RATING CURVE CALIBRATION USING GRID PARTITIONING TECHNIQUE

Rama Mehta and P.K. Bhunia
National Institute of Hydrology, Roorkee
Kumar Garvit and V.K. Dwivedi
NIT, Durgapur

ABSTRACT

At many locations, the discharge is not a unique function of stage; variables such as surface slope or rate of change of stage with respect to time must also be known to obtain the complete relationship in such circumstances. Therefore there is a need to establish the rating curve using soft computing techniques to transform the observed stages into the corresponding discharges. The model has been developed and validated with the data of Champua station in Orrisa.
During the study, models are developed for monsoon period using Grid Partitioning method (GPM). 60% data are used to train/ calibrate the model while rest 40% data to check/validate the model. Results have been compared with empirical method (standard stage –discharge equation). Grid partitioning method is also giving good comparative results with conventional method and observed output.

1.0 INTRODUCTION

The models have been developed using ANFIS technique for Champua guage station in Baitarni basin of Orissa. Several methods have been proposed to improve data fitting, but generally they have not adequately assessed the fundamentals of stage-discharge ratings based on fluid mechanics. As a consequence, several difficulties with stage-discharge ratings have been recognized. The present paper describe the new emerging techniques as soft computing techniques. Soft computing techniques are widely used in the field of hydrology since last three decades.
The term - **soft computing** is introduced by Prof. Zedeh(1973), it is a collection of some biologically-inspired methodologies, such as Neural Network (NN), Fuzzy Logic (FL), Genetic Algorithm (GA) and their different combined forms, in which accuracy is traded for tractability, robustness, ease of implementation and a low cost solution.

Deka and Chandramouli (2003) tested several modern approach to the problem: a neural network model, a modularized neural network model, a conventional curve-fitting approach and a fuzzy neural network model are compared using a case study. Rama et al. (2005) have studied the effect of Membership functions with different number of categories for reservoir operation. Hundecha et al. (2001) developed fuzzy rule based models to simulate the different physical processes involved in the generation of discharge from rainfall, and incorporated them within a conceptual model. Xiong et al. (2001) used a Takagi-Sugeno (TS) fuzzy model combined with a fuzzy c-means clustering algorithm for flood forecasting. Xiong and O'Conner (2002) developed a fuzzy autoregressive threshold updating model for real time river flow forecasting. Nayak et al. (2004, 2005a) employed a fuzzy model based on a back-propagation algorithm for river flow forecasting. Rama et.al.(2009) have developed the optimal operation rules for multi-purpose reservoir using neuro-fuzzy technique.

STUDY AREA

Champua is a census town in Kendujhar district in the state of Orissa , India. Champua is located at 22.08^0 N 85.67^0 E. It has an average elevation of 346 metres (1135 feet). There is only one rain gauge station (at Champua) established in the far upper part of the basin. Due to inadequate rainfall data, the average annual rainfall value for the basin has been estimated as 1500mm from Irrigation Atlas (Plate no.22; Rainfall in India). There are two gauge & discharge stations maintained by Central Water Commission as Champua (1990-onwards) and Anandpur (1971-2003) established on the river Baitarani.

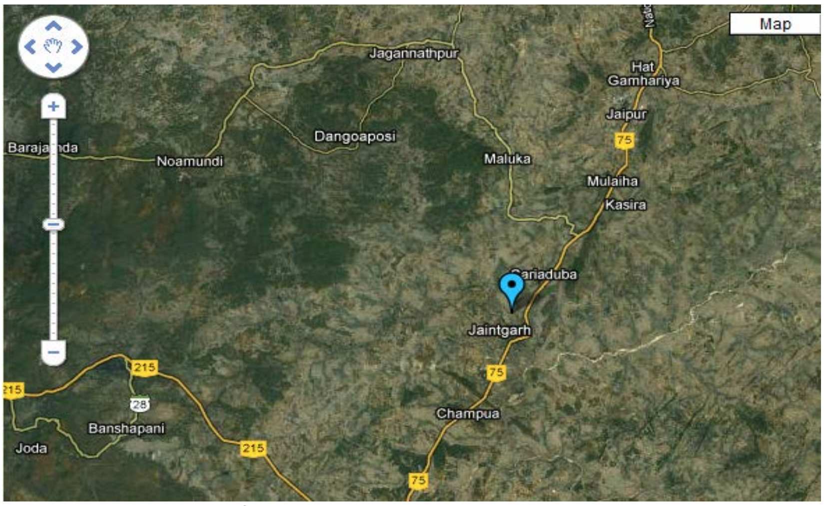


Fig. 1 Location map of Champua

STAGE-DISCHARGE RELATIONSHIP

Discharge rating curves may be simple or complex depending on the river reach and flow regime. These relations are typically developed empirically from periodic measurements of stage and discharge. These data are plotted versus the concurrent stage to define the rating curve for the stream.
For new gauging stations, many discharge measurements are needed to develop the stage discharge relation throughout the entire range of stream flow data. Generally, periodic measurements are needed to validate the underlying stage-discharge relationship and to track changes or shifts in the rating curve. The ISO regulation 1100-2 (ISO, 1998) recommends at least 12-15 discharge measurements during the period of analysis.

CONVENTIONAL METHOD

The conventional method is frequently used to get the relationship between the stage, h_t *(m)* and the corresponding discharge, Q_t *(m^3/sec.)*:

$$Q_t = a(H)^b \quad \text{where} \quad H = (h_t - h_0) \qquad (1)$$

where a and b represent the parameters of the relationship. h_0 is the river stage (m) at which the discharge is nil.

In reality, the stage-discharge relationship may be affected by a number of inherent uncertainties. Both the measurement and the natural uncertainties may lead to a significant scatter in the relationship, and the use of a single-valued curve can cause underestimation and/ or overestimation of discharges. The uncertainties might be substantial if only a few discharge measurements are available and a single-valued curve is used for interpolation and extrapolation of discharges. The uncertainties due to incorrect discharges can cause potentially large errors, influencing flood forecasting, statistical estimation of flood flows for design and decisions to promote flood defence schemes. Therefore soft computing techniques (ANFIS) are preferred to use establish the rating curves.

SOFT COMPUTING TECHNIQUES

ANFIS model allows separating the identification problem into two sub- problem: (i) parameter identification of the consequent part on grid base and; (ii) appropriate partitioning of the state space of interest by clustering. During this study we used only Grid Partitioning Method (GPM).

GRID BASE PARTITIONING

The premise parameters - which describe the shape of the MFs, and the consequent parameters - which describe the overall output of the system are the optimization parameters in ANFIS. The basic learning rule of an adaptive network (backpropagation algorithm (Rumelhart et al., 1986)) based on the gradient descent rule, can be successfully applied to estimate these parameters but Hybrid (combination of the gradient descent method and the least squares estimates (LSE)) is a fast learning algorithm.

An adaptive network is a multi-layered feed forward structure whose overall output behaviour is determined by the value of a collection of modifiable parameters. Consider that the FIS has two inputs x and y and one output z. For the fist order Sugeno fuzzy model, a typical rule set with two fuzzy if-then rules can be expressed as:

Rule1: IF x is A_1 and y is $B_{1,}$ THEN $f_1 = p_1x + q_1y + r_1$ (3)

Rule2: IF x is A_2 and y is $B_{2,}$ THEN $f_2 = p_2x + q_2y + r_2$ (4)

Where A_1, A_2 and B_1, B_2 are membership categories of input parameters x and y respectively. p_1, q_1, r_1 and p_2, q_2, r_2 are constant values.

DATA ANALYSIS

A) RAINFALL

The Baitarani basin receives most of the rainfall from the south-west monsoon during the period from June to October. There is only one rain gauge station (at Champua) established in the far upper part of the basin. Due to inadequate rainfall data, the average annual rainfall value for the basin has been estimated as 1500mm from Irrigation Atlas (Plate no.22; Rainfall in India).

b) Gauge & Discharge Data

There is no guage & discharge site available at the project site. However there are two gauge & discharge stations established on the river Baitarani. The details of availability of data is as follows.

Table - 1: G & D Sites

Sl.No.	Name of site	Maintained by	Years of availability	C.A. km2
1.	Anandpur	CWC	1971-2003	8496 km2
2.	Champua	CWC	1990 onwards	1710 km2

The data collected from Champua station has been used to establish rating curve using conventional method and soft computing techniques as Artificial Neuro Fuzzy Inference Systems (ANFIS). Daily Stage-Discharge data for monsoon period (i.e. July, August and September) from 1990 to 2008 have been used to develop the models. Total four groups of data have been framed from 18 years data as Table 2:

Table 2: Distribution of data length for model analysis.

Group	Length of data
I	1990-1995
II	1990-2000
III	1990-2005
IV	1990-2008

Data has been averaged for all monsoon months considered during the study for each year. Again the monthly averaging have been done over the years to get the input data for training and checking model development.

FUZZIFICATION OF DATA

For fuzzification of input data, the gray values (minimum and maximum values of the premises) have been divided into different classes. The stage parameter has been considered as input parameter while the discharge has been considered as output of the model. Gaussian overlapped membership functions have been used for input parameters. Although membership functions can have any number of categories, but three or four categories per variable seems to be adequate. All the crisp values of premises have their membership values within their specified categories to develop the fuzzified values. Now these fuzzified values developed a database for the fuzzy rules.

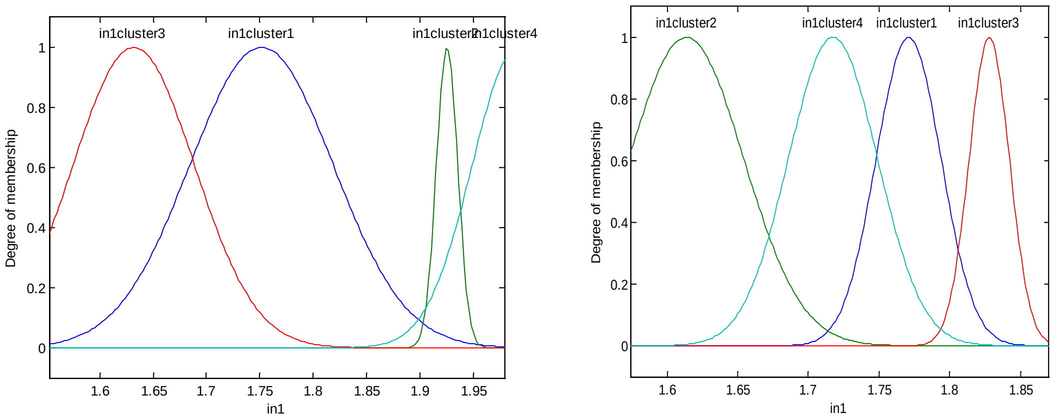


Fig. 2 (a) MF for Champua GPM technique with I (1990-95) data group

(b) MF for Champua GPM technique with II (1990-2000) data group

All input data have been fuzzified with membership functions. Fig. 2 (a, b) show the membership function for group I & II for Grid partitioning method. Both membership functions have four membership categories with Gaussian MF.

MODEL DEVELOPMENT

Conventional Method: Data for selected period of four groups have been averaged over three months (July, August and September) of monsoon period. Results obtained from stage-discharge equation (1) for monsoon period have been shown in Table 3. Power regression analysis has been used to get coefficient of the rating equation. Regression coefficient and Co-relation coefficient are varying between 60 to 80 percent. All the results have been plotted for linear relationship of estimated and observed values in Fig.3. The last plot of Fig. 3 has highest value of co-relation coefficient.

Table 3: Rating equation development with co-relation coefficient.

Group	Year	Rating equation by Power Regression and Regression Coefficient (R'^2).	Co-relation (R^2) between $Q^{observed}$ & $Q^{estimated}$
CI	1990-1995	$Q = 23.681H^{2.3582}$ $R'^2 = 0.8057$	$R^2 = 0.8103$
CII	1990-2000	$Q = 30.986H^{1.9313}$ $R'^2 = 0.6404$	$R^2 = 0.6297$
CIII	1990-2005	$Q = 24.528H^{2.188}$ $R'^2 = 0.7807$	$R^2 = 0.7866$
CIV	1990-2008	$Q = 30.185H^{2.045}$ $R'^2 = 0.845$	$R^2 = 0.8407$

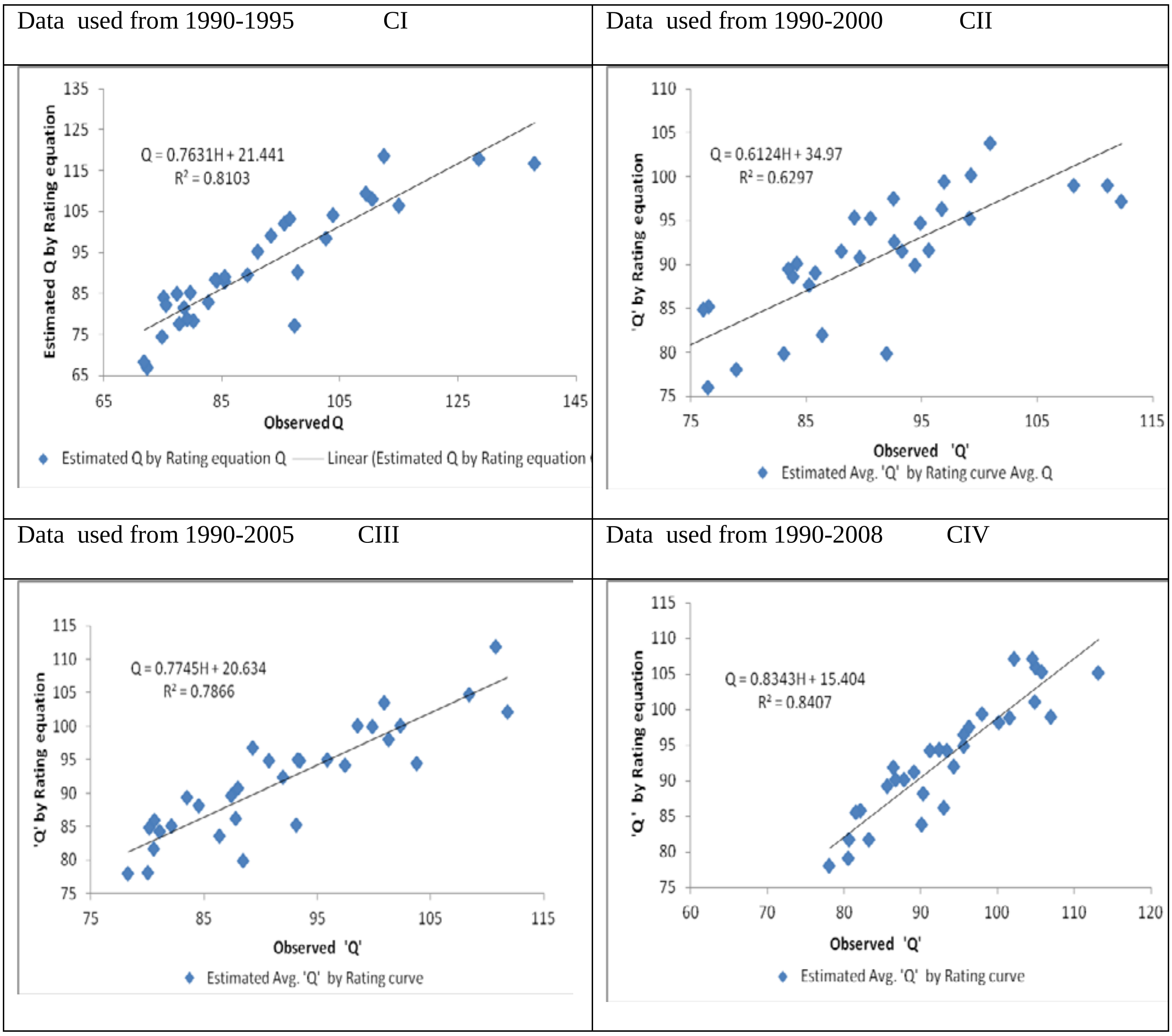


Fig. 3: Linear relationship between observed and estimated values for all four groups of data using stage-discharge equation.

Total four models have been developed using Grid partitioning (GPM I - IV) methods with all four groups of data (Table 2).

A developed FIS file using Grid partitioning method in MATLAB with one input and one output has shown in Fig. 4.

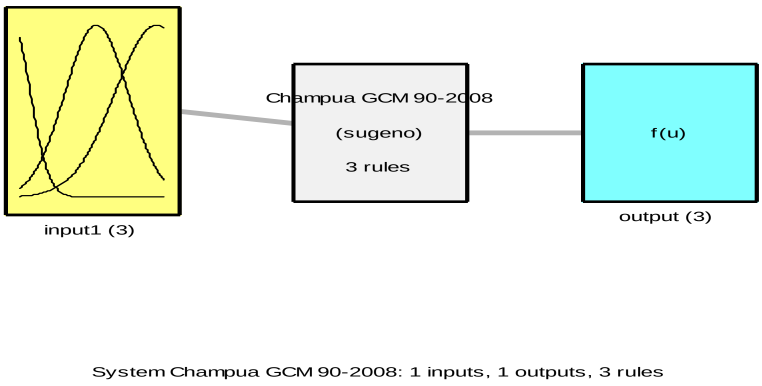


Fig. 4: Developed FIS file for Champua using GPM technique with data group IV (1990-2008) .

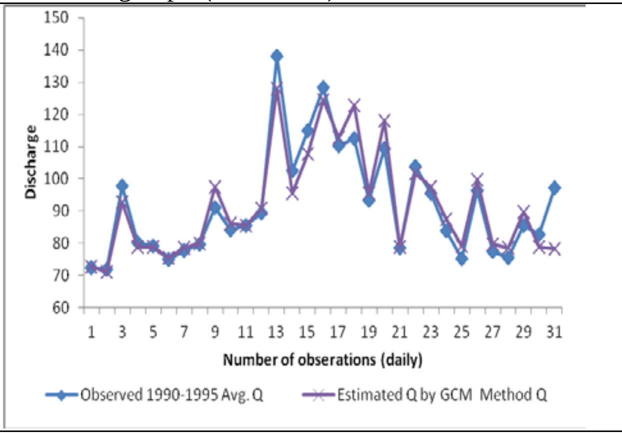
Fuzzy architectures for optimum criteria with minimum training and checking errors have been developed and given in Table 4. for all four models using GPM.

Table 4: Final Architecture of developed models.

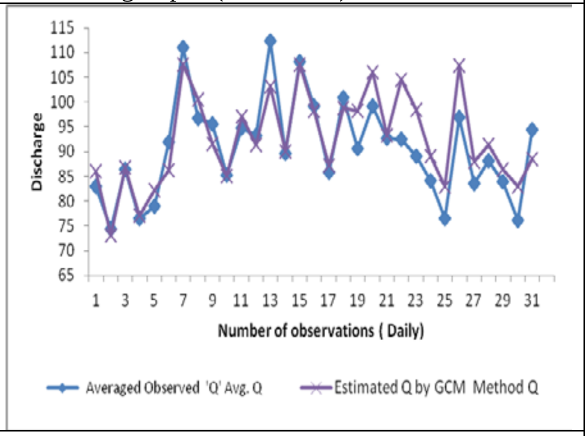
Model no.	Number of nodes	Number of Fuzzy rules	Training error	Checking error
GPM I	20	4	4.87903	6.42931
GPM II	16	3	3.88713	6.90423
GPM III	16	3	3.90127	5.20736
GPM IV	16	3	**2.98162**	**4.56863**

The models for minimum error criteria with optimum output have been saved for validation. Now the validation results using GPM for all four group of data have been compared in Fig. 5. Here in all 31 values first 20 values have been used for calibration/Training the model and last 11 values have been used to validate/checking the model. Model GPM IV is giving better results than GPM I, II and III models because these models developed with less number of data so variability of data is not much. In GPM IV, there is length of data. In which most of the varying conditions are covered within this range.

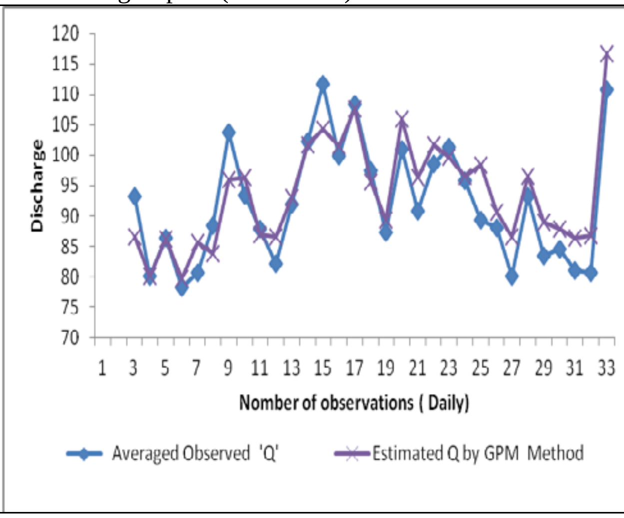
Data from group I (1990-1995) ‘GPM I’



Data from group II (1990-2000) ‘GPM II’



Data from group III (1990-2005) ‘GPM III’



Data from group IV (1990-2008) ‘GPM IV’

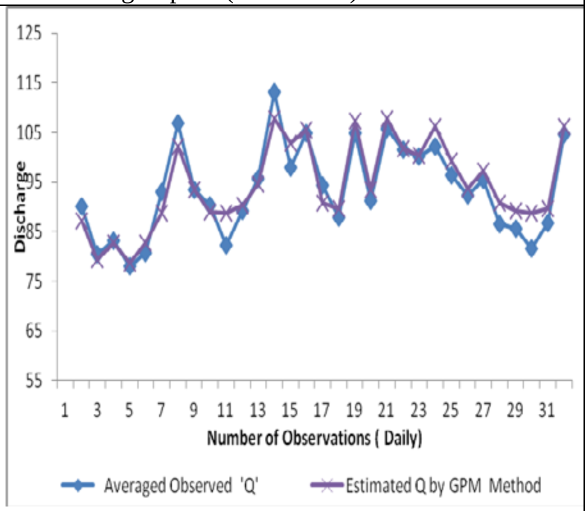
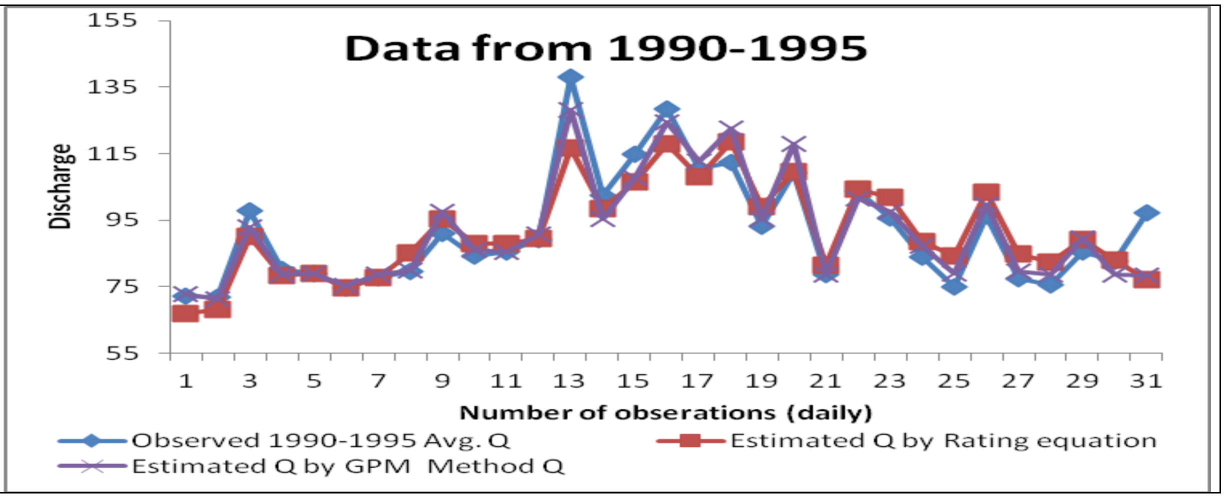


Fig. 5: The results from Grip Partitioning method for all four group of data sets.

RESULTS AND ANALYSIS

Results have been obtained using conventional , subtractive clustering and grid partitioning methods. Total number of observations (using the calibration and validation data) have been plotted for discharge values (Fig. 6). Both the methods are giving very prominent results for stage discharge relationship without using any nearby parameters. Therefore it is concluded that ANFIS technique can be very useful and it could recommend to establish the rating curve with reasonable good length of data.



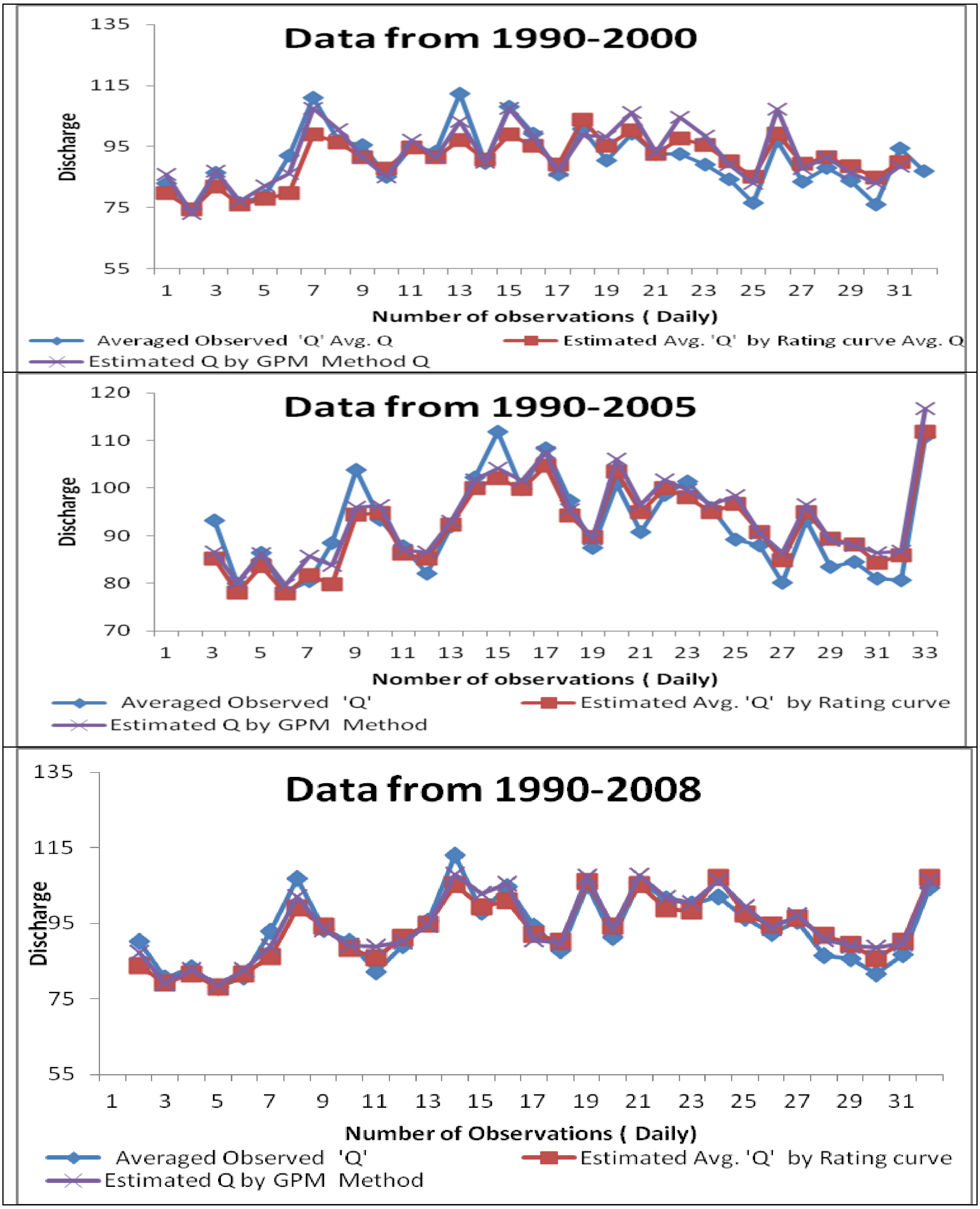


Fig. 6: Comparative graphs for all database using both methods with observed values.

MODEL EVALUATION

Evaluations of model performance utilize a number of different statistics and techniques referred as 'goodness of fit' statistics. To evaluate the model performance, different evaluation measures are considered and the hydrographs resulting from models are statistically examined. Global error statistics, which include the root-mean- square error (RMSE) between the computed and observed runoff, the coefficient of correlation (CORR), and the Nash-Sutcliffe criterion provide relevant information on overall performance. Finally, the error distributions at different threshold levels for all models are compared.

In the Nash-Sutcliffe criterion (Nash & Sutcliffe, 1970), percent variance explained (VAREX) by a model is computed by:

$$VAREX = \left[1 - \frac{\sum_{t=1}^{N} (O_t - P_t)^2}{\sum_{t=1}^{N} (O_t - \overline{O})^2} \right] x100 \tag{6}$$

where N is number of observations; O_t is observed value at time t (m^3/s); $\bar{O}$ is mean of the observed values (m^3/s); and P_t is predicted value at time t (m^3/s). The value of VAREX ranges from 0 (worst performance) to 100 (best performance). Table 5 shows the performance indices of models through validation data sets. The values of three performance evaluation criteria, namely, the root-mean-square error, the coefficient of correlation, and VAREX were found to be very good and consistent for soft computing models. It is shown in Table 5 that ANFIS Grid validation results are having the better performance indices.

Table 5: Performance indices of different models for validation data

S. N.	Models	RMSE	CC	VAREX
1.	Stage-Discharge model C I	2.433274	0.4487	29.109
2.	Stage-Discharge model C II	1.697462	0.7196	27.61286
3.	Stage-Discharge model C III	1.239252	**0.9165**	96.08968
4.	Stage-Discharge model C IV	**1.100476**	0.8907	**98.44247**
5.	GPM I	1.938514	0.5821	55.27914
6.	GPM II	2.081709	0.6299	-8.86826
7.	GPM III	1.539274	0.9055	96.59555
8.	GPM IV	**1.089016**	**0.9484**	99.71093

CC: Correlation coefficient; RMSE: Root Mean Square Error. VAREX: Percent variance

The correlation statistics, which evaluates the linear correlation between the observed and computed results, is consistent during validation period for all models. Slightly better correlation is observed with ANFIS GRID method as compared to conventional method.

CONCLUSIONS

A fuzzy logic algorithm is developed to model the stage-discharge relationship at a gauging site. The satisfactory estimation of the discharges by the proposed models from the four different data sets indicates that the stage-discharge modelling can reliably employ the ANFIS models. The ANFIS models estimate the discharge more accurately than conventional technique at various time steps as it is having least RMSE. The computed discharges have been plotted against conventional and observed discharges with all techniques in figures 6 . Scatter plotting of output show that the results from ANFIS GPM model is more satisfactory than conventional method (as per performance indices). Figure 3. show very good correlation (> 80%) with observed outputs in models C I & C IV. In ANFIS, grid method IV is giving better results with Hybrid optimization as compared to Back-propagation method. The ANFIS models are more flexible than conventional method, with more options of incorporating the fuzzy nature of the real world system. ANFIS (Fuzzy logic algorithm) has the ability to describe the knowledge in a descriptive human like manner in the form of simple rules using linguistic variables. Therefore, the ANFIS grid model may be considered for modelling. The developed models can be used to get related output for future discharge for any stage data or can develop a rating curve for future analysis.

REFERENCES

Deka, P., and V. Chandramouli, 2003, A fuzzy neural network model for deriving the river stage discharge relationship, Hydrol. Sci. J., 48, 197–209.

Eberhart, R. C., and R. W. Dobbins, 1990, Neural Network fC TOOls: A practical Guide, Elsevier, New York.

Hundecha, Y., A. Bardossy, and H.-W. Theisen, 2001, Development of a fuzzy logic based rainfall-runoff model, Hydrol. Sci. J., 46(3), 363–377.

Nash, J. E., and J. V. Sutcliffe, 1970, River flow forecasting through conceptual models: 1. A Discussion of principles, J. Hydrol., 10, 282 – 290.

Nayak PC, Sudheer KP, Rangan DM, Ramasastri KS., 2004. A neuro- fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology* 291: 52 – 66.

Nayak PC, Sudheer KP, Rangan DM, Ramasastri KS., 2005a. Short term flood forecasting with a neurofuzzy model. *Water Resources Research 41: W04*.

Rama Mehta, S.K. Jain & Vipin Kumar, 2005, Fuzzy Technique for Reservoir Operation – effect of Membership functions with different number of categories;, Hydrology Journal, Vol. 28, No. 3-4, Sept. - Dec.

Rama Mehta, and S.K.Jain, 2009, Optimal Operation of a Multi-Purpose Reservoir using Neuro-Fuzzy Technique, Water Resour. Manage, vol. 23, pp.509-529.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986, Learning internal representation by error propagation. Inf: *Parallel Distributed Processing* (ed. by D. E. Rumelhart & J. L. McClelland), vol.1: Foundations. MIT Press, Cambridge, Massachusetts, USA.

Xiong LH, Shamseldin AY, O'Connor KM. 2001. A nonlinear combination of the forecasts of rainfall – runoff models by the first order Takagi – Sugeno fuzzy system. *Journal of Hydrology, 245*: 196 – 217.

Xiong, L. H., and K. M. O'Connor (2002), Comparison of four updating models for real-time river flow forecasting, Hydrol. Sci. J., 47, 621 – 639.

Zadeh, L.A. (1973) Outline of a new approach to the analysis of complex system and decision processes. IEEE transactions on systems, man, and cybernetics, Vol. SMC-3, No. 1.