(3)

GII
2nd Indo-Italian workshop,
On Hydraulic Sciences,
July 19-20, 2012, Venice, Italy.

Reservoir Inflow Modeling Using Artificial Intelligence Techniques and its Use in Real Life

V. Jothiprakash

Department of Civil Engineering
Indian Institute of Technology Bombay, Mumbai, India

R. B. Magar

Department of Civil Engineering
Anjuman I Islam Kalsekar Technical Complex, School of Engineering and Technology, Navi Mumbai, India

Abstract

Operational planning of water resource systems like reservoirs and hydropower plants calls for real-time reservoir inflow forecasting. In spite of large number of works, the underlying phenomena are extremely complex, non-linear and uncertain, as a result of which the outcome may not always yield entirely satisfactory results. With the advancement in computing facilities, field engineers are seeking more efficient tools for predicting daily and hourly real-time reservoir inflow at shorter time intervals. Deterministic and conventional stochastic models lead to less accurate prediction since they don't capture non-linearity of the catchment process with shorter time intervals. Existing conceptual, physical and numerical models might be associated with problems like modeling difficulties and heavy computational demands. Artificial intelligence (AI), as a branch of computer science, is capable of analyzing long-series and large-scale hydrological data. These techniques are more flexible, less assumption-dependent and potentially self adaptive approach, which by their nature are inherently complex, non-linear and dynamic. Moreover, these techniques can be used for modelling the systems on a real-time basis ever for every short time intervals. In recent years, it is one of front issues to apply AI technology to the hydrological forecasting modeling and transfer the knowledge to the field engineers.

Keywords: Soft Computing Technique, Artificial Neural Network, Adaptive Neuro Fuzzy Inference System, Genetic Programming, Model Tree

Introduction

Reservoir inflow forecast is a key component in planning, development, design, operation and maintenance of the available water resources. Inflow forecast models are useful in many water resources applications such as flood control, drought management, optimal reservoir operation, hydropower generation etc [1]. Thus, the identification of suitable generation model for future inflow is necessary for successful planning and management of water resources structures [2]. According to the use of the observational data and to the description of the physical processes large number of empirical, conceptual, physically based and data driven models has been developed and applied to map the rainfall-runoff (RR) relationship [3 & 4]. However, each model has its own advantages and disadvantages [5 & 6]. These models suffer from problems such as identification, assimilability and uniqueness of parameter estimation [7]. Some of the earliest approaches have employed conventional time-series forecasting and modelling [8, 9 & 10] assuming that the data taken over time may have an internal structure, but provided only reasonable accuracy and suffered from the assumptions of stationary and linearity.

Soft Computing Tools in Rainfall-Runoff Modeling

Recently artificially intelligent (AI) based data driven techniques have emerged as important tools to overcome the drawbacks of conventional modelling technique. These data driven approaches are based on extracting and re-using information implicitly contained in hydrological time-series without directly taking into account the physical laws that underline the process [11]. During the last few decades the area of empirical modelling received an important boost due to development in the area of machine learning. It has now entered into a new phase and can be classified separately as AI modelling techniques [4]. These models have been developed with the contributions from artificial intelligent, data mining, knowledge discovery in databases, computational intelligent, machine learning, intelligent data analysis, soft computing, pattern recognition, etc. Data driven modelling is based on the analysis of the data characterizing the system under study. A model can then be defined on the basis of connections between the system state variables (input and output variables) with only a limited number of assumptions about the physical behavior of the system. Soft computing is an emerging field that consists of complementary elements of neural computing, Fuzzy Logic, evolutionary computation, machine learning and probabilistic reasoning. In the last decade, artificial neural network (ANN) has been successfully employed in modelling a wide range of hydrologic processes especially rainfall and runoff due to their ability to model non-linear system efficiently [12, 13, 14, 15, 16]. It is proved and demonstrated that the soft computing techniques are good to model the complex rainfall–inflow process and is better than the conventional modeling techniques. These techniques improve the model performance, help faster model development and calculation times [17, 18, 19, 20]. These methods have the ability to handle large amount of data from dynamic and nonlinear systems. However still there remains a difficult in extracting the exact knowledge of ANN models. The recent AI techniques, model tree (MT) and genetic programming (GP) are overcoming this issue.

Even though many types of reservoir inflow forecasting models are available, the problem of accurate estimation and forecast still persist at field level. All the models developed in the laboratory need to be tested and used in field, where field engineers would also be able to understand, handle and apply the model to predict the real-time reservoir inflow. With the advancement of computer facilities at gross root level, there is a wide scope for developing sitespecific graphical user interface (GUI) interactive software as a decision support system for field engineers.

Artificial Neural Networks (ANN)

ANN is essentially a group of interconnected computing elements, or neurons that has certain performance characteristics resembling biological neural networks of the human brain [21]. ANNs were proposed approximately 75 years ago by McCulloch and Pitts [22] inspired by a desire to understand functions of the human brain and to simulate their functioning [23]. Werbos [24] proposed the first back propagation (BP) ANN in 1974. Nevertheless, the powerfulness of BP networks was not recognized until Rumelhart et al. [25] modified the architecture of BP networks. With all the efforts researchers have developed more sophisticated ANN algorithms [23]. ANN models are able to learn the underlying patterns between the inputs and outputs. This feature makes ANN suitable for modelling natural systems where complex relationship exist between the inputs and outputs, and data are often incomplete or noisy [19, 26 & 27]. ANNs are inherently non-linear, generalisable and noise tolerant. They have been found to be a robust tool for simulating many non-linear hydrologic processes, such as rainfall-runoff, stream flow, water quality, reservoir operation, inflow forecasting, and others [28].

Adaptive Neuro-Fuzzy Inference System (ANFIS)

Jang [29] introduced architecture and learning procedure for the FIS that uses a neural network learning algorithm for constructing a set of fuzzy if-then rules with appropriate membership functions (MFs) from the specified input–output pairs. This procedure of developing a FIS using the framework of adaptive neural networks is called ANFIS. There are two methods that ANFIS learning employs for updating membership function parameters: 1) back propagation for all parameters (a steepest descent method), and 2) a hybrid method consisting of back propagation for the parameters associated with the input membership and least squares estimation for the parameters associated with the output membership functions. As a result, the training error decreases, at least locally, throughout the learning process. Therefore, the more the initial membership functions resemble the optimal ones, the easier it will be for the model parameter training to converge. Human expertise about the target system to be modelled may aid in setting up these initial membership function parameters in the FIS structure. A review of the application of ANFIS to hydrology can be found in [30, 31].

Linear Genetic Programming

In spite of various advantages of applying ANN and ANFIS technique, transferring the knowledge gained through modeling particularly the relationship between input and output to the field engineers is rather limited. Hence, researchers are seeking other data based techniques where the knowledge can be extracted very easily. The Darwin's natural selection theory of evolution based GP, is relatively a new technique and is the member of evolutionary algorithm family [32]. GP is an inductive form of machine learning as it evolves a computer program to perform an underlying process defined by a set of training samples [33]. GP has been successfully applied to complex non-linear problems and its solution describes the input-output relationship.

In the concept of GP introduced by Koza [32], the programs are represented as tree structures and expressed in the LISP functional programming language [34, 35]. Later, researchers regarded it as tree based GP (TGP) due to its tree structural solution. Recently, a subset of GP has emerged, which evolves programs in an imperative programming language (C/C++) and represent the graph-based functional structure, termed as LGP [36]. The imperative program structure of LGP identifies the non-effective instructions efficiently and executes rapidly [37, 38]. Moreover, in LGP, the maximum size of the program was usually restricted to avoid over-growing programs without any condition [37]. The name 'linear' refers to the structure of the (imperative) program representation, and does not stand for functional GP. LGP represents highly non-linear solutions in this meaning [34, 35]. The main advantage of LGP is its ability to produce models that build an understandable structure given that the LGP model exhibits a great potential to screen and prioritize the input variables. The various LGP parameters involved are population size, mutation rate and its different types (block mutation rate, instruction mutation rate and data mutation rate), crossover rate, homologous crossover, function set, number of demes and program size. The parameter selection will affect the model generalization capability of LGP. They were selected based on some previously suggested values [39] and also after trial and error approach. LGPs can be converted into a functional representation by successive replacements of variables starting with the last effective instruction [40].

Study Area

The area selected for the present study is Koyna watershed, situated on the West Coast of Maharashtra, India, lies between the latitude of 17°00'N to 17°59'N and longitude of 73°02'E to 73°35'E. The location of the study area along with nine rain-gauge stations in Koyna watershed is shown in Fig. 1[41]. The Koyna Dam is one among the 23,000 large dams in the world with a gross storage capacity of 2797.4 x $10^6 m^3$. The height of the Koyna dam above foundation level is 103 m and the length of the dam at the crest is about 800 m. The Koyna project is a multi-purpose project, but primarily designed as a hydroelectric project that supplies hydroelectric power to Maharashtra state, India with an installed capacity of 1920 MW. Koyna watershed has an elongated leaf shape, about 64 km in length and about 13 km width with an area of 891.78 km^2. The watershed is bounded by hills and broadly consists of 41% forest, 49% cultivated area, 6% wasteland 4% of others [42]. The water spread area at full reservoir level is 115.36 km^2 which is about 13% of the total catchment area. Nearly 99% of the annual rainfall in this basin occurs during south-west monsoon (June to October) and varies from 2972 mm to 6694 mm annually over the valley.

Application of AI techniques and selection of appropriate method

The available data of Koyna reservoir has been used to develop the AI models. The major aim of the work is to develop AI models and select an appropriate model to implement it in field. The ANN model has been developed by trial and error approach [41]. The LGP model has been developed using Discipulus [38, 39], out of the available 49 years of data 70% of data length is used for model development and 30% for model testing. From the results [43], it is found that both ANN and LGP are performing equally better. Out of ANN and LGP, the latter one had an edge over the ANN, especially in hourly time step real-time reservoir inflow prediction. It is also realized that the extraction of relationship between the input and output is somewhat complicated in ANN model. However, the output of the LGP model in the form of C++ program is found to be very useful to develop a Graphical User Interface (GUI) software to predict daily and hourly real time reservoir inflow prediction. In order to implement the developed reservoir inflow prediction models in real life at Koyna reservoir a GUI has been developed using the best models resulted from this study. Thus in the present study, this input and output relationship given by LGP in the form of C++ program has been used in the development of GUI software to be used in the field.

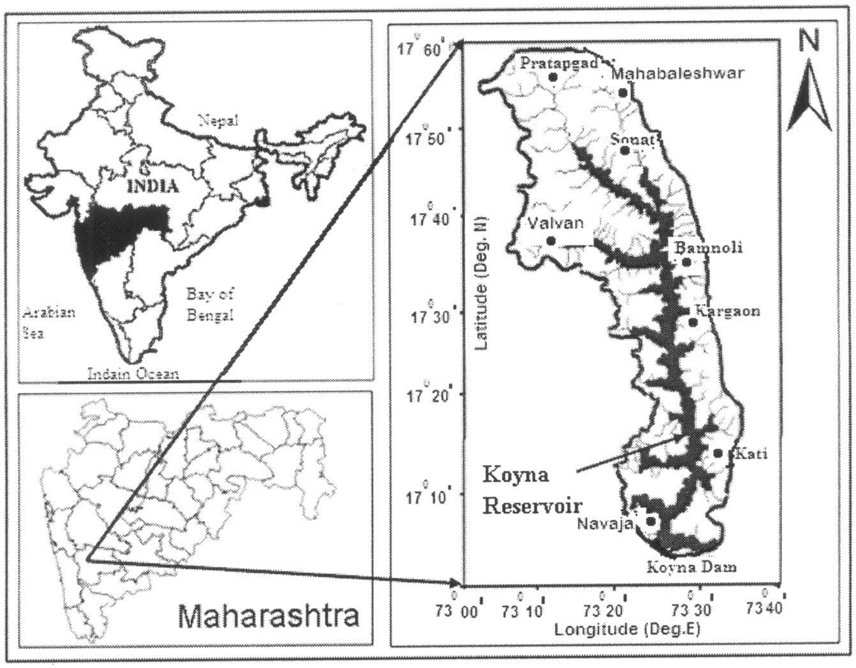

Figure 1: Location of Koyna Watershed[41]

Graphical User Interface (GUI)

This GUI is used as a decision support system and is named as Koyna Reservoir-Inflow Forecasting System "KRIFS". "KRIFS" is a user-friendly graphical interface developed using Microsoft's Visual Studio.NET as front end and Microsoft's Access 2003 as back end. This software is developed with intent to forecast daily and hourly inflow into the Koyna reservoir by using the data from access table and passing it through the developed RR models in this study. The user can enter the input data, forecast the inflow, compare, print or save the results in tabular as well as graphical form in various file formats. Fig. 2 presents the sample screen of the main module of the GUI. This GUI is specifically designed such that even beginners can handle the software.

All the procedures and tools required to prepare the input data, run the model, and visualize the results are included in the module. As the splash screen window is clicked, main window or the functional area of the software begins. It is a menu driven single document interface (SDI), i.e. user can work on one window at a time, thus making the software easy to handle, and the main window appears as shown in Fig. 3.

Figure 2: Splash Screen of KRIFS

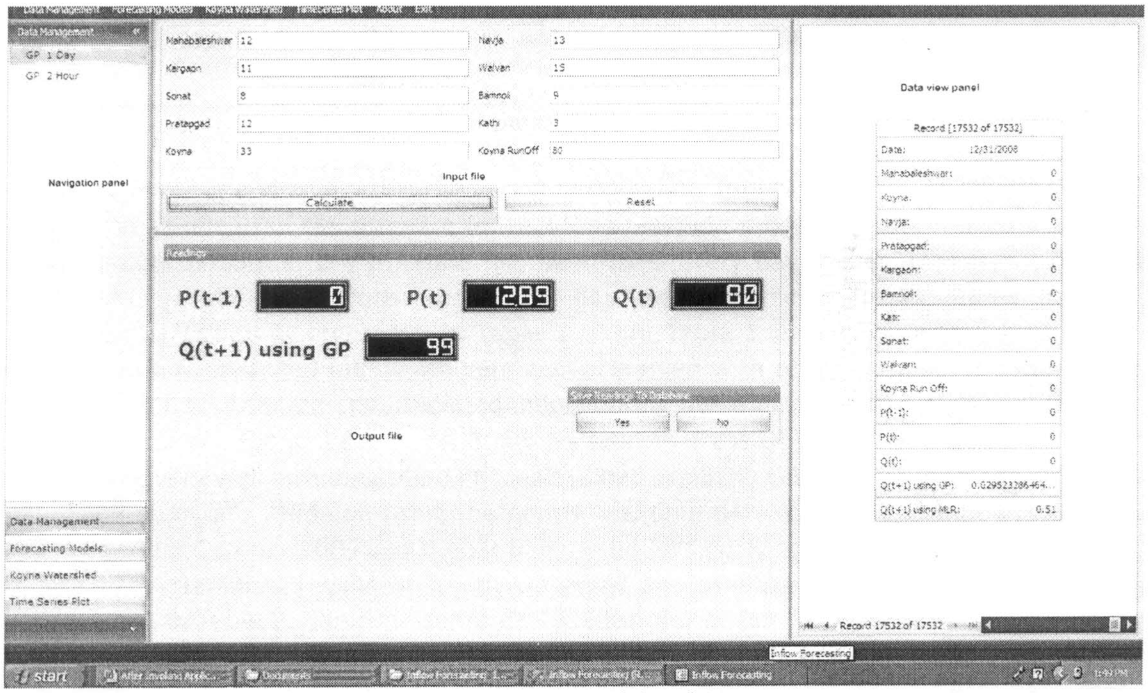

Figure 3: Main module of GUI

Above Main Window is divided into four regions called panels viz. navigation panel, Input Panel, Output Panel and Data-View Panel. It also has two splitters to separate panels from each other and to adjust their size by just moving them. The main data processing of the developed interface system includes input generation, time step selection. The time step of prediction can be selected either as daily or hourly basis and visualized graphically. Depending upon the input data the KRIFS forecast the output using LGP 1 day or 2 hour. Output is displayed at the centre of the layout in output panel (Fig. 4). It also cautions to “Save Readings to Database”.

After developing the GUI, a major step was implementing it in the field. For this purpose, series of discussions were held with the dam authorities in terms of presentation of the research work as well as demo of the software. Hands on experience training program on usage of the software have been given to the dam authorities. This training program helped them to learn about the software, giving the input, saving the input, forecasting for different time steps, saving and visualizing the output in the form of tables and figures. Once convinced, before implementation, the dam authorities them self validated the GUI using 2 years of data that were not used in model development and testing. In other words, the models proposed in this study were further validated by forecasting the real-time reservoir inflow using the unseen data from January 1, 2008 to 31 October, 2010. It is emphasized that this data was neither used in the training nor testing the developed models. For this purpose, the software has been installed at the Koyna dam Executive Engineers office and the field engineers were given training on using the software during February 2010. The field engineers entered the data (both in Table format and day to day entry) processed it and forecasted the inflow for first two years (Jan 1 2008 to Jan 2010).

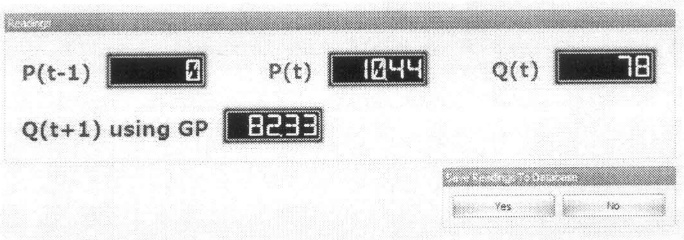

Figure 4: Output panel for GP 1 Day from main window

After visualizing and comparing the encouraging results (predicted inflow) with observed inflow, the field people used the software for real-time reservoir inflow prediction from 1 June 2010 to 31 October 2010 (i.e. the software has been used every day and then every hour also). The statistical results of the observed and 1 day ahead predicted inflow during the real-time forecasting period is presented in Table 1. From Table 1 it can be seen that the statistical properties of predicted inflow are very well preserved with observed one. The time-series and scatter plot of observed and predicted inflow during these three years is shown in Fig. 5 (a) and (b) respectively. From these plots, it can be seen that the observed and predicted inflows are matching very well including the daily peak values. Similarly the hourly models developed in the present study are validated using the hourly real-time data (rainfall and inflow) from 1st June 2008 to 31st October 2009 and then is being used for hourly real-time forecasting from 1st June 2010. The statistical results of hourly observed and predicted inflow for 2 hr lead period is presented in Table 2. The time-series and scatter plot of hourly observed and predicted inflow during validation period is shown in Fig. 6(a) and (b) respectively. From Table 2 and Fig. 6 it can be concluded that the observed and predicted inflows are closely matching with each other. The peak inflows are also perfectly matching with observed inflow values, concluding that that the 2 hour lead model is very much good.

Table 1: Statistical performance of daily observed and 1 day ahead predicted inflow during validation and real-time forecasting period (2008-2010)

Statistical properties	Observed inflow ($10^6 m^3$)	Predicted inflow ($10^6 m^3$)	% error
Mean	9.49	9.62	1.37
Std. Dev.	24.83	24.18	-2.62
Kurtosis	27.45	25.56	-6.88
Skewness	4.58	4.34	-5.24

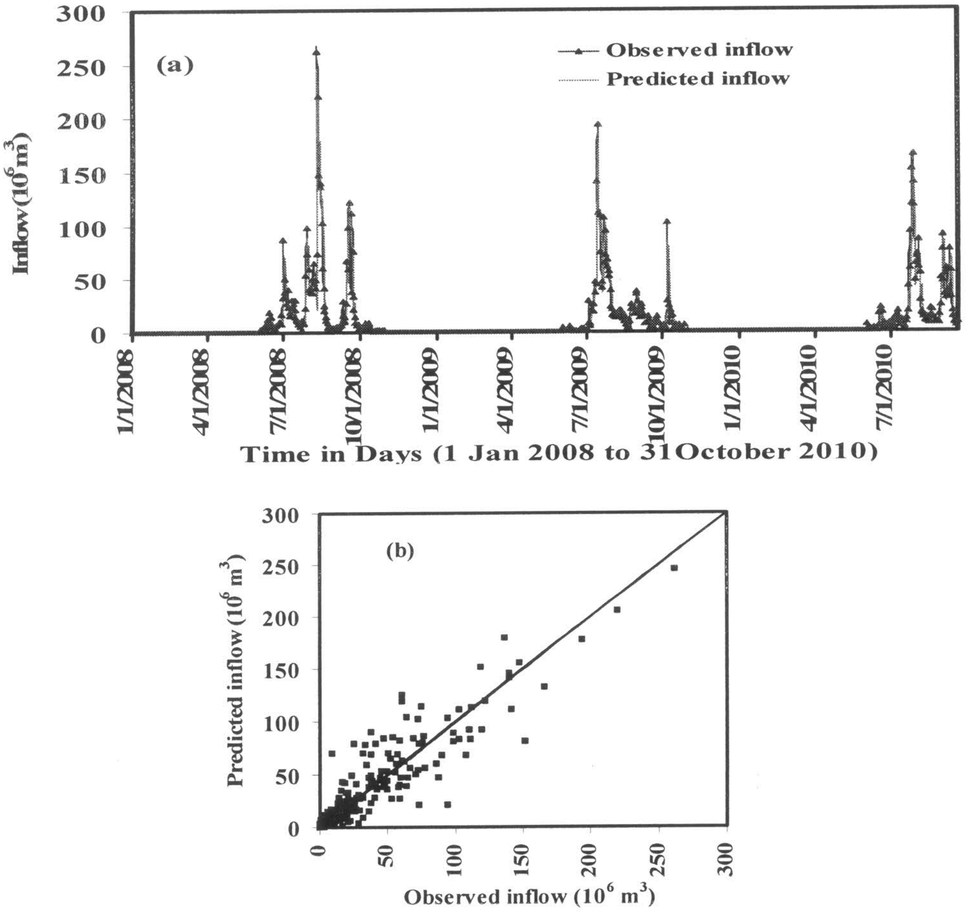

Figure 5: (a) Time-series and (b) scatter plot of observed and predicted inflow during validation and daily real-time forecasting period (2008-2010)

Table 2: Statistical performance of hourly observed and predicted inflow during validation and real-time forecasting period (2 hour ahead)

Statistical Properties	Observed Inflow (m^3/sec)	Predicted inflow (m^3/sec)	% Error
Mean	205.23	206.30	-0.52
Std Dev	354.39	354.24	0.04
Kurtosis	15.02	15.12	-0.66
Skewness	3.36	3.38	-0.59

Conclusion

This study investigated the applicability and capability of artificial intelligent models for inflow forecasting applied to Koyna watershed in Maharashtra, India. Artificial neural network and linear genetic programming techniques has been used to develop the model and their performances are evaluated. These models responded well to most of the fluctuations within the data and have resulted in better prediction of low, medium and peak inflows amongst all the models developed in this study. The LGP model has edge over the ANN models and hence the output of LGP model resulted in the form of C++ program has been used in development of GUI software for real-time reservoir inflow prediction. A drop menu based GUI named as KRIFS has been developed for Koyna watershed to predict hourly and daily real-time reservoir inflow. After installation and training to dam authorities, it has been validated by the dam authorities to predict daily and hourly reservoir inflow, plotting of the observed and predicted inflow for various periods. The field engineers found it very useful and have used it for real-time daily and hourly reservoir inflow forecast. The results of past one year predicted using the software by the dam authorities proves the robustness of the development LGP model as well as GUI software.

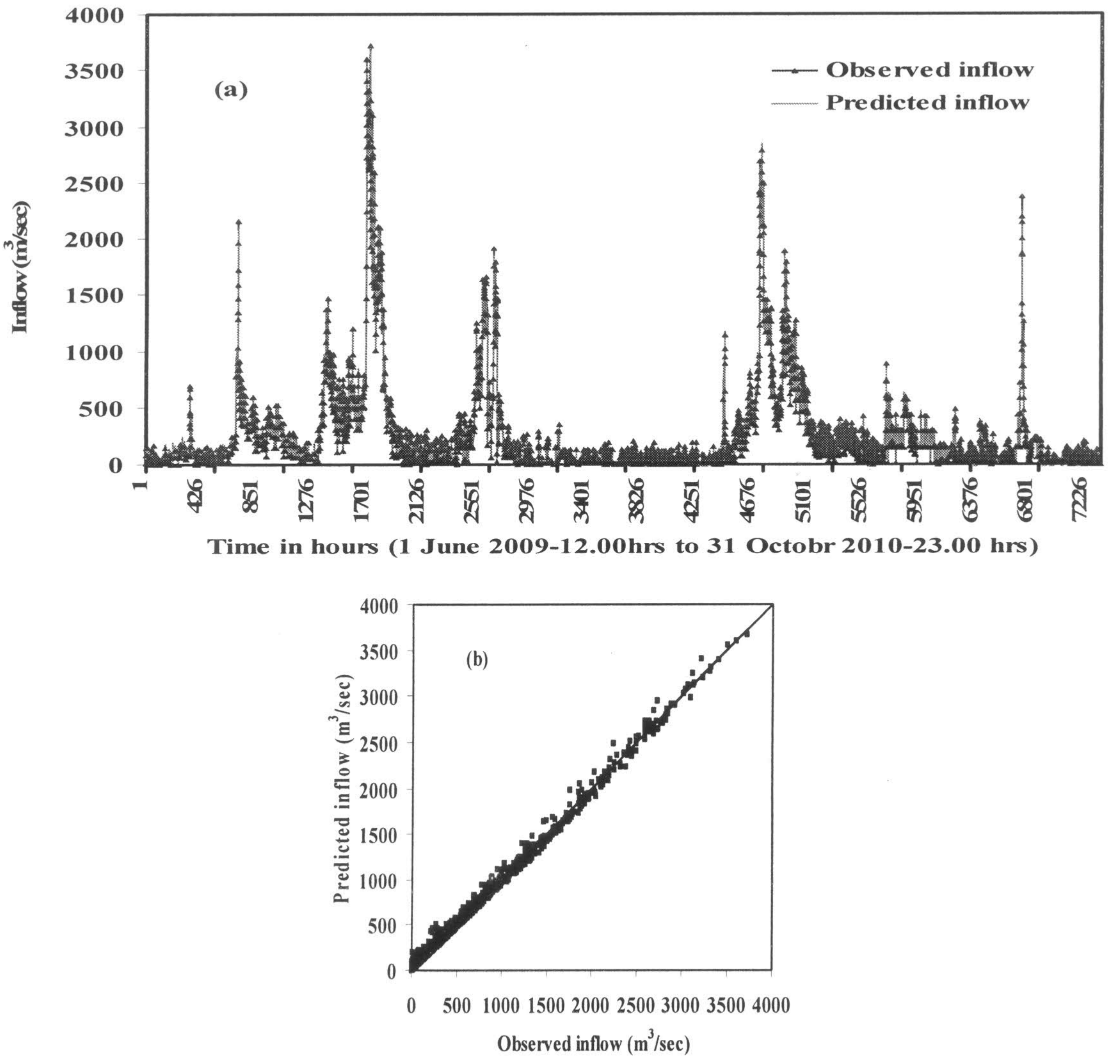

Figure 6: (a) Time series and (b) Scatter plot of observed and predicted inflow during validation and hourly real-time forecasting period (2009-2010)

Reference

[1]. Yeh, W.W-G. Reservoir Management and Operation Models: A state-of-the art review. *Water Resources Research*, Vol. 21/12 (1985), pp. 1797-1818.

[2]. Keskin, M. E., Taylan, D. and Ozlem, T. Adaptive neural –based fuzzy inference system (ANFIS) approaches for modeling hydrological time series. *Hydrological Sciences Journal*, Vol. 51/4 (2006), pp. 588-598.

[3]. Singh, V.P. Hydrologic Systems, Vol 1: Rainfall-Runoff Modeling. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1988.

[4]. Jothiprakash, V. and Magar R. Soft computing tools in rainfall-runoff modeling. *ISH Journal of Hydraulic Engineering*, Vol. 15/(SP-1) (2009), pp. 84-96.

[5]. Sorooshian, S., Duan, Q. and Gupta, V.K. Calibration of rainfall-runoff models: application of global optimization to the Sacramento Soil Moisture Accounting model. Water Resources Research, Vol. 29/4 (1993), pp. 1185-1194.

[6]. Smith , J. and Eli, R. N. Neural network models of the rainfall-runoff process. *Journal of Water Resources Planning and Management*, ASCE, Vol. 121 (1995), pp. 499-508.

[7]. Jain, A. and Indurthy, S.K.V.P. Comparative analysis of event based rainfall-runoff modeling techniques-deterministic, statistical, and artificial neural networks, *Journal of Hydrologic Engineering* ASCE, Vol. 8/2 (2003), pp. 93-98

[8]. Box and Jenkins. Time series analysis forecasting and control. 2nd edition. Colorado State University, Fort Collins, Co., Holden day, SanFrancisco, CA, 1976.

[9]. Thomas, H.A. and Fiering, M.B. Mathematical analysis of streamflow sequences in analysis of river basins by simulation In: A Maass et al. Design of water resources system. Cambridge, Harvard University Press, 1962, p.466.

[10]. Toth, E., Brath and Montanari, A. Comparison of short-term rainfall prediction models for real-time Flood forecasting. *Journal of Hydrology*, Vol. 239/(1-4) (2000), pp. 132 147.

[11]. De Vos, N.J. and Rientjes. Constraints of artificial neural networks for rainfall-runoff modelling: trade-offs in hydrological state representation and model evaluation. *Hydrology and Earth System Sciences*, Vol. 9 (2005), pp. 111-126.

[12]. Tokar, A.S. and Johnson, P.A. Rainfall-runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering*, ASCE, Vol. 4/3 (1999), pp. 232-239.

[13]. Thirumalaiah, K. and Deo, M. C. Hydrological forecasting using neural networks. Journal of Hydrological Engineering, ASCE, Vol. 5/2 (2000), pp. 180-189.

[14]. Sivakumar, B., Jayawardena, A. W. and Fernando, T. M. G. H. River flow forecasting: Use of phase-space reconstruction and artificial neural networks approaches. *Journal of Hydrology*, Vol. 265/(1-4) (2002), pp. 225-245.

[15]. Zhang, J. Cheng, C-T, Li, S., Wu, L-X-Y and Shen, J-J. Daily reservoir inflow forecasting combining QPF into ANNs model. *Hydrology Earth System Sciences Discussions*, Vol. 6 (2009), pp. 121-150.

[16]. Googhari, S.K., Feng, H.Y, Ghazali, A.H.B and Shui, L.T. Neural networks for forecasting Daily Reservoir Inflows. *Pertanika Journal Science and Technology*, Vol. 18/1 (2010), pp. 33-41.

[17].Openshaw, S. and Openshaw, C. Artificial Intelligence in Geography. John Willy and Sons Ltd, Chichester, 1997.

[18].Srinivasulu, S. and Jain, A. A comparative analysis of training methods for artificial neural network rainfall-runoff modeling. *Journal of Applied Soft Computing*, Vol. 6/3 (2006), pp. 295-306.

[19].Maier, H. and Dandy, G. Neural networks for the predictions and forecasting of water resources variables: Review of modeling issues and applications. *Environmental Modeling and Software*, Vol. 15(2000), pp. 101-124.

[20].Venugopal, K.R., Srinivasa, K.G. and Patnaik, L.M. Soft computing for Data Mining Applications. Springer, Illus., ISBN: 978-3-642-00192-5, 2009, 342p.

[21].Haykin, S. *Neural Network: A Comprehensive Foundation*. Prentice-Hall, Upper Saddle River, New Jersey, 1999.

[22].McCulloch, W.S. and Pitts, W. A logic calculus of the ideas immanent in nervous activity. *Bulletin of Mathematics and Biophysics*, Vol. 5 (1943), pp. 115-133.

[23].Govindaraju, R.S. and Rao, A.R. *Artificial Neural Networks in Hydrology*. Kluwer Academic, Dordrecht, The Netherlands, 2000.

[24].Werbos, P. Beyond regression: new tools for prediction and analysis in the behavioral sciences. PhD dissertation, Harvard University, Cambridge, Mass, 1974.

[25].Rumelhart, D.E., Hinton, G.E. and Williams, R.J. Learning internal representations by error backpropagation. Parallel Distributed Processing. In: Rumelhart, D.E., McClelland, J.L. (Eds.), MIT Press, Cambridge, 1986, 318-362

[26].ASCE (American Society of Civil Engineers) Task committee on application of the neural networks in hydrology. Artificial neural networks in hydrology I: Preliminary concepts. *Journal of Hydrologic Engineering* ASCE, Vol. 5/2 (2000a), pp. 115-123.

[27].ASCE (American Society of Civil Engineers) Task committee on application of the neural networks in hydrology. Artificial neural networks in hydrology II: hydrologic applications. *Journal of Hydrologic Engineering* ASCE, 5/2 (2000b), pp. 124-137.

[28].Hsu, K.L., Gupta, H.V. and Sorooshian, S. Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*, Vol. 31/10 (1995), pp. 2517-2530.

[29].Jang, J. S. R. ANFIS: adaptive network based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybenetics*, Vol. 23/3 (1993), pp. 665-683.

[30].Aqil, M., Kita, I., Yano, A. and Nishiyama, S. A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff. *Journal of Hydrology*, Vol. 337/(1-2) (2007), pp. 22-34.

[31].Jothiprakash, V., Magar, R. and Kalkutki, S. Rainfall runoff models using Adaptive Neuro Fuzzy Inference System (ANFIS) for an intermittent river. *International Journal of Artificial Intelligence*, Autumn, 3/A09 (2009), pp. 1-23.

[32].Koza, J. R. Genetic Programming: on the programming of computers by Natural Selection. MIT Press, Cambridge, MA. 1992.

[33].Whigham, P. A. and Crapper, P. F. Modelling rainfall-runoff using genetic programming. *Journal of Mathematical and Computer Modelling*, Vol. 33 (2001), pp. 707-721.

[34].Brameier, M. On Linear Genetic Programming. Ph.D. Thesis, University of Dortmund, 2004.

[35].Guven, A. Linear genetic programming for time-series modeling of daily flow rate. *Journal of Earth System Science*, Vol. 118/2 (2009), pp. 137-146.

[36].Brameier, M., Kantschik, W., Dittrich, P. and Banzhaf, W. SYSGP-AC++ library of different GP variants. Technical Report CI-98/48, Collaborative Research Centre, 531, University of Dortmund. (1998).

[37].Brameier, M. and Banzhaf, W. Evolving teams of predictors with linear genetic programming. *Genetic Programming and Evolvable Machines*, Vol. 2/4 (2001), pp. 381–407.

[38].Foster, J. A. Discipulus: A commercial genetic programming system. *Genetic Programming and Evolvable Machines*, Vol. 2 (2001), pp. 201-203.

[39].Francone, F. Discipulus Lite™ Owner's Manual, Version 4.0. Register Machine Learning Technologies, 2004.

[40].Oltean, M., Grosan, C. A Comparison of several linear genetic programming techniques. *Complex-Systems*, 14/4 (2003), pp. 285-313.

[41].Magar, R. B. *Real-time reservoir inflow prediction using soft computing techniques*. Ph.D Thesis Report, Indian Institute of Technology Bombay, Mumbai, 2011.

[42].CDO., *Final report on revised flood study for koyna dam*, Government of Maharashtra, Irrigation department, India, 1992.

[43].Jothiprakash, V., and R.B. Magar. Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent technique using lumped and distributed data. *Journal of Hydrology*, Vol. 450-451 (2012), pp. 293-307.