The Hybrid Method for seasonal streamflow forecasting

Huynh Ngoc Phien* **and Jean Wei-Haw**
Asian Institute of Technology, P.O. Box 2754, Bangkok 10501, Thailand.

Abstract

The amount of water stored in a reservoir can be utilized more efficiently if it is possible to forecast future inflows. However, not many techniques for seasonal streamflow forecasting have been devised. Recently, a particular scheme has been introduced, in which regression analysis is used in model identification and parameter estimation, and a time series approach is used in providing forecasts with long lead times. Because of this combination, the scheme has been referred to as the Hybrid Method.

In this study, the Hybrid Method is evaluated along with its two simplified versions through practical applications to monthly streamflow forecasting. It is found that the Hybrid Method and these simplified versions produce satisfactory forecasts for stations with large drainage areas. However, for stations with small drainage areas they seem to perform very poorly.

Introduction

Although the forecasting of seasonal streamflows is very important in the efficient use of the amount of water stored in reservoirs, it has not attracted due attention from researchers. Apart from a few publications which appeared irregularly in the literature (e.g. Hino and Ishikawa, 1975; Rao and Kashyap, 1973, 1974), three general surveys have recently been conducted by Dyhr-Nielson (1982), O'Connell (1983) and Phien (1983a). These surveys mainly addressed the important techniques which are likely to be useful for the forecasting of seasonal streamflows and simple applications were given for illustrative purposes. The attempt made by Phien (1983b) may also be considered as an illustration of the use of the Kalman filter and Group Method of Data Handling techniques in this regard. Therefore, it is desirable to have a systematic evaluation of the most promising methods in seasonal streamflow forecasting.

The present study evaluates a particular scheme introduced by O'Connell (1983). In this scheme, known as the Hybrid Method, two useful techniques, regression analysis and time series, are combined in order to make the maximum use of the available data. Moreover, two simplified versions which are based on the models developed by Thomas and Fiering (1962) and Sen (1978) for data generation are also considered to see whether simpler methods can be used as well.

As in the paper by Phien and Twu (1984), a monthly basis is adopted here. Accordingly, forecasts with lead times from one to six months are of importance. With a lead time greater than one month, the Thomas-Fiering and Sen models cannot be readily used and hence suitable extensions are presented. They are then applied along with the model obtained from the Hybrid Method to the data from the Mekong and Tachia basins, whereby conclusions can be made.

The Hybrid Method

Basically, the Hybrid Method employs the stepwise regression scheme in the selection of the explanatory variables and in the estimation of the parameters of the resulting equations. Then in the forecasting stage, the Box-Jenkins approach is used to produce forecasts corresponding to lead times greater than one season by replacing unknown values of the explanatory variables by their forecast values when they become available. This approach can readily be adopted to make the Thomas-Fiering and Sen Models capable of serving the same purpose.

The Hybrid Model

General considerations

In the model development stage, several factors affecting monthly streamflows are considered as explanatory variables. The resulting equations obtained by means of stepwise regression analysis are referred to as the Hybrid Model in this study for brevity. Typically, the following problems are considered:

Problem 1: Forecast the flow in month (m + 1) from:
- monthly flows in months m, m − 1, m − 2, . . .
- explanatory variable 1 in months m, m − 1, . . .
- explanatory variable 2 in months m, m − 1, . . .
- . . .

Problem 2: Forecast the flow in month (m + 2) from:
- monthly flows in months m + 1, m, m − 1, . . .
- explanatory variable 1 in months m, m − 1, . . .
- explanatory variable 2 in months m, m − 1, . . .
- . . .

Problem 3: Forecast the flow in month (m + 3) from:
- monthly flows in months m + 2, m + 1, m, m − 1, . . .
- explanatory variable 1 in months m, m − 1, . . .
- explanatory variable 2 in months m, m − 1, . . .
- . . .

In these problems, the flow in the preceding month is explicitly listed as an explanatory variable. This is done because for a given month, the flow in the preceding month can always be viewed as the integrating factor of all other explanatory variables which influence the monthly flows until that month. Clearly, the performance of the Hybrid Model depends heavily on this variable. The other variables may include monthly rainfall, last day's flow, etc. As the lead time increases, the latter variables would have decreasing influences.

It should be noted that the explanatory variables other than the monthly flow are considered up to the present month, m. Otherwise, their forecasts are needed before a forecast for the flow can be made. The forecasting of these variables is in most cases even more difficult than the forecasting of monthly flows.

*To whom all correspondence should be addressed.
Received 24 May 1985.

Once a mode has been developed for a month, the forecasting of monthly flows in the succeeding months is carried out as follows:

(1) To forecast the flow in month (m + 1) from data in month m, use the model obtained in Problem 1.

(2) To forecast the flow in month (m + 2) from data in month m, use the model obtained in Problem 2, where the flow in month (m + 1) is substituted by its forecast computed in (1).

(3) To forecast the flow in month (m + 3) from data in month m, use the model obtained in Problem 3, and substitute the forecasts in (1) and (2) for the flows in months (m + 1) and (m + 2), respectively. And so on until the maximum lead time required is reached.

Mathematical development

From the above considerations, the model corresponding to Problem 1 can be readily obtained from (stepwise) regression analysis. For lead times larger than one month, one may apply regression analysis to Problem 2, Problem 3, etc. to establish the required models. This approach works under the assumption that the flows in months (m + 1), (m + 2), . . . are known. In actual situations, these flows are to be replaced by their forecasts. Consequently, it is more realistic to express the forecasting equations in terms of these forecasts.

A mathematical derivation according to this approach is as follows. Suppose, for simplicity, there is only one explanatory variable X besides the monthly flow Y both with the corresponding means removed. The mathematical equation stipulated in the forecasting stage can be written as

$$\hat{Y}_{m|m-1,m-k} = \sum_{i=1}^{\infty} a_{i,m}^{(k)} Y_{m-i} + \sum_{i=1}^{\infty} b_{i,m}^{(k)} X_{m-(k-1)-i} \qquad (1)$$

for k = 1, 2, 3, . . . In eq. 1, $\hat{Y}_{m|m-1,m-k}$ denotes the *predictor* of Y based on values of Y up to month (m − 1) and on values of X up to month (m − k). To allow for seasonality, the subscript m as included in the regression coefficients on the right hand side of eq. 1 with the following equality:

$$a_{i,m}^{(k)} = a_{i,m\pm 12}^{(k)} \quad ; \quad b_{i,m}^{(k)} = b_{i,m\pm 12}^{(k)}$$

The infinite upper limit to the summations is used only for simplicity. Actually there would be only a finite number of coefficients which are different from zero. More generally, for any lead time k, one can write (with the same simplification in writing the upper limit):

$$\hat{Y}_{m|m-k,m-k} = \sum_{i=1}^{\infty} \alpha_{i,m}^{(k)} Y_{m-(k-1)} + \sum_{i=1}^{\infty} \beta_{i,m}^{(k)} X_{m-(k-1)-i} \qquad (2)$$

According to the Hybrid Method, the forecast values $\hat{Y}$ are substituted for the unknown values of Y:

$$\hat{Y}_{m|m-k,m-k} = \sum_{i=1}^{\infty} a_{i,m}^{(k)} \hat{Y}_{m-i|m-k,m-k} + \sum_{i=1}^{\infty} b_{i,m}^{(k)} X_{m-(k-1)-i} \qquad (3)$$

where

$$\hat{Y}_{m-i|m-k,m-k} = Y_{m-i} \quad \text{if} \quad m > k$$

If the coefficients α and β are known for lead times up to k − 1, the necessary coefficients to substitute for $\hat{Y}_{m-i|m-k,m-k}$ are available. Upon this substitution:

$$\hat{Y}_{m|m-k,m-k} = \sum_{i=1}^{k-1} a_{i,m}^{(k)} \left\{ \sum_{j=1}^{\infty} \alpha_{j,m-i}^{(k-i)} Y_{m-(k-1)-j} + \sum_{j=1}^{\infty} \beta_{j,m-i}^{(k-i)} X_{m-(k-1)-j} \right\} + \sum_{i=1}^{\infty} b_{i,m}^{(k)} X_{m-(k-1)-i} \qquad (4)$$

By identifying terms between eqs. 2 and 4, one arrives at the following result:

$$\alpha_{i,m}^{(k)} = \sum_{j=1}^{k-1} a_{j,m}^{(k)} \alpha_{i,m-j}^{(k-j)} + a_{i+k-1,m}^{(k)} \qquad (5)$$

$$\beta_{i,m}^{(k)} = \sum_{j=1}^{k-1} b_{j,m}^{(k)} \beta_{i,m-j}^{(k-j)} + b_{i,m}^{(k)} \qquad (6)$$

By computing the coefficients a and b through regression using eq. 1, one computes α and β recursively using eqs. 5 and 6, with starting values given by

$$\alpha_{i,m}^{(1)} = a_{i,m}^{(1)} \quad ; \quad \beta_{i,m}^{(1)} = b_{i,m}^{(1)}$$

An alternative expression for eq. 2 where the current month is explicitly written as m, is:

$$\hat{Y}_{m+k|m,m} = \sum_{i=1}^{\infty} \alpha_{i,m+k}^{(k)} Y_{m+1-i} + \sum_{i=1}^{\infty} \beta_{i,m+k}^{(k)} X_{m+1-i} \qquad (7)$$

Confidence interval

Analytical derivation for the confidence interval for each forecast may be complicated. O'Connell (1983) proposed the following empirical scheme: For a given lead time, the derived model is applied to all the applicable data points of the historical record, whereby all forecast errors are computed. Let $s_{m,k}$ denote its standard deviation (i.e. the root mean squared error), then the 95% confidence interval is:

$$(\hat{Y}_{m+k|m,m} - 1{,}96 s_{m,k} \quad , \quad \hat{Y}_{m+k|m,m} + 1{,}96 s_{m,k}) \qquad (8)$$

The Thomas-Fiering Model

Let $Q_{i,j}$ denote the flow in month j (j = 1, 2, . . . , 12) of the i^{th} year, with mean $\bar{Q}_j$ and standard deviation S_j, then in terms of the standardized variable

$$Y_{i,j} = (Q_{i,j} - \bar{Q}_j)/S_j \qquad (9)$$

the Thomas-Fiering Model can be expressed as:

$$Y_{i,j+1} = r_{j+1} Y_{i,j} + (1 - r_{j+1}^2)^{1/2} e_{i,j+1} \qquad (10)$$

where r_{j+1} = the correlation coefficient between the flows in months j + 1 and j,

and

$e_{i,j+1}$ = a random variable with mean = 0 and variance = 1.

While additional conditions are imposed upon e to preserve other properties of the historical flows besides the mean, standard deviation and correlation (Phien and Ruksasilp, 1981), this variable is neglected in the forecasting context. As a direct consequence, eq. 12 can provide a forecast of $Y_{i,j+1}$, with lead time equal to one month, made at time j as follows:

$$\hat{Y}_{i,j+1} = r_{j+1} Y_{i,j} \qquad (11)$$

In view of eq. 9, the forecast for $Q_{i,j+1}$ is obtained as:

$$\hat{Q}_{i,j+1} = \bar{Q}_{j+1} + r_{j+1} S_{j+1} (Q_{i,j} - \bar{Q}_j)/S_j \qquad (12)$$

In the form expressed by eq. 10, the Thomas-Fiering model cannot be used to produce forecasts with lead time greater than one month. However, two immediate extensions of the model are readily possible. The first extension is to obtain the forecast with lead time equal to k, made at time j, as:

$$\hat{Y}_{i,j+k} = r_{j+1}^{k} Y_{i,j}$$

or

$$\hat{Q}_{i,j+k} = \bar{Q}_{j+k} + r_{j+1}^{k} S_{j+k} (Q_{i,j} - \bar{Q}_j)/S_j \qquad (13)$$

Since $|r_{j+1}| \leq 1$, r_{j+1}^{k} approaches zero as k increases. Thus, in view of eq. 13, the forecast made several months ahead will be almost equal to the monthly mean $\bar{Q}_{j+k}$.

It is clear that the first extension does not take the seasonality of the correlation structure into account, because r_{j+1} is used for all values of k. The second extension includes the seasonality by adopting the following Hybrid Method which applies first to the standardized variable:

- The forecast with lead time equal to one month is obtained using eq. 11.
- The forecast with lead time = 2 months is obtained by writing

  $$\hat{Y}_{i,j+2} = r_{j+2} Y_{i,j+1}$$

  and then replacing the unknown value $Y_{i,j+1}$ by its forecast:

  $$\hat{Y}_{i,j+2} = r_{j+1} r_{j+2} Y_{i,j} \qquad (14)$$

- The forecast with lead time = 3 months is obtained by the same way, i.e. by writing

  $$\hat{Y}_{i,j+3} = r_{j+3} Y_{i,j+2}$$

  and replacing $Y_{i,j+2}$ by its expression in eq. 14:

  $$\hat{Y}_{i,j+3} = r_{j+1} r_{j+2} r_{j+3} Y_{i,j} \qquad (15)$$

- The forecast with any larger lead time is obtained similarly. Expressing eqs. 14 and 15 using the flow itself gives:

  $$\hat{Q}_{i,j+k} = \bar{Q}_{j+k} + r_{j+1} r_{j+2} \cdots r_{j+k} S_{j+k} (Q_{i,j} - \bar{Q}_j)/S_j \qquad (16)$$

  for k = 1, 2, . . .

Since the second extension can take the seasonality in the correlation into account, it is adopted in this study.

The Sen Model

This model is not as popular as that of Thomas and Fiering. It is written (Sen, 1978) as:

$$W_{i,j} = a_j W_{i,j-1} + b_j W_{i-1,j} + e_{i,j} \qquad (17)$$

where

$$W_{i,j} = Q_{i,j} - \bar{Q}_j \qquad (18)$$

a_j, b_j = parameters which reflect the relationships between successive months of the same year and between successive years for the same month, respectively, and

$e_{i,j}$ = random variable, independent of $W_{i,j-1}$ and $W_{i-1,j}$ with mean = 0.

As in the case of the Thomas-Fiering Model, the random variable e is dropped in forecasting applications. Thus the forecast with lead time equal to one month, made at time j, is given by

$$\hat{W}_{i,j+1} = a_{j+1} W_{i,j} + b_{j+1} W_{i-1,j+1} \qquad (19)$$

or

$$\hat{Q}_{i,j+1} = \bar{Q}_{j+1} + a_{j+1} (Q_{i,j} - \bar{Q}_j) + b_{j+1} (Q_{i-1,j+1} - \bar{Q}_{j+1}) \qquad (20)$$

Forecasts with lead times greater than one month can be obtained by using eq. 19 repeatedly and by replacing the unknown values by their forecasts. Doing so will give:

$$\hat{W}_{i,j+k} = \prod_{m=1}^{k} a_{j+m} W_{i,j} + \prod_{m=1}^{k-1} \left(\prod_{n=1}^{k-m} a_{i+m+n} \right) b_{j+m} W_{i-1,j+m} + b_{j+k} W_{i-1,j+k} \qquad (21)$$

for k = 1, 2, . . .

Although the resulting equation looks rather complicated, the forecasting can be conveniently carried out as follows in practical situations:

- Compute the forecast with lead time equal to one month using eq. 19.
- Use that forecast value to compute the forecast with lead time equal to two months:

  $$\hat{W}_{i,j+2} = a_{j+2} \hat{W}_{i,j+1} + b_{j+2} W_{i-1,j+2} \qquad (22)$$

- Use the forecast with lead time equal to two months to obtain the forecast with lead time equal to three months:

  $$\hat{W}_{i,j+3} = a_{j+3} \hat{W}_{i,j+2} + b_{j+3} W_{i-1,j+3} \qquad (23)$$

  and so on.

However, if the intermediate forecasts are not needed, one can simply bypass by using eq. 21 directly. This seems to be the only advantage which may compensate for its rather complicated form.

Remarks

- The Thomas-Fiering Model only employs the flow in the immediately preceding month; it is therefore a particular case of the Sen Model, where $a_j = r_j S_j / S_{j-1}$ and $b_j = 0$.
- The confidence intervals for the Thomas-Fiering and Sen Models can be derived analytically. However, the empirical scheme of O'Connell (1983) is also adopted in this study.

Applications to actual data

In order to evaluate the performance of the three models under consideration, data from five stations in the Mekong River Basin (Fig. 1) and two stations in the Tachia River Basin (Fig. 2) were employed.

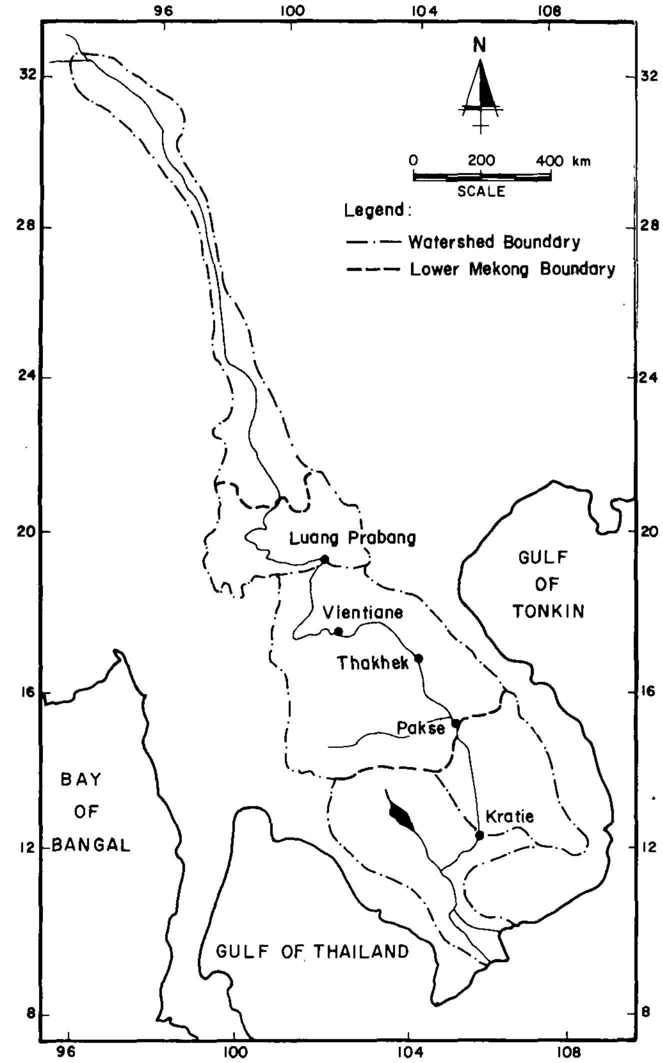

Figure 1
The Mekong River Basin and station locations.

The Mekong River

The Mekong is an international river which starts at an elevation of about 5 000 m in the snow-covered mountain ranges of Tang Ku La on the Great Tibet Plateau. It flows southward for about 1 600 km between the mountain ranges of Yunnan province in China, enters the Indochina Peninsula at the common border of China, Burma and Laos. From this border, it continues to flow to the south for an additional length of 2 400 km and finally discharges into the Bien Dong (South China Sea). The lower section of the river forms part of the boundaries between Burma and Laos, between Thailand and Laos, flows across Cambodia (Kampuchea) and the South of Vietnam.

South-West and North-East Monsoons exert a dominant influence on the climate and rainfall of the Mekong Basin, resulting in the marked wet and dry seasons. The South-West Monsoon is the rainy season, lasting from May to September, especially in the Lower Mekong Basin. During this season, high amounts of moisture are present in the air throughout this lower basin. The magnitude of the rainfall is determined by mechanisms for releasing this moisture. Local wind circulations of small magnitude may produce showers over small areas, while typhoons may set off heavy showers with prodigious amounts of rain over areas as large as 200 000 km².

The data used in this study were obtained from the Mekong Secretariat whose main responsibilities are to develop the water resources of the Lower Mekong Basin. For the five stations listed in Table 1, daily discharge records were available for the periods employed.

The Tachia River

Normally, it is anticipated that seasonal streamflow forecasting is hardly satisfactory for small basins. To emphasize this point, data for two stations in the Tachia River Basin, Taiwan were used.

The Tachia River Basin is located in the central part of Taiwan (Fig. 2), with a drainage area of 1 236 km², which is about 3,4 per cent of the total area of Taiwan. Of the whole basin area, 90 per cent is occupied by mountains with some peaks over 3 000 m high. The river is about 140 km long and flows at an average bed slope of about 1:60.

The mean annual rainfall over the river basin is about 2 340 mm, of which 79 per cent is brought by typhoons during the months May to October while the remaining 21 per cent is brought by the North-East Monsoon lasting from November to March.

The data employed in this study were provided by the Taiwan Power Company. All the data are available on a 10-day basis for the two stations considered (Table 2).

For all the stations under consideration, the data in the final year of the record employed are used for comparison between the observed and forecast values, while all the data prior to that year are used in the model building and parameter estimation (simply referred to as the calibration stage).

General observations

In the following, the three models are considered according to their order of complexity: the Thomas-Fiering Model, the Sen

TABLE 1
STREAMFLOW STATIONS ON THE MEKONG RIVER BASIN

Station	Location	Drainage area (km²)	Period of record used
1. LP	Luang Prabang	268 000	1950 – 1981
2. VT	Vientiane	299 000	1946 – 1981
3. TK	Thakhek	373 000	1946 – 1981
4. PS	Pakse	545 000	1945 – 1981
5. KT	Kratie	646 000	1945 – 1968*

*Data at this station were not available after 1968

TABLE 2
DATA FROM THE TACHIA RIVER

Station	Sung-Mao	Tien-Lun
Drainage Area (km²)	417	950
Streamflow	1954 – 1976	1954 – 1976
Rainfall	1952 – 1976	1952 – 1976
Evaporation	1957 – 1976	1957 – 1976
Local inflow	– not available –	1954 – 1976

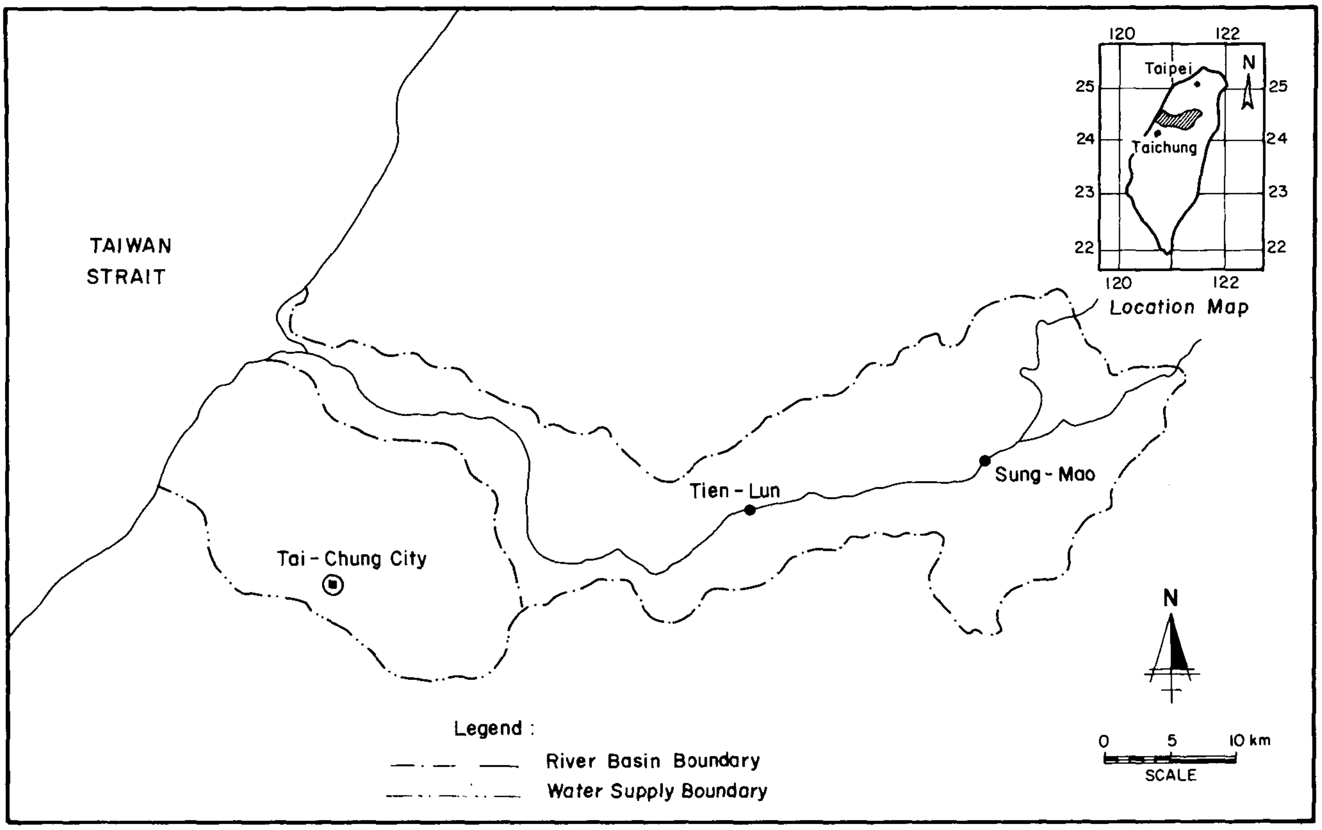

Figure 2
Tachia River Basin and station locations.

Model, and the Hybrid Model.

In the Thomas-Fiering Model, the correlation coefficient of the flows in any two successive months plays the most important role, as revealed by eq. 10. The values of this coefficient are collected in Table 3 for all the stations considered.

For the Sen Model, besides the correlation coefficient of the flows in two successive months, the lag-one serial correlation between the flows of the same month in two successive years is involved. This is quite similar to the seasonal Box-Jenkins model (Phien and Twu, 1984). In addition, the lag-one cross correlation between month j in one year and month j-1 in the preceding year is also included. Nevertheless, their contribution can be summarized in two parameters a and b, whose values are collected in Table 4 for easy reference. From this table, it is seen that in almost all cases, a has a much larger magnitude than does b, indicating that the contribution of the flow in the immediately preceding month is much more dominant than that of the flow in the same month in the preceding year. This should be considered to be obvious when one refers to the basin storage concept.

TABLE 3
CORRELATION COEFFICIENT OF SUCCESSIVE MONTHLY FLOWS

Station	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Luang Prabang	0,89	0,89	0,85	0,73	0,62	0,52	0,43	0,64	0,61	0,54	0,59	0,75
Vientiane	0,88	0,88	0,89	0,76	0,77	0,47	0,46	0,64	0,62	0,55	0,42	0,73
Thakhek	0,82	0,90	0,82	0,69	0,30	0,31	0,49	0,52	0,42	0,65	0,27	0,51
Pakse	0,69	0,93	0,86	0,72	0,34	0,25	0,55	0,41	0,24	0,53	0,56	0,59
Kratie	0,84	0,89	0,90	0,70	0,66	0,54	0,71	0,30	0,37	0,41	0,65	0,81
Sung-Mao	0,53	0,51	0,63	0,77	0,55	0,24	−0,04	−0,02	0,18	0,53	0,02	0,46
Tien-Lun	0,57	0,39	0,65	0,81	0,62	0,34	0,11	0,06	0,31	0,45	0,15	0,33

Notes (1) for the correlation between January and December (of the preceding year).
(2) for the correlation between February and January.
(3) for the correlation between March and February.
.
.
.
(12) for the correlation between December and November.

TABLE 4
VALUES OF THE PARAMETERS IN THE SEN MODEL

Station		Month j											
		1	2	3	4	5	6	7	8	9	10	11	12
Luang Prabang	a	0,40	0,61	0,71	0,70	1,23	1,41	0,89	1,03	0,56	0,26	0,55	0,37
(LP)	b	0,05	−0,26	−0,23	−0,16	−0,07	−0,11	0,06	0,01	0,00	0,24	−0,14	0,06
Vientiane	a	0,42	0,57	0,75	0,86	0,71	1,12	0,95	0,89	0,69	0,29	0,28	0,34
(VT)	b	0,16	−0,14	−0,23	−0,19	−0,19	−0,07	0,13	0,01	0,05	0,24	0,08	0,10
Thakhek	a	0,47	0,62	0,62	0,87	0,83	1,05	0,69	0,64	0,46	0,43	0,12	0,23
(TK)	b	0,26	−0,16	−0,05	−0,25	0,04	0,01	0,29	−0,06	0,00	0,27	0,06	0,08
Pakse	a	0,41	0,61	0,49	0,92	1,18	0,82	0,86	0,49	0,26	0,44	0,20	0,24
(PS)	b	0,23	−0,12	−0,11	−0,29	0,00	−0,01	0,19	−0,03	−0,14	0,30	0,28	0,11
Kratie	a	0,39	0,59	0,55	0,72	1,18	1,51	1,13	0,29	0,33	0,43	0,26	0,42
(KT)	b	0,01	−0,10	−0,15	−0,02	−0,35	−0,05	−0,02	0,09	−0,06	0,12	0,40	−0,21
Sung-Mao	a	0,66	1,12	0,72	0,66	0,61	1,09	0,01	−0,12	0,21	0,23	0,01	0,17
(SM)	b	−0,08	−0,03	−0,02	0,00	0,01	−0,15	−0,23	−0,13	0,01	0,04	−0,11	−0,04
Tien-Lun	a	0,59	0,93	0,76	0,75	0,78	1,84	0,04	−0,09	0,32	0,16	0,07	0,18
(TL)	b	−0,04	−0,05	−0,04	−0,02	−0,11	−0,15	−0,11	−0,19	0.09	−0,02	0,08	0,02

Notes: 1 = January, 2 = February, ..., 12 = December

Results for calibration stage

The Thomas-Fiering Model

Using the empirical scheme suggested by O'Connell (1983), the root mean squared error was computed. Although this statistic is useful in determining the confidence intervals of forecasts, it does not give much information on the accuracy of the forecasts. It is more appropriate then to express it as a percentage of the mean monthly flow for each lead time k:

$$d_{j,k} = 100 s_{j,k}/\bar{Q}_j \quad , \quad j = 1, \ldots, 12 \qquad (24)$$

This quantity may be, for convenience, referred to as the *relative error with respect to the mean*. Typical values of this relative error are collected in Table 5 for Luang Prabang for lead times from one to six months. Except for some irregular cases, the relative error increases with the lead time as commonly expected. When the mean of the twelve month values is taken, all the irregularities disappear. This is clearly seen from Table 6 where the means are collected for all stations and lead times considered.

TABLE 5
RELATIVE ERROR WITH RESPECT TO THE MEAN MONTHLY FLOW (IN PERCENTAGE) FOR THE THOMAS-FIERING MODEL AT LUANG PRABANG

Month	Lead time (months)					
	1	2	3	4	5	6
Jan	8,4	12,7	15,5	17,8	18,4	18,5
Feb	7,2	10,2	12,0	13,0	15,0	15,7
Mar	8,4	10,4	11,2	13,2	13,9	15,2
Apr	9,6	10,5	11,1	11,6	12,5	13,0
May	15,7	18,4	18,7	18,4	18,1	19,0
Jun	20,8	23,8	23,6	23,2	23,3	20,7
Jul	22,9	23,5	25,2	25,0	25,1	25,0
Aug	18,4	21,8	23,6	24,1	23,8	23,7
Sep	19,9	22,0	23,5	24,3	24,5	24,6
Oct	18,4	21,8	21,0	20,9	21,2	21,4
Nov	25,3	30,2	30,6	29,8	30,6	30,7
Dec	18,2	23,2	26,4	26,4	26,4	26,9
Mean	16,1	19,0	20,2	20,6	21,1	21,2

TABLE 6
MEAN OF THE RELATIVE ERROR WITH RESPECT TO THE MEAN MONTHLY FLOW (IN PERCENTAGE) FOR THE THOMAS-FIERING MODEL

Station	Lead time (months)					
	1	2	3	4	5	6
Luang Prabang	16,1	19,0	20,2	20,6	21,1	21,2
Vientiane	16,9	19,9	21,1	21,6	22,0	22,3
Thakhek	16,5	18,9	19,5	19,8	20,0	20,1
Pakse	16,5	19,3	19,8	20,0	20,2	20,3
Kratie	14,3	17,4	18,1	18,3	18,5	18,8
Sung-Mao	57,2	61,3	63,0	63,8	64,1	64,4
Tien-Lun	47,0	51,3	52,7	53,2	53,2	53,6

The Sen Model

Using the empirical scheme, the root mean squared error was computed. The relative error with respect to the mean was then obtained using eq. 24. Typical values are collected in Table 7 for Luang Prabang. Except for some irregular cases, one may say that d has an increasing magnitude with an increasing lead time. As for the Thomas-Fiering Model, when the mean of the monthly values is taken, the irregularities no longer exist. The values collected in Table 8 support this claim.

The Hybrid Model

For the Hybrid Model, data other than the monthly flow are required, because this model can incorporate several explanatory variables. Prior to the application of the model, the explanatory variables must be selected. This selection could be carried out efficiently using the autocorrelation function of the monthly flow and the cross correlation functions of it with other variables. All the variables including the monthly flow may be subject to different transformations as needed. In this study, no suitable transformations were found (to increase the values of the autocorrelation and cross correlation functions). The variables considered (partly due to the availability of required data) are listed in Table 9 for the Mekong River and Table 10 for the Tachia River. For the Mekong River, other variables such as the maximum daily flow and minimum daily flow with different lags were also considered. However, these were found to be irrelevant due to low values of the cross-correlation coefficient. The same remark

TABLE 7
RELATIVE ERROR WITH RESPECT TO THE MEAN MONTHLY FLOW (IN PERCENTAGE) FOR THE SEN MODEL AT LUANG PRABANG

Month	Lead time (months) 1	2	3	4	5	6
Jan	8,3	12,4	15,8	17,3	17,8	17,8
Feb	8,4	11,7	13,8	15,5	17,1	17,7
Mar	8,6	12,3	13,7	15,9	17,1	18,5
Apr	9,5	10,5	12,1	13,2	14,4	15,3
May	15,7	18,6	19,1	19,3	19,4	20,8
Jun	20,6	23,7	23,9	23,4	23,9	21,5
Jul	22,9	23,5	25,3	25,0	25,0	24,8
Aug	18,4	21,6	23,6	24,1	23,9	23,7
Sep	19,9	22,0	23,5	24,3	24,6	24,6
Oct	17,6	20,6	19,9	19,8	20,1	20,3
Nov	25,6	29,8	30,4	29,9	30,6	30,8
Dec	17,9	23,4	25,9	25,9	25,9	26,4
Mean	16,1	19,2	20,6	21,1	21,7	21,8

TABLE 8
MEAN OF THE RELATIVE ERROR WITH RESPECT TO THE MEAN MONTHLY FLOW (IN PERCENTAGE) FOR THE SEN MODEL

Station	Lead time (months) 1	2	3	4	5	6
Luang Prabang	16,1	19,2	20,6	21,1	21,7	21,8
Vientiane	16,7	19,7	21,1	21,5	21,9	22,2
Thakhek	16,3	18,6	19,2	19,5	19,6	19,8
Pakse	16,3	19,0	19,5	19,7	19,9	20,0
Kratie	14,0	17,3	18,2	18,6	19,0	19,4
Sung-Mao	56,8	60,9	62,7	63,4	63,7	64,0
Tien-Lun	46,7	51,3	52,7	53,2	53,4	53,5

TABLE 9
INPUT VARIABLES CONSIDERED FOR THE STATIONS IN THE MEKONG RIVER BASIN

Variable	Description
1	Monthly flow lagged by one month
2	Average flow of the last three days in a month lagged by one month
3	Flow of the last day in a month lagged by one month
4	Monthly flow lagged by two months
5	Average flow of the last three days in a month lagged by two months
6	Flow of the last day in a month lagged by two months
7	Monthly flow lagged by three months
8*	Monthly rainfall lagged by one month
9*	Monthly rainfall lagged by two months

*Applicable only for Vientiane

applies to the Tachia River as well, in which case, even more input variables such as the monthly evaporation, last 10-day evaporation, etc. were considered.

With the input variables so selected, the stepwise regression procedure was used to select the explanatory variables. As mentioned earlier, the Hybrid Model (like the other two models) incorporates the seasonality of the monthly flow. Consequently, for each lead time, there is a corresponding equation for each month. It was found that in almost all cases, not more than three explanatory variables were finally selected. The results expressed in terms of the portion of variance explained by the model, R^2 and the relative error with respect to the mean, d, are summarized in Table 11 for a lead time equal to one month.

From the results obtained for the Thomas-Fiering, Sen and Hybrid Models, one can draw the following observations:

- When only monthly streamflow data are employed, i.e. in the Thomas-Fiering and Sen Models, the relative error with respect to the mean is less than 23% for all the stations in the Mekong Basin. This indicates that both models perform satisfactorily at stations with large drainage areas. Inspection of individual values of this relative error reveals that the performance is even more satisfactory for the months January to April, i.e. the months of the dry season. For the two stations in the Tachia Basin, the relative error is large, being greater than 50% in most cases, indicating that the performance of these models is unsatisfactory for small drainage areas.

- A closer examination of Tables 6 and 8 reveals that no significant improvement is achieved when the Sen Model is used in place of the Thomas-Fiering Model.

- For the five stations in the Mekong Basin, the relative error with respect to the monthly standard deviation, in almost all cases, is less than 100% for all the lead times considered (Table 12). Thus monthly forecasts with mean squared error less than the historical monthly variance would be expected for lead times up to six months. This compares very well with the result obtained by Rao and Kashyap (1973) or that by McKerchar and Delleur (1972).

- When the Hybrid Model is applied to the Tachia River Basin, meaningful equations cannot be established for a few months, regardless of the fact that more input data are available. The results obtained by the Hybrid Model for this basin show no improvement upon those obtained by the previous two models (Tables 6 and 8).

  For the Mekong River Basin, it seems that significant improvement has been made.

- When the lead time is greater than one month, to conform

TABLE 10
INPUT VARIABLES CONSIDERED FOR THE STATIONS ON THE TACHIA RIVER BASIN

Variable	Description
1	Monthly flow lagged by one month
2	Monthly flow lagged by two months
3	Monthly flow lagged by three months
4	Last 10-day flow in a month lagged by one month
5	Last 10-day flow in a month lagged by two months
6	Monthly rainfall lagged by one month
7	Monthly rainfall lagged by two months
8	Last 10-day rainfall in a month lagged by one month
9	Last 10-day rainfall in a month lagged by two months
10*	Last 10-day local flow in a month lagged by one month

*Applicable only for Tien-Lun

with the formulation in Problem 2, Problem 3, . . . force-fitting is needed in order to explicitly incorporate the first explanatory variable (i.e. monthly flow in preceding months).

- Inspection of the values of d in Tables 5, 7 (for lead time = one month) and 11 reveals that the Hybrid Model is better than both the Thomas-Fiering and Sen Models in all the months of the year, except May. In terms of the average values of d, the Hybrid Model is capable of reducing it from 16,1 (for both simplified models) to 13,5, i.e. a reduction of 16%.

- From Tables 6, 8 (with lead time = one month) and 11 the following points should be noted:
  - For Vientiane, the Hybrid Model reduces the value of d from 16,9 in the Thomas-Fiering Model, or 16,7 in the Sen Model, to 12,9.
  - For Thankhek, the reduction is from 16,5 (Thomas-Fiering) or 16,3 (Sen) to 14,0.
  - For Pakse, the reduction is from 16,5 (Thomas-Fiering) or 16,3 (Sen) to 14,5.
  - For Kratie, it is from 14,3 (Thomas-Fiering) or 14,0 (Sen) to 13,3.

  In all these cases, a reduction of at least 5 per cent has been made.

Results for verification stage

As mentioned previously, the data in the final year of the record at each station were used in the comparison between the forecast and actual monthly flows, where the maximum lead time was set equal to 3 months. In the evaluation of the models' performance, the following statistics were used:

- Maximum relative error:

$$RE_{max} = \max_{1 \leq j \leq 12} \left| \frac{Y_j - \hat{Y}_j}{Y_j} \right| \qquad (25)$$

- Mean relative error:

$$MRE = \frac{1}{12} \sum_{j=1}^{12} \frac{Y_j - \hat{Y}_j}{Y_j} \qquad (26)$$

- Mean absolute relative error:

$$MAD = \frac{1}{12} \sum_{j=1}^{12} \left| \frac{Y_j - \hat{Y}_j}{Y_j} \right| \qquad (27)$$

- Mean of squared relative errors:

$$MSE = \frac{1}{12} \sum_{j=1}^{12} \left(\frac{Y_j - \hat{Y}_j}{Y_j} \right)^2 \qquad (28)$$

In eqs. 25 to 28, Y_j and $\hat{Y}_j$ denote the actual monthly flow and the forecast flow in the final year of the record used.

- For the five stations in the Mekong Basin, the results collected in Table 13 show that all the models perform satisfactorily. However, for the two stations in the Tachia Basin, all the statistics considered have very large magnitudes. The maximum relative error always exceeds 1, indicating that relative errors exceeding 100% may very well occur. In other words, all three models considered perform very badly for this river.

- For the Mekong Basin (Table 13), a rather clear improvement may be seen as one moves from a simpler model to a more complicated model. However, this improvement may not be very significant to compensate for the additional effort exerted in data collection, in model identification, and in parameter estimation, especially for the case of the Hybrid Model. For the Tachia Basin, no such improvement is found (Table 14).

- It seems from Tables 13 and 14 that a longer period (three to four years) may be needed to have more reliable estimates of the statistics used. There are quite a large number of irregularities (such as the decreasing value of a statistic corresponding to an increase in lead time) in these two tables. Consequently, the results collected in this verification stage should not be generalized.

TABLE 11
SUMMARIZED RESULTS FOR THE HYBRID MODEL CORRESPONDING TO A LEAD TIME EQUAL TO ONE MONTH

Station		Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Mean
Luang Prabang	(a)	88,6	87,9	71,0	66,7	28,9	77,5	61,7	49,4	66,7	56,3	76,1	77,0	67,3
	(b)	6,6	5,6	8,2	7,7	25,7	14,7	17,2	16,9	14,3	16,2	15,2	13,3	13,5
Vientiane	(a)	86,0	81,8	83,2	73,6	46,9	52,3	58,9	41,5	67,1	64,3	80,5	72,8	67,4
	(b)	6,9	6,3	6,1	8,4	19,8	20,1	18,5	16,7	13,9	13,6	12,7	12,6	12,9
Thakhek	(a)	85,9	77,1	69,3	54,0	30,3	31,3	46,0	37,8	58,7	57,7	49,4	65,2	55,2
	(b)	5,7	6,6	7,4	10,4	24,2	26,7	18,7	15,3	13,2	14,9	15,5	10,3	14,0
Pakse	(a)	82,7	90,4	74,7	65,8	26,4	32,2	40,5	28,3	40,1	71,0	53,9	67,1	56,1
	(b)	7,3	5,1	6,7	9,3	26,1	27,9	21,3	18,2	14,8	13,3	13,7	9,0	14,5
Kratie	(a)	86,2	86,2	89,6	55,2	49,7	13,8	63,7	22,2	39,0	67,9	39,1	88,9	58,5
	(b)	6,1	5,8	4,5	8,6	23,9	31,6	16,1	16,6	11,4	13,2	15,9	6,5	13,3
Sung-Mao	(a)	4,4	18,9	41,8	48,2	25,7	8,9	3,3	51,2	9,0	10,2	31,7	29,7	25,6
	(b)	37,6	67,7	46,9	45,1	82,3	90,3	63,7	55,2	68,9	71,7	58,1	35,3	60,1
Tien-Lun	(a)	34,1	11,9	43,6	50,3	36,8	24,7	3,0	25,3	19,5	21,2	31,9	67,8	30,8
	(b)	29,9	63,2	41,3	39,6	45,2	72,3	42,4	60,9	60,6	47,2	35,3	17,9	46,3

Notes – (a): Portion of variance explained by the selected model, R^2 (percentage)
(b): Relative error with respect to the mean, d (percentage)

Conclusions

The Hybrid Method introduced by O'Connell (1983) was used in this study for forecasting the monthly streamflows at five stations in the Mekong River Basin and two stations in the Tachia River Basin. The first five stations have large drainage areas while the remaining two have small drainage areas. Basically, the Hybrid Method is just a combination of regression analysis and time series approach, the former being used for model identification and parameter estimation and the latter for providing forecasts with lead time greater than one month. The Thomas-Fiering Model (which can be treated as a regression problem in which the flow in a month is expressed in terms of the flow of the intermediately preceding month) and the Sen Model (in which the flow in a month is regressed upon the flow of the intermediately preceding month and the flow of that month in the intermediately preceding year) were also considered in the Hybrid content by introducing suitable extensions. From the applications, the following conclusions may be drawn:

- for streamflow stations with large drainage areas, satisfactory forecasts can be obtained by the use of the Thomas-Fiering,

TABLE 12
MEAN OF THE RELATIVE ERROR WITH RESPECT TO THE MONTHLY STANDARD DEVIATION (IN PERCENTAGE) FOR THE STATIONS IN THE MEKONG RIVER BASIN

Station		Lead time (months) 1	2	3	4	5	6
Luang Prabang	(T)	68,7	81,6	87,1	90,1	92,5	93,6
	(S)	68,9	82,8	98,9	93,5	96,4	97,9
Vientiane	(T)	66,5	78,9	84,3	86,8	88,9	90,0
	(S)	65,6	78,1	84,1	86,5	88,4	89,5
Thakhek	(T)	79,2	91,9	96,2	99,0	100,1	100,9
	(S)	78,0	90,7	94,6	97,1	98,2	99,0
Pakse	(T)	76,3	91,0	95,0	96,4	97,3	97,6
	(S)	75,6	89,8	93,4	94,9	95,9	96,3
Kratie	(T)	68,6	84,5	88,9	90,8	91,7	93,3
	(S)	67,5	83,6	89,0	92,1	93,9	96,1

Notes: (T) for Thomas-Fiering Model
(S) for Sen Model

TABLE 13
VALUES OF THE STATISTICS USED IN EVALUATING THE MODELS' PERFORMANCE (VERIFICATION STAGE FOR THE MEKONG RIVER)

Station		REmax (1)	REmax (2)	REmax (3)	MRE (1)	MRE (2)	MRE (3)	MAD (1)	MAD (2)	MAD (3)	MSE (1)	MSE (2)	MSE (3)
Luang Prabang	(T)	0,50	0,50	0,50	0,08	0,13	0,14	0,11	0,18	0,18	0,08	0,06	0,06
	(S)	0,50	0,50	0,49	0,07	0,11	0,12	0,11	0,17	0,18	0,03	0,06	0,06
	(H)	0,39	0,47	0,47	0,06	0,12	0,13	0,11	0,17	0,19	0,03	0,05	0,06
Vientiane	(T)	0,36	0,40	0,41	0,07	0,10	0,12	0,13	0,18	0,19	0,03	0,05	0,05
	(S)	0,37	0,41	0,44	0,07	0,10	0,12	0,13	0,20	0,21	0,03	0,05	0,06
	(H)	0,34	0,38	0,38	0,05	0,07	0,09	0,11	0,14	0,21	0,02	0,04	0,06
Thakhek	(T)	0,30	0,38	0,38	0,08	0,11	0,13	0,13	0,15	0,16	0,03	0,04	0,04
	(S)	0,30	0,38	0,38	0,09	0,12	0,13	0,14	0,16	0,17	0,03	0,04	0,04
	(H)	0,38	0,37	0,38	0,15	0,11	0,12	0,15	0,13	0,16	0,04	0,03	0,04
Pakse	(T)	0,45	0,48	0,49	0,09	0,12	0,14	0,14	0,19	0,20	0,03	0,05	0,06
	(S)	0,45	0,48	0,49	0,09	0,13	0,15	0,13	0,19	0,21	0,03	0,05	0,06
	(H)	0,22	0,45	0,46	0,08	0,11	0,14	0,09	0,15	0,19	0,01	0,04	0,05
Kratie	(T)	0,62	0,51	0,69	– ,09	– ,13	– ,17	0,14	0,19	0,21	0,05	0,06	0,08
	(S)	0,63	0,52	0,75	– ,10	– ,11	– ,13	0,16	0,19	0,19	0,05	0,06	0,08
	(H)	0,64	0,64	0,64	– ,06	– ,14	– ,15	0,13	0,16	0,18	0,05	0,06	0,07

Notes: (1) Lead time = 1 month, (2) Lead time = 2 months, (3) Lead time = 3 months
(T): Thomas-Fiering; (S): Sen; (H): Hybrid

TABLE 14
VALUES OF THE STATISTICS USED IN EVALUATING THE MODELS' PERFORMANCE (VERIFICATION STAGE FOR THE TACHIA RIVER)

Station		REmax (1)	REmax (2)	REmax (3)	MRE (1)	MRE (2)	MRE (3)	MAD (1)	MAD (2)	MAD (3)	MSE (1)	MSE (2)	MSE (3)
Sung-Mao	(T)	2,39	2,44	2,45	– ,59	– ,63	– ,63	0,75	0,82	0,81	0,97	1,06	1,01
	(S)	2,39	2,39	2,43	– ,52	– ,61	– ,59	0,67	0,79	0,78	0,89	1,00	0,96
	(H)	2,17	2,17	2,17	– ,48	– ,61	– ,64	0,69	0,77	0,79	0,81	0,93	0,94
Tien-Lun	(T)	1,89	1,97	1,90	– ,37	– ,37	– ,38	0,53	0,55	0,55	0,58	0,54	0,52
	(S)	1,85	1,78	1,82	– ,34	– ,32	– ,33	0,52	0,51	0,52	0,56	0,45	0,46
	(H)	4,62	1,11	1,68	– ,55	– ,27	– ,40	0,70	0,55	0,53	1,92	0,26	0,46

Notes: as in Table 13

Sen and Hybrid Models. Significant improvement in the forecasts may be expected when the first two models are replaced by the last model.

- for all these models, very good forecasts can be expected for the dry months, January to April, for the Mekong River.

- for the stations with small drainage areas (Sung-Mao and Tien-Lun), the relative error is always large.

It seems that the forecasting of monthly flows for small drainage areas is a difficult task.

Acknowledgements

Sincere thanks are extended to the Mekong Secretariat and the Taiwan Power Company for the provision of the data employed. Deep gratitude is expressed to the Asian Institute of Technology for having allowed the use of its computer facilities.

References

DYHR-NIELSON, M. (1982) Long-range water-supply forecasting. World Meteorological Organization, WMO, No. 578, Geneva.

HINO, M. and ISHIKAWA, K. (1975) Long- and short-term forecasting of river discharge. *Proc. Japan Soc. Civil Eng.* **236** 59-70.

McKERCHAR, A.I. and DELLEUR, J.W. (1972) Evaluation of seasonal time series model. Application to Mid-West River flow data. Symp. on Uncertainties in Hydrology, Water Resources Systems, Tucson.

O'CONNELL, P.E. (1983) An assessment of seasonal streamflow forecasting techniques. Seminar on Seasonal Streamflow Forecasting, Vientiane, Laos.

PHIEN, H.N. (1983a) Introduction to seasonal streamflow forecasting techniques. Seminar on Seasonal Streamflow Forecasting, Vientiane, Laos.

PHIEN, H.N. (1983b) Seasonal streamflow forecasting with Kalman filter and Group Method of data handling. Seminar on Seasonal Streamflow Forecasting, Vientiane, Laos.

PHIEN, H.N. and RUKSASILP, W. (1981) A review of single-site models for monthly streamflow generation. *Journal of Hydrology* **52** 1-12.

PHIEN, H.N. and TWU, C.T. (1984) Forecasting of seasonal streamflows with Box-Jenkins models. *Water SA* **10**(3) 127-135.

RAO, R.A. and KASHYAP, R.L. (1973) Analysis, construction and validation of stochastic models for monthly river flows. School of Civil Eng., Purdue Univ., West Lafayette, Indiana.

RAO R.A. and KASHYAP, R.L. (1974) Stochastic modeling of river flows. *IEEE Trans. Automatic Control*, AC-19.

SEN, Z. (1978) A mathematical model of monthly flow sequences. *Hydrological Science Bulletin* **23** 223-229.

THOMAS, Jr., H.A. and FIERING, M.B. (1962) Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In: A.A. Mass *et al.* (Editors) *Design of Water Resource Systems.* Chap. 12. Harvard University Press, Cambridge, Mass., pp. 459-493.