ADVANCESIN FLOOD FORECASTTECHNIQUES

Anup K Srivastava & Dr. R N Sankhua
Directors, National Water Academy, CWC, Pune

ABSTRACT

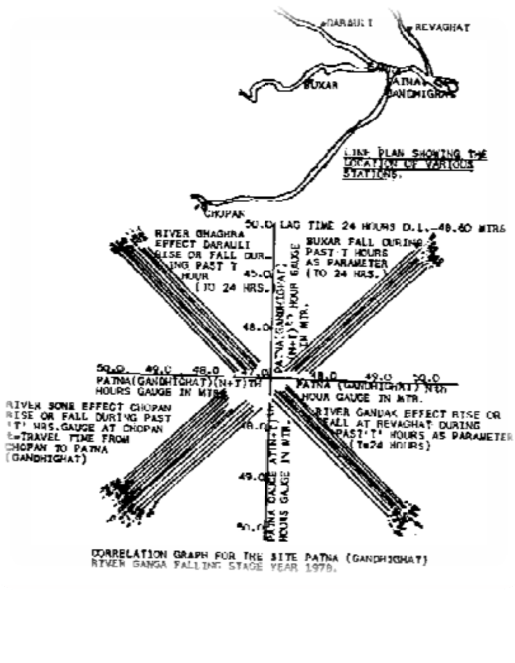
It is common in India to witnesslarge-scale submergence particularly during wet season which spans between June and mid-October each year. During this part of the year, protracted rainfall activities cause almost every river to swell and overflow its bank. A breach in bund and/or a river over-spilling its bank frequently flood nearby plains. In a few scattered cases, flooding happens because of cloud burst over an area or by spurt of water from a natural obstruction in the river stream formed by landslide. Even if, India has constructed hundreds of kms of embankment besides other form of structural measures; and operates a well-organized flood forecasting network across the country, losses on account of flood do occur every year. There is also increased perception that climate change may induce an increased probability of extreme precipitation and, hence, of flooding. This, therefore, calls for major emphasis on the improvement of operational flood forecasting systems.

This article discusses the modelling of rainfall-run-off and flow-routing processes in river systems within the context of real-time flood forecasting. It is important to emphasize that the hydrological models considered in this note discusses lumped-parameter variety (they consist of the linear and nonlinear transfer functions that are the discrete-time equivalents of differential equations and describe the temporal behaviour only at selected spatial nodes within the catchment system). The alternative `distributed-parameter' models, which involve spatiotemporal aspects of the catchment and are described by models such as partial differential equations in time and space (or some equivalent of these), are not considered.

FLOOD FORECAST TECHNIQUES

In order to contain losses and defend people against flood to the extent possible, India puts stress on blended approach of structural and non- structural measures. Among several non-structural measures, flood forecast is widely in use in India. Succeeding paragraphs seek to illustrate a set of traditional as well as a few software based techniques which help forecaster issue advance warning of the magnitude of flood and its time of occurrence. According to the methods commonly employed in arriving at forecast values, these methods can be classified into five categories.

a. **Methods based on correlation/coaxial diagrams between two variables or even more;**
b. **Mathematical equations developed using regression/multiple linear regression techniques which combines independent variable with one or more than one variable;**
c. **Hydrological models**
    - ✓ **Rainfall run-off model**
        - **Lumped models**
        - **Quasi-distributed**
        - **Distributed models**
    - ✓ **Routing techniques**
      - Lumped routing
      -distributed
d. **Hydraulic models**
    - **Distributed hydraulic routing**
e. **Data driven hydrological models**

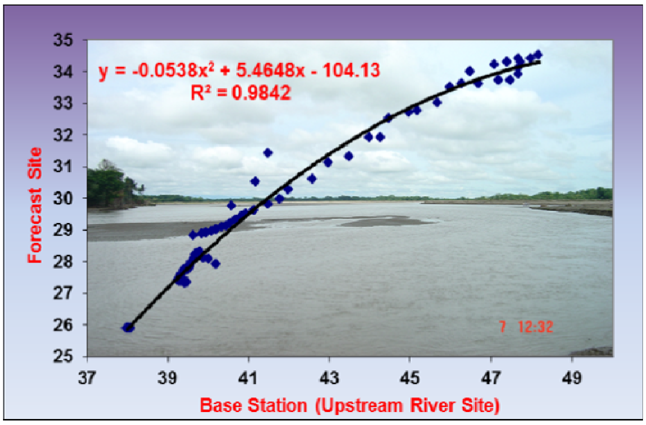


a. CORRELATION/CO-AXIAL DIAGRAMS

Forecasters in India have developed a large set of correlation diagrams which display the pattern of correlation exhibited by two or more variables. Such charts are relatively less complex; and are quite popular among its users. This apart, they need periodical updating to account for constant alteration in catchment characteristics and river regime. However, these charts provide only peak flow or water level information, given a set of input data; and thus, fail to plot entire flood hydrograph.One of such diagrams used in India is shown next.With advances in numerical modeling and use of software, reliance on these techniques is now fast diminishing.

b. MATHEMATICAL EQUATIONS

Unlike previous method, this method defines relationship among variables mathematically by regression/multiple regression techniques' and thus rid the forecaster of reading a chart manually each time to formulate forecast. The strength of such pattern is determined by correlation coefficient, 'r'. Mathematical equations offer much ease in calculation of dependent variable, and in turn speed up the process. Adjacent chart along with an equation estimates y, water level at downstream location, Mahemdabad in Gujarat with change in x, water level recorded at upstream site. A respectable degree of R^2 as 0.9854 is achieved by introducing a time lag/shift of 4 hrs between two sets of data. The arrival of this time lag is based on output obtained through cross-correlogram technique.Another relationship derived by multiple regression technique determines the change in water level at forecast site bases on the variations recorded at two upstream sites, commonly known as base stations. While preceding equation relates water levels of two sites, the following equation correlates variation in water levels at different sites.



$$G_{(t+1).t} = 0.87G_{t.(t-1)} + 1.78\ H_{(t+1).t} - 3.38\ H_{t.(t-1)} - 1.12\ H'_{(t+1).t} + 2.55\ H'_{t.(t-1)}$$

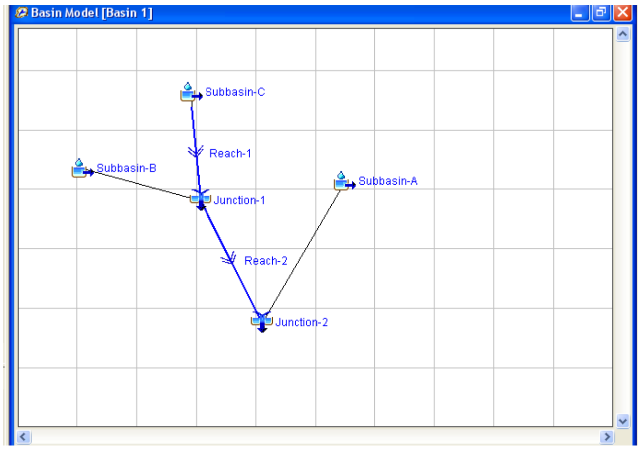
Where, $G_{(t+1).t}$means change in water level at forecast station between time (t+1) and t. H and H' are water levels at two base stations, and their subscript indicates variation in water level between two time interval. Mathematical equations involving contribution of rainfall in upland area, API (antecedent moisture index) etc. are also used in India.

c. HYDROLOGICAL MODELS

- ***Lumped & Quasi-distributed***

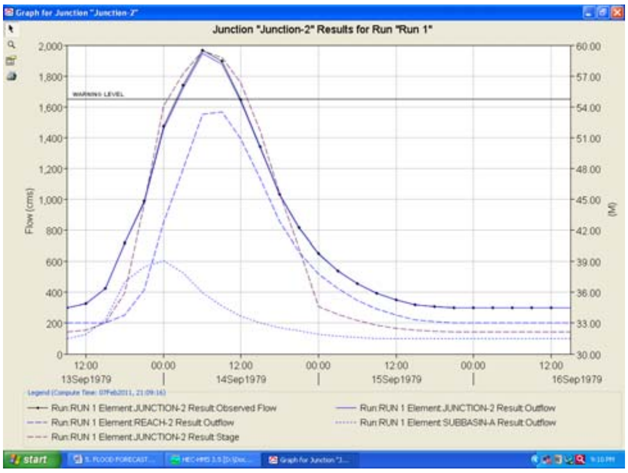
In preceding methods, water level or discharge observation at one or more than one site forms the basis for forecast formulation except a few cases where rainfall is considered as an additional input. This approach works well where sites conditions present enough lead time, as in case of large rivers like the Ganges or the Brahmaputra. However, where catchment is small or river is flashy, forecaster and relief team are left with little time to prepare for and respond to approaching flood waves. Such scenario demands inclusion of rainfall data as one of the essential input followed by convolution of Unit Hydrograph of concerned basin/catchment to obtain flood hydrograph at terminus/outlet or at a location of forecaster's choice.

A HEC-HMS set-up of three basins with two reaches shown here demonstrates application of UH in conjunction with MUSKINGUM routing method to estimate magnitude of flood and time of its occurrence using computational ease offered by HEC-HMS software. Here, three basins receive rainfall as input, and shed water at respective basin outlet. Two reaches route this water upto junction 2 which represents forecast station. Muskingum method needs two fitted parameters, such as K & X, and number of steps to output outflow hydrograph. Using a rating curve, HEC-HMS can also plot a variation in water



level – an information often looked for by forecaster (as seen in the figure right). Distributed hydrologic routing methods, such as Muskingum-Cunge, Kinematic-wave can be used in HEC-HMS provided the availability of few river x-sections in between base and forecasting stations.

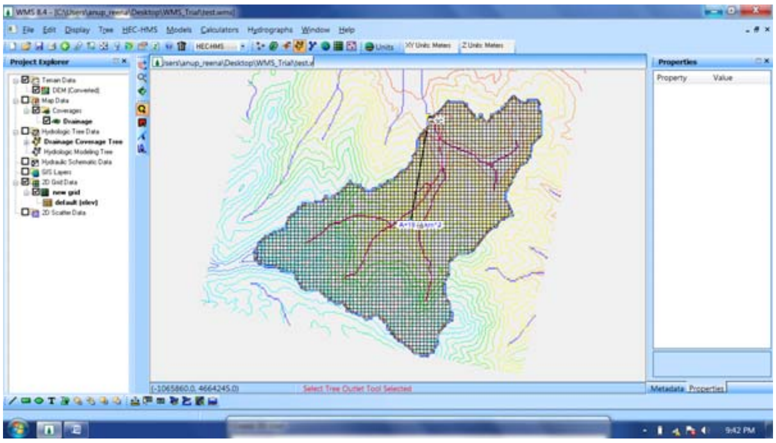
While setting up this model, catchment and reach details are distilled from GIS based HEC-GeoHMS add-on package, and were imported in HEC-HMS. Other details, such as Muskingum parameters were estimated based on historical data. With proper selection of a series of models to account for effective rainfall, rainfall-runoff model, base flow contribution, and routing along the river reaches, HEC-HMS generates a flood hydrograph and corresponding water level resulting from a burst of rainfall over the catchment. This basin model is an example of lumped and single event model and does not account for fluctuation in soil moisture between two rainfall events. To account for rainless period between two spells of rain, soil moisture accounting model can be used instead.



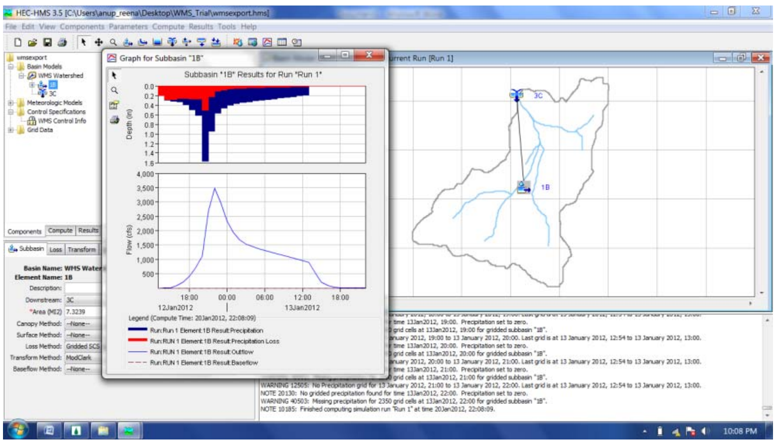
A flood hydrograph, as seen here, not only shows the peak and its time of occurrence but also indicates the period when water will flow above warning level- a period of heightened risk in the forecast area nearby.

- ***Distributed Hydrological Model***

Forecast estimated by applying hydrological models such as one presented above tends to vary widely from real values where assumptions in unit hydrograph or routing models are greatly violated by prevailing hydro-meteorological conditions over catchment. For example, rainfall is non-uniform over the basin; it is not stationary and moving across the basin; rainfall is concentrated in one pocket and leaving holes elsewhere. Apart from this, soil type and land-use pattern also vary over the catchment/basin that govern the rising and falling limb of resulting hydrograph. In lumped model, these characteristics are represented by a single SCS CN(curve number). Such scenario demands application of distributed model to accurately capture the basin response.



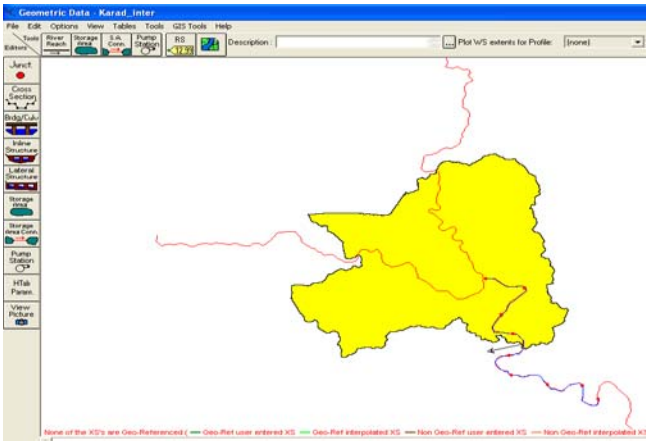
Here, we present a distributed model developed and analyzed using Water Modeling Software (WMS) and HEC-HMS. In this model, WMS software first creates grid; and thereafter determines CN value of each according to its soil type and land-use cover.Additionally, this software also prepares a gridded precipitation database. A set of these gridded information aresubsequently exported to HEC-HMS for simulation run. HEC-HMS calculates the rainfall excess and route it to outlet by MODCLARK method. A couple of screenshots shown here display model set-up and results obtained at the end.



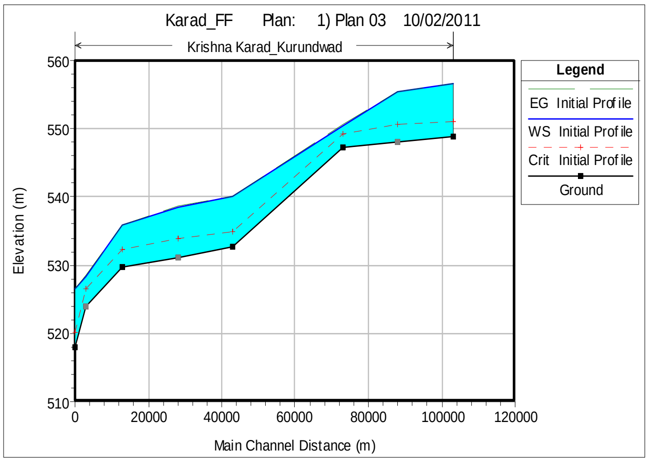
Forecasters in India are quite familiar with rest of methods included in this paper. And, depending upon data and resources availability, they elect appropriate method to reach forecast value.

d. HYDRAULIC ROUTING TECHNIQUE

Hydraulic routing technique employs the full dynamic wave (St. Tenant) equations, i.e. the continuity and the momentum equation which takes the place of the storage-discharge relationship used in hydrologic routing. In order to solve these equations to deduce water level and corresponding flow at several locations, tool such as HEC-RAS is used in the study shown here.



The view shows a schematic of Krishna & its tributary Koyna river with the study area. The red tick marks and corresponding numbers denote the location of individual cross-sections. This aside, software also require a set of other information, such as variation in roughness coefficient - across and along the river reach, inflow hydrograph, boundary conditions, data regarding in-line and/or off-line hydraulic structures. A successful run of the model, HEC-RAS generates a number of useful outputs that can be picked up by forecasters to design their flood bulletin. A plot here displays a water surface profile along a 10km long reach on a particular date and time. Results in this real case study came within a range of ± 10 cumecs of observed discharge.



Unlike hydrological routing techniques which work either on fitted parameters and/or on certain assumptions, hydraulics routing process solely depends on the physical properties of the reach, its geometry and is therefore more accurate and robust almost in all river conditions including when river flow is unsteady or is influenced by backwater effect. For similar kind of analysis, MIKE 11 software is also used at few forecast centres in India.

e. DATA DRIVEN HYDROLOGICAL MODELS

Sometimes, it is argued that deterministic, reductionist models are inappropriate for real-time forecasting because of the inherent uncertainty that characterizes river catchment dynamics and the problems of model over parameterization. The advantages of alternative, efficiently parameterized data-based mechanistic models, identified and estimated using statistical methods, are discussed.

Neuromorphicmodelling techniques are now well established methods for describing physical processes occurring in the aquatic environment. The development in information technology over the last decade has presented opportunities of extended computational ability together with improved data manipulation, storage and retrieval. As a result, the Neuromorphic models are now being used more extensively in the management, design and operation of water based assets. The reason behind this is that in many areas of applications pertaining to the complex flow systems, the demands on computing time are of such a magnitude, which is far from acceptable. An elementary brief part of ANN has been added here in the distance learning course for easy insight, though there are advanced architectures like recurrent neural network (RNN), radial basis function (RBF), self-organizing map (SOM) and others used in flood forecasting.

i) Artificial Neural Networks

An artificial neural network is nothing but a collection of interconnected processing elements (PEs). The connection strengths, also called the network weights, can be adapted such that the network's output matches a desired response.

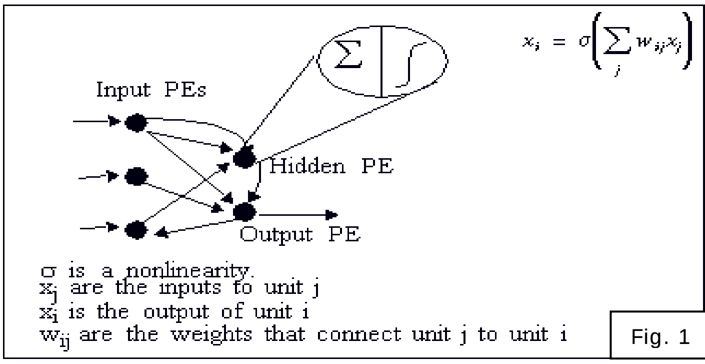


Fig. 1

Typical architecture of single hidden layer neural network

Fig. 1 depicts a typical multilayer perceptron, which has been used in this research, which resembles a black box model, where a set of a data like x_1, x_2, $x_3 \ldots x_n$ are fed directly to the network through the input layer, and subsequently produces expected result y in the output layer. The output is determined by the architecture of the network.

In multi-layered perceptron, hidden layer means a third layer of processing elements or units in between the input and output layers that increases computational power. In principle, the hidden layer can be more than one layer. In practice the number of neurons in this layer is evaluated by trial and error. Hornik et al. (1989) proved that a single hidden layer containing a sufficient number of neurons can be used to approximate any measurable functional relationship between the input data and the output variable to any desired accuracy. In addition, De Villars and Barnard (1993) showed that an ANN comprising of two hidden layers tends to be less accurate than its single hidden layer counterpart. In this, a single hidden layer ANN has been used.

Each input x_i (i =1,…,n) is attenuated by a factor w_{ij}, more commonly called a weight of the network, which is associated with the connection linking input x_i to hidden neuron j (j = 1,….,k), where, k is the number of neurons in the single hidden layer. The weighted sum of the incoming signals entering a neuron is fed via an activation function, which is non–linear, producing a value that in turn, act as an input signal sent to the output layer. This is repeated for the output weights. The following expression gives the output value of the network.

$$y = \psi\left[\sum_{j=1i}^{N} \varphi\left\{\sum_{i=1}^{N} x_i w_{ij}\right\} a_j\right] \qquad (1)$$

Where, the sigmoidal activation function ψ is given by

$$\psi(x) = \frac{1}{1+e^{-\alpha x}} \qquad (2)$$

This function given at eq. 2 is a continuous function that varies gradually between asymptotic values 0 and 1 or –1 and +1. Where, α is the slope parameter, which adjusts the abruptness of the function as it changes between the two asymptotic values. Sigmoid functions are differentiable, which is an important feature of neural network theory.

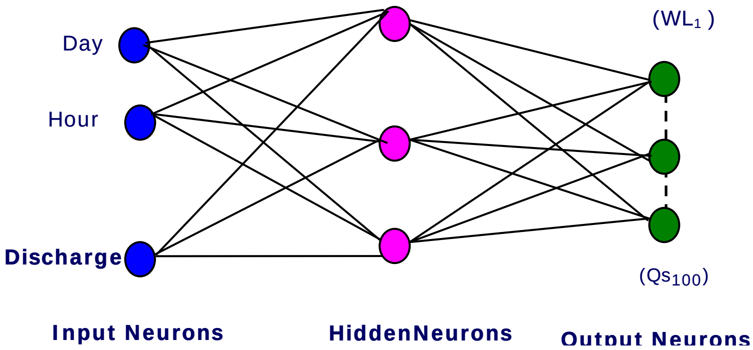
To obtain the best approximations, it is needed to determine the optimum set of weights w_{ij} and $a_j$that will yield the least mean square value of the desired response. Thus the following performance criterion needs to be satisfied.

$$\min_{a_j \ w_{ij}} \frac{1}{2} E\left[(y - y_D)^2\right] \qquad (3)$$

Where, E is the statistical expectation operator, the factor ½ is included for convenience of presentation, y and y_D are observed and desired parameters. A complete description of the back propagation can be found in Hagan et al. (1996).

The configuration chosen for the ANN models are shown in Fig. 2, where the bias inputs have the effect of lowering or increasing the net result of the activation function. The activation used is the sigmoidal function, which has the purpose of limiting the permissible amplitude range of the output values to some finite value.

Fig. 2



Normalization of the data

It is mentioned that the sigmoidal function can take the values ranging in the (0, 1) domain, a normalisation of the values of the input variables are done. The standard procedure in neural network theory gives the normalisation equation used for this purpose.

$$SF = (SR_{max} - SR_{min}) / (X_{max} - X_{min}) \quad (4)$$

$$Xp = SR_{min} + (X - X_{min}) * SF \quad (5)$$

Where: X - actual value of a numeric column, X_{min} - minimum actual value of the column, X_{max} - maximum actual value of the column, SR_{min} - lower scaling range limit , SR_{max} - upper scaling range limit and SF - scaling factor, Xp–pre-processed value.

Interpretation of results

Numerous goodness of fit statistical criteria are proposed in the literature for evaluating hydrological modelling results. Here, only two of these are considered in this study, namely RMSE, and Nash - Sutcliffe coefficient (1970). RMSE can take any positive value but the closer it is to zero, the better the model performs. When the Nash value is between 0 and 1, the forecast model does better than simply forecasting using y_o. The closer the Nash index is to one, the better. These performance criteria are used as basis of comparison to select the best model.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i}^{N}(Y_o - Y_P)^2} \quad (6)$$

Nash-Sutcliffe coefficient is defined by

$$\text{Nash} = 1 - \frac{\sum_{i}^{N}(Y_o - Y_p)^2}{\sum_{i}^{N}(Y_o - \overline{Y}_o)^2} \quad (7)$$

Where, Y_o= Observed daily gauge of the catchment on day i;
Y_p= Predicted daily gauge of the catchment on day i;

As could be seen from Fig 3, the model proved their capability in predicting the data, especially the stage data, which shows a high correlation of the observed and predicted data.

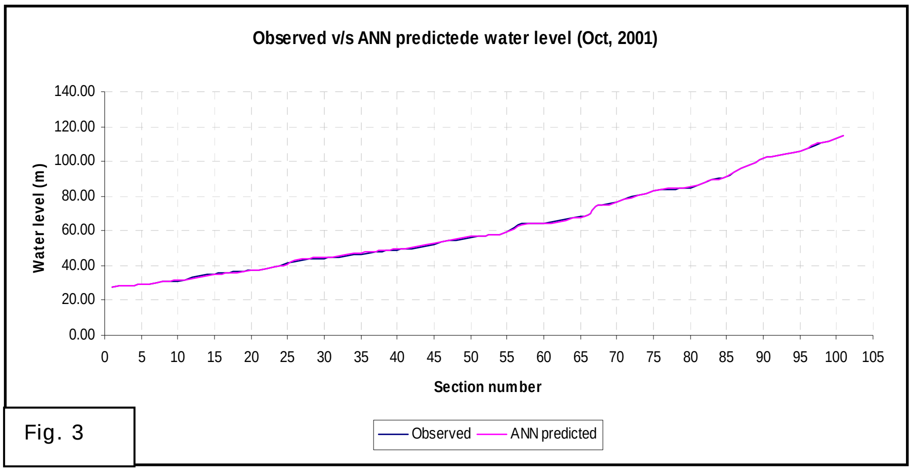


Fig. 3

ii) Fuzzy expert system design for flood forecasting

Linguistic terms are chosen to describe the input variable stage and the results. Further refinement of the models could not be achieved by adding extra membership functions. Gaussian membership functions (the function is generally suited for Indian rivers) can be used. Applying a similar method of data classification, membership functions are determined for the output variable discharge.

Rule definition

Some years of average hourly stage data and expert knowledge are used to create a rule base for the fuzzy logic model. Rules are defined for both the high and low extreme conditions, with regard to actual occurrences, because of the physical nature of the relationships. Depending on number of membership functions for each input variable; the minimum rule base is created. For each data point, all rules are evaluated.

Fuzzy model construction

The platform selected for the fuzzy logic expert system is MATLAB and MATLAB'S Fuzzy Logic Toolbox. The variables are combined into rules using the concept of 'AND'. The fuzzy operator 'minimum' is applied as the 'AND' function to combine the variables. No weightings are applied, which means no rule is emphasized as more important in respect to estimating the discharge. Implication is performed with the minimum function, and aggregation is performed with the maximum function. The centre of gravity method is applied as a means of defuzzification of the output membership functions to determine a crisp set. Based on this structure a baseline model fuzzy logic expert system for stage-discharge relationship is constructed for the G&D stations. Alternate functions for the expert system are investigated through sensitivity analysis.

Sensitivity analysis

A sensitivity analysis is performed for the fuzzy logic operator AND, and for methods of implication, aggregation and defuzzification. The results of changing a single operator or method while the rest of the model is held constant are compared with the results from the baseline model. The results are evaluated on the basis of correct linguistic matches. Based on this sensitivity analysis, the AND operator 'minimum' and the implication method 'minimum' are found to perform better than the product method. The fuzzy logic and ANN models are evaluated based on their ability to predict the discharge.

iii) ANFIS (Adaptive Neuro-Fuzzy Inference System) models

The hybrid system of learning has been attempted at combining ANN and fuzzy logic for developing the stage-discharge relationship to achieve a faster rate of convergence by controlling the learning rate parameter with fuzzy rules. The objective is to get a minimizer, which has a low computing cost and a large convergence domain. This learning ability is achieved by presenting a training set of different examples to the network and using learning algorithm, which changes the weights in such a way that the network reproduces a correct output with the correct input values. The main dissimilarity between fuzzy logic system (FLS) and neural network is that FLS uses heuristic knowledge to form rules and tunes these rules using sample data, whereas NN forms "rules" based entirely on data. Sugeno type ANFIS can be used. Gaussian membership functions can be used with rule bases, because of their low computational requirements and as it has the important properties of smooth mapping, universal approximation, normal distributions can be approximated well by this type of functions. Learning rate control by fuzzy logic has been depicted at fig 4. (FLC - fuzzy logic controller, MLP - multilayer perceptron)

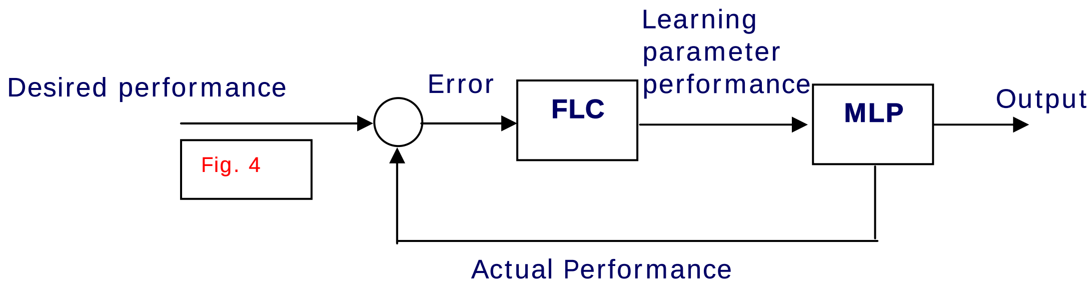


Fig. 4

iv) ***Validation and comparison of results***

The ANN, fuzzy and neuro-fuzzy models thus developed is validated and compared with the observed data points and the statistical measures of goodness-of-fit of the neuromorphic models. Numerous goodness of fit statistical criteria are proposed in the literature for evaluating hydrological modelling results. Goodness of fit can be tested from standard statistics literature as has been shown in the aforesaid ANN paragraph.Fig.5 shows the validation and comparison of models with observed data.

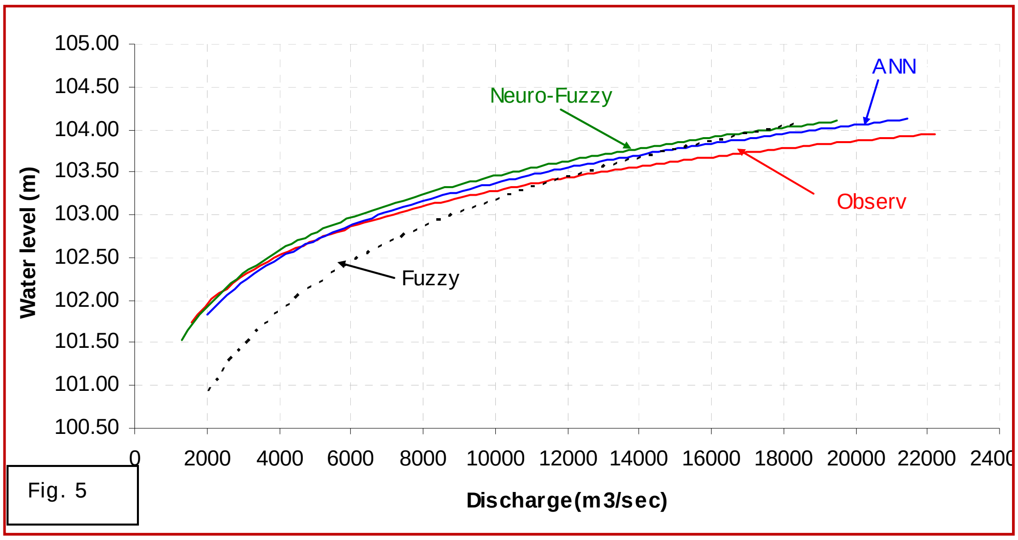


Fig. 5

CONCLUSIONS

As could be seen in preceding paragraphs, advance warning about the incoming flood peak and its probable time of occurrence can be achieved by several models. However, selection of a particular method or model, and its accuracy for a given site is largely governed by three factors - data availability; forecaster's knowledge of, and his experience with the basin; and forecaster's familiarity with software to be used in the forecast process.

The illustrated texts mentioned in this module are just the trail of a beginning, and for more of the subject and in-depth precision knowledge base, the readers are suggested to refer to advanced literature layouts.

REFERENCES

1. Flood forecasting Manual, (1989), Central Water Commission, New Delhi
2. Raghunath, H M., (2006), Book on Hydrology principles analysis and design, New Age International
3. Technical references Manual, (2004), HEC-HMS, USACE
4. Technical references Manual, (2004), HEC-RAS, USACE

5. Wheater, H. S., Jakeman, A. J. & Beven. K. J. (1993) Progress and directions in rainfall-run-of modelling. In *Modelling change in environmental systems* (ed. A. J. Jakeman, M. B. Beck &M. J. McAleer), pp. 101-132, Wiley.