "Construction and evaluation of models of environmental systems," *In: Beck, M.B., McAleer, M.J. (Eds.), Modelling Change in Environmental Systems.John Wiley and Sons,* pp. pp. 3–35, 1995.

[19] J. D. Kalma and M. Sivapalan, *Scale issues in hydrological modelling*: John Wiley and Sons, 1995.

[20] D. K. Borah, "Hydrologic procedures of storm event watershed models: a comprehensive review and comparison," *Hydrological Processes,* vol. 25, pp. 3472-3489, 2011.

[21] N. R. Pradhan, C. W. Downer, and B. E. Johnson, "A physics based hydrologic modeling approach to simulate non-point source pollution for the purposes of calculating TMDLs and designing abatement measures," in *Practical Aspects of Computational Chemistry III*, ed: Springer, 2014, pp. 249-282.

[22] S. Kim, D.-J. Seo, H. Riazi, and C. Shin, "Improving water quality forecasting via data assimilation–Application of maximum likelihood ensemble filter to HSPF," *Journal of Hydrology,* vol. 519, pp. 2797-2809, 2014.

[23] A. K. Sajjan, Y. Gyasi-Agyei, and R. H. Sharma, "Modeling Grass-Cover Effects on Soil Erosion on Railway Embankment Steep Slopes," *Journal of Hydrologic Engineering,* 2014.

[24] I. B. Karlsson, T. O. Sonnenborg, J. C. Refsgaard, and K. H. Jensen, "Significance of hydrological model choice and land use changes when doing climate change impact assessment," 2014.

[25] J. jeanne Huang, X. Lin, J. Wang, and H. Wang, "The precipitation driven correlation based mapping method (PCM) for identifying the critical source areas of non-point source pollution," *Journal of Hydrology,* 2015.

[26] R. A. Letcher, A. J. Jakeman, W. S. Merritt, L. J. McKee, B. D. Eyre, and B. Baginska, "Review of techniques to estimate catchment exports," 1999.

[27] C. Perrin, C. Michel, and V. Andréassian, "Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments," *Journal of Hydrology,* vol. 242, pp. 275-301, 2001.

[28] M. Thorsen, J. Refsgaard, S. Hansen, E. Pebesma, J. Jensen, and S. Kleeschulte, "Assessment of uncertainty in simulation of nitrate leaching to aquifers at catchment scale," *Journal of Hydrology,* vol. 242, pp. 210-227, 2001.

[29] C. Adami, *Introduction to artificial life*: Springer, 1998.

[30] W. Banzhaf, "Evolutionary Computation and Genetic Programming," 2012.

[31] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," *Progress in Energy and Combustion Science,* vol. 34, pp. 574-632, 2008.

[32] A. S. Tokar and P. A. Johnson, "Rainfall-runoff modeling using artificial neural networks," *Journal of Hydrologic Engineering,* vol. 4, pp. 232-239, 1999.

[33] J. Lin and F. L. Lewis, "Two-time scale fuzzy logic controller of flexible link robot arm," *Fuzzy sets and systems,* vol. 139, pp. 125-149, 2003.

[34] P. Vas, *Artificial-intelligence-based electrical machines and drives: application of fuzzy, neural, fuzzy-neural, and genetic-algorithm-based techniques* vol. 45: Oxford University Press, 1999.

[35] A. Aytek and O. Kisi, "A genetic programming approach to suspended sediment modelling," *Journal of Hydrology,* vol. 351, pp. 288– 298, 2008.

World Academy of Science, Engineering and Technology
International Journal of Geological and Environmental Engineering
Vol:9, No:9, 2015

[36] J. Smith and R. N. Eli, "Neural-network models of rainfall-runoff process," *Journal of water resources planning and management,* vol. 121, pp. 499-508, 1995.

[37] D. Ömer Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction," *Engineering Applications of Artificial Intelligence,* vol. 23, pp. 586-594, 2010.

[38] J. Lloret, "Underwater sensor nodes and networks," *Sensors,* vol. 13, pp. 11782-11796, 2013.

[39] C. Sivapragasam, R. Maheswaran, and V. Venkatesh, "Genetic programming approach for flood routing in natural channels," *Hydrological processes,* vol. 22, pp. 623-628, 2008.

[40] T. Mulvaney, "On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment," *Proceedings of the institution of Civil Engineers of Ireland,* vol. 4, pp. 18-33, 1851.

[41] Y. M. Chiang and F. J. Chang, "Integrating hydrometeorological information for rainfall-runoff modelling by artificial neural networks," *Hydrological Processes,* vol. 23, pp. 1650-1659, 2009.

[42] F. Anctil, C. Perrin, and V. Andreassian, "ANN output updating of lumped conceptual rainfall/runoff forecasting models1," ed: Wiley Online Library, 2003.

[43] V. P. Singh and D. A. Woolhiser, "Mathematical modeling of watershed hydrology," *Journal of hydrologic engineering,* vol. 7, pp. 270-292, 2002.

[44] D. Solomatine and A. Ostfeld, "Data-driven modelling: some past experiences and new approaches," *Journal of hydroinformatics,* vol. 10, pp. 3-22, 2008.

[45] K. Beven, A. Calver, and E. Morris, "The Institute of Hydrology distributed model," 1987.

[46] D. Yang, S. Herath, and K. Musiake, "Development of a geomorphology-based hydrological model for large catchments," *Annual Journal of Hydraulic Engineering, JSCE,* vol. 42, pp. 169-174, 1998.

[47] S. Bergström and V. Singh, "The HBV model," *Computer models of watershed hydrology.,* pp. 443-476, 1995.

[48] V. P. Singh, *Computer models of watershed hydrology*: Water Resources Publications, 1995.

[49] B. Yegnanarayana, *Artificial neural networks*: PHI Learning Pvt. Ltd., 2009.

[50] J. Dorado, J. R. RabuñAL, A. Pazos, D. Rivero, A. Santos, and J. Puertas, "Prediction and modeling of the rainfall-runoff transformation of a typical urban basin using ANN and GP," *Applied Artificial Intelligence,* vol. 17, pp. 329-343, 2003.

[51] J. Shiri and Ö. Kişi, "Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations," *Computers & Geosciences,* vol. 37, pp. 1692-1701, 2011.

[52] G. Tayfur, *Soft computing in water resources engineering: artificial neural networks, fuzzy logic and genetic algorithms*: WIT Press/Computational Mechanics, 2011.

[53] F. Modaresi and S. Araghinejad, "A Comparative Assessment of Support Vector Machines, Probabilistic Neural Networks, and K-Nearest Neighbor Algorithms for Water Quality Classification," *Water Resources Management,* vol. 28, pp. 4095-4111, 2014.

[54] B. Scholkopf, A. Smola, R. C. Williamson, and P. Bartlett, "New support vector algorithms," *Neural Computation,* vol. 12, pp. 1207–1245, 2000.

[55] R. S. Govindaraju and A. R. Rao, *Artificial neural networks in hydrology*: Springer Publishing Company, Incorporated, 2010.

[56] J. D. Salas, *Applied modeling of hydrologic time series*: Water Resources Publication, 1980.

[57] B. Sivakumar and R. Berndtsson, "Nonlinear Dynamic and Chaos in Hydrology," in *Advances in data-based approaches for hydrologic modeling and forecasting*, B. Sivakumar and R. Berndtsson, Eds., ed Singapore: World Scientific, 2010, pp. 411-461.

[58] C. Wu, "Hydrological predictions using data-driven models coupled with data preprocessing techniques," The Hong Kong Polytechnic University, 2010.

[59] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*: John Wiley & Sons, 2013.

[60] R. Poli and J. Koza, *Genetic Programming*: Springer, 2014.

[61] K. Ahmadaali, A. Liaghat, O. B. Haddad, and N. Heydari, "Estimation of Virtual Water Using Support Vector Machine, K-nearest neighbour, and Radial Basis Function Neural Network Models," *International Journal of Agronomy and Plant Production,* vol. 4, pp. 2926-2936, 2013.

[62] A. Danandeh Mehr, E. Kahya, and E. Olyaie, "Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique," *Journal of Hydrology,* vol. 505, pp. 240-249, 2013.

[63] J. Sreekanth and B. Datta, "Comparative evaluation of genetic programming and neural network as potential surrogate models for coastal aquifer management," *Water resources management,* vol. 25, pp. 3201-3218, 2011.

[64] J. R. Koza, *Genetic programming: on the programming of computers by means of natural selection* vol. 1: MIT press, 1992.

[65] R. Poli, W. W. B. Langdon, N. F. McPhee, and J. R. Koza, *A field guide to genetic programming*: Lulu. com, 2008.

[66] E. K. Burke and G. Kendall, *Search methodologies: introductory tutorials in optimization and decision support techniques*: Springer, 2005.

[67] W. Banzhaf, P, R. E. Keller, and F. D. Francone, *Genetic programming: an introduction*. San Francisco (CA): Morgan Kaufmann, 1998.

[68] O. Giustolisi, "Using genetic programming to determine Chezy resistance coefficient in corrugated channels," *Journal of Hydroinformatics,* vol. 6, pp. 157-173, 2004.

[69] A. P. Mitra, A. A. Almal, B. George, D. W. Fry, P. F. Lenehan, V. Pagliarulo*, et al.*, "The use of genetic programming in the analysis of quantitative gene expression profiles for identification of nodal status in bladder cancer," *BMC cancer,* vol. 6, p. 159, 2006.

[70] A. Makkeasorn, N.-B. Chang, and J. Li, "Seasonal change detection of riparian zones with remote sensing images and genetic programming in a semi-arid watershed," *Journal of Environmental Management,* vol. 90, pp. 1069-1080, 2009.

[71] R. Nunkesser, T. Bernholt, H. Schwender, K. Ickstadt, and I. Wegener, "Detecting high-order interactions of single nucleotide polymorphisms using genetic programming," *Bioinformatics,* vol. 23, pp. 3280-3288, 2007.

[72] L. Zhang, L. B Jack, and A. K. Nandi, "Fault detection using genetic programming," *Mechanical Systems and Signal Processing,* vol. 19, pp. 271-289, 2005.

[73] D. A. Savic, G. A. Walters, and J. W. Davidson, "A genetic programming approach to rainfall-runoff modelling," *Water Management,* vol. 13, pp. 219-231, 1999.

[74] J. A. Zyserman and J. Fredsøe, "Data analysis of bed concentration of suspended sediment," *Journal of Hydraulic Engineering,* vol. 120, pp. 1021-1042, 1994.

[75] V. Babovic, M. Keijzer, D. Aguilera, and J. Harrington, "Automatic discovery of settling velocity equations," *D2K Technical Rep,* p. 1, 2001.

[76] S. Y. Liong, T. R. Gautam, S. T. Khu, V. Babovic, M. Keijzer, and N. Muttil, "GENETIC PROGRAMMING: A NEW PARADIGM IN RAINFALL RUNOFF MODELING1," *JAWRA Journal of the American Water Resources Association,* vol. 38, pp. 705-718, 2002.

[77] E. Harris, V. Babovic, and R. Falconer, "Velocity predictions in compound channels with vegetated floodplains using genetic programming," *International Journal of River Basin Management,* vol. 1, pp. 117-123, 2003.

[78] A. Johari, G. Habibagahi, and A. Ghahramani, "Prediction of soil–water characteristic curve using genetic programming," *Journal of Geotechnical and Geoenvironmental Engineering,* vol. 132, pp. 661-665, 2006.

[79] J. Rabunal, J. Puertas, J. Suarez, and D. Rivero, "Determination of the unit hydrograph of a typical urban basin using genetic programming and artificial neural networks," *Hydrological processes,* vol. 21, pp. 476-485, 2007.

[80] O. Kisi and J. Shiri, "A comparison of genetic programming and ANFIS in forecasting daily, monthly and daily streamflows," in *Proceedings of the international symposium on innovations in intelligent systems and applications*, 2010, pp. 118–122.

[81] J. Shiri, Ö. Kişi, G. Landeras, J. J. López, A. H. Nazemi, and L. C. Stuyt, "Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain)," *Journal of Hydrology,* vol. 414, pp. 302-316, 2012.

[82] A. S. Kizhisseri, D. Simmonds, Y. Rafiq, and M. Borthwick, "An evolutionary computation approach to sediment transport modeling," in *Fifth International Conference on Coastal Dynamics*, 2005, pp. 4-8.

[83] K. Ozgur and S. Jalal, "River suspended sediment estimation by climatic variables implication:Comparative study among soft computing techniques," *Computers & Geosciences,* vol. 43, pp. 73-82, 2012.

[84] A. Guven and Ö. Kişi, "Estimation of suspended sediment yield in natural rivers using machine-coded linear genetic programming," *Water resources management,* vol. 25, pp. 691-704, 2011.

[85] V. Garg and V. Jothiprakash, "Evaluation of reservoir sedimentation using data driven techniques," *Applied Soft Computing,* vol. 13, pp. 3567–3581, 2013.

[86] O. Kisi, A. H. Dailr, M. Cimen, and J. Shiri, "Suspended sediment modeling using genetic programming and soft computing techniques," *Journal of Hydrology,* vol. 450-451, pp. 48-58, 2012.

[87] O. Kisi and A. Guven, "A machine code-based genetic programming for suspended sediment concentration estimation," *Advances in Engineering Software,* vol. 41, pp. 939-945, 7// 2010.

[88] J. McDermott, D. R. White, S. Luke, L. Manzoni, M. Castelli, L. Vanneschi*, et al.*, "Genetic programming needs better benchmarks," in *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference*, 2012, pp. 791-798.

[89] M. O'Neill, L. Vanneschi, S. Gustafson, and W. Banzhaf, "Open issues in genetic programming," *Genetic Programming and Evolvable Machines,* vol. 11, pp. 339-363, 2010.

[90] T. R. Naik and V. K. Dabhi, "Improving Generalization Ability of Genetic Programming: Comparative Study," *arXiv preprint arXiv:1304.3779,* 2013.

[91] R. Poli, "Exact Schema Theory for Genetic Programming and Variable-length Genetic Algorithms with One-Point Crossover " *Genet. Program. Evol. Mach. ,* vol. 2, p. 163, 2001.

[92] R. Poli, L. Vanneschi, W. B. Langdon, and N. F. McPhee, "Theoretical results in genetic programming: the next ten year?," *Genet Program Evolvable,* pp. 285–320, 2010.

[93] C. Ryan, J. Collins, and M. O. Neill, "Grammatical evolution: Evolving programs for an arbitrary language," in *Genetic Programming*, ed: Springer, 1998, pp. 83-96.

[94] A. J. Owens, M. J. Walsh, and L. J. Fogel, *Artificial intelligence through simulated evolution*, 1966.

[95] J. F. Miller, "An empirical study of the efficiency of learning boolean functions using a cartesian genetic programming approach," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 1999, pp. 1135-1142.

[96] V. Nourani, R. G. Ejlali, and M. T. Alami, "Spatiotemporal groundwater level forecasting in coastal aquifers by hybrid artificial neural network-geostatistics model: a case study," *Environmental Engineering Science,* vol. 28, pp. 217-228, 2011.

[97] T. Rajaee, V. Nourani, M. Zounemat-Kermani, and O. Kisi, "River suspended sediment load prediction: Application of ANN and wavelet conjunction model," *Journal of Hydrologic Engineering,* vol. 16, pp. 613-627, 2010.

[98] O. Kisi and J. Shiri, "Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models," *Water resources management,* vol. 25, pp. 3135-3152, 2011.