HYDROLOGICAL PROCESSES
Hydrol. Process. **13**, 123–138 (1999)

A self-organization algorithm for real-time flood forecast

Fi-John Chang* and Yuan-Yih Hwang
Department of Agricultural Engineering, National Taiwan university, Taipei, Taiwan, Republic of China

Abstract:

The group method of data handling (GMDH) algorithm presented by A. C. Ivakhnenko and colleagues is an heuristic self-organization method. It establishes the input–output relationship of a complex system using a multilayered perception-type structure that is similar to a feed-forward multilayer neural network. This study provides a step towards understanding and evaluating a role for GMDH in the investigation of the complex rainfall–runoff processes in a heterogeneous watershed in Taiwan. Two versions of the revised GMDH model are implemented: a stepwise regression procedure and a recursive formula. Eleven typhoon events in the Shen-cei Creek watershed, Taiwan, are used to build the model and verify its usefulness. The prediction results of the revised GMDH models and the instantaneous unit hydrograph (IUH) model are compared. Based on the criteria of forecasting precision and the rate and time of peak error, a much better performance is obtained with the revised GMDH models. Copyright © 1999 John Wiley & Sons, Ltd.

KEY WORDS self-organization algorithm; flood forecasting; GMDH; rainfall–runoff; stepwise regression; recursive formula

INTRODUCTION

Real-time flood forecasting is one of the most challenging and important tasks for hydrologists in Taiwan. It is a challenge because of the high mountains and steep upstream channels of all watersheds in the island. Floods generated by typhoons from July to October can reach the downstream cities within a few hours, while the temporal and spatial variability in both watershed characteristics and rainfall make its very difficult to model the within-basin processes. Thus, flood forecasting is important not only because a flood can cause tremendous damage to the downstream cities, but also because the massive stream-flow during the flood period is the major water input to upstream reservoirs so that a failure to operate the reservoirs during a flood can lead to drought in the following year.

Most of mathematical models used for flood forecasting in Taiwan are deterministic methods, which use quasi-physical conceptual models, because of the lack of rainfall–runoff data and the preferences of most scientists in Taiwan. Although conceptual models with optimized parameters may match an observed discharge record more satisfactorily in representing the non-linearity in a watershed, there is a great danger of over parameterization, as has been frequently pointed out (e.g. Jakeman and Hornberger, 1993; Beven, 1989). It seems that building a conceptual rainfall-runoff model would only involve wild guesses at these variables and their interaction; consequently, they can hardly be expected to provide a reliable model for forecasting purposes. The best way of building a model for flood forecasting for Taiwan's watersheds is more likely to be just to look at the data and nothing else.

* Correspondence to: Professor Fi-John Chang, Department of Agricultural Engineering, National Taiwan University, Taipei, Taiwan 10617, Republic of China.

Contract grant sponsor: National Science Council, Republic of China.
Contract grant number: NSC87-2313-B–002-049.

CCC 0885–6087/99/020123–16$17·50

Received 3 December 1997
Accepted 24 April 1998

A useful tool is Ivakhnenko's group method of data handing (GMDH) (Ivakhnenko, 1970, 1989; Ivakhnenko *et al*., 1990). As cited by many scientists, GMDH is a useful data process for identifying complex systems, especially when only a small amount of collected data is available (Tamura and Kondo, 1980; Farlow, 1984). Its most remarkable advantage is that it self-selects the structure of the model without using *a priori* information of the relationship between input and output variables, which one might call 'regression without models — directions in the search for structure'. However, this method has received very little attention in the water resource literature despite important developments as long as two decades ago and successful use in broad areas such as signal processing, control theory, system identification, etc. There are only a few applications of GMDH to modelling of environmental and ecological systems and most of them were performed in the Soviet Union and Japan (Ikeda *et al*., 1976; Abdullayev *et al*., 1988; Krotov and Kozubovskiy, 1987; Mamedov and Ivakhnenko, 1987).

It is our hope that this study can provide not only a direct contribution to the methodology of operational forecasting, but also an impetus for continued re-examination of rainfall–runoff methodology for engineering purposes. In this study, the GMDH algorithm was represented, modified and used to produce one hour ahead time forecasts of stream flows in several typhoon events. The results will also be compared with those obtained from the conceptual model. First, the original GMDH algorithm and its modified versions are described, then the GMDH application experiment and its results are presented. Finally, we present the conclusions and some closing remarks.

BASIC GMDH ALGORITHMS

General concept

GMDH is an heuristic self-organization method that models the input–output relationship of a complex system using a multilayered Rosenblatt's perception-type network structure, which is similar to a feed-forward multilayer neural network. The basic ideas of this algorithm are as follows. We know nothing about the system; let us generate and compare all possible input–output combinations, while each element in the network implements a non-linear function of two inputs and its coefficients are determined by a regression technique. Self-selection thresholds are employed at each layer in the network to filter out those elements that are useless in predicting the correct output. Only those elements whose performance indices exceed the threshold are allowed to pass to succeeding layers, where more complex combinations are formed. The procedure is performed until a satisfactory result is reached. Figure 1 shows the structure of the overall input–output transformation. In general, the advantageous characteristics of the GMDH algorithm for modelling or problem solving can be summarized as follows: (1) a small training set of data is required; (2) the computational burden is reduced; (3) the procedure automatically filters out input properties that provide little information about the location and shape of the decision hypersurface; and (4) a multilayers structure is a computationally feasible way to implement multinomials of very high degree. In the following section, a brief description of the GMDH algorithms used in this work is given.

Main processes

Let us assume the output variable Y is a function of the input variables (x_1, x_2, x_3, …, x_n), as in the following equation

$$Y = f(x_1, x_2, \dots, x_n) \tag{1}$$

The Kolmogorov–Gabor polynomial

$$\hat{Y} = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots \tag{2}$$

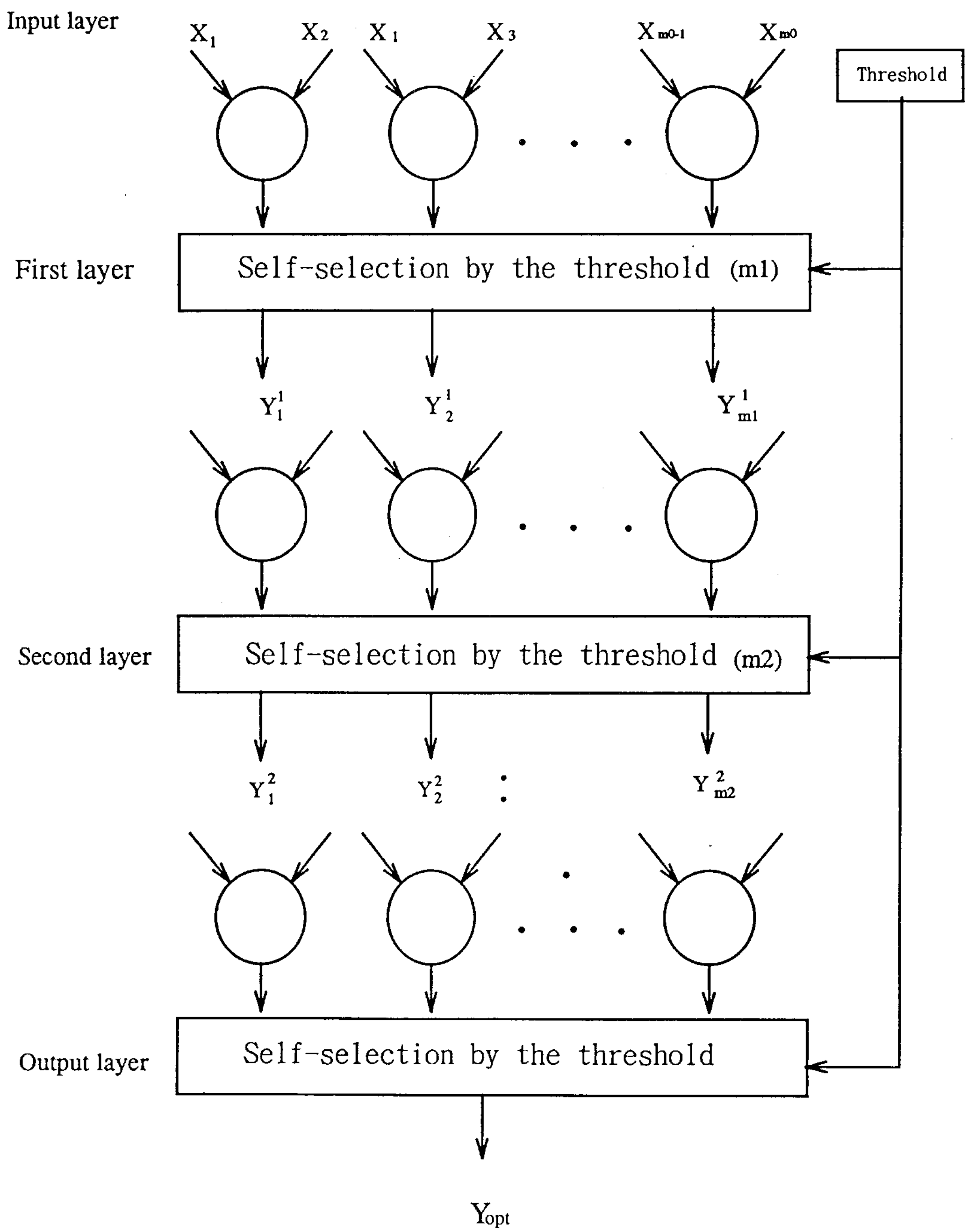

Figure 1. The structure of the GMDH algorithm

can simulate the input–output relationship perfectly and has been widely used as a complete description of the system model. However, if the complete polynomial is pursued directly with the method of least squares to estimate the coefficients, several problems may arise, e.g. insufficient data, ill conditioned data structure, etc. By combining the so-called partial polynomial of two variables in the multilayers, the GMDH algorithm can easily solve these problems. The main process is summarized in the following sequence.

Step 1. Select input variables. N useful input variables are chosen according to selection rules such as correlation criteria. In the case of the rainfall–runoff model, the rainfall and/or stream flow amount in the previous stages may be chosen. Let Q and R represent the stream flow and rainfall amount, respectively. The model is set as $Q(t) = f[Q(t-1), Q(t-2), \ldots, R(t-1), R(t-2), \ldots]$, or in terms of vector form, as Equation (1): $Y = f(x_1, x_2, \ldots, x_n)$, where Y and x_i represent vectors of successive states of the output and ith input, respectively.

Step 2. Divide the original data into a training and a checking set.

Step 3. Construct new intermediate variables.

In this step, all of the independent variables are taken two at a time to construct the partial polynomial equation

$$\hat{Y}_i = f_j(x_j, x_k) = a_{0i} + a_{1i}x_j + a_{2i} + x_k + a_{3i}x_j^2 + a_{4i}x_k^2 + a_{5i}x_j x_k$$
$$i = 1, \ldots, q; \; j = 1, \ldots, n; \; k = 1, \ldots, n-1; \; q = \frac{n(n-1)}{2} \tag{3}$$

The method of least squares is used to estimate the coefficients so that the equation will best fit the observed stream flow record, Y.

Step 4. Select the new variables.

Evaluate the root mean square of all residual errors in the preceding step. A self-selected threshold, i.e. residual error greater than a specific value, is set to filter out the least effective intermediate variables. Only the best of these intermediate variables are allowed to pass to the succeeding layer. These new variables can be interpreted as new improved variables that have better predictability powers than the previous generation.

Step 5. Truncate the multilayered interative computation.

Compare the best result of the present layer with the preceding layer; if the result is not improved then stop the process; otherwise go to step 3.

Step 6. Compute the predicted value.

The prediction model can be obtained as the intermediate variables remaining in the final layer.

REVISED GMDH ALGORITHMS

Two modified versions of GMDH, i.e. a stepwise regression procedure and a recursive (or sequential) formula, have been implemented in this study.

Stepwise regression procedure

In the original GMDH algorithm, the intermediate variables at each layer are generated from the previous level with their values fitted to the same output by adjusting their coefficients; consequently, they are highly correlated, especially at higher levels of the algorithm. To solve the 'multicollinearity' problem arising from the original GMDH algorithm, the stepwise-regression procedure is implemented in all the processes in computing the intermediate variables. The stepwise regression is capable of filtering out individual terms that are 'collinear' with other terms by mens of a significant test in its computation procedure. Furthermore, the number of terms contained in a final model can be dramatically reduced by using only an optimal partial polynomial. Several studies have shown successful results by implementing this procedure to the original GMDH algorithm (e.g. Duffy and Franklin, 1975; Tamura and Kondo, 1980)

Recursive formula

For real-time forecasting it would be more suitable if the model could be operated with an adaptive mode; that is, outputs at the current time-step are related to previously observed outputs. The GMDH algorithms can easily be modified to have the 'feedback' structure. A recursive formula can be found in Catlin (1989) and is implemented in this study to shorten the computation time in sequentially changing the model coefficients at each time-step.

APPLICATION OF THE GMDH ALGORITHMS

Description of watersheds

The GMDH algorithms are applied to the Shen-cei Creek, Taiwan, as shown in Figure 2. Hourly rainfall and runoff records of typhoon events for the watershed are generally good, with two rainfall stations used to represent the average watershed rainfall (using the Theissen weighting method).

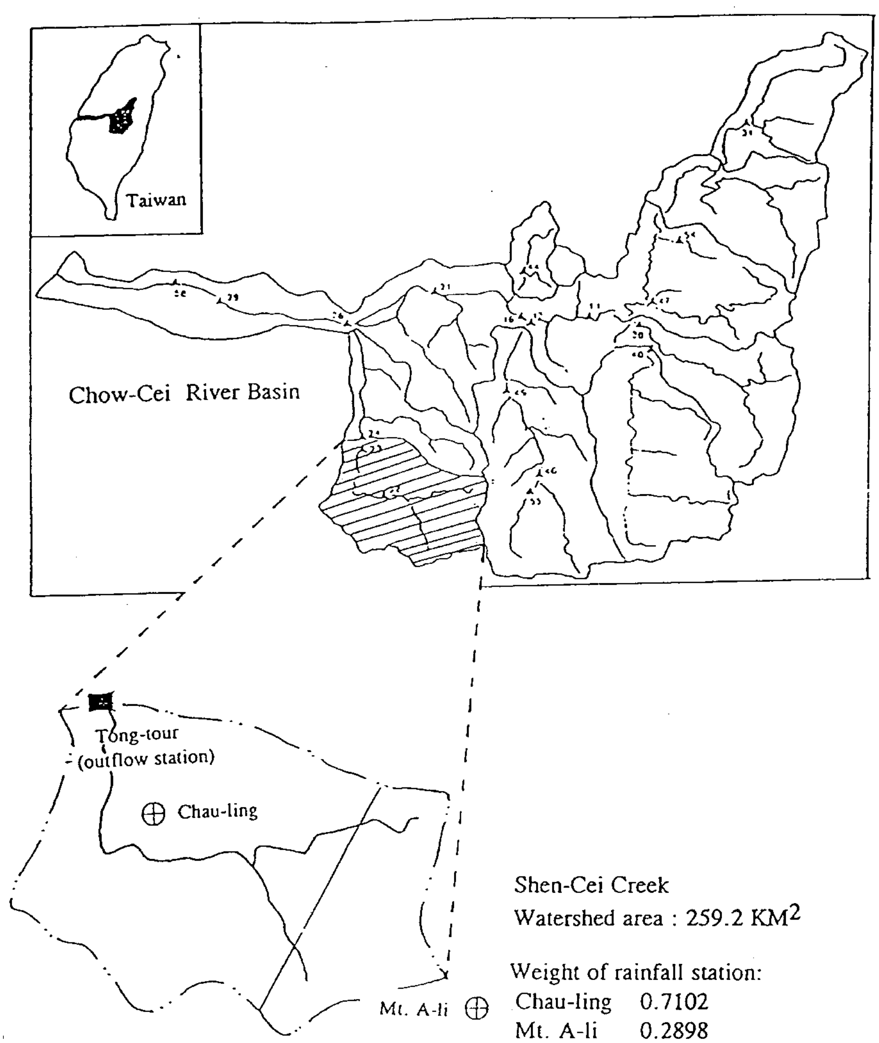

Figure 2. Location of hydrometric stations on Shen-cei Creek, Taiwan

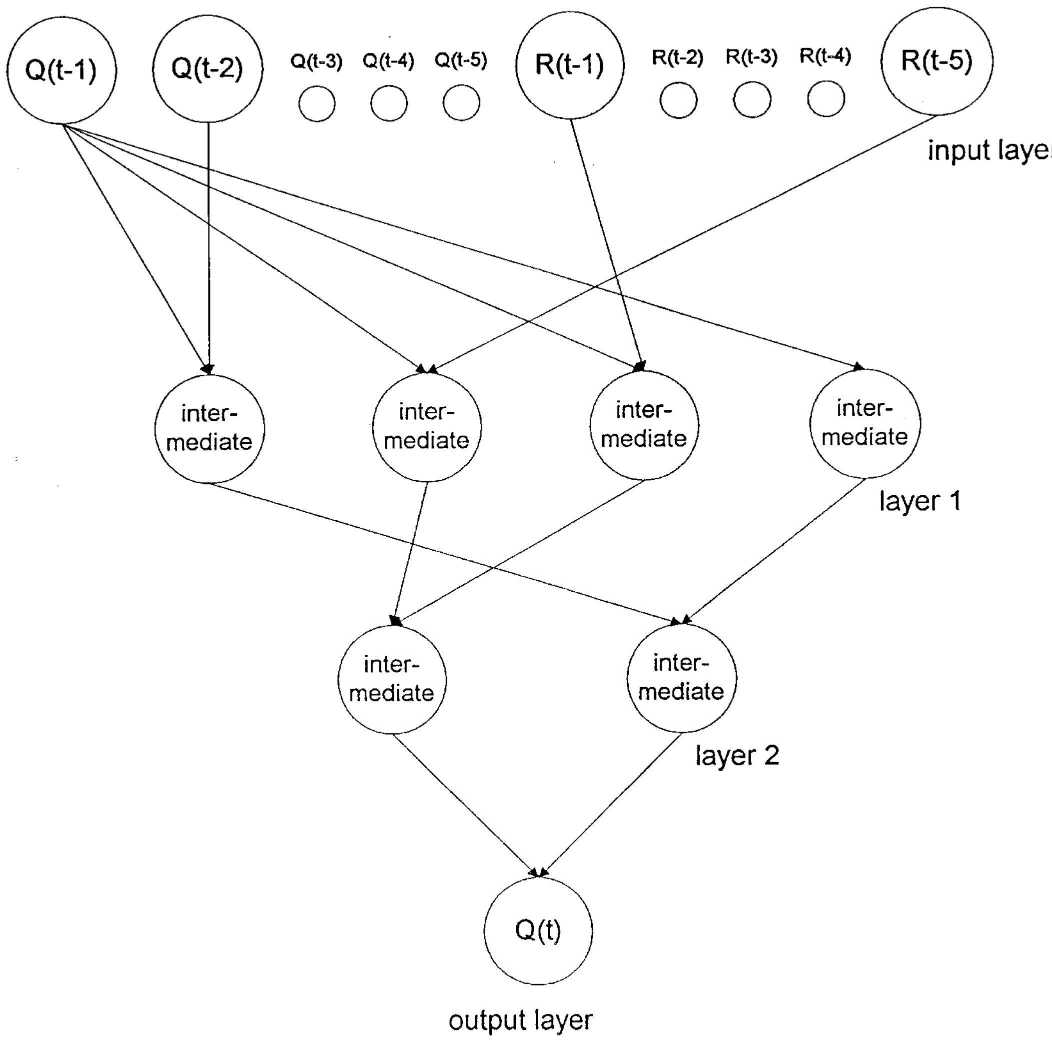

Figure 3. The forecasting model constructed by the second typhoon event

Table I. The selected flood events at Shen-cei Creek

Event Number	Date	Peak flow (cms)
1	6–8 September 1970	2680
2	16–18 September 1971	1270
3	22–24 September 1971	1390
4	5–7 June 1972	1110
5	25–26 July 1972	1160
6	18–19 June 1974	1610
7	3–4 August 1975	3170
8	1–3 June 1977	1340
9	27–28 August 1980	3400
10	22–27 July 1981	1010
11	12–16 August 1988	2810

Shen-cei Creek has a catchment area of 259 km^2 and is located upstream of the Chow-cei River Basin in central Taiwan. The Chow-cei River is the largest river in Taiwan, and the river basin has had the largest fertile plain for agricultural products during the last several decades and for industrial manufacturing in recent years. The Chow-cei Basin has a subtropical climate with pronounced wet and dry seasons. Catastrophic flood events sometimes occur during the wet season. These events are difficult to predict, even with non-linear and/or adaptive models. The flood events were recorded at the outlet of Shen-cei Creek. Only events that occur at this station and have a flow of over 1000 $m^3 s^{-1}$ have been taken into account in this study. This produced 11 events during the period 1970–1988.

Rainfall–runoff modelling

Stepwise GMDH algorithm (SGMDH). Based on hurly rainfall–runoff data, the river flow $Q(t)$ at present time t can generally be estimated by a function of previous stages of rainfall and runoff amounts as follows

$$Q(t) = f[Q(t-1), Q(t-2), \ldots, Q(t-5), \ldots R(t-1), R(t-2) \ldots, R(t-5)] \tag{4}$$

Because the area of the catchment is small, where flow from upstream can reach downstream in just a few hours, only the most recent five previous stages of rainfall and runoff data are used to construct the forecast model. From our experience, these data can fully represent the water impulse and hydrological persistence to the present flow. To determine the unknown non-linear structure from the available past data, the stepwise GMDH algorithms are applied. The number of input variables of the first layer is then set to 10, while the number of input variables of the rest of the layers is limited to eight. The revised GMDH algorithm is applied to build forecasting models by using a data set chosen from typhoon events used as a calibration. The revised algorithm converges very quickly, generally in no more than three layers, and it does provide relatively simple forecasting models. For example, if the data set of the second typhoon event is used, the forecasting model can be structured as shown in Figure 3, and its output function is

$$\begin{aligned} Q(t) = {} & 0{\cdot}1417 \times Q(t-1) + 0{\cdot}8335 \times Q(t-2) + 11{\cdot}5040 \times R(t-1) + 4{\cdot}6390 \times R(t-5) - 0{\cdot}0003 \\ & \times Q(t-2)^2 - 0{\cdot}1141 \times R(t-5)^2 - 0{\cdot}3019 \times R(t-1)^2 + 0{\cdot}0203 \times R(t-1) \\ & \times Q(t-2) - 0{\cdot}0017 \times Q(t-1) \times R(t-5) \end{aligned} \tag{5}$$

IUH. The instantaneous unit hydrograph (IUH) is the hydrograph of runoff from unit depth of effective rainfall in an instant time on a catchment. It has a significant meaning from a mathematical point of view and has found wide-ranging application for both the design and estimation of actual floods. A catchment, for an IUH, may be represented by a series of N identical linear reservoirs which have the same storage constant, K. The IUH can be obtained by a storage function and continuity equation and is shown as:

$$U(t) = \frac{1}{K\Gamma(N)} e^{-t/k} \left(\frac{t}{k}\right)^{N-1} \tag{6}$$

where $U(t)$ is the instantaneous unit hydrograph, K is the reservoir storage constant and $\Gamma(N)$ is the gamma function, where the parameters K and N can be obtained by the method of moments. When N is integer, $\Gamma(N) = (N-1)!$

The excess rainfall hydrograph (ERH) and the direct runoff hydrograph (DRH) are obtained from the recorded rainfall–runoff data of the catchment. The related equations of N and K can also be computed from the first and second moment of IUH. N and K values can therefore be derived by combining the information in these related equations (Chow *et al.*, 1988).

The IUH can be transformed into a unit hydrograph $U(T,\mathrm{t})$ with excess rainfall time lag, T. The unit hydrograph with one hour time lag $U(1,t)$, in this study, can hence be derived. The $U(1,t)$ can then be

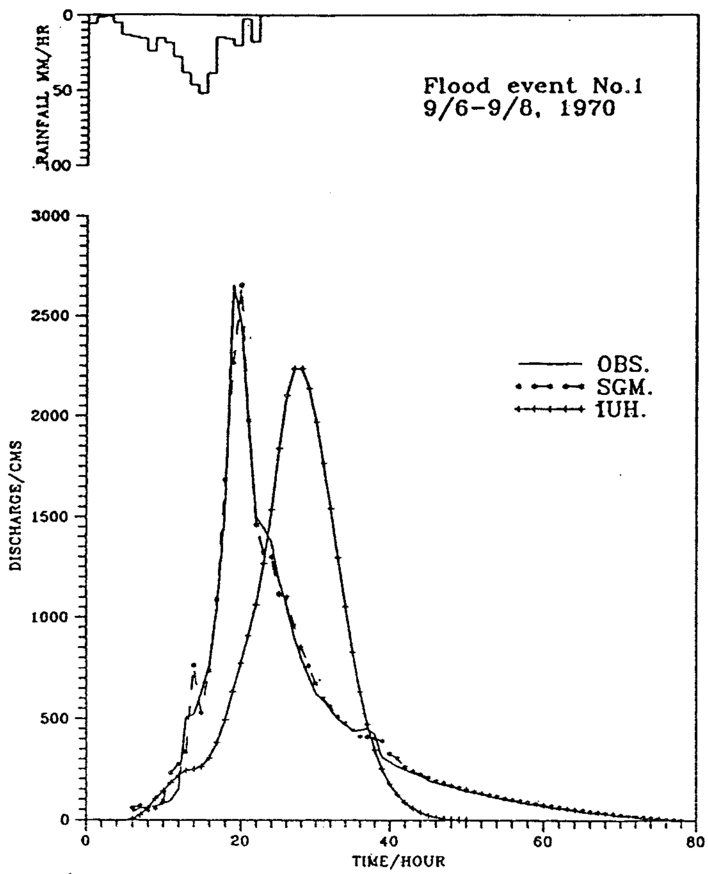

Figure 4. Observed vs. simulated responses at Shen-cei Creek for event no. 1

combined with the excess rainfall, $R(t)$, to obtain the direct runoff hydrograph, $Q(t)$, by the method of the convolution integral, as follows

$$Q(t) = \sum_{j=0}^{t} R_j U_{t-j} \Delta T \tag{7}$$

The IUH model is used for comparison with the GMDH models.

Comparison of model performances

When one deals with the flood forecasting performance of a model, several criteria should be taken into account: the forecasting precision, the rate of peak error and the time at peak error. In general, the best model according to one criterion may give a poor performance according to others. A multi-criteria approach is considered to rank the studied models according to objective criteria. The criteria of comparison used in this study are as follows.

1. Coefficient of efficiency

$$CE = 1 - \frac{\Sigma(Q_{\text{obs}} - Q_{\text{est}})^2}{\Sigma(Q_{\text{obs}} - \bar{Q}_{\text{obs}})^2} \tag{8}$$

where Q_{obs} is the observed flow, $\bar{Q}_{\text{obs}}$ is the mean of the observed flow and Q_{est} is the estimated flow.

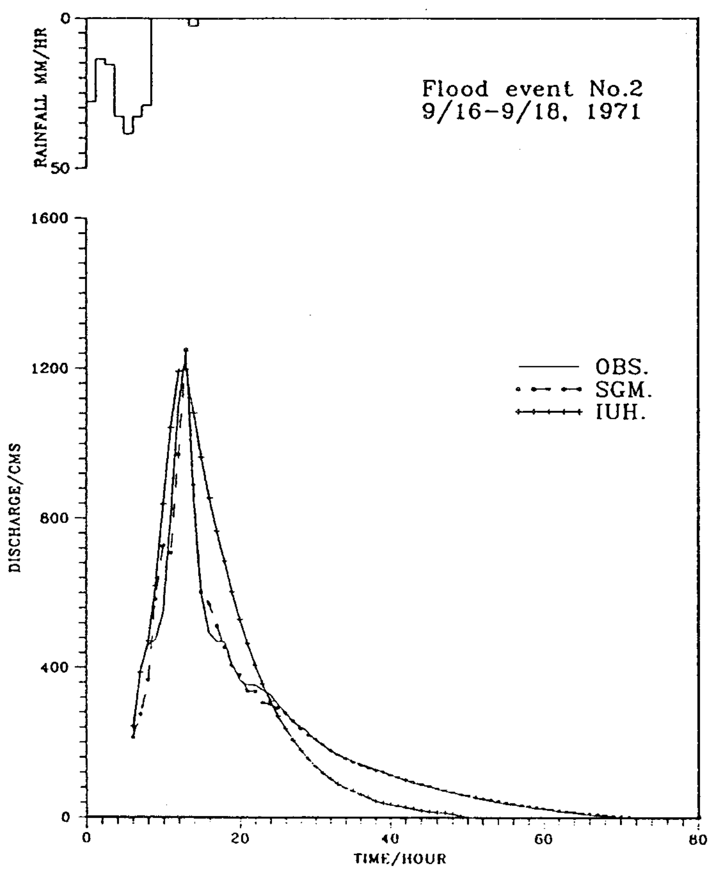

Figure 5. Observed vs. simulated responses at Shen-cei Creek for event no. 2

2. Error rate of peak flow

$$EQ_{\mathrm{p}} = \frac{Q_{\mathrm{p_{est}}} - Q_{\mathrm{p_{obs}}}}{Q_{\mathrm{p_{obs}}}} \tag{9}$$

where $Q_{\mathrm{p_{est}}}$ is the peak of estimated flow and $Q_{\mathrm{p_{obs}}}$ is the peak of observed flow.

3. Time error of peak flow

$$ET_{\mathrm{p}} = T_{\mathrm{p_{est}}} - T_{\mathrm{p_{obs}}} \tag{10}$$

where $T_{\mathrm{p_{est}}}$ is the time at which estimated peak flow arrived and $T_{\mathrm{p_{obs}}}$ is the time at which observed peak flow arrived.

A multi-criteria of objective function is set as follows

$$\mathrm{Min}(10^3 \times |ET_{\mathrm{p}}| + 10^3 \times |1 - CE| + 0{\cdot}1 \times |ET_{\mathrm{p}}|)$$

where the weighted coefficients for each criterion are allotted so that the criteria have about same weight.

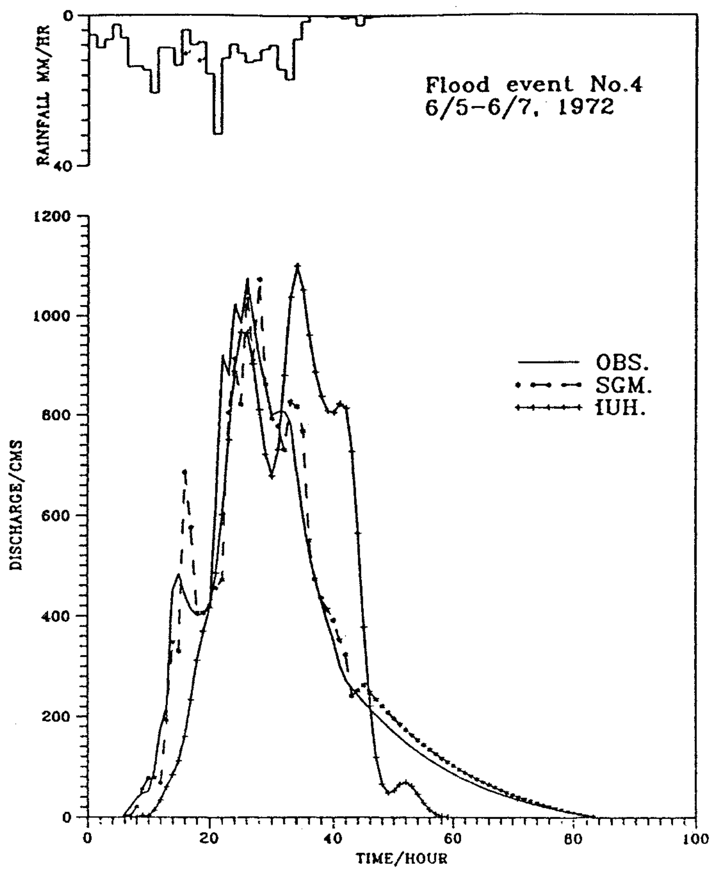

Figure 6. Observed vs. simulated responses at Shen-cei Creek for event no. 4

Summary of results

The IUH and SGMDH models have been applied to 11 events to simulate or forecast the flow at the outlet of Shen-cei Creek with the same hourly rainfall and stream flow data. Three events, 1, 2, and 4, have been chosen to train or build the model and calibrate the model parameters; and events 3, 6 and 9 have been used to check or verify the forecast performance of the above models. The models built by using calibrated events with the best performances in the verification events were then used to forecast the rest of the events, i.e. 5, 7, 8, 10 and 11, and to measure their forecast performances. The actual data for these flood events are shown in Table I.

In calibration events, the SGMDH model can fit the flow hydrographs almost perfectly in all three training sets, while the IUH model can only match the second event and does not fit the other two events (Figures 4–6). The models built without their formulae and/or parameters being further changed are then used to execute a one hour ahead, on-line flood forecast for three verification events. The results are shown in Table II. Apparently, both models have the best performance if the second event is used to build their formulae and/or parameters, while the SGMDH model has better performance, in terms of forecasting precision and rate and time of peak errors, than the IUH model in almost all the cases. The calibrating results demonstrate that the SGMDH model is much more flexible in fitting or simulating the rainfall–runoff hydrographs than the IUH model, while the verificative results of flood forecasting show that the SGMDH model is more robust than the IUH model.

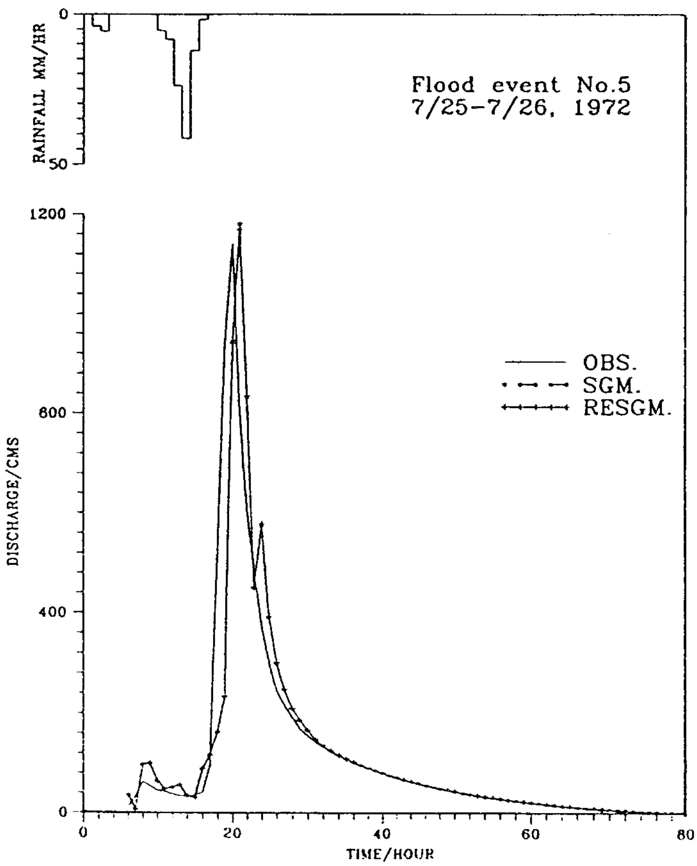

Figure 7. Observed vs. one hour ahead forecast at Shen-cei Creek for event no. 5

Table II. The results of SGMDH and IUH to calibrate and verify flood events

Purpose	Event number	Performances					
		Coefficient of efficiency (%)		Error rate of peak flow (%)		Time error of peak flow (hr)	
		SGMDH	IUH	SGMDH	IUH	SGMDH	IUH
Calibration	1	98·28	76·84	−0·03	−15·75	−1	3
Verification	3	93·24	84·83	1·27	11·57	4	0
	9	75·82	55·85	−32·54	−62·41	2	2
	6	80·64	70·70	48·96	−13·84	0	2
Calibration	2	99·25	86·10	−0·15	−4·31	0	0
Verification	3	95·05	95·83	8·44	−12·00	−1	−1
	9	84·96	48·20	7·20	−68·97	0	4
	6	83·82	89·70	9·86	−33·52	−1	1
Calibration	4	93·09	80·41	−0·31	2·11	−2	3
Verification	3	90·46	57·04	−9·43	27·18	−2	3
	9	9·44	50·12	−49·27	−50·39	−4	4
	6	39·48	43·74	−40·02	10·90	−4	5

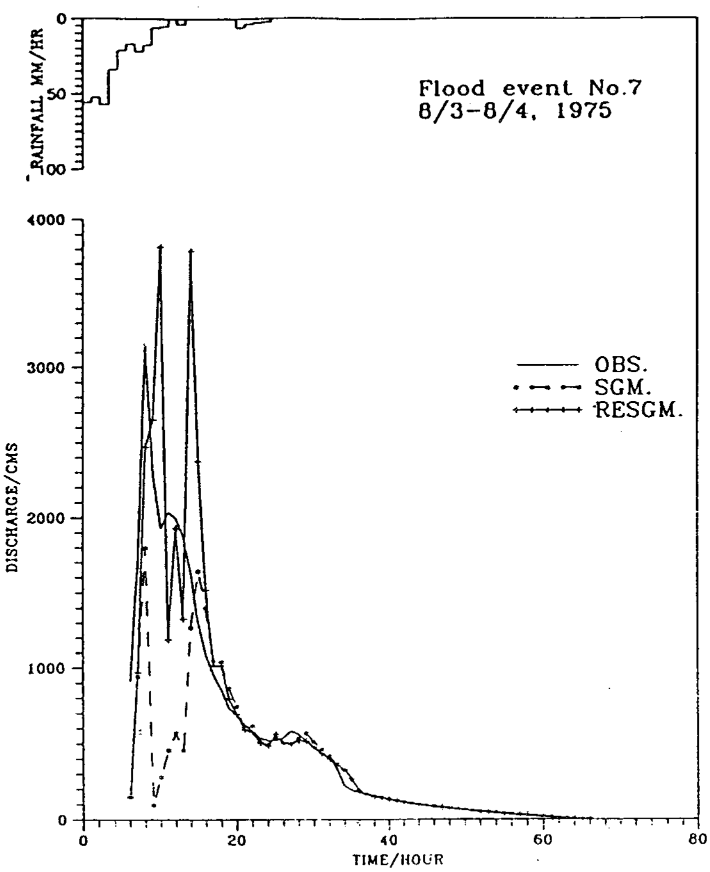

Figure 8. Observed vs. one hour ahead forecast at Shen-cei Creek for event no. 7

The SGMDH model built by using the data set from the second event was then further used to make a one hour ahead forecast for the rest of the events. The results are shown in Figures 7–11. The model can forecast the stream flow reasonably well for events 5, 8 and 10, but does not match the stream flow records for events 7 and 11. To find the reason, we might look through all the rainfall–runoff histograms. obviously, events 7 and 11 have a relatively large rainfall intensity in their very early stages, while the second event, which is used to build the forecast model, does not have large rainfall intensity in its early stage.

These results point out the limitations in applicability of the model, where, if the rainfall histograms between the calibration event (model built) and verification event (model used) are significantly different, the model might not be used properly. When we inspect the model, i.e. Equation (5), and the rainfall–runoff histograms of events 7 and 11, it is found that Equation (5) is significantly affected by $R(t-5)$, while $R(t-5)$ is set as zero before $t = 6$. Events 7 and 11 have a very large rainfall intensity and peak flow in their very early stages and consequently a forecast error does not seem to be a surprise in this circumstance.

In order to improve the real-time forecast, a recursive formula is implemented to the SGMDH model, which is called RESGMDH. The results show that the RESGMDH provides some improvement of model forecast performance (Table III and Figures 7–11), especially for events 7 and 11.

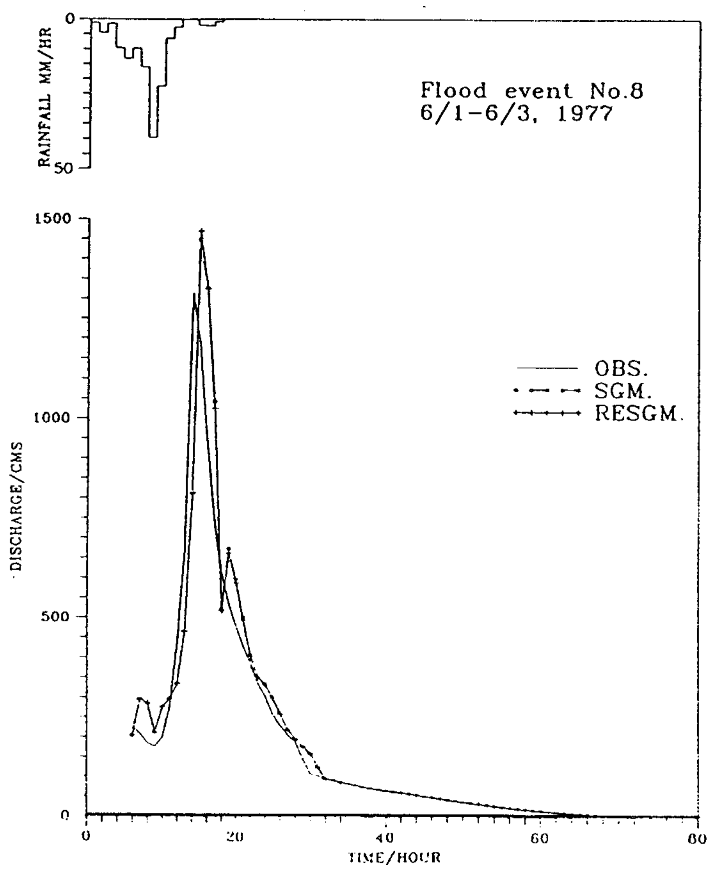

Figure 9. Observed vs. one hour ahead forecast at Shen-cei Creek for event no. 9

Table III. The results of SGMDH and RESGMDH models built by the second event

Purpose	Event number	Performances		
		Coefficient of efficiency (%)	Error rate of peak flow (%)	Time error of peak flow (hr)
Calibration	2	99·25	−0·15	0
Verification	3	95·05	8·44	−1
	9	84·96	7·20	0
	6	88·32	9·86	−1
Forecast (SGMDH)	5	72·34	3·49	−1
	7	40·06	−42·91	0
	8	86·15	10·03	−1
	10	88·94	10·86	−1
	11	35·01	−32·52	1
Forecast (RESGMDH)	5	72·49	2·49	−1
	7	57·95	21·01	−2
	8	86·28	11·55	−1
	10	92·81	4·24	−1
	11	76·79	13·83	−1

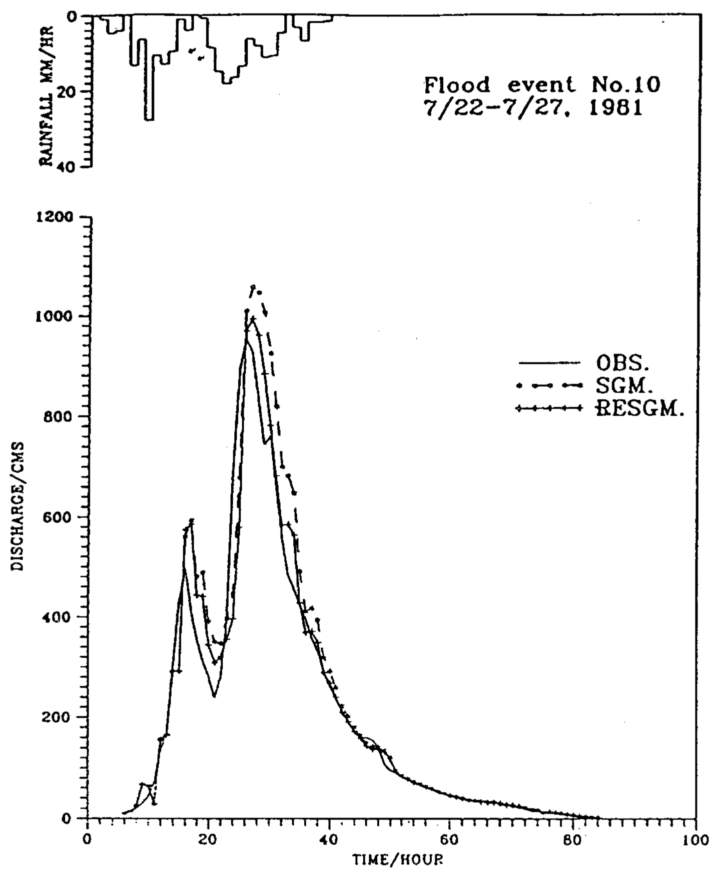

Figure 10. Observed vs. one hour ahead forecast at Shen-cei Creek for event no. 10

CONCLUSIONS

The GMDH is an heuristic, computer-oriented method that provides the foundation for construction of high-order regression models of complex systems and has been applied with great success in a wide variety of fields, ranging from economics to biology (Farlow, 1984); nevertheless, until now it has not been considered important in the field of water resources. This study provides a step towards understanding and evaluating a role for GMDH in the investigation of the complex rainfall–runoff processes in a heterogeneous watershed. A stepwise regression procedure and recursive formula were implemented to revise the GMDH algorithm, to simplify the model formulation and to enhance its efficiency in parameter estimation. Our study also shows that both technologies can easily be implemented in the original GMDH algorithm.

Comparing the prediction results, based on the criteria of forecasting precision and the rate and time of peak error, of the revised GMDH model with the IUH model by using 11 typhoon events in the Shen-cei Creek, Taiwan, a much better performance was obtained with the revised GMDH models. In the calibration events, the results demonstrate that the SGMDH model is much more flexible and easy to fit the rainfall–runoff histogram than the IUH model. In the verification events, the results support the finding that the SGMDH model is more robust than the IUH model. When we enquire further into the SGMDH-built model, it is found that a limitation to applicability does exist, whereby, if the rainfall histograms between the calibration event and the verification event are very different, then the model cannot be used properly. In this

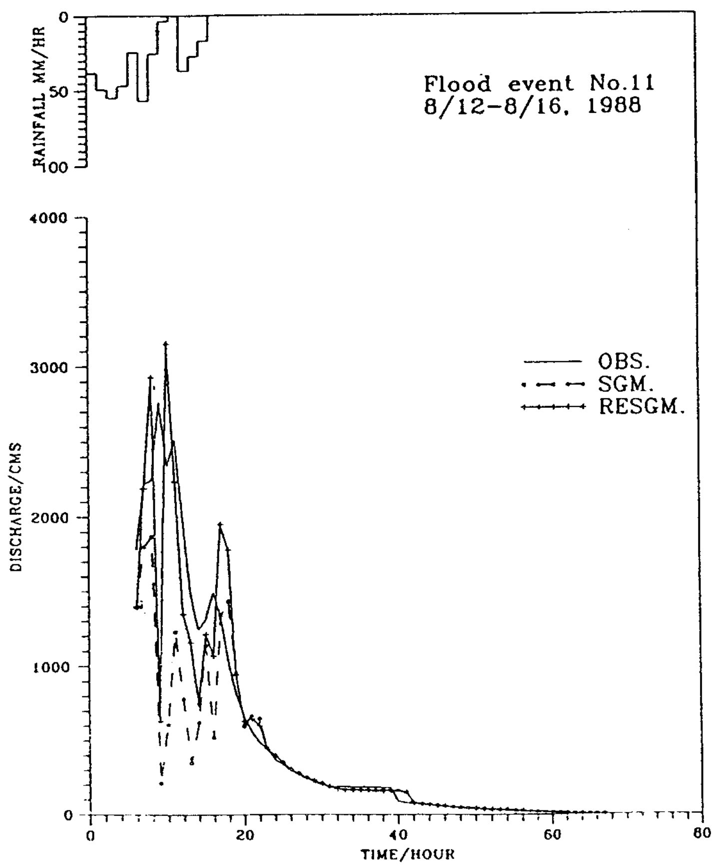

Figure 11. Observed vs. one hour ahead forecast at Shen-cei Creek for event no. 11

circumstance, the recursive formula of the model's parameter adjustment, i.e. RESGMDH, was found to be useful for improving the model forecast performance.

ACKNOWLEDGEMENT

Research support from the National Science Council, ROC (Grant no. NSC87-2313-B-002-049) is gratefully acknowledged.

REFERENCES

Abdullayev, F. A., Akhmedov, Sh A., and Mamedov, M. I. 1988. 'GMDH algorithm for forecasting the optical state of the atmosphere', *Sov. J. Autom. Inf. Sci.*, **21**, 58–61.

Beven, K. 1989. 'Changing ideas in hydrology — the case of physically-based models', *J. Hydrol*, **105**, 157–172.

Catlin, D. E. 1989. 'Estimation, Control, and the Discrete Kalman Filter.' SpringerVerlag, New York.

Chow, V. E., Maidment, D. R., and Mays, L. W. 1988. *Applied Hydrology*. McGraw-Hill, New York. pp. 260–262.

Duffy, J. J., and Franklin, M. A. 1975. 'A learning identification algorithm and its application to an environmental systems', *IEEE Trans. System. Man. Cybernet.*, **SMC-5**, 226–239.

Farlow, J. S. 1984. 'Self-Organizing Methods in Modeling GMDH Type Algorithms'. Marcel Dekker, New York.

Ikeda, S., Fugishige, S., and Sawaragi, Y. 1976. 'Nonlinear prediction model of river flow by self-organization method', *Int. J. Syst. Sci.*, **7**, 165–176.

Ivakhnenko, A. G. 1970. 'Heuristic self-organization in problem of engineering Ivakhnenko cybernetics', *Automatica*, **6**, 207–219.
Ivakhnenko, A. G. 1989. 'Nonparametric GMDH predicting models. Part 2. Indicative systems for selective modeling, clustering, and pattern recognition', *Sov. J. Autom. Inf. Sci.*, **22**, 2, 1–10.
Ivakhnenko, A. G., Fateyeva, Ye N., and Ivakhnenko, N. A. 1989. 'Nonparametric GMDH forecasting models. Part 1. Sorting the Bayes or Wald formulas', *Sov. J. Autom. Inf. Sci.*, **22**, 1, 1–8.
Jakeman, A. J., and Hornberger, G. H. 1993. 'How much complexity is warranted in a rainfall-runoff model?', *Wat. Resour. Res.*, **29**, 2637–2649.
Krotov, G. I., and Kozubovskiy, S. F. 1987. 'Verification of dendroscale forecasting by a multiplicative GMDH algorithm', *Sov. J. Autom Inf. Sci.*, **20**, 1–7.
Mamedov, M. I., and Ivankhnenko, N. A. 1987. 'Forecasting model of air pollution of an industrial city', *Sov. J. Autom. Inf. Sci.*, **20**, 90–92.
Tamura, H., and Kondo, T. 1980. 'Heuristics free group method of data handling algorithm of generating optimal partial polynomials with application to air pollution prediction', *Int. J. Syst. Sci.*, **11**, 1095–1111.
Taiwan Provincial Government Water Conservancy Bureau (TPWCB), Taiwan, 1991. 'Studies on the application of computerized hydrologic files in Taiwan'. [In Chinese.]
Wang, R. Y., and Lee, G. D. 1987. 'Study on the optimum rainfall–runoff models for the upstream water sheds of Shih-men reservoir', *The Hazards Mitigation Programm*, No. 75–34. Granted by the National Science Council, Taiwan. pp. 45–46. [In Chinese.]