Weather Radar Information and Distributed Hydrological Modelling (Proceedings of symposium HS03 held during IUGG2003 at Sapporo, July 2003). IAHS Publ. no. 282, 2003.

Application of models with different types of modelling methodologies for river flow forecasting

H. A. PRASANTHA HAPUARACHCHI, LI ZHIJIA
College of Water Resources and Environment, Hohai University, Nanjing 210098, China
prasanth@jlonline.com

ALBERT WOLFGANG FLÜGEL
International Water Management Institute, No. 127, Battaramulla, Sri Lanka

Abstract In the present study, a conceptual watershed model, a distributed watershed model, and an artificial neural network (ANN) have been applied to river flow forecasting in the Kalu River upper catchment in Sri Lanka. The Xinanjiang watershed model has been used as a conceptual watershed model and the SWAT model (Neitsch, 2000) has been used with spatial data as a distributed model. Two types of ANN architectures, namely multi-layer perceptron network (MLP) and radial basis function network (RBF) have been implemented as "black box" type modelling methodology. Based on the application results, it seems that the conceptual watershed model could perform slightly better than the distributed model and the ANN for this watershed. It was clearly noted that the performance of distributed models strictly depends on the quality of input data (Arnold *et al.*, 1998) whereas the performance of conceptual models depends on the calibration (Duan *et al.*, 1992, 1993).

Key words artificial neural networks; conceptual watershed models; distributed watershed models; multi-layer perceptron network; radial basis function; Xinanjiang model

INTRODUCTION

At present, a number of modelling methodologies are available for watershed modelling and river flow forecasting. Basically they can be divided into three categories as conceptual, distributed, and "black box" type modelling. The application and performance of these types of models depend on the data existence, data quality, hydrological characteristics of the catchment, etc. Therefore the selection of an appropriate model should be done carefully, analysing these dependents. The main difficulty of using conceptual watershed models is the calibration. In general, evidence from previous conceptual rainfall–runoff (CRR) model calibration studies has indicated that the calibration problem should be solved by global optimization techniques, since inferior optimal solutions have been obtained by various local optimization techniques. On the other hand, distributed watershed models require more input data, which may not be readily available for every catchment. Generally, the performance of distributed models is strictly dependent on the quality of data (Arnold *et al.*, 1998). While conceptual or physically-based models are of importance in the understanding of hydrological processes, there are many practical situations where the main concern is with making accurate predictions at specific locations with less available data. In such situations it is preferred to implement a simple "black box" (data driven, or machine learning) model to identify a direct mapping between the input and output without detailed consideration of the internal structure of the physical process. The artificial neural network (ANN) is probably the most successful machine learning technique with a flexible mathematical structure which is capable of identifying complex nonlinear relationships between input and output data, without attempting to reach an understanding of the nature of the phenomena. Neural networks are particularly useful in situations where the underlying physical process relationships are not fully understood, or where the nature of the event being modelled (i.e. a flood) may display chaotic properties (Dibike *et al.*, 1999).

In the present study, the Xinanjiang model (Zhao *et al.*, 1980, 1995) as a conceptual watershed model, the SWAT model (Soil Water Assessment Tool, Arnold *et al.*, 1998) as a distributed watershed model and two types of ANN architecture named multi layer perceptron (MLP) network and radial basis function (RBF) network have been applied for predicting daily flow of the Kalu River upper catchment in Sri Lanka for comparing the behaviours of different types of models.

CATCHMENT CHARACTERISTICS

The catchment is located to the southwest and south of the central highlands of Sri Lanka, and lies between 80.40–80.60°N latitude and 6.53–6.80°E longitude. Generally elevation varies from 20 to 2225 m a.m.s.l. and the total area is about 603 km^2. Mountain ranges, high peaks, dissected plateaus, escarpments, cover a greater part of the area. On average (depending on their lithology and structure) slopes vary from 10° to 35° in the upland ridges. The slopes and the drainage system have largely been governed by its geological structure. Annual precipitation is above 4000 mm and there are eight precipitation-gauging stations in and around the catchment. The average annual temperature of the area ranges from 26.9°C to 27.8°C. Dominant soil types visible in this area are red brown earths and low humic grey soils, reddish brown earths and immature brown loams, red yellow podzolic soils, bog and half-bog soils, and alluvial soils. But nearly 86% of the area is covered with Red Yellow Podzolic soils (Karunanayake *et al.*, 1990). Two main vegetation types, tropical rain forests and mountain forests, can be seen in this area. Basically there are 13 types of land-use patterns. About 30.2% of the land has been used for Chena cultivation (Department of Survey, Sri Lanka, land-use map 1992, 1:50 000), which is a causative factor for soil erosion and deforestation. Tea and rubber are other major cultivations.

MODELS

Xinanjiang watershed model

The Xinanjiang rainfall–runoff model is a conceptual watershed model developed at Hohai University, China in the 1970s. It provides an integral structure, statistically describing the non-uniform distribution of runoff producing areas, which features it as one of the conceptual, semi-distributed hydrological models. By comparing with the

Table 1 Parameters of the Xinanjiang model.

Parameter	Abbreviation	
Evapotranspiration	*K*	Ratio of potential evapotranspiration to the pan evaporation
	WUM	Tension water capacity of upper layer
	WLM	Tension water capacity of lower layer
	C	Evapotranspiration coefficient of deeper layer
Runoff generation	*WM*	Areal mean tension water capacity
	B	Exponential of the distribution of tension water capacity
	IM	Ratio of impervious area to the total area of the basin
Runoff separation	*SM*	Free water storage capacity
	EX	Exponential of distribution water capacity
	KG	Out flow coefficient of free water storage to the ground water flow
	KI	Out flow coefficient of free water storage to the inter flow
Runoff concentration	*CG*	The recession constant of ground water storage
	CI	The recession constant of lower interflow storage
	CS	The recession constant of channel network storage
Flow routing	*XE*	Muskingum coefficient

Pitman model of South Africa, the Sacramento model of USA, the NAM model of Europe and the SMAR model of Ireland, Gan *et al.* (1996) proved that the Xinanjiang model did consistently better, even in dry catchments. The model consists of 15 parameters (Table 1) and the application results are best for the humid and semi-humid catchments.

All the parameters of Xinanjiang model should be calibrated before application. The hydrological data inputs of the model are areal mean rainfall *P*, and measured pan evaporation *EM*. Besides these, subcatchment areas, and initial state of the catchment are necessary for the calculations. The outputs are the discharge from the whole basin *TQ*. State variables are the areal mean free water storage *S* and the areal mean tension water storage *W*. Basically *W* is divided into three layers as upper, lower, and deeper having the tension water storage components *WU*, *WL*, *WD* and evaporation components *EU*, *EL*, *ED*, respectively (see Fig. 1). *RD* is the direct runoff from the impervious area and *FR* is the runoff contributing area factor, which is related to *W*. The runoff (*R*) produced from pervious area is divided into three components *RS*, *RI* and *RG* referred to as surface runoff, interflow and groundwater flow, respectively. The three components are further transferred into *QS*, *QI* and *QG* and together form the total inflow to the channel network of the basin. The outflow from each sub-basin is first simulated and then routed down through channels to the main basin outlet.

Artificial neural networks (ANN)

Neural nets are now widely used in many different areas of science to model complex problems as substitutes for, and in association with, conventional mathematical and statistical models. Normally, conventional models require a great deal of detailed data, for example, topographical maps, river networks and characteristics, soil rainfall and

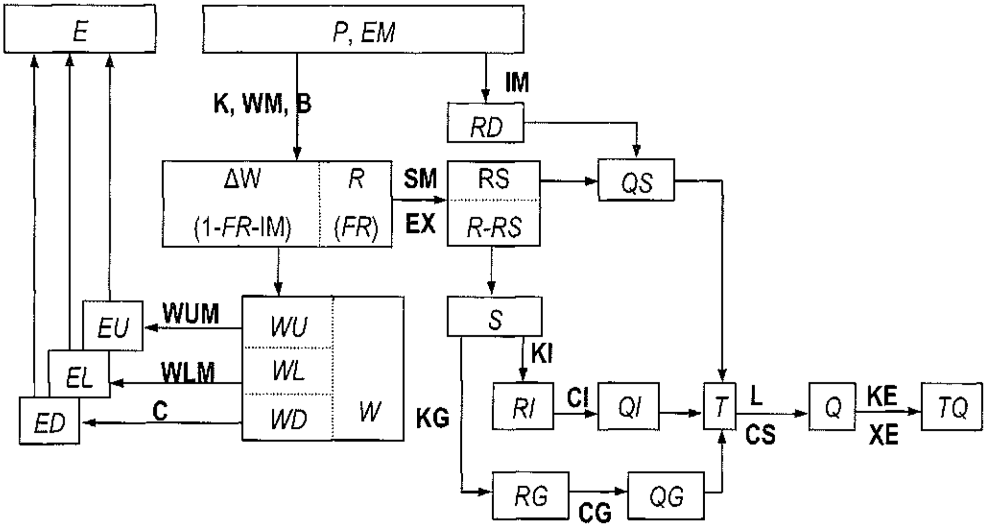

Fig. 1 Flowchart of the Xinanjiang model (All the symbols outside the blocks are parameters).

runoff data. Often, these data may not available and pose a great difficulty for model calibration. Neural networks, on the other hand, offer real prospects for a cheaper, more flexible, less assumption-dependent and adaptive methodology well suited to model flood processes, which by their nature are inherently complex, nonlinear, sometimes hard to predict because of nonlinearity and chaotic effects and sometimes life critical. In this study, two types of feedforward ANN architectures, multi-layer perceptron network (MLP) and radial basis function network (RBF) were implemented for downstreamflow forecasting. Details of these two methods can be found in Dibike *et al.* (1999) and Mason *et al.* (1996).

SWAT model

The Soil and Water Assessment Tool (SWAT) (Neitsch *et al.*, 2000) was developed by the US Department of Agriculture (USDA), Agricultural Research Service (ARS). It was developed to predict the impact of land management practices on water, sediment and agricultural chemical yields in large complex watersheds with varying soils, land-use and management conditions over long periods of time. Application of the model requires specific information about weather, soil properties, topography, vegetation, and land management practices in a watershed. SWAT is capable of handling with water movement, sediment movement, crop growth, and nutrient cycling with the above information. Therefore, watersheds with no monitoring data can be modelled, and the effect of changes in input data can be quantified. In the present study, only the hydrological processes embedded in SWAT are considered. A detailed description of the SWAT model and user manual can be found in Neitsch *et al.* (2000) or at the web site http://www.brc.tamus.edu/swat/.

APPLICATION AND RESULTS

Nine years of daily hydrological data (1987–1995) have been used for the calculations such that seven years of data (1987–1993) are used for calibrating or training and two years of data (1994–1995) for validating the models. The watershed was delineated into seven subcatchments and flow computations were done for each subcatchment.

Statistical indices

Two statistical indices have been selected to compare the results obtained from calibration and verification. They are the Nash-Sutcliffe coefficient D_y and the percentage of total error $\%Err$ in each year as:

$$D_y^2 = 1 - \frac{\sum_{i=1}^{m} w_i^2 \left[Q_{Obs,i} - Q_{Cal,i}\right]^2}{\sum_{i=1}^{m} w_i^2 \left[Q_{Obs,i} - \overline{Q}_{obs}\right]^2} \tag{1}$$

$$\%Err = \left[\frac{\sum_{i=1}^{m} Q_{Obs,i} - \sum_{i=1}^{m} Q_{Cal,i}}{\sum_{i=1}^{m} Q_{Obs,i}}\right] \times 100 \tag{2}$$

Where $\overline{Q}_{obs}$ is daily mean observed discharge, Q_{obs} is the observed discharge, Q_{cal} is the calculated discharge, m is the number of time steps in each year and w_i is a weighting factor.

Application of ANN

The precipitation data set was prepared by averaging the total daily precipitation of seven gauging stations in and around the basin. Evaporation was assumed to be the same in the whole catchment. Firstly, the networks were trained with concurrent rainfall and evaporation data as input and discharge as output. The efficiency was checked by trial-and-error process, with various combinations of precipitation, evaporation and discharge data. After a series of experiments, the best performance was obtained using two antecedent runoff values (observed), in addition to three rainfall and one evaporation value, as input to the network (some of the validation results are shown in Table 2). The number of nodes in the hidden layer was varied between two and seven and the one with three gave the highest efficiency in the verification stage. However, when additional antecedent runoff and rainfall data were added to the input pattern, it was noted that the performance of the network for the training period was improved, but its performance for validation period was deteriorated.

Table 2 Validation results of ANNs.

Network combination	Year	Q_{obs} (mm)	Multi-layer perceptron (MLP): Q_{cal} (mm)	%*Err*	D_y	Radial basis function (RBF): Q_{cal} (mm)	%*Err*	D_y
R10	1994	2105.63	2101.650	0.189	0.861	2210.279	–4.970	0.801
	1995	3039.80	2870.665	5.564	0.842	2941.006	3.250	0.836
P2, R1, E1	1994	2105.63	2510.121	–19.213	0.849	2224.598	–5.650	0.819
	1995	3039.80	3286.601	–8.119	0.873	2986.968	1.738	0.831
P3*, R2†, E1‡	1994	2105.63	2167.240	–2.926	0.876	2099.813	0.276	0.847
	1995	3039.80	2979.118	1.996	0.875	2967.643	2.374	0.816
P5, R2, E1	1994	2105.63	2310.697	–9.739	0.835	2215.038	–5.196	0.823
	1995	3039.80	3022.590	0.566	0.896	2990.585	1.619	0.844
P5, R3, E1	1994	2105.63	2087.489	0.862	0.854	2077.381	1.342	0.832
	1995	3039.80	2927.813	3.684	0.860	2987.477	1.721	0.822
P7, R5, E3	1994	2105.63	1918.734	8.876	0.853	2196.530	–4.317	0.814
	1995	3039.80	2688.946	11.542	0.872	2948.697	2.997	0.824
P9, R9, E9	1994	2105.63	2125.022	–0.921	0.595	2216.280	–5.255	0.457
	1995	3039.80	2823.184	7.126	0.556	2917.782	4.014	0.570

*Three antecedent precipitation data sets; †Two antecedent runoff data sets; ‡One evaporation data set.

Table 3 Calibration and validation results of the Xinanjiang model.

From	To	Q_{obs} (mm)	Q_{cal} (mm)	%*Err*	D_y
1987.01.01	1987.12.31	1820.686	1865.743	–2.475	0.853
1988.01.01	1988.12.31	3230.345	3130.724	3.084	0.848
1989.01.01	1989.12.31	3050.908	2951.325	3.264	0.946
1990.01.01	1990.12.31	2260.360	2215.434	1.988	0.901
1991.01.01	1991.12.31	2095.514	2082.390	0.626	0.862
1992.01.01	1992.12.31	2451.887	2389.625	2.539	0.919
1993.01.01	1993.12.31	3404.041	3340.351	1.871	0.908
1994.01.01	1994.12.31	2105.631	2070.130	1.686	0.894
1995.01.01	1995.12.31	3039.802	2980.708	1.944	0.885

Application of the Xinanjiang model

Hapuarachchi *et al.* (2001) calibrated all the parameters of the Xinanjiang model using the SCE-UA method (Duan *et al.*, 1992, 1993, 1994). The calibration and verification results of the Xinanjiang model are shown in Table 3. According to the results, %*Err* is considerably small ($-3 < \%Err < 4$) and the Nash coefficient D_y is grater than 84% for each year, in both the calibration and verification stages.

Application of the SWAT model

Firstly, the watershed was delineated into 15 subcatchments using a DEM (200 × 200 m^2 cell size) and a digitized river network. The number of subcatchments

Table 4 Calibration and validation results of the SWAT model.

From	To	Q_{obs} (mm)	Q_{cal} (mm)	*%Err*	D_y
1987.01.01	1987.12.31	1820.683	1483.143	18.539	0.720
1988.01.01	1988.12.31	3230.340	2552.167	20.994	0.703
1989.01.01	1989.12.31	3050.903	2748.810	9.902	0.762
1990.01.01	1990.12.31	2260.356	1882.359	16.723	0.774
1991.01.01	1991.12.31	2095.511	1774.940	15.298	0.759
1992.01.01	1992.12.31	2451.883	2197.655	10.369	0.782
1993.01.01	1993.12.31	3404.035	2409.872	29.205	0.763
1994.01.01	1994.12.31	2105.631	1470.104	30.182	0.715
1995.01.01	1995.12.31	3039.802	2374.725	21.879	0.705

generated depends on the threshold limit of flow accumulation. The SWAT was applied for the period 1987–1995 for daily streamflow generation. However, we are only interested in comparing results of two years, 1994 and 1995, that are common to the validation periods of ANNs and the Xinanjiang model. The application results are shown in Table 4. The *%Err* (*%Err* > 9) is considerably high for each year indicating the water imbalance of the catchment. Furthermore, the highest Nash coefficient that could obtain was 78% for the data year of 1992.

DISCUSSION AND CONCLUSIONS

The compiled results of performance of the three models in the validation stage, are shown in Table 5. A careful inspection revels that the results of ANNs and the Xinanjiang model are acceptable. Based on the results (Table 5), the performance of Xinanjiang model is better than ANN and the SWAT model for daily river flow predictions of the Kalu River upper catchment. Thereby it is possible to conclude that the Xinanjiang model can be successfully applied in humid or semi humid catchments in Sri Lanka for computing river flow. The input data requirement for the Xinanjiang model are only precipitation, pan evaporation, and observed discharge. Therefore it is suitable for modelling many catchments where sophisticated data are not readily

Table 5 Compiled results of different models.

Model		Xinanjiang	ANN		SWAT
Criteria	Year	Model	MLP	RBF	Model
Q_{obs}	1994	2105.631	2105.63	2105.63	2105.631
(mm)	1995	3039.802	3039.80	3039.80	3039.802
Q_{cal}	1994	2070.130	2167.240	2099.813	1470.104
(mm)	1995	2980.708	2979.118	2967.643	2374.725
%Err	1994	1.686	−2.926	0.276	30.182
	1995	1.944	1.996	2.374	21.879
D_y	1994	0.894	0.876	0.847	0.715
	1995	0.885	0.875	0.816	0.705

available. The main difficulty of using the Xinanjiang model is its calibration. The manual calibration requires detailed understanding of the model's behaviour, which can only be obtained through long-term calibration experience and data. Autocalibration is efficient, but it is difficult to incorporate the expert knowledge in the calibration process.

When considering the two ANNs, MLP and RBF, the performance of the MLP network is slightly better than the RBF network (Table 5). The *%Err* of the Xinanjiang model, MLP and RBF networks are acceptable (<5%) proving that the water utilization and losses balance in the catchment. However, the main disadvantage of using ANNs is the absence of criteria that can optimize the network parameters, dominant model inputs and, input data combinations for making the model output optimum. In this case study, we have carried out many trials (Table 2) for finding the optimum combination. It was observed that the performance of the networks in the validation period was reduced when only runoff data (R10) or more additional runoff and rainfall data are added (P9, R9, E9) to the input pattern (Table 2). However, implementing the ANN is easier and the input data requirement is small.

In this case study the performance of the SWAT model is poor compared to the other two models. Although the efficiency D_v is acceptable (Table 5), the *%Err* is high indicating that the model is unable to satisfy the water balance in the catchment. It may be partly attributable due to the poor quality of the data since the output of distributed models strictly depends on the quality of data. The main disadvantage of using such distributed models for river flow predictions is the higher requirement of data. In addition, many people in Sri Lanka use well water for their domestic use. When considering a catchment as a whole, it is normally a very large area, and therefore it is not possible to record or count all the individual minor scale water utilizations in detail such as small irrigation, animal husbandry in minor scale, industrial water utilizations in minor scale, etc. The cumulative value of such water utilizations might be large. The absence of these data may especially affect the distributed models in water balancing. But the conceptual watershed models and ANNs are capable of adjusting their parameters while calibrating or training, according to the situation, since most of their parameters have no physical background. As a result, a conceptual watershed model and ANNs show better performance than a distributed model where the catchment characteristics and model inputs are limited or incomplete. However, distributed watershed models are excellent for understanding the physical processes of the catchment.

Acknowledgements This project was funded by the International Water Management Institute (IWMI). The support received from the IWMI in data collection and preparation is gratefully acknowledged.

REFERENCES

Arnold, J. G., Srinivasan, R., Muttiah, R. S. & Williams, J. R. (1998) Large area hydrologic modelling and assessment part I: model development. *J. Am. Water Res. Assoc.* 34(1), 73–89.

Dibike, Y. B., Solomatine, D. P. & Abbott, M. B. (1999) On the encapsulation of numerical-hydraulic models in artificial neural network. *J. Hydraul. Res.* 37(2), 147–161

Duan, Q. Y., Gupta, V. K. & Sorooshian, S. (1993) Shuffle complex evolution approach for effective and efficient global minimization. *J. Opti. Theory Appl.* **76**(3), 501–521.

Duan, Q., Sorooshian, S. & Gupta, V. K. (1992) Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resour. Res.* **28**(4), 1015–1031.

Duan, Q., Sorooshian, S. & Gupta, V. K. (1994) Optimal use of SCE-UA global optimization method for calibrating watershed models. *J. Hydro.* **158**, 265–284.

Gan, T. Y. & Biftu, G. F. (1996) Automatic calibration of conceptual rainfall–runoff models: optimization algorithms, catchment conditions and model structure. *Water Resour. Res.* **32**, 3513–3524.

Hapuarachchi, H. A. P., Li, Z. J, Ranjit, M. & Wang, Q. J. (2001) Application of global optimization technique for calibrating the Xinanjiang watershed model. *Lowland Tech. Int.* **3**, 43–57

Karunanayake, M. M. & Katupotha, J. (1990) An environmental profile of the Ratnapura district. Tech. report, Central Environmental Authority & NORAD, Sri Lanka.

Mason, J. C., Price, R. K. & Tem'me, A. (1996) A neural network model of rainfall–runoff using radial basis functions. *J. Hydraul. Res.* **34**, 537–548

Neitsch, S. L, Arnold, J. G., Kiniry, J. R., Srinivasan, R. & Williams, J. R. (2000) *SWAT User Manual 2000*. Grassland, Soil and Water Research Laboratory, Agricultural Research Service, USA.

Zhao, R. J., Zhuang, Y. L., Fang, L. R., Liu, X. R. & Zhang, Q. S. (1980) The Xinanjiang model. In: *Hydrological Forecasting* (Proc. Oxford Symp., April, 1980), 351–356. IAHS Publ. no. 129.

Zhao, R. J. & Liu, X. R. (1995) *Computer Models of Watershed Hydrology: The Xinanjiang Model* (ed. by V. P. Singh), 215–232. Water Resources Publications, Highlands Ranch, Colorado, USA.