A hybrid model for daily flow forecasting

Huynh Ngoc Phien* and Nguyen Tan Danh
Asian Institute of Technology, PO Box 4, Klong Luang, Pathumthani 12120, Thailand

Abstract

Having a simple structure, the Tank Model has been widely used in modelling the rainfall-runoff process of many watersheds. In this study, the model was slightly modified to make it suitable for forecasting daily discharges. Two basins in Vietnam, namely the Da Nhim and La Nga Basins were considered. The daily discharges were forecast one day ahead using daily rainfall and daily evaporation. It was found that forecast discharges for the smaller basin (Da Nhim) were not as good as those for the larger basin (La Nga).

In order to improve the forecasting performance of the Tank Model, a first-order autoregressive model, AR(1) was introduced to represent the error time series. The resulting hybrid model is the combination of the Tank Model and the AR(1) Model. Thus the forecast discharge for any day is obtained as the forecast value obtained from the Tank Model plus the forecast value of the error for that day. Application of the hybrid model to the above two basins was found to provide more accurate forecast values even though the overall performance of the hybrid model largely depends on the efficiency of the Tank Model.

Introduction

Forecasting of daily discharges has been an important problem in reservoir operations and to some extent, in flood protection. In this connection rainfall data are commonly used along with evaporation data and past discharge records.

There have been many models developed to model the runoff process using rainfall and evaporation as key factors. These are roughly classified as conceptual models and the Tank Model (Sugawara, 1961) has been widely used thanks to its simple structure and the automatic calibration procedure (Sugawara, 1979). In this study it was selected to forecast discharges for two basins in Vietnam, the Da Nhim and La Nga Basins. However, due to the fact that the model itself was not meant for forecasting, a slight modification in the treatment was made to render it suitable for forecasting purposes.

As is shown, the Tank Model alone did not perform very satisfactorily. Therefore a stochastic model was used to correct the discrepancies in the computed discharges of a deterministic rainfall-runoff model. The stochastic model used in this study was the first-order autoregressive model, AR(1), and the combined model becomes the hybrid model.

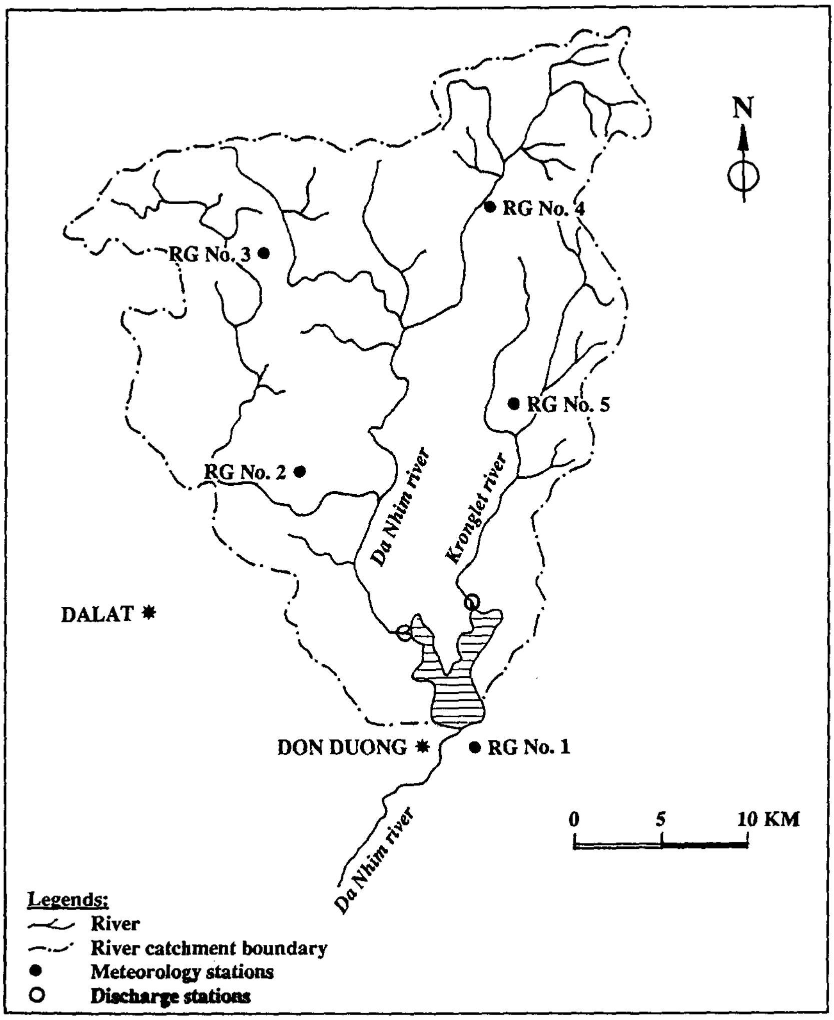


Figure 1
Location map of Da Nhim Basin and its monitoring stations

Selected basins

Two basins, Da Nhim and La Nga, situated between latitude 11^{0} -12^{0} 20' north and longitude 107^{0} - $108^{0}30'$ west, in the tropical

---

* To whom all correspondence should be addressed.
☎(66-2) 516-5701; fax (66-2) 524-5721; e-mail hnp@cs.ait.ac.th

Received 12 December 1996; accepted in revised form 14 March 1997.

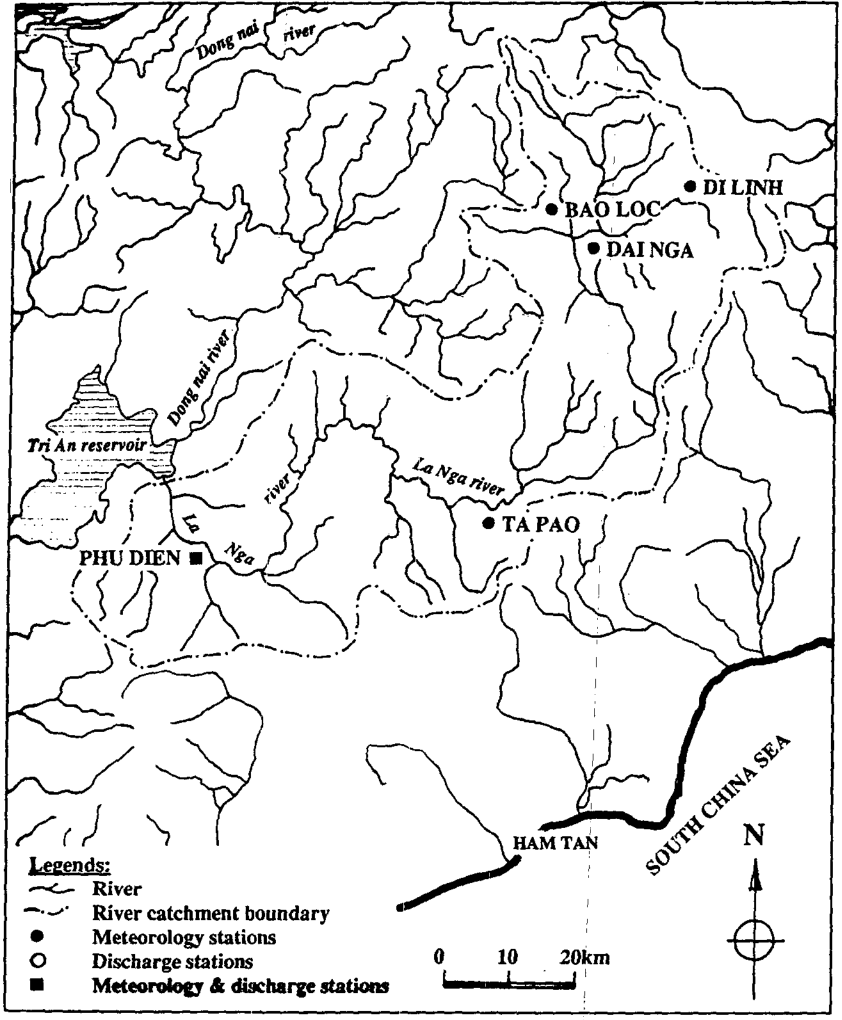


Figure 1
Location map of Da Nhim Basin and its monitoring stations

monsoon zone, were selected for this study. Both basins are part of the Dong Nai River Basin where a comprehensive development program is being planned for its exploitation.

Da Nhim Basin

Originating from the Lang Biang Plateau at 2 000 m above mean sea level (amsl), the Da Nhim River flows between mountains and hills at an elevation of 1 500 m down to 700 m amsl. The river has two main tributaries, namely the Da Nhim River and the Kronglet River, with a total drainage area of 775 km². Within the basin, there are five meteorological stations; two streamflow gauging stations, one on each river (Fig. 1). The collected data for the period 1982 to 1992 include mean areal daily rainfall, mean areal daily evaporation, and total daily discharge to the reservoir (but not the data at each individual station). The data in the first 7 years (1982 to 1988) were employed in the calibration of the models considered, and the data in the first 4 years (1989 to 1992) were used for model validation.

La Nga Basin

The La Nga River originates from the Blao Plateau at 580 m amsl, on the border between Lam Dong and Thuan Hai Provinces. The river flows through a vast area of forest and mountains with many waterfalls and chutes. Meteorological data were collected at five stations in the basin, and discharge data were collected at Phu Dien hydrological station. The drainage area at Phu Dien Station is 3 060 km² (Fig. 2). The data in the first 3 years (1987 to1989) were used for model calibration, and the data in the last year (1994) were used for model validation

Description of the tank model

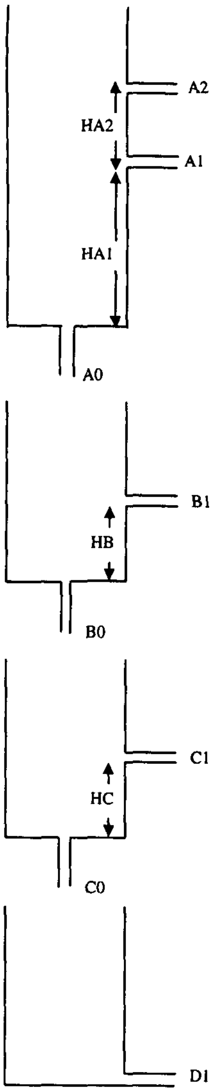
The Tank Model used in this study was the simple tank for a humid region as described in Phien and Pradhan (1983). It consists of 4 tanks laid vertically (Fig. 3). The top tank has two side outlets and one bottom outlet corresponding to the conceptual structure of surface discharge and infiltration, respectively. The second tank and the third tank have two outlets (side and bottom) each, while the fourth tank has only one side outlet. Water in the second tank partly moves to the stream channel through the side outlet corresponding to interflow; the bottom outlet of the second tank provides percolation to the third tank. The bottom outlet of the third tank provides deep percolation to the fourth tank; the side outlets of the third and the fourth tank provide the baseflow.

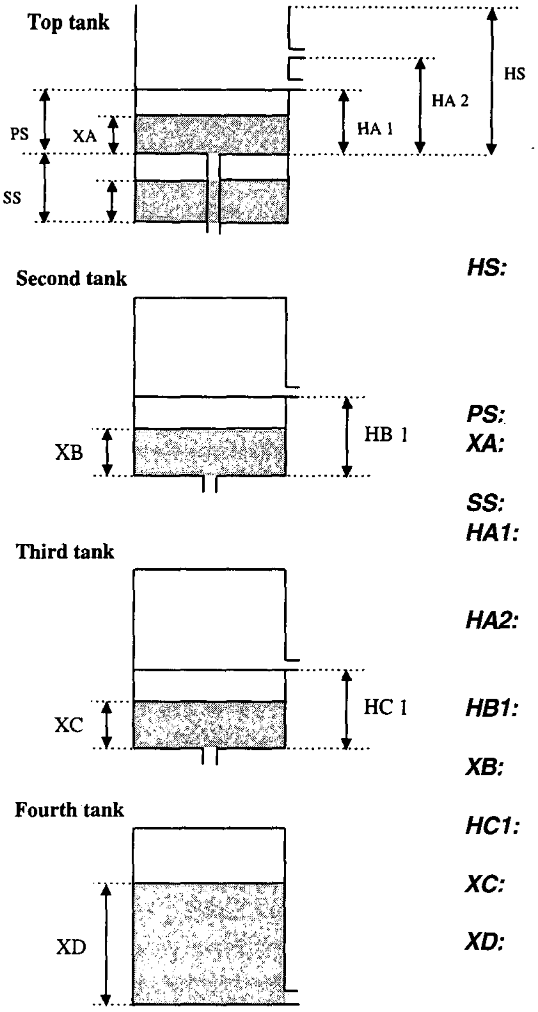
The parameters of the tank model are grouped into two types. The first type consists of the side and bottom outlet coefficients (A_0, A_1, A_2, B_0, B_1, C_0, C_1, D_1). The second type are the storage parameters (HA_1, HA_2, PS, SS, HS, XA, XS for the top tank; HB_1, XB for the second tank; HC_1, XC for the third tank; XD for the fourth tank). An illustration of the parameters is shown in Fig. 4.

The calibration of the Tank Model includes two stages. At Stage 1, parameters of Type 1 are assumed with values in "an initial model" suggested by Sugawara (1961), parameters of Type 2 are adjusted one by one by trial and error until the best agreement is achieved. At Stage 2, parameters of Type 2 are kept unchanged, while parameters of Type 1 are adjusted by means of an automatic calibration procedure. The adjustment of the parameters is carried out using the following guidelines:

- The parameters of the top tank are adjusted according to the shape and volume of the computed hydrograph in a period of high discharge resulting from high rainfall.
- The parameters of the second tank are adjusted by examining the hydrograph of the period that immediately follows the peak discharge.
- The parameters of the last two tanks are adjusted by examining the part of the hydrograph which corresponds to the base flow.

The mean areal rainfall over the catchment of each discharge station was calculated for the Tank Model as a weighted mean following the procedure introduced by Sugawara (see Sugawara et al., 1980). At the start of the calibration all stations have the





HS: *The maximum storage above the bottom of the primary soil moisture structure in which the outflow through the bottom outlet becomes constant.*
PS: *Primary soil moisture.*
XA: *Initial storage in the primary soil moisture.*
SS: *Secondary soil moisture.*
HA1: *The storage between the lower side outlet and the bottom of the primary soil moisture.*
HA2: *The storage between the upper side outlet and the bottom of the primary soil moisture.*
HB1: *The storage from the bottom to the side outlet.*
XB: *Initial storage above the bottom.*
HC1: *The storage from the bottom to the side outlet.*
XC: *Initial storage above the bottom.*
XD: *The initial storage above the bottom.*

Figure 3 (left)
Tank Model for humid regions

Figure 4 (top)
Parameters of Type 2

same weight equal to 1; after fairly good results are obtained for the computed runoff values, modification of the weight will start. If rainfall data at a station are found to cause larger differences between computed and observed values of runoff, they should be less representative, i.e. should have less weight. Consequently, the weight is decreased from 1 to 0.5, or from 0.5 to 0.25, or even from 0.25 to 0.0, where zero weight means that the rainfall station concerned is neglected.

Originally, the Tank Model was used for simulating the rainfall-runoff process. In this case, the output (daily discharge) from the model may be expressed as:

$$Q_t = f(R_t, E_t, \Theta) \qquad (1)$$

where Q_t, R_t and E_t denote respectively the discharge, rainfall and evaporation on day t, Θ denotes symbolically the vector of model parameters, and f denotes a functional relationship. To make the Tank Model suitable for forecasting with lead time equal to one day, the following form should be adopted:

$$Q_{t+1} = f(R_t, E_t, \Theta) \qquad (2)$$

This means that the forecast value of the discharge on day t+1 made on day t is expressed as a function of the rainfall and evaporation on day t. Unlike other non-physical based models where past values of rainfall, evaporation and even discharge on preceding days t, t-1, t-2, ... can also be used in forecasting the discharge on day t+1, the Tank Model can accommodate the values of rainfall and evaporation on day t only.

It should be noted that forecasting with other larger lead times (two or three days, for example) can also be formulated. But for the two basins considered, which are rather small and steep forecasting with one day lead time is appropriate.

Results

Station weights for rainfall data

For the Da Nhim Basin, the collected rainfall data consist of the mean areal values, therefore there was no need for the determination of rainfall station weights. In the case of the La Nga Basin, the determination was made by changing (with values of 0.25, 0.5) the weight of a station one at a time while keeping the others unchanged. At each change, the efficiency index (EI) (to be defined later) was noted for comparison. The best set of weights was chosen corresponding to the highest value of EI. Table 1 shows the results of the calculation for station weights.

TABLE 1
VALUES OF EI IN THE CALCULATION OF STATION WEIGHTS FOR RAINFALL IN LA NGA BASIN

Values of weights for each station					EI
Bao Loc	Di Linh	Dai Nga	Phu Dien	Ta Pao	
1	1	1	1	1	0.91
0.5	1	1	1	1	0.89
1	0.5	1	1	1	0.89
1	1	0.5	1	1	0.89
1	1	1	0.5	1	0.89
1	1	1	1	0.5	0.89
0.25	1	1	1	1	0.68
1	0.25	1	1	1	0.88
1	1	0.25	1	1	0.61
1	1	1	0.25	1	0.69
1	1	1	1	0.25	0.67

TABLE 2
SELECTED VALUES OF TYPE 1 PARAMETERS OF THE TANK MODEL

Basin	A_0	A_1	A_2	B_0	B_1	C_0	C_1	D_1
Da Nhim	5.27E-2	0.03	0.06	2.73E-3	0.20	2.98E-5	3.0	0.0001
La Nga	4.49E-4	0.03	0.09	7.22E-3	0.01	5.2[illegible]E-4	0.1	0.0001

TABLE 3
SELECTED VALUES OF TYPE 2 PARAMETERS OF THE TANK MODEL

Basin	HA_1	HA_2	HB_1	HC_1	XA	XB	XC	XD	XS	PS	SS	HS
Da Nhim	1	80	40	1000	22.2	40	1000	2000	30.9	5	100	100
La Nga	5	60	40	620	26.5	40	620	5000	16.3	5	50	150

Values of parameters of the Tank Model

Following the combined approach consisting of a trial-and-error procedure and an automatic calibration, the parameters of the Tank Model were obtained and shown in Table 2 and Table 3 for Type 1 and Type 2, respectively.

Performance assessment

To evaluate the forecasting capability of the proposed approach, the following statistics were used:

Efficiency index (EI): The efficiency of a model can be measured by the efficiency index defined by Nash and Sutcliffe (1970):

$$EI = \frac{ST - SSE}{ST} \tag{3}$$

where ST is the total variation:

$$ST = \sum_{i=1}^{N} (Q_i - \bar{Q})^2 \tag{4}$$

and SSE is the sum of squared errors:

$$SSE = \sum_{i=1}^{N} E_i^2 = \sum_{i=1}^{N} (Q_i - F_i)^2 \tag{5}$$

In both Eqs. 4 and 5, N is the total number of data used, E is the error, Q_i and F_i denote respectively the observed and forecast discharges on day i, while $\bar{Q}$ is the mean value taken over N values of Q.

Mean absolute deviation (MAD): The mean absolute deviation is defined as the mean of the absolute value of the used in Eq. 5. This statistic is used frequently in connection with forecasting evaluation.

Mean squared error: The mean squared error MSE is obtained as:

$$MSE = \frac{SSE}{N} \tag{6}$$

This statistic, like the MAD, is useful to compare the performance of the different methods on the same set of data. It does not give

TABLE 4
PERFORMANCE STATISTICS FOR THE TANK MODEL

Basin	Stage	EI	RMSE	RMSEM	MAD	RMSES
Da Nhim	Calibration	0.69	14.36	0.78	6.59	0.58
	Validation	0.65	12.52	0.73	5.45	0.65
La Nga	Calibration	0.91	43.12	0.34	26.00	0.29
	Validation	0.93	45.11	0.28	29.57	0.25

a clear indication as to which value of MSE the performance of a method is acceptable. In this study, we introduce the root mean squared error with respect to the mean (RMSEM) and root mean squared error with respect to the standard deviation (RMSES):

$$\text{RMSEM} = \sqrt{\text{MSE} / \bar{Q}} \qquad (7)$$

$$\text{RMSES} = \sqrt{\text{MSE} / S}$$

where S is the standard deviation of the discharge.

The results obtained for the two basins are shown in Table 4. For the Da Nhim Basin, these performance statistics show that the Tank Model did not forecast the discharges satisfactorily as evidenced by the low values for EI and high values of RMSEM and RMSES. The error is above 70% of the mean (indicated by RMSEM) and above 58% of standard deviation (indicated by RMSES) of the observed discharge. For the La Nga Basin, the Tank Model performs reasonably well with the EI value up to 0.91 in calibration stage and 0.93 in validation stage. The error is less than 35% of the mean and 30% of the standard deviation.

When examining the forecast-observed discharge hydrographs plotted for each year in the calibration and validation stages of both basins, it was found that for the Da Nhim Basin the baseflow and medium peaks were well reproduced but high peaks were underestimated. Typical cases are illustrated in Fig. 5 for the year 1987 (calibration stage) and in Fig. 6 for the year 1990 (validation stage). For the La Nga Basin, baseflow and medium peaks were a little bit overestimated while high peaks were satisfactorily reproduced. Typical cases are illustrated in Fig. 7 for the year 1988 (calibration stage) and in Fig. 8 for the year 1994 (validation stage).

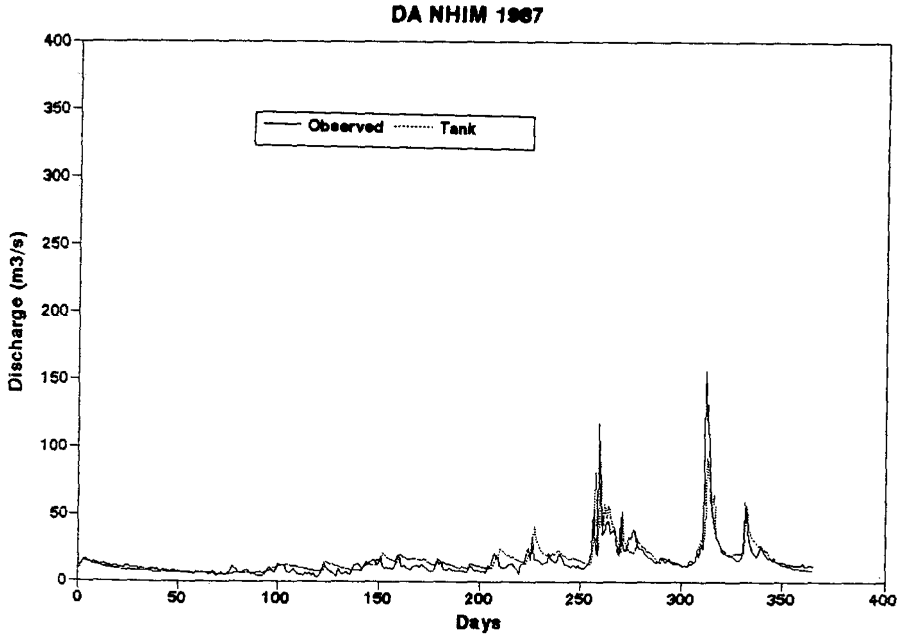


Figure 5
Comparison between observed and forecast discharges by the Tank Model for Da Nhim, 1987 (calibration)

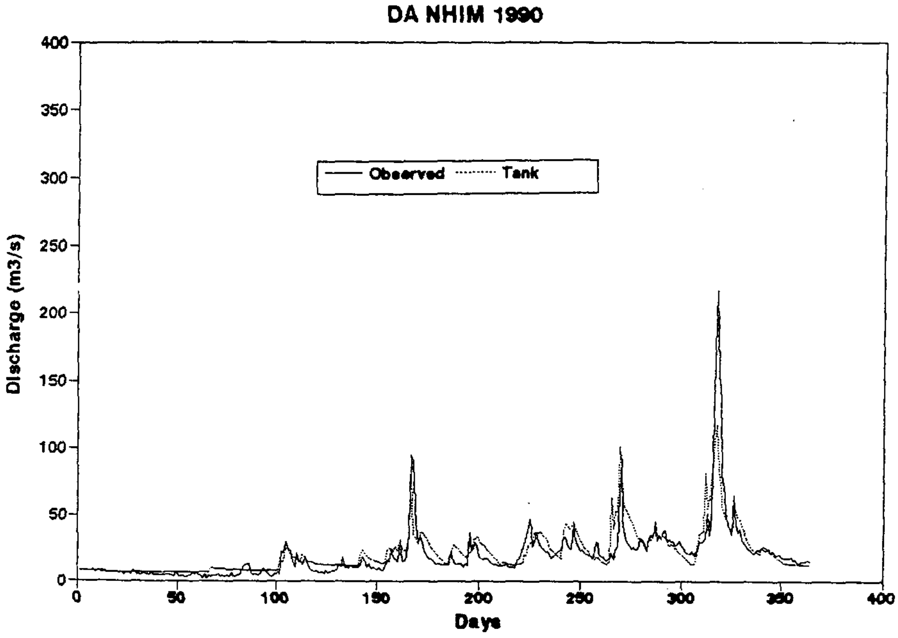


Figure 6
Comparison between observed and forecast discharges by the Tank Model for Da Nhim, 1990 (validation)

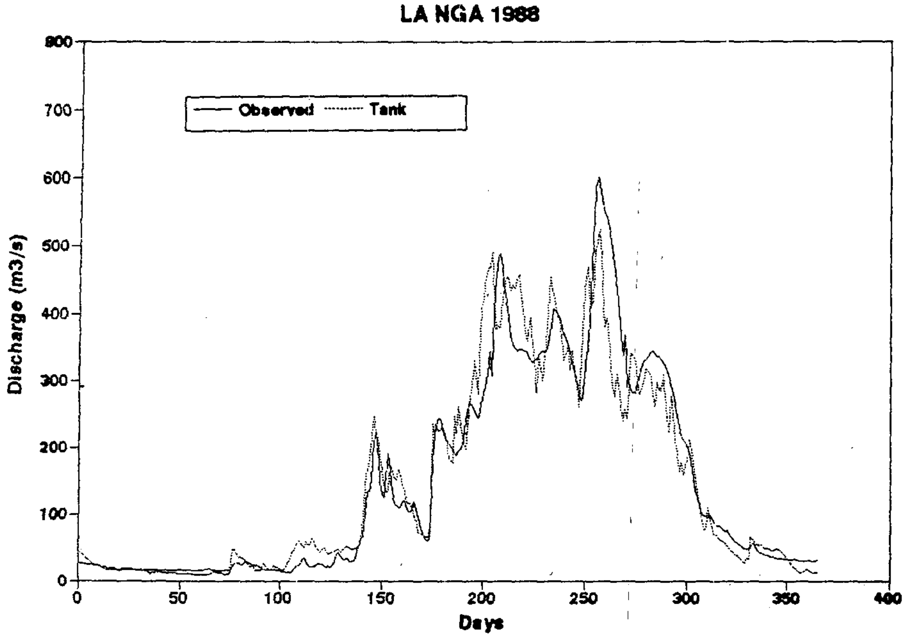


Figure 7 (left)
Comparison between observed and forecast discharges by the Tank Model for La Nga, 1988 (calibration)

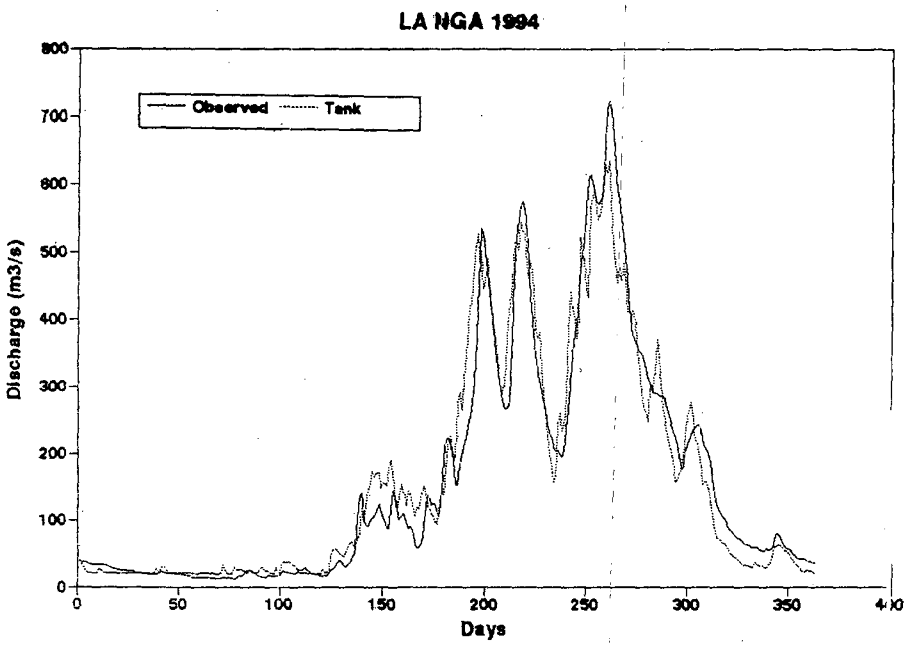


Figure 8
Comparison between observed and forecast discharges by the Tank Model for La Nga, 1994 (validation)

The hybrid model

As seen, the Tank Model did not perform very well. It needed further improvement. For this reason, a first-order autoregressive model, AR(1), was used to represent the time series of the errors. It was combined with the Tank Model to form a hybrid model: a forecast value by the hybrid model is the sum of the forecast value obtained by the Tank Model and the forecast value of the error obtained by the AR(1) Model:

$$Fc(t) = Ft(t) + Fe(t) \qquad (8)$$

where Ft(t) is the forecast discharge by the Tank Model (denoted as before), Fe(t) is the forecast of the error, and Fc(t) is the combined (hybrid) forecast value, all for time t. The forecast value of the error by the AR(1) Model is obtained as:

$$Fe(t) = \Phi * e_{t-1} \qquad (9)$$

where Φ is the autoregressive coefficient at lag 1. It should be recalled that:

$$e_{t-1} = Q_{t-1} - F(t-1) \qquad (10)$$

The value of Φ was obtained as the autocorrelation coefficient of the first-order of the error term during the calibration stage. Its value is 0.274 for Da Nhim and 0.861 for La Nga, respectively.

This Hybrid Model was applied to the basins and the performance statistics obtained are shown in Table 5.

Comparing the values of the performance statistics in Table 5 with those in Table 4, we see that the values

TABLE 5
PERFORMANCE STATISTICS FOR THE HYBRID MODEL

Basin	Stage	EI	RMSE	RMSEM	MAD	RMSES
Da Nhim	Calibration	0.72	13.26	0.72	5.09	0.63
	Validation	0.68	10.79	0.63	4.29	0.56
La Nga	Calibration	0.98	21.89	0.17	11.33	0.15
	Validation	0.98	20.68	0.14	9.79	0.18

of EI of the Hybrid Model are higher, and the values of RMSEM and MAD are lower. Particularly for the La Nga Basin, the values of EI in calibration and validation stages are as high as 0.98; and the values of RMSES are as low as 0.15 and 0.18, respectively. For the Da Nhim Basin, even though there has been some improvement, the performance has not been very satisfactory.

The hydrographs of observed and forecast discharge by the Hybrid Model were plotted for each year during both the calibration and validation stages.

For the Da Nhim Basin, the forecast discharges were generally closer to the observed ones. This applies to baseflow and medium peaks as well as to high peaks. Typical cases are illustrated in Fig. 9 for the year 1987 (calibration stage) and Fig. 10 for the year 1990 (validation stage). Comparing Fig. 9 with Fig. 5 and Fig. 10 with Fig. 6, we see clearly the improvement in the forecast values. This clearly shows that the Hybrid Model has a much better forecasting performance as compared to the Tank Model even though the performance statistics do not indicate likewise. It should be noted that the data obtained for the Da Nhim Basin as used in the present study are just the average values for which the appropriateness of the method used to calculate them is not known. It seems that the average values may not be representative.

For the La Nga Basin, the forecast discharges by the Hybrid Model were extremely good. Typical cases are shown in Fig. 11 for the year 1988 (calibration stage) and Fig. 12 for the year 1994 (validation stage). Comparing Fig. 11 with Fig. 7, and Fig. 12 with Fig. 8, we see that the forecast discharges by the Hybrid Model are very close to the observed discharges. However, when the error at the previous time step, e_{t-1} is big, then the forecast error by the auto-regressive model for the next time step is also large. Moreover, if the forecast discharge Ft(t) by the Tank Model is close to the corresponding observed discharge, the forecast discharge by the Hybrid Model may consequently overshoot the observed value. This means the forecast value is much higher or lower than the observed value. For example, considering the highest peaks in the hydrograph of the La Nga Basin forecast by the Tank Model (Fig. 7) and by the Hybrid Model (Fig. 11), we see that the error along the rising limb of the peak when forecast by the Tank Model is large, the forecast discharge by the Hybrid Model comes closer to the observed discharge. But when the difference between the observed discharge and the forecast discharge by the Tank Model is not much (just after the top of the peak) while the forecast error is still large due to the large error of the previous time step, the forecast discharge by the Hybrid Model then becomes quite big compared to the observed discharge. This observation also suggests that the efficiency of the Hybrid Model depends very much on the efficiency of the Tank Model.

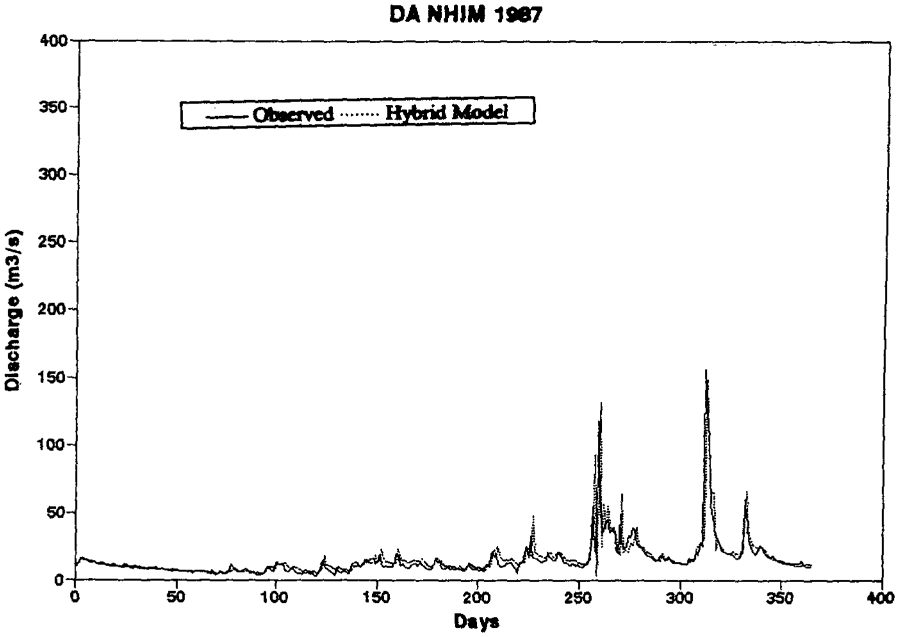


Figure 9
Comparison between observed and forecast discharges by the Hybrid Model for Da Nhim, 1987 (calibration)

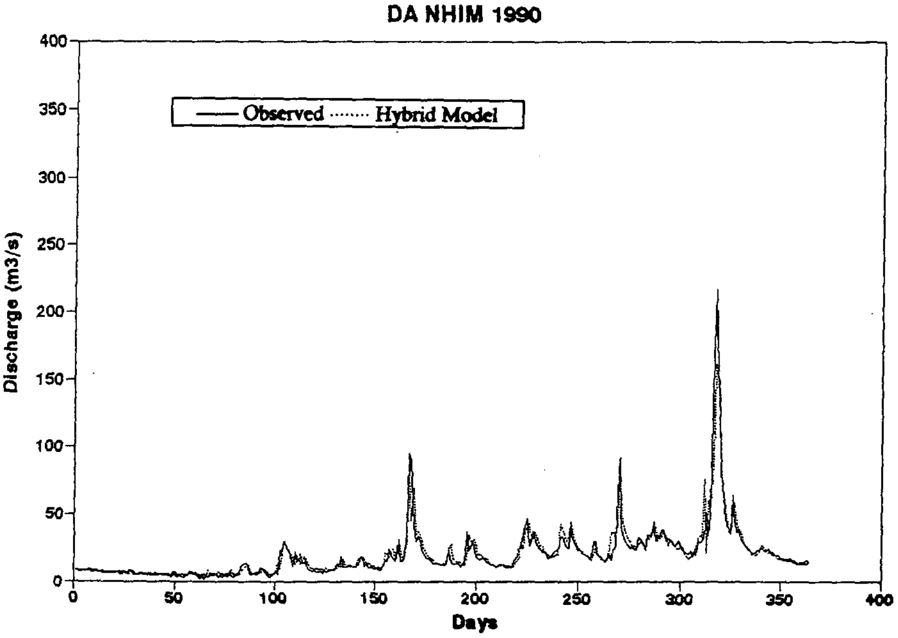


Figure 10
Comparison between observed and forecast discharges by the Hybrid Model for Da Nhim, 1990 (validation)

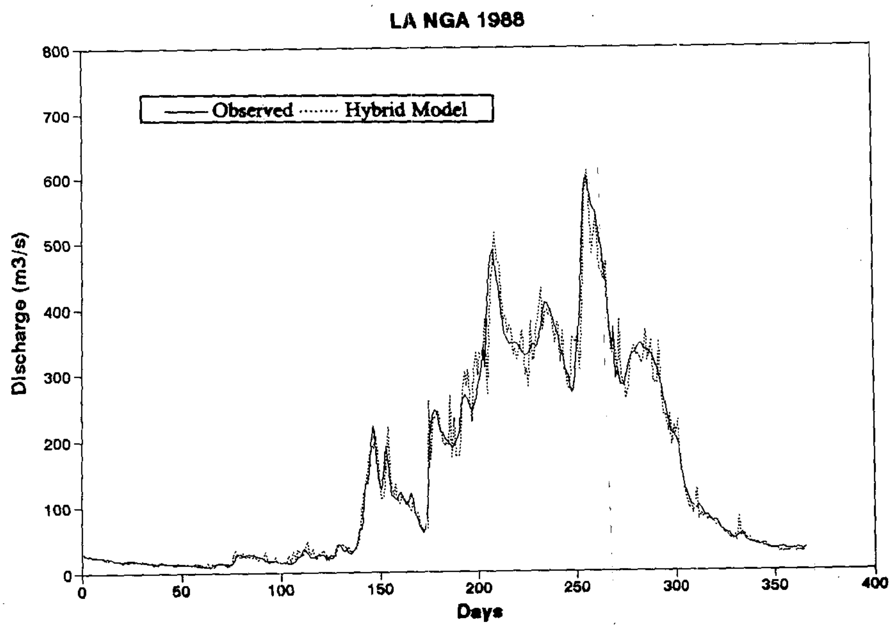


Figure 11

Comparison between observed and forecast discharges by the Hybrid Model for La Nga, 1988 (calibration)

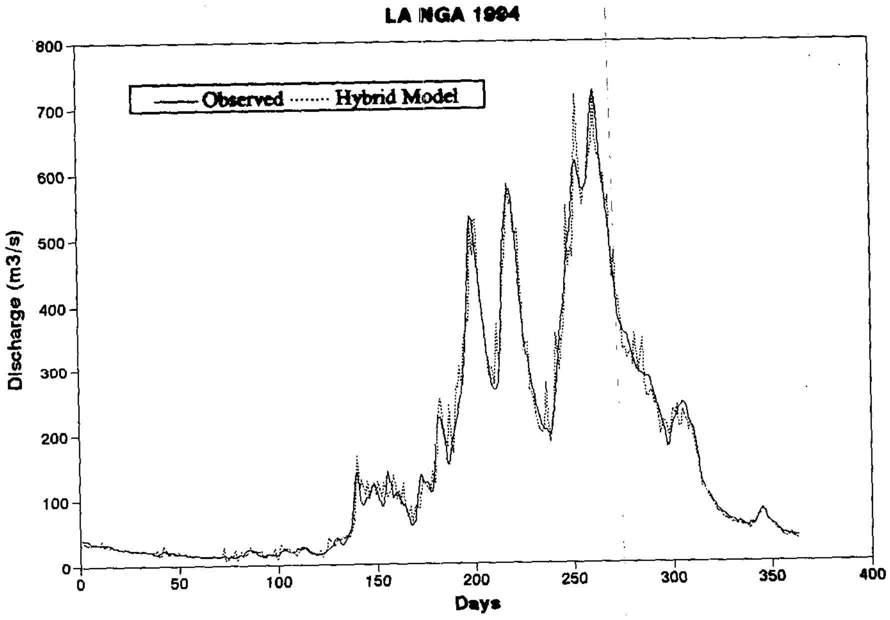


Figure 12

Comparison between observed and forecast discharges by the Hybrid Model for La Nga, 1988 (validation)

It should be noted that the value of the first-order autoregressive coefficient for the Da Nhim Basin is very low (0.274). This indicates that the errors of the forecast values by the (deterministic) Tank Model for this basin are not highly dependent. As such the improvement by adding the (stochastic) autoregressive component is not much. But in the case of the La Nga Basin, the value of the first-order autoregressive coefficient is large (0.861), indicating that the errors are highly dependent. As such, the additional stochastic component can still improve the performance further.

Conclusion

In this study, the Tank Model was slightly modified to make it applicable to forecast daily discharges from daily rainfall and daily evaporation one day ahead for two basins, namely the Da Nhim and La Nga Basins in Vietnam. The results showed that the Tank Model provided better forecast values for the larger basin (La Nga) as expected. A first-order autoregressive model was then used to represent the errors that are the differences between the observed discharges and the forecast discharges by the Tank Model. The Hybrid Model, obtained as a combination of the Tank Model and the first-order autoregressive model, showed further improvement in forecasting performance for both cases, particularly in terms of the discharge hydrograph.

References

PHIEN HN and PRADHAN PSS (1983) The Tank Model in rainfall-runoff modelling. *Water SA* **9**(3) 93-102.

NASH JE and SUTCLIFFE JV (1970) River flow forecasting through conceptual models. *J. of Hydrol.* **10** 282-290.

SUGAWARA M (1961) On the analysis of runoff structure about several Japanese rivers. *Japanese J. of Geophys.* **2** (4) 1-76.

SUGAWARA M (1979) Automatic calibration of the Tank Model. *Hydrol. Sci. Bull.* **24** (3) 375-388.

SUGAWARA M, WATANABE I, OZAKI E and KATSUYAMA Y (1980) *Tank Model for Snow Component.* National Research Center for Disaster Prevention.