

Multi-step-ahead neural networks for flood forecasting

FI-JOHN CHANG¹, YEN-MING CHIANG¹ & LI-CHIU CHANG²

¹ Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, Taiwan
changfj@ntu.edu.tw

² Department of Water Resources & Environmental Engineering, TamKang University, Taipei, Taiwan

Abstract A reliable flood warning system depends on efficient and accurate forecasting technology. A systematic investigation of three common types of artificial neural networks (ANNs) for multi-step-ahead (MSA) flood forecasting is presented. The operating mechanisms and principles of the three types of MSA neural networks are explored: multi-input multi-output (MIMO), multi-input single-output (MISO) and serial-propagated structure. The most commonly used multi-layer feed-forward networks with conjugate gradient algorithm are adopted for application. Rainfall–runoff data sets from two watersheds in Taiwan are used separately to investigate the effectiveness and stability of the neural networks for MSA flood forecasting. The results indicate consistently that, even though the MIMO is the most common architecture presented in ANNs, it is less accurate because its multi-objectives (predicted many time steps) must be optimized simultaneously. Both MISO and serial-propagated neural networks are capable of performing accurate short-term (one- or two-step-ahead) forecasting. For long-term (more than two steps) forecasts, only the serial-propagated neural network could provide satisfactory results in both watersheds. The results suggest that the serial-propagated structure can help in improving the accuracy of MSA flood forecasts.

Key words neural networks; multi-step-ahead; flood forecasting; serial-propagated structure; Taiwan

Réseaux de neurones à échéances multiples pour la prévision de crue

Résumé Un système d'annonce de crue performant s'appuie sur une technologie de prévision efficace et précise. Nous présentons une analyse systématique de trois types communs de réseaux de neurones artificiels (RNAs) pour la prévision de crue à échéances multiples. Les mécanismes et principes opératoires des trois types de réseaux de neurones à échéances multiples sont explorés: multi-entrées multi-sorties (MEMS), multi-entrées sortie unique (MESU) et structure de propagation sérielle. Les réseaux multi-couches progressifs les plus utilisés, avec un algorithme de gradient conjugué, ont été adoptés pour l'application. Les jeux de données pluie-débit de deux bassins versants Taiwanais sont utilisés séparément pour étudier l'efficacité et la stabilité des réseaux de neurones pour la prévision de crues à échéances multiples. Les résultats indiquent que, même si l'architecture MEMS est la plus couramment utilisée pour les RNAs, elle est moins précise parce que ses multiples objectifs (la prévision à échéances multiples) doivent être optimisés simultanément. Les réseaux MESU et à propagation sérielle sont autant capables de produire des prévisions précises à court terme (à échéances de un ou deux pas de temps). Pour les prévisions à long terme (plus de deux pas de temps), seul le réseau à propagation sérielle peut produire des résultats satisfaisants dans les deux bassins versants. Les résultats suggèrent que la structure de propagation sérielle peut aider à améliorer la précision des prévisions de crue à échéances multiples.

Mots clefs réseaux de neurones; multi-échéances; prévision de crue; structure de propagation sérielle; Taïwan

INTRODUCTION

Streamflow forecasting during thunderstorm and/or typhoon periods is extremely important for efficient water resource management and effective flood warning systems. The further ahead that the forecast can be made, the more beneficial it is in terms of allowing time for adjustment and reduction of damage due to flooding. With the high mountains and steep slopes of Taiwan, heavy rainfall, especially during a typhoon event, can cause downstream flooding within a few hours. Nevertheless, such typhoon events are the most valuable yearly water resource. An accurate prediction of upstream flow rates is crucial for both flood control and optimizing the management of water resources. Due to spatial and temporal variations in rainfall distribution and the inordinately complex and highly nonlinear nature of the rainfall–runoff relationship, flood forecasting remains one of the most challenging and important tasks of operational hydrology.

Flood forecasting represents not only a very complex nonlinear problem, but also one that is extremely difficult to model. For decades, a great deal of research has been carried out in describing the rainfall–runoff processes based on either deterministic/conceptual models (e.g. HEC, 1990), or stochastic models (e.g. Salas *et al.*, 1985). Although conceptual models are important in understanding hydrological processes and are reliable in forecasting the most important features of the hydrograph, the implementation and calibration of such a model can typically present various difficulties, requiring sophisticated mathematical tools, a large number of parameters for modelling and some degree of experience with the model (Duan *et al.*, 1992; Grayson *et al.*, 1992). The problem with conceptual models is that empirical regularities are not always evident and the corresponding mathematical relationships are not easily found. A stochastically-based approach to flood forecasting has been undertaken traditionally by using a linear technique, such as autoregressive (AR), autoregressive moving average with exogenous inputs (ARMAX) (Salas *et al.*, 1985; Mishra *et al.*, 2004), and also nonlinear regression (Chang & Hwang, 1999). The applicability of these models may be expected to be limited to basins for which streamflow has been measured for long periods and where no significant change in watershed conditions has occurred (Baratti *et al.*, 2003).

Over the last decade, the artificial neural network (ANN) has evolved as a branch of artificial intelligence that is now a recognized tool for modelling the underlying complexities in any artificial or physical system. An ANN is described as an information process system which consists of many nonlinear and densely interconnected processing units. With this parallel-distributed processing architecture, ANNs have been proven to be an efficient alternative to traditional methods for hydrological modelling, such as rainfall forecasting (Hsu *et al.*, 1999; Luk *et al.*, 2000; Toth *et al.*, 2000), streamflow forecasting (Dawson & Wilby, 1998; Atiya *et al.*, 1999; Sajikumar & Thandaveswara, 1999; Zealand *et al.*, 1999; Chang & Chen, 2001, 2003; Coulibaly *et al.*, 2001; Sivakumar *et al.*, 2002; Cigizoglu, 2003; Hu *et al.*, 2005) and groundwater modelling (Lallahem & Mania, 2003). A comprehensive review of the application of ANNs to hydrology can be found in the ASCE Task Committee report (2000a,b) and in a specialized publication (Govindaraju & Rao, 2000).

In this study, a systematic investigation of three common strategies used in ANNs for multi-step-ahead (MSA) forecasting was carried out. In the remainder of this paper, first, the MSA forecasting methodologies are presented, including an overview, the conjugate gradient algorithm for the ANN and the architectures of the MSA. In the next section, the details of the Da-Chia and Kee-Lung river basins are presented and neural networks with different architectures are applied to streamflow data of these basins. The results are discussed in the subsequent section and, finally, conclusions are drawn.

THE MULTI-STEP-AHEAD FORECASTING METHODOLOGIES

Overview

Multi-step-ahead prediction using a neural network is a challenging and complex task which tries to make predictions several time steps into the future without the output measurements. In the past several years, this work has attracted an increasing number

of scientific studies, such as weather forecasting, electric load forecasting, reservoir operation and flood warning systems. It is undoubtedly very beneficial and desirable that the models provide accurate MSA predictions. However, MSA predictors have been difficult to develop, because closely related input information could not be determined and observations in the prediction horizon for re-adjusting the model performance could not be obtained in time. A small single-step prediction error at the beginning could accumulate and propagate to the future, often resulting in poor prediction accuracy.

In many engineering problems, one-step-ahead prediction using neural networks has been performed and reported with satisfactory results. However, one-step-ahead prediction may not provide enough information, especially in the situation in which it is desirable to understand the behaviour of multiple steps in the future, such as signal processing and time series prediction. Some earlier studies have used neural networks successfully for MSA prediction: one of the earliest researches, reported by Su *et al.* (1992), was based on the back-propagation through time for long-term predictions of chemical processes; Schenker & Agarwal (1995) reported long-range prediction for poorly known systems through training two distinct networks; a neural network-based long-term hydropower forecasting system was proposed by Coulibaly *et al.* (2000); Parlos *et al.* (2000) developed dynamic recurrent neural networks for multi-step prediction in complex processes; Gao *et al.* (2002) proposed temporal difference method-based multi-step-ahead prediction of deep fading in mobile networks; and Chang *et al.* (2004) developed a two-step-ahead recurrent neural network for streamflow forecasting.

Architecture of neural network-based multi-step-ahead forecasting

It is noted that most neural network approaches to the problem of time series forecasting use the standard multi-layer perceptron trained with the back-propagation (BP) algorithm. Even though the well-known steepest descent method is a widely used training algorithm, it often yields sub-optimal solutions and its slow convergence characteristic has encouraged research into faster algorithms. The conjugate gradient (CG) algorithm has now become more popular and widely used, since it represents a good compromise between simplicity of the steepest descent algorithm and the fast quadratic convergence of Newton's method (Battit, 1991). Therefore, in this study, this algorithm is implemented for training the MSA neural networks. The CG method is a numerical technique for solving various optimization problems. It differs from the standard BP algorithm in updating directions of weights. The CG algorithm searches a system of conjugate directions on the error surface and updates the weights along these directions. In practice, the process makes good uniform progress toward the solution at every time step and has been found to be effective in finding better optimization than the standard BP algorithm (Chiang *et al.*, 2004). The CG algorithm can be found in Ham & Kostanic (2001) and can be easily implemented by using mathematical software (e.g. MATLAB).

Two common strategies with three popularly used neural network architectures for multi-step-ahead forecasting are investigated. The first relies on the model capability to provide direct MSA forecasts, whereas the second uses a serial-propagated method. In this study, the direct method is further divided into two types of structure, namely

multi-input multi-output (MIMO) and multi-input single-output (MISO), to provide a comparison with the serial-propagated structure. We assume a time series $y(1), \dots, y(t)$ to forecast $\hat{y}(t+1)$ to $\hat{y}(t+n)$, where $n > 1$. The architectures, their advantages and disadvantages are described below.

MIMO Multi-input multi-output is the most common and basic neural network architecture for MSA forecasting (Campolo *et al.*, 2003; Coulibaly, 2003). The neurons of the input and output layers will be t and n , respectively. Because the model objective function is often defined as a weighted sum squared error of multi-output, the model can produce only a compromising optimization of all outputs and cannot guarantee to have minimum error for each forecasting step. As stated by Huang & Lian (2000), the difficulty of MIMO systems is to overcome the coupling effects between the degrees of freedom. Furthermore, MIMO architectures take a much greater number of initial settings to reduce the chance of sub-optimal solutions and need considerable training time (Selvaraj *et al.*, 1995), because the weights to be updated for appropriate prediction are $(t + n) \times k$ (where k represents the number of hidden neurons). Generally speaking, the accuracy of MIMO architectures is relatively low, because of the great and complex interactions of the weights for different outputs. Owing to the multi-objective solutions, the searching process in the error domain is more complicated than that for a single objective.

MISO In this model, n independent networks are constructed to forecast $\hat{y}(t+1)$ to $\hat{y}(t+n)$, respectively. The MISO architecture for each desired step-ahead forecast must be determined and trained. Even though n networks for n -step-ahead forecasting would need to be individually constructed, compared with MIMO, MISO networks would efficiently reduce the training time and increase the accuracy. Moreover, a standard search algorithm, such as BP or CG, can be easily implemented into all the n constructed networks, and a parallel computation algorithm (Chang & Chang, 2002) for simultaneously training these networks is also quite straightforward. The major reason for the accuracy of the MISO architecture being better than MIMO may be due to the number of parameters. For an individual MISO network, the number of weights between hidden and output layers is much less than that of MIMO, therefore reducing the interaction and complexity. The drawback of the MISO architecture is that the network mapping capability gradually weakens as the predictive time steps increase.

Serial-propagated Basically, the concept of a serial-propagated structure is to combine n identical structures of a network, which are all one-step-ahead predictors, to produce a series of predictions. Figure 1 shows the architecture of such a network. The serial-propagated structure not only has the advantages of the MISO network, but also it uses previous predictions of the network as additional input information for further predictions into the future. The main concern of this serial-propagated structure is whether it can produce a sequence of successful predictions to prevent error or noise from propagating. Consequently, we should design a detecting threshold for the predictions at each step to avoid the problem of poor prediction accuracy that can occur when a small prediction error at the beginning accumulates and propagates to the future. To the best of our knowledge, threshold selection in a serial-propagated structure remains problem-dependent and subjective.

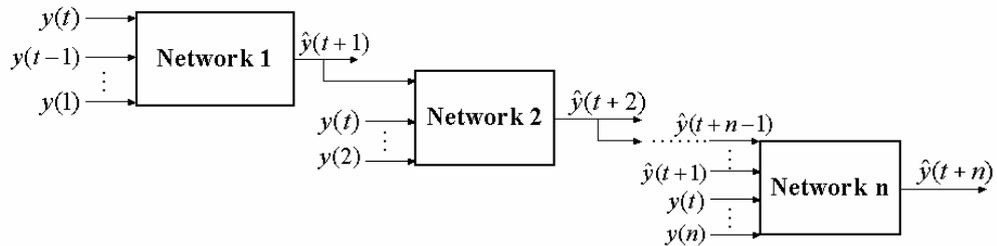


Fig. 1 The architecture of a serial-propagated structure.

APPLICATION

Study area

The main reason for using two case studies is to investigate the generalization capability and stability of the constructed models in different sizes of watersheds. The models are applied to the upstream part of the Da-Chia and Kee-Lung river basins for MSA flood forecasting. The Da-Chia River basin is located in central Taiwan and has a total drainage area of approximately 1236 km², is about 140 km long, and has an average channel slope of 1/39 (which is the steepest channel in Taiwan). A series of hydraulic structures has been constructed for power generation and water supply for the Tai-Chung metropolitan area. The Kee-Lung River basin is located in northern Taiwan surrounding Taipei City and has a drainage area of approximately 204 km². The metropolitan area of Taipei (population 6.5×10^6), is located on the Tanshui River and its tributaries in northern Taiwan. The Kee-Lung River merges into the downstream section of the Tanshui River and is the major tributary of the Tanshui River. The two study basins are shown in Fig. 2, where the S1 represents the streamflow gauging station and R1, R2 and R3 represent the three rainfall gauging stations in both watersheds. The Song-Mao streamflow gauging station (S1, Fig. 2(a)) was established

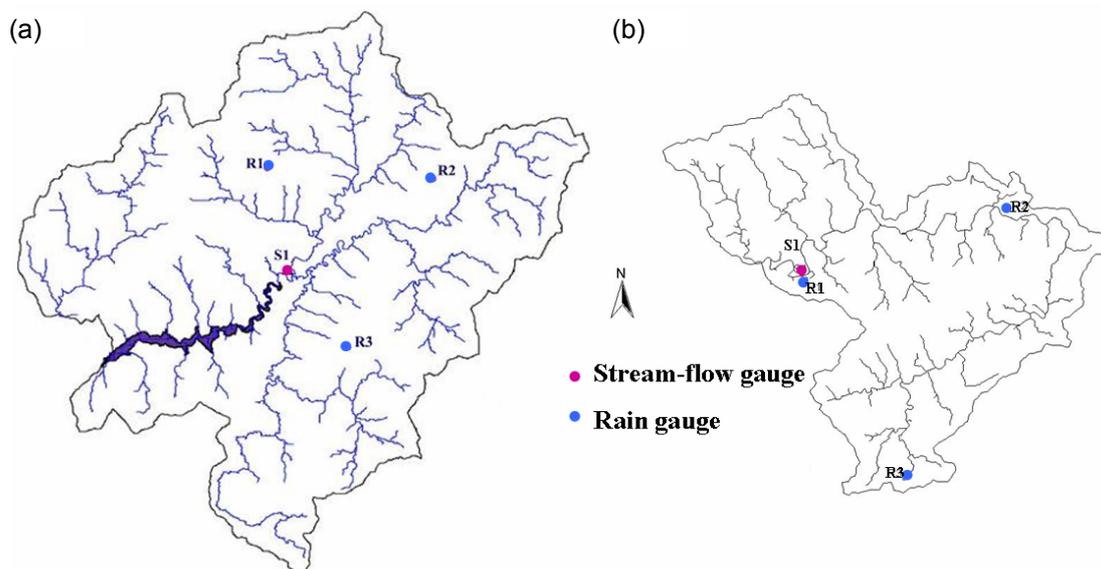


Fig. 2 Location of gauging stations in (a) Da-Chia River basin and (b) Kee-Lung River basin.

to measure the inflow of the De-Chi Reservoir, which is the pivotal reservoir in the Da-Chia River. Flash floods in these two basins often occur suddenly (in less than four hours) and can cause serious damage. For instance, on 17 September 2001, Typhoon Nari struck northern Taiwan and brought record-breaking rain to flood Taipei. The flooding seriously damaged the Taipei MRT (Mass Rapid Transit system): a total of 16 metro stations, the Nangang depot, the heavy-capacity system control centre and administration building were completely submerged. The whole MRT system was shut down immediately and was only partly operational again three months later, during which time the city experienced traffic chaos and suffered tremendous losses.

Description of data and ANN model set-up

Clearly, accurate MSA flood forecasting is extremely important to provide time for reservoir operation (De-Chi Reservoir) and for rescuing human life and property. The observed streamflow (m^3/s) and precipitation (mm/h) data used here are gathered from the Water Resources Agency, Taiwan. The hourly precipitation and streamflow data were collected from 1988 to 2000, and consist of 13 and 17 storm events in Da-Chia and Kee-Lung basins, respectively. For the most common strategy used in building the neural network, the data sets should be divided into three independent sub-sets, for training, validation and testing. The data of training sub-sets are repeatedly used to adjust the connected weights of the constructed networks; the data of validation sub-sets are implemented to check the suitability of the above trained networks and the best network is selected for later use. The testing data set is then used for final evaluation of the selected network performance. In this study, we arbitrarily divided the storm event data sets for both watersheds into training, validation and testing sub-sets, consisting of seven, three and three events and 11, three and three events in the Da-Chia and Kee-Lung data sets, respectively.

The three-layer feed-forward neural network with one hidden layer is used in this study because many previous works (e.g. Zealand, *et al.*, 1999) have shown that it should be better to increase the number of hidden neurons rather than to add another layer to the network. A trial-and-error procedure is used for identifying the number of neurons in the hidden layer, since no general guidelines exist for the identification of ANN structure for specific applications. The hidden neurons were varied from 1 to 12. This is because too many hidden neurons may reduce the generalization capabilities and increase the training time, but without significant improvement (Ranjithan *et al.*, 1993); also Minns & Hall (1996) indicated that the appropriate number of hidden neurons is roughly half of the number of input dimensions. Therefore, we believe that hidden neurons within the range (1, 12) are enough for the models, since the maximum designed input dimensions are 12. For the selection of activation function, a sigmoid function and a linear function are adopted in the hidden and output layers, respectively. Furthermore, the procedure of data pre-processing is also important for the model performance. In this study, the streamflow data are normalized within the range of (0.1, 0.8). This range would allow model prediction of streamflow higher than values used in the training phase. Prior to training, the connected weights and the biases in the model are initialized with random values in the range (-1, 1) and the stop criterion is dependent on either minimum mean squared error obtained, or training iterations reached.

To construct MSA flood forecasting for a particular gauging station, its streamflow data and the rainfall information for three nearby gauging stations were selected as the input vector. Because the drainage area of the study basin is small and the basin slope is quite steep, the time of concentration is short—no more than three hours (as mentioned in Chang & Chen, 2001). Therefore, the input information for a desired output should be not earlier than three hours. For example, if the output is streamflow at time $t + 2$, the time delay of input pattern should be not earlier than $t - 1$. Consequently, the input pattern is focused on the current, previous one-hour and previous two-hour information only, and the desired outputs are focused on one-step-ahead to four-step-ahead predictions. Therefore, the input and output dimensions of the MIMO network would be 12 and 4, respectively.

In contrast to the MIMO network, we constructed two types of MISO networks to forecast the next one to three steps. The impact of lag time from the upstream rainfall

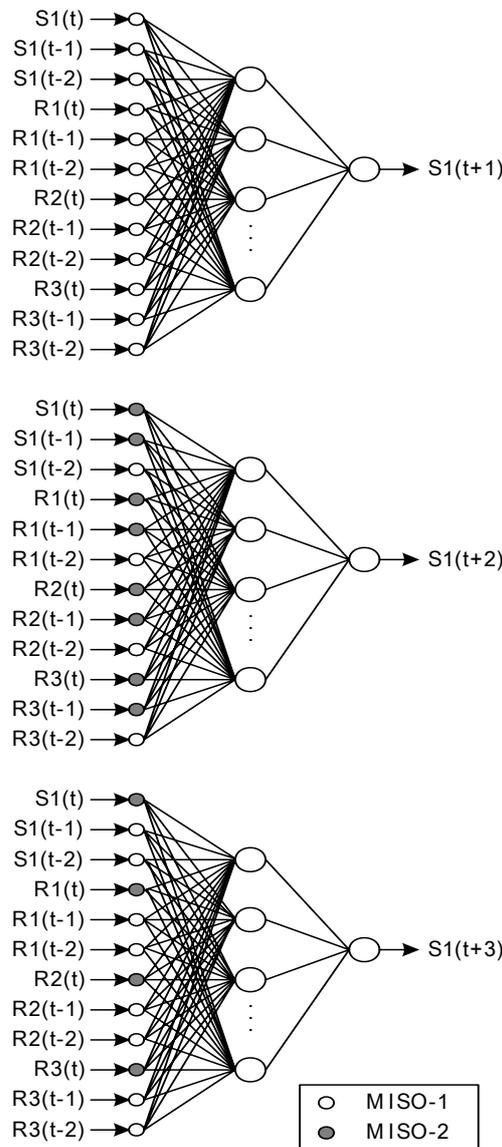


Fig. 3 Architecture of the MISO network for multi-step-ahead streamflow forecasting.

gauging stations to the downstream streamflow gauging station is explored in these two MISO networks. The architecture of the MISO networks is illustrated in Fig. 3. It appears that there is only one desired output for each network, while different input dimensions are designed. For the first MISO networks (MISO-1), the input nodes are fixed for one- to three-step-ahead forecasts. For the second MISO networks (MISO-2), the input dimensions for one-step-ahead forecast are 12, which are same as those of MISO-1. Nevertheless, the input dimensions of MISO-2 were reduced to 8 and 4 for two- and three-step-ahead forecasts, respectively. The MISO-1 networks include redundant information ($S_1(t-2)$, $R_1(t-2)$, $R_2(t-2)$ and $R_3(t-2)$ for two-step-ahead forecasting and $S_1(t-2)$, $S_1(t-1)$, $R_1(t-2)$, $R_1(t-1)$, $R_2(t-2)$, $R_2(t-1)$, $R_3(t-2)$, and $R_3(t-1)$ for three-step-ahead forecasting) as inputs, which might be noise or large bias, while the MISO-2 networks only involve contracted inputs (8 and 4 inputs for two- and three-step-ahead forecasting, respectively). Finally, the serial-propagated neural network forecast is similar to that of the MISO-2 network, except for the additional inputs which are the forecasting values in the previous steps. Figure 4 shows the architecture of the serial-propagated neural network, which involves the effect of lag time of precipitation.

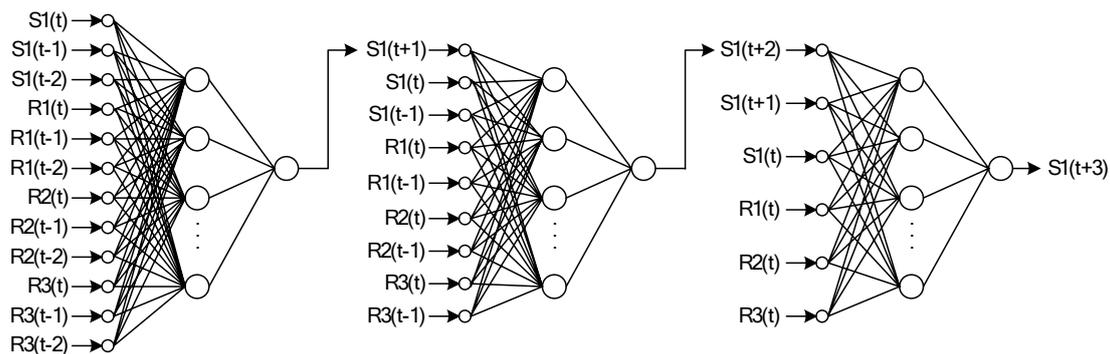


Fig. 4 Neural network with serial-propagated structure for multi-step-ahead streamflow forecasting.

As mentioned above, this study provides an idea of using a threshold to check the model predictions and to prevent error propagation in the serial-propagated structure. As mentioned by Sivakumar (2005), the use of a threshold provides the opportunity of sharpening the model's ability for data learning, parameter optimization and future extrapolation, which could emerge as a new research and teaching philosophy in hydrology. However, how to select the threshold appropriately is still highly problem-dependent and no definitive methodology exists to set up the value. Therefore, the threshold used in this case study is the goodness-of-fit with respect to the benchmark G_{bench} (Nash & Sutcliffe, 1970; Seibert, 2001). If the G_{bench} value of model predictions is higher than 0.5, then the predictions represent a successful series and could be the additional input for the next network. In order to measure the ability and to compare the prediction accuracy of the MIMO, MISO and serial-propagated structures, the performance of these three MSA forecasting methods is evaluated based on the criteria of root mean squared error (RMSE) and G_{bench} , defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^N (\mathcal{Q}_{\text{sim}}(t) - \mathcal{Q}_{\text{obs}}(t))^2}{N}} \quad (1)$$

$$G_{\text{bench}} = 1 - \frac{\sum_{i=1}^n (\mathcal{Q}_{\text{obs}}(t) - \mathcal{Q}_{\text{sim}}(t))^2}{\sum_{i=1}^n (\mathcal{Q}_{\text{obs}}(t) - \tilde{\mathcal{Q}}_{\text{obs}}(t))^2} \quad (2)$$

where \mathcal{Q}_{sim} is the forecast value, \mathcal{Q}_{obs} is the observed value, and $\tilde{\mathcal{Q}}_{\text{obs}}$ is a previously observed value, i.e. $\tilde{\mathcal{Q}}_{\text{obs}}(t) = \mathcal{Q}_{\text{obs}}(t - n)$ for n -step-ahead prediction.

RESULTS AND DISCUSSION

Experiment I

Table 1 compares the results of the MIMO, MISO and serial-propagated structures for MSA flood forecasting in Da-Chia River basin. For one-hour-ahead forecasting, owing to the identical structure of MISO and serial-propagated structures, their performances are identical. Figure 5(a) shows a comparison of observed vs forecast hourly streamflow for a flood event of the testing period, so one can easily differentiate the variations in model performance. It appears that the MIMO structure can only present acceptable one-step-ahead forecasting. In contrast, the MISO and serial-propagated structures can appropriately predict the trend of observations and produce a quite satisfactory result. One can also readily see that the MISO (also serial-propagated) network performed better in terms of small RMSE and high G_{bench} , compared to the MIMO network in all three stages (sub-sets).

Table 1 Comparison of different network architectures for multi-step-ahead forecasting in Da-Chia River basin (value of RMSE is in m^3/s).

	Architecture	Training: RMSE	G_{bench}	Validation: RMSE	G_{bench}	Testing: RMSE	G_{bench}
<i>One-hour-ahead:</i>							
MIMO	12-6-4	19.2	0.72	16.6	0.46	21.1	0.41
MISO	12-6-1	17.4	0.77	13.2	0.66	18.4	0.55
Serial-propagated	12-6-1	17.4	0.77	13.2	0.66	18.4	0.55
<i>Two-hour-ahead:</i>							
MIMO	12-6-4	28.6	0.81	29.9	0.50	38.8	0.38
MISO-1	12-6-1	28.5	0.81	29.4	0.52	38.5	0.40
MISO-2	8-6-1	25.7	0.85	23.5	0.69	34.9	0.50
Serial-propagated	9-6-1	29.2	0.80	25.8	0.63	33.0	0.56
<i>Three-hour-ahead:</i>							
MIMO	12-6-4	34.5	0.86	44.6	0.46	58.7	0.28
MISO-1	12-8-1	35.5	0.85	44.7	0.45	54.6	0.37
MISO-2	4-8-1	36.6	0.84	45.8	0.43	53.6	0.40
Serial-propagated	6-6-1	39.6	0.81	36.5	0.63	46.4	0.55

MISO-1: model with all inputs; MISO-2: model with contracted input.

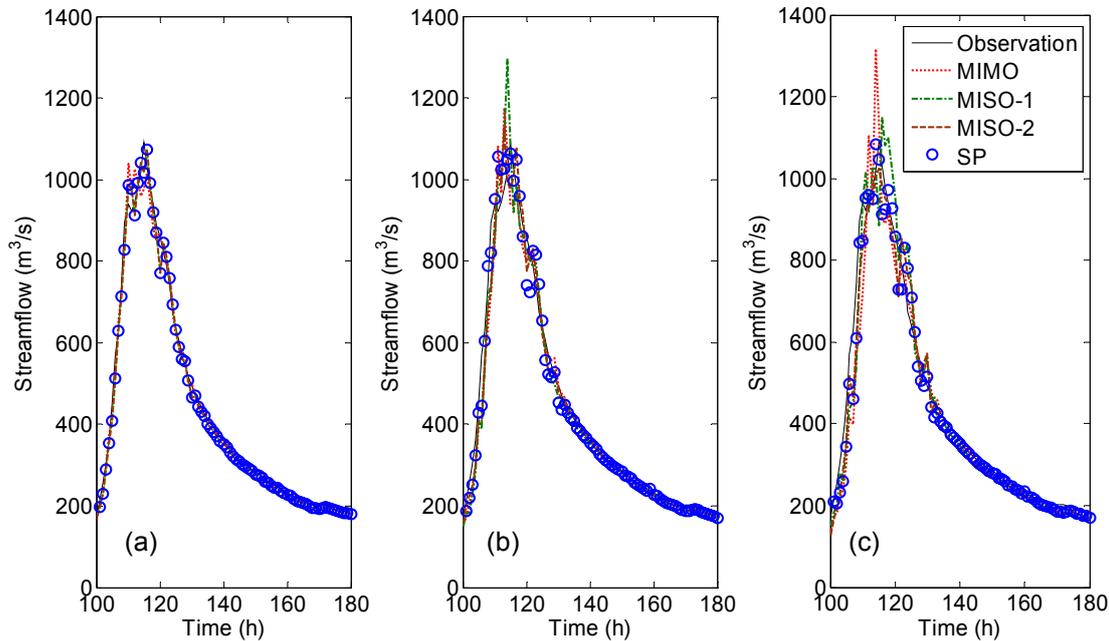


Fig. 5 Observed vs forecast streamflow for a flood event of the testing period in Da-Chia River basin: (a) one-hour-ahead (b) two-hour-ahead (c) three-hour-ahead.

The results of all methods for two- and three-hour-ahead flood forecasting are also presented in Table 1. Figure 5(b) and (c) shows observations vs predictions of MIMO, MISO and serial-propagated structures in the testing phase for two- and three-hour-ahead forecasting, respectively. It appears that the serial-propagated network performed much better than the MIMO and MISO networks. Another important aspect of the results of MISO-1 and MISO-2 networks (Table 1) is that the latter performed better than the former even though it has fewer inputs. This is mainly because the time of concentration of precipitation in the study basin is quite short; therefore, the rainfall information before two hours seems to be redundant input (a kind of noise). These redundant inputs might cause the over-fitting of the weights when training a neural network. Therefore, the redundant and/or noise inputs could result in poor performance in the validation and testing phases.

Furthermore, the MISO network performed slightly better than the serial-propagated network in the training phase of two- and three-hour-ahead forecasting; however, the serial-propagated neural network performed much better than the MISO network in terms of smaller RMSE and higher G_{bench} in the testing phases. The testing phase results demonstrate that the serial-propagated structure is relatively stable and performs consistently well in MSA forecasting ($G_{\text{bench}} > 0.5$ in all testing phases), whereas the accuracy of the MISO network gradually decreased as the predictive time steps increased. The forecasting skill of the serial-propagated structure is reduced as the lead time increases, but the decreasing rate is much lower than MIMO and MISO networks.

In addition to three prediction steps, we also produced four-step-ahead forecasts using the serial-propagated and MIMO networks. The main motivation was to test the stability of the serial-propagated structure. Because of the reducing inputs strategy used in the serial-propagated structure, the input the four-step-ahead forecast will have three dimensions and will consist only of the three forecast values in previous steps.

Table 2 Comparison of MIMO and serial-propagated structures for four-hour-ahead forecasting in Da-Chia River basin (value of RMSE is in m^3/s).

	Architecture	Training:		Validation:		Testing:	
		RMSE	G_{bench}	RMSE	G_{bench}	RMSE	G_{bench}
MIMO	12-6-4	45.0	0.84	60.9	0.36	77.2	0.23
Serial-propagated	3-6-1	51.4	0.79	49.9	0.57	66.7	0.43

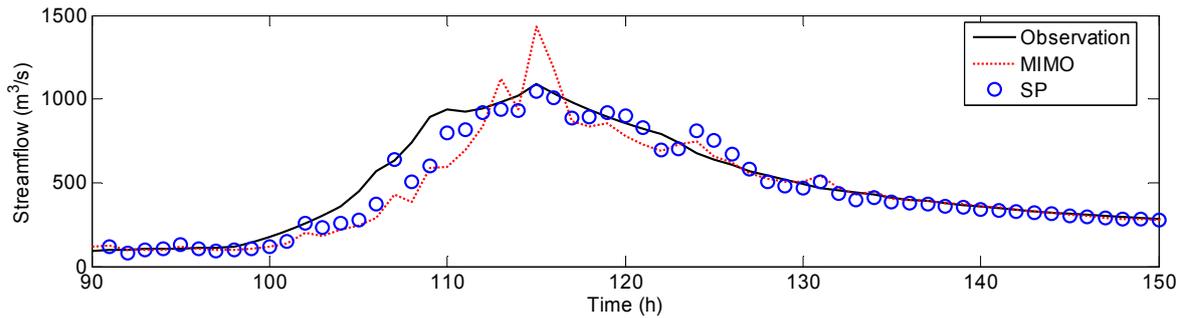


Fig. 6 Observed vs four-hour-ahead forecast streamflow for a flood event of the testing period in Da-Chia River basin.

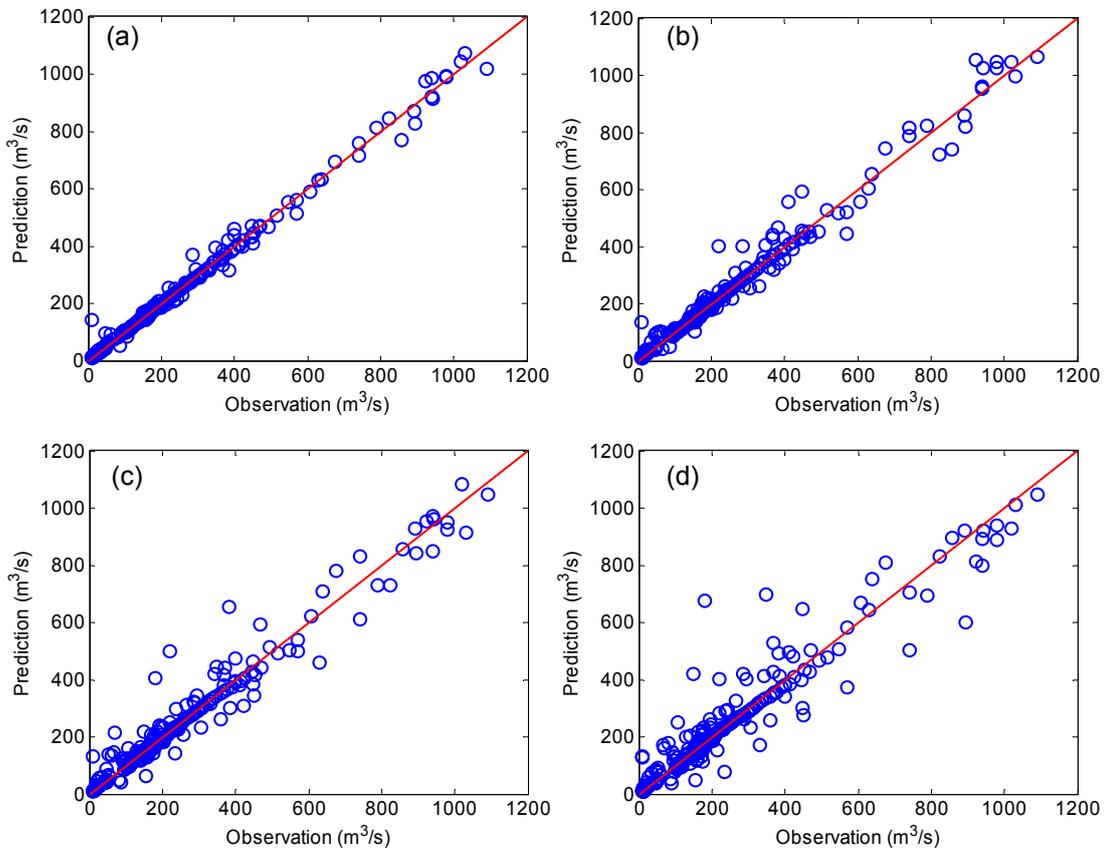


Fig. 7 Comparison of observed and forecast streamflow by the serial-propagated structure in Da-Chia River basin: (a) one-hour-ahead; (b) two-hour-ahead; (c) three-hour-ahead; and (d) four-hour-ahead.

The results of both methods for four-hour-ahead forecasting are presented in Table 2. Figure 6 shows the observed *vs* forecast streamflow by the MIMO and serial-propagated structures for a flood event in the testing period. As expected, the MIMO network produced better results than the serial-propagated structure in the training phase, whereas its predicting skill in the validation and testing phases dramatically declined and was worse than that of the serial-propagated structure. Again, the results indicate that the serial-propagated structure could produce more reliable prediction than the MIMO, even though the model consists only of previously predicted information and is without real observational data. Finally, Fig. 7 depicts scatter plots of the serial-propagated structure for four time steps of predictions in the testing phase. In this figure, all the predictions are close to the ideal line, especially for high streamflow.

Experiment II

A second study area was used to verify whether the proposed models can provide accurate and stable hydrological responses in different sizes of watersheds. Table 3 provides the performance statistics for one- to four-hour-ahead forecasts for all models. As in Experiment I, the MIMO network produced the worst forecasting, the results of MISO-2 are better than those of MISO-1, and the serial-propagated structure performed best in the testing phase of all forecasting time steps. Figure 8 shows observations *vs* predictive error of all the models for a flood event in the testing phase for forecasting time steps from one to four hours. It clearly shows that the serial-propagated structure can capture the main trend of the observed values, even in the four-hour-ahead forecasting; however, the other three networks seem to fail to catch the trend and have oscillations in the vicinity of peak flow for lead times longer than

Table 3 Comparison of different network architectures for multi-step-ahead forecasting in Kee-Lung River basin (value of RMSE is in m^3/s).

	Architecture	Training:		Validation:		Testing:	
		RMSE	G_{bench}	RMSE	G_{bench}	RMSE	G_{bench}
<i>One-hour-ahead:</i>							
MIMO	12-7-4	24.1	0.618	30.4	0.570	33.4	0.450
MISO	12-5-1	20.2	0.741	24.2	0.700	23.4	0.748
Serial-propagated	12-5-1	20.2	0.741	24.2	0.700	23.4	0.748
<i>Two-hour-ahead:</i>							
MIMO	12-7-4	34.1	0.762	52.7	0.630	66.5	0.371
MISO-1	12-6-1	31.5	0.795	49.7	0.654	57.1	0.599
MISO-2	8-6-1	35.9	0.735	44.4	0.725	47.8	0.718
Serial-propagated	9-5-1	32.4	0.783	42.5	0.734	41.4	0.780
<i>Three-hour-ahead:</i>							
MIMO	12-7-4	41.6	0.817	74.2	0.629	96.3	0.321
MISO-1	12-8-1	42.9	0.803	67.0	0.683	89.3	0.518
MISO-2	4-4-1	57.7	0.646	72.6	0.630	72.9	0.679
Serial-propagated	6-5-1	43.9	0.795	66.0	0.693	65.2	0.743
<i>Four-hour-ahead:</i>							
MIMO	12-7-4	52.0	0.817	88.2	0.662	112.4	0.390
Serial-propagated	3-5-1	58.2	0.770	85.2	0.693	83.3	0.720

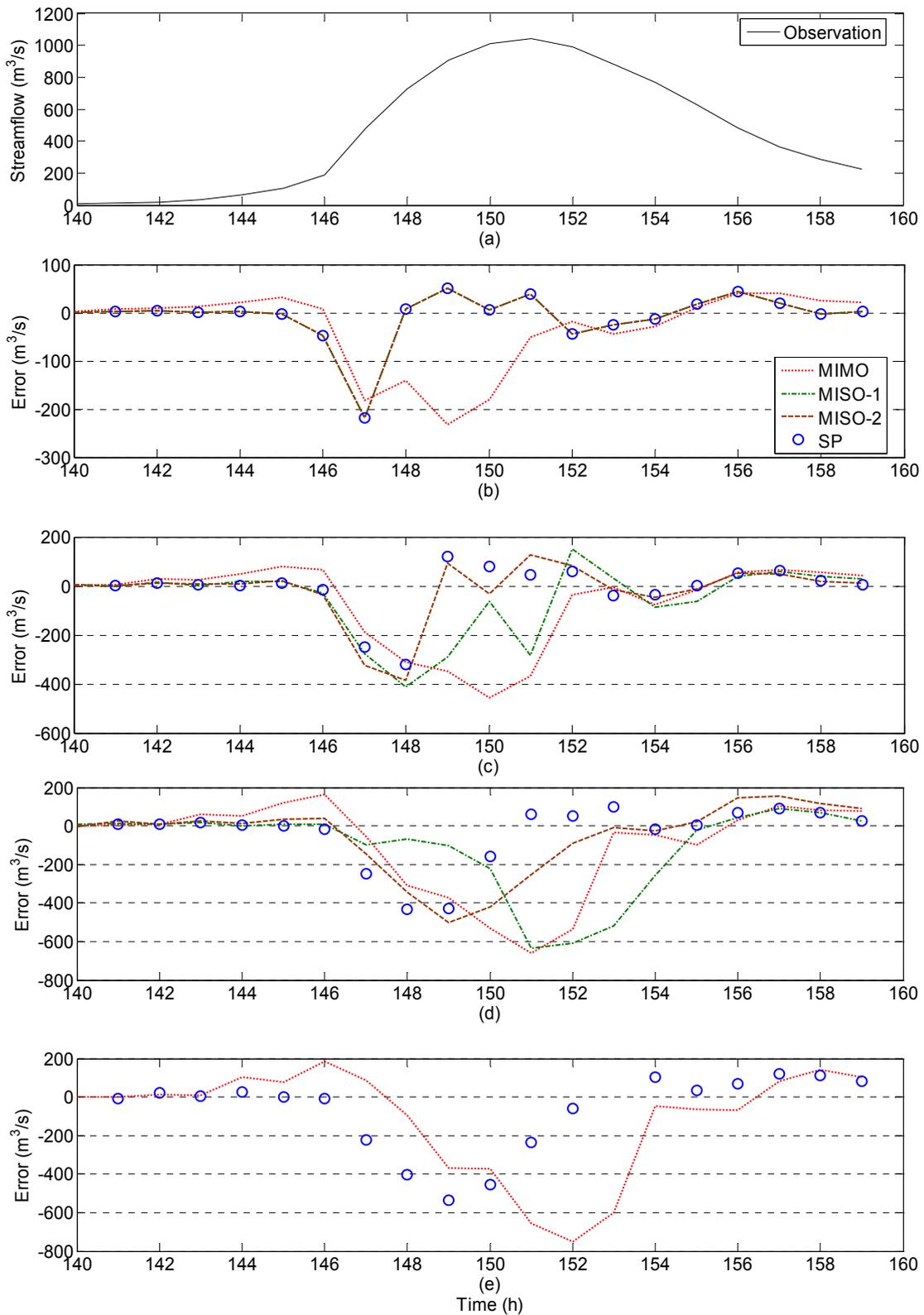


Fig. 8 Observed streamflow vs error time series for a flood event of the testing period in Kee-Lung River basin: (a) observations; error (forecast minus observed value) from all models in: (b) one-hour-ahead; (c) two-hour-ahead; (d) three-hour-ahead; and (e) four-hour-ahead.

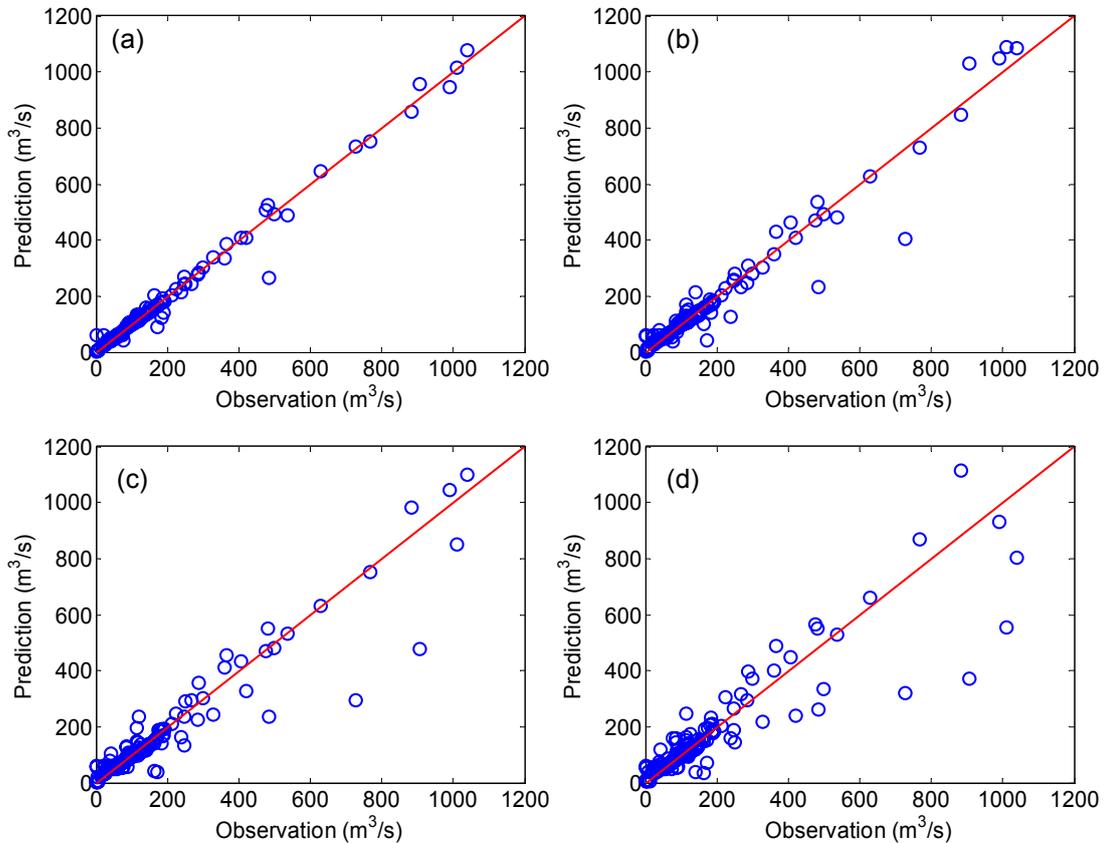


Fig. 9 Comparison of observed and forecast streamflow by serial-propagated structure in Kee-Lung River basin: (a) one-hour-ahead; (b) two-hour-ahead; (c) three-hour-ahead; and (d) four-hour-ahead.

two hours. Figure 9 depicts the scatter plots of the serial-propagated structure forecasts for four lead times in the testing phase. It should be noted that all the predictions are close to the observations, except for the lead time of four hours; the predictions during high flow are different from the observations.

Overall, the serial-propagated structure performed better for all validation (except for two-hour-ahead forecasting in Da-Chia basin) and testing sets in both watersheds. Table 1 clearly demonstrates that the serial-propagated structure was more stable (consistent) than other structures as the model performance was consistently better for all three phases. For example, for three-hour-ahead forecasting, the error shows an increase of $6.8 \text{ m}^3/\text{s}$ from training (39.6) to testing (46.4) in terms of RMSE for the serial-propagated structure, which is much better (less) than those of the MISO-2 ($17 \text{ m}^3/\text{s}$), MISO-1 ($19.1 \text{ m}^3/\text{s}$) and MIMO ($24.2 \text{ m}^3/\text{s}$) structures.

Discussion

Firstly, we would like to clarify that the optimal structure of a neural network is highly problem-dependent, since it could be dominated by many factors, such as number of inputs, hidden and output neurons and learning rates. Secondly, it is common to see that neural network models show almost perfect fitting in the training phase, but their

performance rapidly decreases in both validation and testing phases. This is mainly because the structure of the ANN is very flexible, and it is not difficult to find a number of suitable structures that could well fit the input–output patterns of training data. Unfortunately, these trained networks are not necessarily all suitable for the validation and/or testing set, especially in the complex cases that occur in nature. Consequently, the model with the greatest reliability should be the one we expect to be more reliable for future use.

Finally, it appears that the results obtained from both watersheds are consistent and demonstrate that the serial-propagated structure is better than the direct search neural networks (MIMO and MISO). These results are similar to those of Coulibaly (2003), who investigated the direct and serial-propagated structures for a daily spring low-flow forecast for lead times of up to seven days. It is interesting to note for how many steps ahead the forecasts could still be reliable. The specific time step is highly dependent on how fast a flash flood from heavy rainfall could concentrate at the downstream gauging station. In our study cases, this is about three to four hours. A reliable weather prediction might be an effective method for increasing the forecasting time step; in the cases studied, the rainfall estimating technique for a specific time and area during the typhoon period is being undertaken.

CONCLUSIONS

Most of the present neural networks for MSA forecasting belong to the direct method, which might produce acceptable performance under special cases. In this study, we investigated three common strategies: MIMO, MISO and serial-propagated architectures, used in neural networks for MSA flood forecasting in two different watersheds and performed a systematic comparison between these structures. Multi-layer feed-forward networks with a conjugate gradient algorithm for optimizing the connected weights were adopted. Based on the extensive experimental results, the following suggestions and remarks can be made.

1. The major difference between serial-propagated and other structures is that the former use predictive outputs as additional inputs. This concept not only provides the possibility of finding a better solution for multi-step-ahead forecasting, but enhances the reliability. The technique also helps in improving the prediction accuracy at several time steps in the future, if successful predictions could be iteratively used as additional inputs for the network. In this study, a detecting threshold (i.e. $G_{\text{bench}} > 0.5$) was implemented for the predictions at each step to avoid the problem of small prediction error at the beginning accumulating and propagating to the future, resulting in poor prediction accuracy.
2. Generally speaking, the accuracy of the MIMO architectures is relatively low, because of the complex interactions of the weights for different outputs and the complicated searching process in the error domain.
3. To investigate and demonstrate the impacts of input lag time on desired outputs, the input dimensions of MISO-1 are 12 (including some redundant inputs) for both two- and three-step-ahead forecasting, while the input dimensions of MISO-2, which take into account the impact of lag time, are only 8 and 4 for two-step and three-step-ahead forecasting, respectively. The results obtained from the MISO

network indicate that the model with reduced inputs (MISO-2) does perform better (in the testing sets of both watersheds) than the model with fixed inputs (MISO-1). This suggests that using an amount of unvalued information to construct a network might result in a lack of stability and accuracy in the testing phase, especially since, in most cases, a large number of training and validation data sets is not available.

4. The testing performances based on two experimental areas show consistent results for all models. Basically, all models have the ability to generate acceptable forecasts ($G_{\text{bench}} > 0$) in the two study basins. However, the serial-propagated neural network can produce the most accurate and stable predictions for all forecasting time steps. According to these two experimental results, it is also confirmed that the serial-propagated network has the capability to forecast up to four-hour-ahead of a flood, even in the flashy basins of Taiwan.
5. The MISO network performed slightly better than the serial-propagated network in the training phase; however, the latter performed much better than the MISO network in terms of smaller RMSE and higher G_{bench} in the testing phase. The results indicate that the accuracy of the MISO network gradually decreases as the predictive time steps increase, whereas the performance of serial-propagated structure decreases at a lower rate in MSA forecasting. Consequently, we suggest that a neural network with a serial-propagated technique could be applied for MSA streamflow forecasting during flood events.

Acknowledgements This paper is partially based on work supported by National Science Council, Republic of China. (Grant no. NSC91-2313-B-002-343). In addition, the authors are indebted to the reviewers for their valuable comments and suggestions.

REFERENCES

- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a) Artificial neural networks in hydrology I: Preliminary concepts. *J. Hydrol. Engng ASCE* **5**(2), 115–123.
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b) Artificial neural networks in hydrology II: Hydrologic applications. *J. Hydrol. Engng ASCE* **5**(2), 124–137.
- Atiya, A. F., El-Shoura, S. M., Shaheen, S. I. & El-Sherif, M. S. (1999) A comparison between neural-network forecasting techniques—case study: river flow forecasting. *IEEE Trans. Neural Networks* **10**(2), 402–409.
- Baratti, R., Cannas, B., Fanni, A., Pintus, M., Sechi, G. M. & Toreno, N. (2003) River flow forecast for reservoir management through neural networks. *Neurocomputing* **55**(3/4), 421–437.
- Battit, R. (1991) First and second order methods of learning: between the steepest descent and Newton's Method. *Neural Computation* **4**(2), 141–166.
- Campolo, M., Soldati, A. & Andreussi, P. (2003) Artificial neural network approach to flood forecasting in the River Arno. *Hydrol. Sci. J.* **48**(3), 381–398.
- Chang, F. J. & Chen, Y. C. (2001) A counterpropagation fuzzy-neural network modeling approach to real time stream flow prediction. *J. Hydrol.* **245**, 153–164.
- Chang, F. J. & Chen, Y. C. (2003) Estuary water-stage forecasting by using Radial Basis Function neural network. *J. Hydrol.* **270**, 158–166.
- Chang, F. J. & Hwang, Y. Y. (1999) A self-organization algorithm for real-time flood forecast. *Hydrol. Processes* **13**(2), 123–138.
- Chang, L. C. & Chang, F. J. (2002) An efficient parallel algorithm for LISSOM neural network. *Parallel Computing* **28**, 1611–1633.
- Chang, L. C., Chang, F. J. & Chiang, Y. M. (2004) A two-step ahead recurrent neural network for streamflow forecasting. *Hydrol. Processes* **18**, 81–92.
- Chiang, Y. M., Chang, L. C. & Chang, F. J. (2004) Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling. *J. Hydrol.* **290**, 297–311.
- Cigizoglu, H. K. (2003) Estimation, forecasting and extrapolation of river flows by artificial neural networks. *Hydrol. Sci. J.* **48**(3), 349–361.

- Coulibaly, P. (2003) Impact of meteorological predictions on real-time spring flow forecasting. *Hydrol. Processes* **17**, 3791–3801.
- Coulibaly, P., Anctil, F. & Bobée, B. (2000) Neural network-based long-term hydropower forecasting system. *Computer-Aided Civil and Infrastructure Engng* **15**, 355–364.
- Coulibaly, P., Anctil, F. & Bobée, B. (2001) Multivariate reservoir inflow forecasting using temporal neural networks. *J. Hydrol. Engng ASCE* **6**(5), 367–376.
- Dawson, C. W. & Wilby, R. L. (1998) An artificial neural network approach to rainfall–runoff modelling. *Hydrol. Sci. J.* **43**(1), 47–67.
- 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.
- Gao, X. Z., Ovaska, S. J. & Vasilakos, A. V. (2002) Temporal difference method-based multi-step-ahead prediction of long term deep fading in mobile networks. *Computer Communications* **25**, 1477–1486.
- Govindaraju, R. S. & Rao, A. R. (2000) *Artificial Neural Networks in Hydrology*. Kluwer, Dordrecht, The Netherlands.
- Grayson, R. B., Moore, I. D. & McMahon, T. A. (1992) Physically based hydrologic modelling, 2: is the concept realistic? *Water Resour. Res.* **28**(10), 2659–2666.
- Ham, F. M. & Kostanic, I. (2001) *Principles of Neurocomputing for Science and Engineering*. McGraw-Hill, New York, USA.
- Hsu, K. L., Gupta, H. V., Gao, X. G. & Sorooshian, S. (1999) Estimation of physical variables from multichannel remotely sensed imagery using a neural network: application to rainfall estimation. *Water Resour. Res.* **35**(5), 1605–1618.
- Hu, T. S., Lam, K. C. & Ng, S. T. (2005) A modified neural network for improving river flow prediction. *Hydrol. Sci. J.* **50**(2), 299–318.
- Huang, S. J. & Lian, R. J. (2000) A combination of fuzzy logic and neural network controller for multiple-input multiple-output systems. *Int. J. Systems Sci.* **31**, 343–357.
- HEC (Hydrological Engineering Center) (1990) *HEC-1 Flood Hydrograph Package. Program Users Manual*. US Army Corps of Engineers, Davis, California, USA.
- Lallahem, S. & Mania, J. (2003) Evaluation and forecasting of daily groundwater outflow in a small chalky watershed. *Hydrol. Processes* **17**, 1561–1577.
- Luk, K. C., Ball, J. E. & Sharma, A. (2000) A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *J. Hydrol.* **227**, 56–65.
- Minns, A. W. & Hall, M. J. (1996) Artificial neural networks as rainfall–runoff models. *Hydrol. Sci. J.* **41**(3), 399–417.
- Mishra, A., Hata, T. & Abdelhadi, A. W. (2004) Models for recession flows in the upper Blue Nile River. *Hydrol. Processes* **18**, 2773–2786.
- Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models, Part I: a discussion of principles. *J. Hydrol.* **10**, 282–290.
- Parlos, A. G., Rais, O. T. & Atiya, A. F. (2000) Multi-step-ahead prediction using dynamic recurrent neural networks. *Neural Networks* **13**, 765–786.
- Ranjithan, S., Eheart, J. W. & Garrett, J. H., Jr (1993) Neural network-based screening for groundwater reclamation under uncertainty. *Water Resour. Res.* **29**(3), 563–574.
- Sajikumar, N. & Thandaveswara, B. S. (1999) A non-linear rainfall–runoff model using an artificial neural network. *J. Hydrol.* **216**, 32–55.
- Salas, J. D., Delleur, J. W., Yevjevich, V. & Lane, W. L. (1985) *Applied Modeling of Hydrologic Time Series*. Water Resources Publications, Littleton, Colorado, USA.
- Schenker, B. & Agarwal, M. (1995) Long-range prediction for poorly-known systems. *Int. J. Control* **62**, 227–238.
- Seibert, J. (2001) On the need for benchmarks in hydrological modeling. *Hydrol. Processes* **15**, 1063–1064.
- Selvaraj, R., Deshpande, P. B., Tambe, S. S. & Kulkarni, B. D. (1995) Neural networks for the identification of MSF desalination plants. *Desalination* **101**, 185–193.
- Sivakumar, B. (2005) Hydrologic modeling and forecasting: role of thresholds. *Environ. Model. Software* **20**(5), 515–519.
- Sivakumar, B., Jayawardena, A. W. & Fernando, T. M. K. G. (2002) River flow forecasting, use of phase-space reconstruction and artificial neural networks approaches. *J. Hydrol.* **265**, 225–245.
- Su, H. T., McAvoy, T. J. & Werbos, P. (1992) Long-term prediction of chemical processes using recurrent neural networks: a parallel training approach. *Indust. Appl. Chem. Engng Res.* **31**, 1338–1352.
- Toth, E., Brath, A. & Montanari, A. (2000) Comparison of short-term rainfall prediction models for real-time flood forecasting. *J. Hydrol.* **239**, 132–147.
- Zealand, C. M., Burn, D. H. & Simonovic, S. P. (1999) Short term stream flow forecasting using artificial neural networks. *J. Hydrol.* **214**, 32–48.

Received 25 September 2005; accepted 8 August 2006