# FLOOD FORECASTING IN THE KALU GANGA BASIN USING NUMERICAL FILTERING TECHNIQUES

by

**P.N. Wickramanayake**

## Introduction

The need for reliable flood forecasting arises wherever there is a possibility of damage due to the rapid increase in the discharge of a river. If the discharge (and therefore the level of water) in a river can be forecast with reasonable accuracy then the resulting damages can be avoided by taking suitable precautionary measures. This study presents the results of forecasting the flow at the gauging station at Putupaula, which is the lowest station on the Kalu Ganga (see Fig. 1), for three historical floods. It is a continuation of an earlier study of flood forecasting in the Ratnapura catchment in the upper reaches of the Kalu Ganga (Amirthanathan (1)).

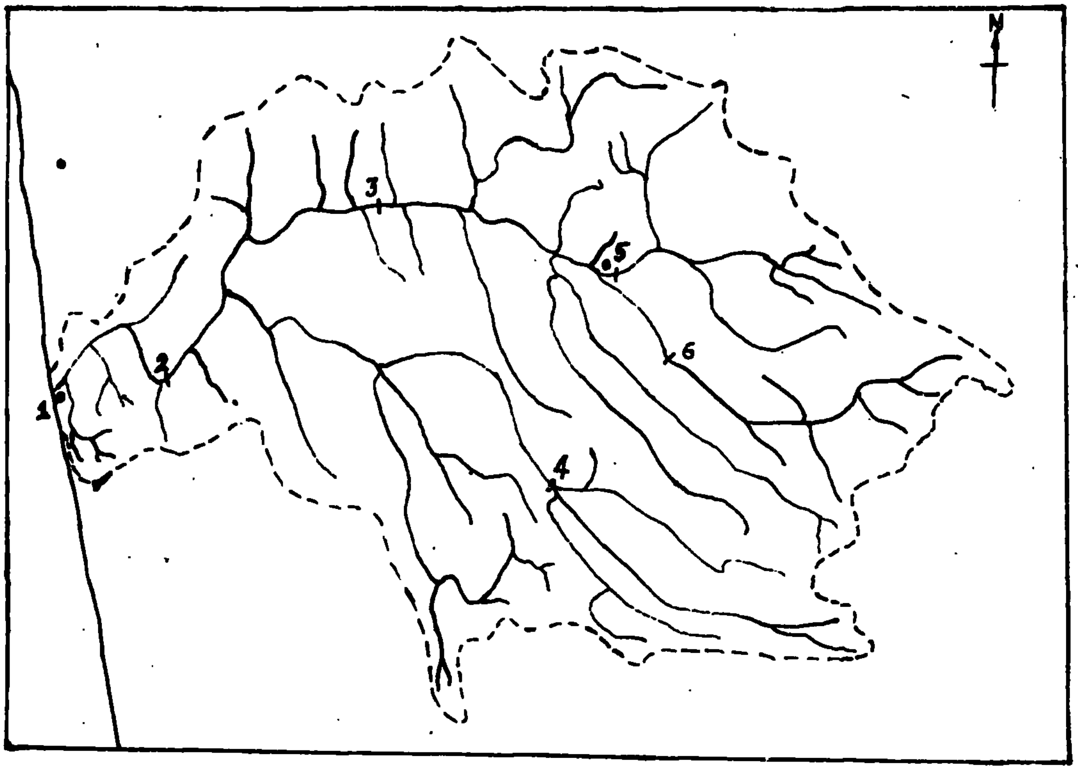
The Kalu Ganga begins in the area around Adam's Peak and discharges into the sea at Kalutara. The average annual rainfall in the basin is about 4100 mm (160") of which more than half falls during May and June. Though only 80 miles long, with a drainage area of 1023 sq. miles, it discharges more water than any other river in Sri Lanka. The river has steep upper reaches and very gentle lower reaches. For instance, at Ratnapura which is over 50 miles from the sea, the river is only about 50 feet above sea level. As a result, flooding is frequent particularly in the lower reaches (Arumugam (2)).

Forecasting is usually done using a mathematical model of the river basin which uses inputs such as rainfall and river flows at various points. Theoretically, it is possible to use a hydrologic model which takes into account all the relevant basin characteristics for this purpose. However, the large number of parameters and data required for a river basin and the fact that not all the processes involved are well understood make this approach impracticable. The movement of a flood wave down a river can also be forecast by solving the dynamic equations governing its motion but this method requires a detailed knowledge of the river geometry and is time consuming and costly. A simpler method is to adopt a time series approach where the discharge at a certain time step is related to other relevant parameters by a simple linear equation. The problem then is to estimate the coefficients of this equation as accurately as possible.

Earlier methods using this approach were based on the last observed value. Later, methods such as moving averages and exponential smoothing were used. A further development were the Box-Jenkins models which include the Auto-regressive Moving Average (ARMA) model as a special case. These provide a more flexible approach to the problem (Kottegoda (3)). However, they have the drawback

Mr. P.N. Wickramanayake graduated with First Class Honours in Civil Engineering and worked as an Assistant Lecturer at Department of Civil Engineering, University of Peradeniya.

He is now doing his post graduate studies at MIT in the U.S.A.




- Gauging Station - Town

1 - Kalutara 2 - Putupaula 3 - Ellagawa

4 - Kukulegama 5 - Ratnapura 6 - Dela

FIGURE 1

Map of the Kalu Ganga Basin

of requiring fixed model coefficients. In this study an ARMA model is used with a method that permits these coefficients to vary.

For flood forecasting, where it is required to predict an event during its occurrence (known as on-line forecasting), short term performance is much more important than that in the long term. On-line forecasting should use the flood measurements as they become available to improve the model. Therefore allowing for variation in the model coefficients enables the model to evolve under the flood conditions. In this way the model can cope with short term non-stationary behaviour. These models are known as adaptive models.

In this study the Recursive Least Squares (RLS) method and the Kalman Filter method were used for forecasting. The Kalman Filter has the added advantage that it provides an idea of the accuracy of the parameter estimation.

## Data

Figure 1 gives a plan of the Kalu Ganga basin with the five flow gauging stations and the town of Kalutara. The hourly gauge height records at these stations are available at the Hydrology Division of the Irrigation Department which maintains these stations. Even though records

| Flood Number | Date of first reading 0000 hrs. on | Date of last reading 1800 hrs. on | Number of Readings | Peak Flow (m/s) |
|---|---|---|---|---|
| 1 | 24-09-1986 | 12-10-1986 | 68 | 842 |
| 2 | 09-11-1985 | 19-11-1985 | 44 | 793 |
| 3 | 29-09-1985 | 12-10-1985 | 56 | 926 |
| 4 | 15-05-1985 | 04-07-1985 | 204 | 908 |
| 5 | 01-07-1984 | 24-07-1984 | 96 | 968 |

TABLE 1

Floods chosen for the study

have been maintained at Putupaula for over 40 years it was decided to select only the floods given in Table 1 for the study. The reasons for this were that prior to 1984 no data was available for Kukulegama and the recording of flood stages at some of the other stations was incomplete and in some cases unreliable. The data for floods 4 and 5 were also incomplete as the readings from 1700 hours to 0500 hours had not been recorded at Putupaula.

The time step chosen for the study was 6 hours. The 6-hourly flow readings (in steps from 0000 hours) for all five stations during the selected floods were obtained from the Irrigation Department and the gaps filled using polynomial interpolation.

The flood flows were calculated from the gauge heights using the given rating curves. These curves have been established at dates ranging from 1976 (for Putupaula) to 1984 (for Kukulegama). Only occasional checks have been carried out. Furthermore, the calibration readings have been taken mostly in the middle of the flow range with the resulting curve being extended for high and low flows. Therefore due to the error present in establishing the curve and due to the presence of two different rating curves for rising and falling stages during floods (loop rating curve) appreciable error can be present in the observed flows. In other words the actual flows are observed in the presence of a high degree of noise.

## Model Formulation

Consider the catchment shown in Figure 2. Q(t) is the observed flow at the downstream outlet of the basin and is due to previous rainfall over the basin. Adopting a decomposition approach to the problem we can consider Q(t) as made up of separate contributions due to the rainfall over each of the sub-basins 1, 2 ....n. It is assumed that unit impulse rainfall over the i th sub-basin will result in a flow at the outlet given by the impulse response function $h_i(t)$. Therefore a general expression for Q(t) could be written as

$$Q(t) = \sum_{i=1}^{n} \int_{0}^{t-b_i} h_i(\tau)*p_i(t-\tau-b_i)d\tau + N(t) \qquad (1)$$

where,

- $Q(t)$ - measured discharge at outlet
- $h_i(t)$ - impulse response function for rainfall in sub-basin i
- $p_i(t)$ - lumped rainfall for sub-basin i
- $b_i$ - time lag for first significant value of $h_i$
- $N(t)$ - total noise
- n - no. of sub-basins

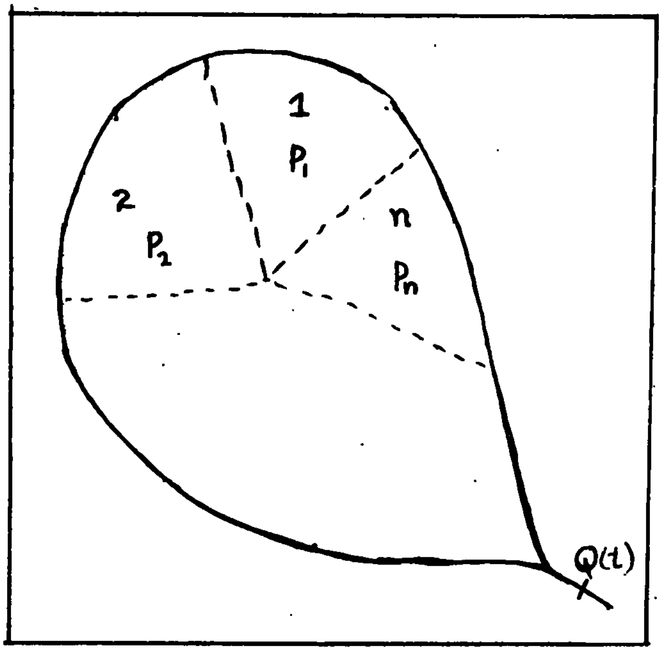



FIGURE 2

The noise term accounts for the error introduced in assuming the model to represent the real behaviour as well as the measurement error.

If all the functions in (1) are sampled at a discrete time interval $\Delta t$ the equation for discharge at time step k could be written as

$$Q(k) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} h_i(j)*p_i(k - j - b_i)* \Delta t + N(k) \qquad (2)$$

where $m_i$ is the number of significant impulse responses for rainfall in sub-basin i.

If instead of the rainfall we have the outflow from each sub-basin a different form of the model can be developed by considering Q(t) as made up of these outflows (see Fig. 3). A unit impulse outflow from sub-basin i produces an impulse response function $g_i(t)$ at the outlet. The general expression for Q(t) is

$$Q(t) = \sum_{i=1}^{n} \int_{0}^{t-c_i} g_i(\tau)*q\,(t - \tau - c_i)d\tau + N(t) \qquad (3)$$

where,

- $g_i(t)$ - impulse response function for outflow from sub-basin i
- $q_i(t)$ - outflow from sub-basin i
- $c_i$ - time lag for first significant value of $g_i$

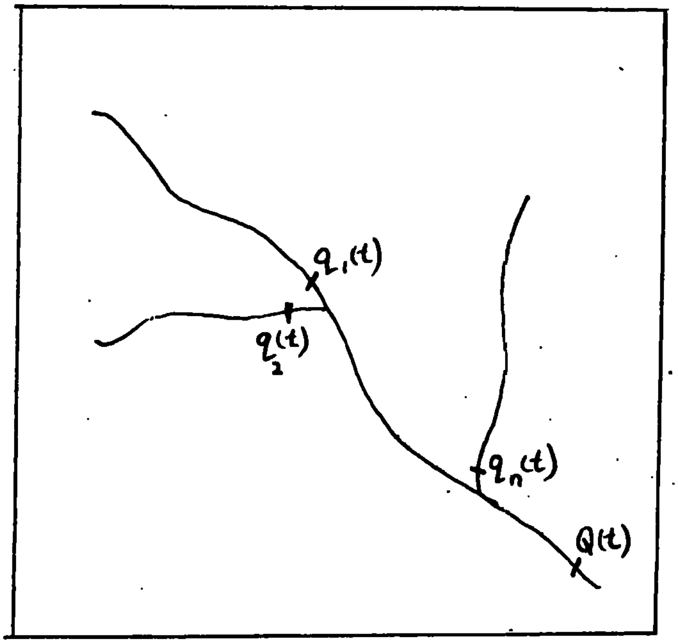



FIGURE 3

It will be seen that in contrast to (1) the outflow from each sub-basin is considered separately so that $q_i(t)$ replaces the integration of rainfall impulse response function in (1). $g_i(t)$ represents the motion of the flood wave from the sub-basin outlet to the basin outlet. For the decrete case

$$Q(k) = \sum_{i=1}^{n}\sum_{j=1}^{r_i} g_i(j)*q_i(k - j - c_i)*\Delta t + N(k) \qquad (4)$$

where $r_i$ is the number of significant values of $g_i$.

In many practical cases the information available may consist of the rainfall for some sub-basins and the outflows for others. In these cases a mixed model can be used and the expression for Q(t) in the discrete case is

$$Q(k) = \sum_{i=1}^{n_1}\sum_{j=1}^{m_i} h_i(j)*p_i(k - j - b_i)*\Delta t + \sum_{i=1}^{n_2}\sum_{j=1}^{r_i} g_i(j)*q_i(k - j - c_i)*\Delta t + N(k) \qquad (5)$$

where $n_1$ and $n_2$ are number of sub-basins of which the rainfall and outflow are known respectively. It should be remembered that this derivation has assumed that the impulse response functions and the lags for each sub-basin are invariant from flood to flood. In reality they can vary.

Usually the impulse response functions $h_i(t)$ and $g_i(t)$ rise quickly to a peak but have a long recession curve. Therefore the successful use of equations (2), (4) or (5) will require a large Kernel length ($m_i$ and $r_i$) and therefore a large number of parameters. The estimation of these from limited data may be subject to significant errors. The long recession curve of these functions is usually due to storage effects within the channel. It has been shown (Amirthanathan (4) that such storage effects can be accounted for by an auto-regressive (AR) term.

Therefore, since economy of parameters is a desirable feature, the AR terms will be used to represent all the decaying portions of the impulse response functions. This restricts $m_i$ and $r_i$ to a small value and the resulting model could be called an ARMA model even though in this case the moving average terms are not a series of random shocks. The ARMA representation of the simplified linear model is

$$Q(k) = \sum_{l=1}^{L} \delta_l*Q(k - 1) + \sum_{i=1}^{n_1}\sum_{j=1}^{m_i} \gamma_i*p_i(k - j - b_i) + \sum_{i=1}^{n_2}\sum_{j=1}^{r_i} w_i(j)*q_i(k - j\ c_i) + v(k) \qquad (6)$$

where $\delta_l$ (l=1, L), $\gamma_i$(i=1, $n_1$) and $w_i$(i=1, $n_2$) are the model parameters and v(k) is the total noise. The numbers $m_i$ and $r_i$ indicate the length to the peak of the impulse response functions. In practice the numbers $m_i$, $r_i$ and L rarely exceed 2. In this study only river flows are used in ARMA model and therefore $n_1$ is zero.

The ARMA model has the further advantage that the outflow or rainfall for all the sub-basins is not required for successful use of the model. This situation occurs frequently in practice where there are gauging stations only on the important tributaries of a river system. In this case the AR term takes into account the ungauged flow entering the river and will improve the forecast. During a flood the parameters $r_i$ and $w_i$ will be important during the rising limb of the hydrograph and the AR terms during the flood recession.

## Forecasting Methods

### Recursive Least Squares Method

Consider the case where a variable $x_0$ is related to n other linearly independent variables $x_1$, $x_2$, ...$x_n$ by the relationship

$$x_o = a_1*x_1 + a_2*x_2 + \dots\dots\dots + a_n*x_n \quad (7)$$

where $a_i$ (i = 1,n) are unknown constant parameters. However, $x_o$ can only be observed in the presence of noise $\varepsilon$. The observed value is y. Suppose that there are k such observations y (j = 1,k) and let the value of the variable $x_i$ (i = 1,n) associated with the j th observation be $x_{ji}$. Let be the noise term for the j th observation. Then all the observations can be given in matrix form as

$$Y = X*A + \varepsilon \quad (8)$$

where A - parameter vector
X - given by $X_{ij} = x_{ij}$
Y - observation vector
$\varepsilon$ - noise vector

Now an estimate of the parameters can be made from these observations using the method of least squares (i.e. by minimising the square of the difference between the observed value of y and that calculated from (7). The resulting estimate for A is given by

$$A = M*S \quad (9)$$

where, $M = (X*X^T)^{-1}$
$S = (X*Y)$

It is convenient to develop this into a recursive form so that with every new observation the parameter vector will be updated without repeating the above calculation each time. The estimate of A after k observations will be a linear sum of the estimate after (k-1) observations and a correction based on the k th observation. Let suffixes k and k-1 denote values after k and k-1 observations respectively. Let x(k) be the values of $x_i$ (i = 1,n) associated with the k th observation y(k). Then from the definition of M and S

$$M_k^{-1} = M_{k-1}^{-1} + x(k)*x(k)^T \quad (10)$$

$$S_k = S_{k-1} + x((k)*y(k) \quad (11)$$

After some matrix manipulations the RLS algorithm can be obtained as (Young (5)).

Output prediction: $\hat{y}(k) = x(k)*A \quad (12)$

Update Parameters: $A_k = A_{k-1} - F_k*[x(k)*A_{k-1} - y(k)]$ .... (13)

Update M matrix: $M_k = M_{k-1} - F_k*x(k)^T*M_{k-1}$ .....(14)

where

$F_k = M_{k-1}*x(k)*[1 + x(k)^T*M_{k-1}*x(k)]$ .... (15)

It will be seen that in (13) the term in brackets is the forecast error for step k (i.e. the difference between the value forecast for the k th step using parameter values at the k-1 th step). If the error is positive (forecast too large) then the parameters are decreased and vice versa. The matrix M stores all the information from the observed values of the variables $x_i$ (i = 1,n).

If the error sequence $\varepsilon_j$ is assumed to be serially uncorrelated and have zero mean with variance $\sigma^2$ it can be shown that the covariance matrix of the error in the parameter estimation (P) is given by

$$P = \sigma^2*M \quad (16)$$

**Kalman Filter**

Filtering has been defined as "A mathematical operation which utilizes past data (measurements) of a dynamic system in order to make more accurate statements about the present, future or past state than could have been made using information from a single direct measurement". [Burn and McBean (6)].

The Kalman Filter was originally developed to deal with problems in communications and controls [Kalman (7)]. However, it has since been successfully used elsewhere and is now a popular technique for flood forecasting. It is based on a linear discrete dynamic model which is given by the following equations.

State equation $Z_k = \phi_k * Z_{k-1} + w_k$ (17)

Observation equation

$$y_k = H_k * Z_k + v_k \quad (18)$$

Both equations are affected by white noise (errors) $w_k$ and $v_k$ which have covariance matrices Q and R respectively. $\phi_k$ and $H_k$ are transition matrices. The filter is designed to recursively estimate the state vector $Z_k$ using the new observations $y_k$ as they become available. The estimation is done such that the square of the error in the forecast of the next value of y and the covariance matrix of the parameter estimation error are minimised.

In the usual formulation of the problem for flood forecasting the model parameters are treated as the state variables and the ARMA model is written as the observation equation. Then y, $v_k$ and R are scalars. For the flood problem $\phi$ is put equal to 1 and the $H_k$ is a vector consisting of the values of the variables in the ARMA model at the relevant step. Now the noise term in (19) will account for both the error due to the inadequacy of the model as well as that due to the measurement error. The term $w_k$ in (17) represents the variation of the model parameters. This formulation will not distinguish between the sources of error and will forecast the observed flow and not the true flow. However, this is not a drawback in on-line forecasting as it is the measured flow (which is related to flood levels) that is of interest [Ngan & Russel (8)].

Using this formulation the Kalman filter algorithm is as follows with suffixes k-1 and k refer to values after k and k-1 time steps and k/k-1 refers to the prediction of values at step k from step k-1.

State prediction $A_{k/k-1} = A_{k-1}$ (19)

Output prediction $\hat{y}_k = H_k * A_{k/k-1}$ (20)

After the measured value of $y_k$ is available the model parameters are updated as follows.

Error covariance matrix

$$P_{k/k-1} = P_{k-1} + Q_{k-1} \quad (21)$$

Gain $G_k = P_{k/k-1} * H_k^T * (H_k * P_{k/k-1} * H_k^T + R_k)$ (22)

Updating parameters

$$A_k = A_{k/k-1} - G_k(\hat{y}_k - y_k) \quad (23)$$

Update covariance matrix

$$P_k = P_{k/k-1} - G_k * H_k * P_{k/k-1} \quad (24)$$

Here too the parameter correction depends on the forecast error for that step.

The value Q represents the dynamics of the system model and is an indication of how much the model parameters are allowed to vary. In the RLS method the M matrix and therefore from (16) the P matrix becomes smaller as more readings are used. This means that the estimate converges towards the real values. In the Kalman filter since the parameters are allowed to vary the P matrix can never become zero (the parameters can never be known exactly) and this is ensured by (19). A larger value for Q means a larger average value for P (more uncertainity in the estimates). R represents the effect of model and measurement error. When R is large less weight is given to the new measurement (small gain) while for small R the parameters will follow the observations closely.

Comparing (13) - (15) with (19) - (21) it will be seen that the Kalman filter reduces to the RLS algorithm when Q is zero and R is constant (unknown). It could be said that the RLS method brackets a stationary target when estimating the parameters while the Kalman filter tracks a moving target. In general the RLS method gives a biased estimate and it has been shown that when ordinary least squares algorithms are used the parameters are always underestimated. [Burn & McBean (6)].

In the practical implementation of the RLS method all that is required is a sufficient number of readings so that (9) and (10) can be solved. The Kalman filter requires initial values of A and P as well as estimates of Q and R. Initially it is possible to begin with zero parameter values and a high value for P as the values converge rapidly. However, it has been found that the values obtained from an RLS analysis of a previous flood are a good initial estimate. [Amirthanathan (1), (4)].

Various methods have been used to estimate R and Q for the Kalman filter. Ideally, estimates of Q should be based on the physical characteristics of the flow generation process while those of R should be based on the expected model and measurement accuracy. This would result in a method that can be used for any basin. Due to poor understanding of the processes, however, such a method is not available so that various arbitrary methods have to be used. Therefore the success of the filter depends on the judgement of the forecaster.

Q is usually assumed to be a diagonal matrix in this formulation. This means that it is assumed that the errors in the model parameters are independent. The simplest method is to assume a single constant value for all the diagonal terms. Different diagonal terms could also be used. More advanced methods allow these terms to vary from step to step. Two possible methods are to assume them proportional to the H vector [Q (I,I) $\propto$ H(I)] or to the diagonal elements of the P matrix [Q (I,I) $\propto$ P(I,I)].

In the formulation used R is a scalar. The simplest method is to assume a constant value. However, the measurement error usually increases for high flows due to rating curve errors. Therefore it is reasonable to make R proportional to the flow at that time step.

In all these methods it is required to optimize either a value for Q and R or some constant of proportionality. This is usually done using data from previous floods either mathematically [Ngan & Russel (8)] or by trial and error [Amirthanathan (1), (4)]. Using data from a previous flood can introduce errors if there has been some change in the system between calibration and application of the model. An algorithm for sequential estimation of Q and R outside the filter has been proposed [Sage & Husa (9)] which avoids this problem. This algorithm is given by the following equations.

$$R_k = (k-1)*R_{k-1}/k + (v_k*v_k^T - H_k*P_{k/k-1}\,H_k^T)\,/k \qquad (25)$$

$$Q_k = (k-1)*Q_{k-1}/k + (w_k*w_k^T + P_k - P_{k-q})\,/k \qquad (26)$$

where $v_k = y_k - H.*A$

$w_k = G_k*v_k$

## Application

It was attempted to forecast the flow at the Putupaula gauging station (see Fig. 1). The flow at the five gauging stations in the basin were selected as possible terms for the ARMA model given by (6). After considering the length of the river and the spacing of the stations a time step of 6 hrs. was selected.

Initially floods 1 and 2 of Table 1 were chosen for identification and calibration of the model. The other floods were used to test the calibrated model. Model identification involved selecting the number of flow terms used from each station in the ARMA model and their lags. In the notation of (6) the order and lag of the AR term are L and l while those of the other flow terms (Moving Average terms) are $r_i$ and $c_i$ respectively.

The selection of optimum values for order and lag can be based on any suitable error indicator. Some common indicators are the Mean Square Error (MSE), the Relative Root Mean Square Error, the largest relative error, the error at the flood peak and the number of relative errors greater than some prescribed value. In this study the MSE was chosen as the error indicator. This and the other indicators above were found for various combinations of orders and lags using the RLS method for floods 1 and 2.

It was found that for one step ahead forecasts the inclusion of the flow terms from Ratnapura and Dela did not improve the MSE. This was to be expected as all the information from these stations will be contained in the flow term from Ellagawa which is downstream from them. The best AR model was a second order one for both floods 1 and 2 with MSE values of 71 and 152 respectively. These values were improved (to 52 and 113 respectively) when flow terms from Ellagawa and Kukulegama were included but a different optimum model was obtained for each flood. Even when the number of terms was the same the best lag was different - values of 3 and 2 for Ellagawa and 1 and 2 for Kukulegama being obtained for floods 1 and 2 respectively.

These differences are due to the assumption made in deriving the model that the impulse response functions and lags are invariant. Actually they vary with the outflow from the sub-basin as well as with the fraction of the total flood given by that outflow. Since different floods are made up of the sub-basin flows in different proportions, the orders and lags can change from flood to flood. Also the difference in the lags could be because the real lag is not a multiple of 6 hours. The lag for Ellagawa is greater than that for Kukulegama even though the former is closer to Putupaula. This could be because Kukulegama is at a higher elevation (650 feet as against 40 feet).

If the random noise term in (8) is uncorrelated with zero mean and constant variance the forecast errors (residuals) should also be uncorrelated (white). The first five Auto-correlation coefficients of the residuals of the models with a low MSE were calculated. They were found to be insignificant at the 95% confidence level. The residuals were taken as uncorrelated and there was no need to consider error correlation in the model.

The reasonable results obtained using a second order AR model is due to the nature of the river basin. More than 70% of each flood at Putupaula comes down the reach from Ellagawa which has a very gentle gradient. Therefore this reach will be dominated by storage terms which are well represented by the AR terms.

Two models were selected for calibration and testing. Model 1 was the second order AR model. Model 2 was an ARMA model with an order 2 AR component and one term each from Ellagawa and Kukulegama at lags of 3 and 1 respectively. Model 2 had MSE values of 68 and 113 for floods 1 and 2 from the RLS analysis.

The flows for floods 3, 4 and 5 were forecast using both models using the following eight methods.

1) Basic RLS method
   Methods 2 to 8 used the Kalman filter with various means of estimating R and Q.
2) Both R and Q constant
3) R constant and Q(I,I) proportional to H(I)
4) R constant and Q(I,I) proportional to P(I,I)
5) R proportional to the flow and Q(I,I) to H(I)
6) R constant and Q from (26) (adaptive algorithm for Q)
7) Adaptive algorithm for both R and Q (25 and 26)
8) R proportional to the flow and Q adaptive (26)

The initial values needed for the Kalman Filter were obtained as follows. The parameter vector A was taken to be the mean parameter value from the RLS analysis for flood 1. The initial P matrix was found from (16) using the final value of M from flood 1. The best noise covariances (constant values of R and Q or constants of proportionality) were found by trial and error using the Kalman Filter for flood 2. It was found that using different diagonal elements in Q did not improve the forecasts.

The noise covariances thus obtained should be realistic. Otherwise the forecasting procedure will be artificial and arbitrary. The values obtained for constant R ranged from 5000 to 20000. The coefficient for methods 5 and 7 would give an R range of 3000 to 25000. A value of 20000 for R implies that the standard deviation of the measurement error is about 140 which is about 15% of the maximum flow. This is probably higher than the actual value.

The value for constant Q was 0.0005. The coefficient from methods 3 and 5 was 0.1E-5 and that from method 4 was 0.05. These values give imply that the standard deviation of the parameter variation is from 0.015 to 0.035. This range agrees with the parameter variation observed during the forecasts.

Method 8 gave unreliable forecasts. It was found that negative values of R were being generated by (25). Since R is a variance this has no physical meaning. Other researchers [Burn & McBean (6)] who encountered this problem used a modified version of (25). However, for this study it was decided to discard method 8. The results from the other 7 methods are given in Tables 2 and 3.

The results show that model 1 always performs better than model 2. This differs from the earlier results. This could be due to the AR terms being dominant. Also if the noise level is very high the new information given by any additional reading could be outweighed by the extra noise they introduce. The Kalman filter always performs better than the RLS method. Methods 2 to 7 perform similarly with those allowing R and Q to vary giving slightly better results on the average.

The forecasts for flood 3 were much better than those for floods 4 and 5. This could be because the latter had gaps in the gauge records and are therefore less accurate. The difference between the RLS and Kalman filter methods are greatest for flood 1. This agrees with earlier results [Amirthanathan (1)] showing

| Flood Number | Method 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 3 | 102 | 63 | 65 | 62 | 62 | 64 | 63 |
| 4 | 188 | 176 | 178 | 198 | 174 | 170 | 171 |
| 5 | 349 | 331 | 333 | 335 | 331 | 324 | 327 |

TABLE 2

Mean Square Error of Forecasts using Model 1

| Flood Number | Method 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 3 | 188 | 70 | 70 | 70 | 70 | 95 | 76 |
| 4 | 256 | 201 | 189 | 213 | 187 | 219 | 227 |
| 5 | 392 | 334 | 331 | 343 | 332 | 324 | 330 |

TABLE 3

Mean Square Error of Forecasts using Model 2

that the advantage of the Kalman filter over the RLS method is less when the noise level is very high. Fig. 4 gives a comparison of the forecasts from the RLS and Kalman filter methods for flood 5 (which gave the worst results).

When the Kalman filter is used with estimated noise covariances it may not operate optimally. For optimal filtering the residuals should be a white gaussian sequence [Mehra (10)]. The first five Auto-correlation coefficients were found for the forecasts. Since they were not significant at the 95% level of confidence it could be said that the filtering was near optimal.

The values used for Q and R were not always the optimum values. For example in method 2 the best R value for all three test floods was greater than that used. The optimum constant for Q in method 3 was also different for flood 1 (0.1E-6 instead of 0.1E-5). This demonstrates the danger of using prior estimates of Q and R.

As a comparision with these methods forecasting was done with constant parameters using model 1. When the parameters from flood 1 were used the performance was as good as the Kalman filter. However, this result should be considered accidental because when the parameter set from flood 2 was used the performance was worse than the RLS method.

Since in practice 6 hours is too short a time step forecasts were made up to 6 steps (36 hours) ahead using model 1. This was done by making 6 successive forecasts at each step using the forecast one step ahead as data for the second forecast and so on. As expected the results for flood 3 are the best. These are given in Fig. 5. The MSE values for the forecasts from 1 to 6 steps ahead are 60, 230, 559, 1147, 2127 and 3537 respectively. This agrees with the expected behaviour for an ARMA model where the error variance increases exponentially with the number of steps.

This method of making long range forecasts is very crude. A direct forecast 6 steps ahead that used all the information from the upstream stations would be much better. This can be done in a further study.

## Conclusions

The study shows that satisfactory forecasts 6 hours ahead can be made using the given techniques. These techniques give more reliable forecasts than using an ARMA model with constant coefficients. The Kalman filter is superior to the RLS method. All the methods used to estimate noise covariances give similar results with those that allow R and Q to vary being marginally better. It is seen that the forecasts are not much affected when the noise covariances are not the optimum values.

For one step ahead forecasts for the station considered a simple second order AR model gave the best results. The forecasts made 24 and 36 hours ahead using this model were also reasonable showing that good forecasts at these ranges might be obtainable with a more sophisticated model.

## References


1) Amirthanathan, G.E. - "Short Term Flood Forecasting using Sequential Optimal Estimation Techniques" - 4th Congress, Int. Assoc. for Hydr. Res. : Asia Pacific Division - Thailand, Sept. 1984
2) Arumugam, S. - "Water Resources of Ceylon" - Water Resources Board Publications (1969) - p.62
3) Kottegoda, N.T. - "Stochastic Water Resources Technology" - Macmillan Press (1980) - pp. 111-166
4) Amirthanathan, G.E. - "Contribution des Techniques de Filtrage Optimal a Quelques Problems Hydrologiques" - Docteur Ingenieur Thesis, Universite des Sciences et Techniques du Langdoc, Montpellier, France (1982) - pp. 23-48, 99-113.
5) Young, P.C. - "A Recursive Approach to Time Series Analysis" - Bull. Inst. of Math. and its Applications (IMA) Vol. 10 No. 5/6 pp. 209-224
6) Burn, D.H. and McBean;, E.A. - "River Flow Forecasting Model for the Sturgeon River" - ASCE J. of Hydr. Eng. Vol. III No. 2 Feb. 1985 - pp. 316-333
7) Kalman, R.E. - "A New Approach to Linear Filtering and Prediction Problems" - Trans. of the ASME, J. of Basic Eng. Mar. 1960 pp. 35 - 45
8) Ngan, P. and Russel, S.O. - "Example of Flow Forecasting with Kalman Filter" - ASCE J. of Hydr. Eng. Vol. 112 No. 9 Sept. 1986 pp. 818-832
9) Sage, A.P. and Husa, G.W. - "Adaptive Filtering with Unknown Prior Statistics" - Proc. Joint Automatic Control Conference (1969) pp. 760-769
10) Mehra, R.K. - "On the Identification of Variances and Adaptive Kalman Filtering" - IEEE Trans. on Automatic Control Vol. AC-15 Apr. 1970 pp. 175-184

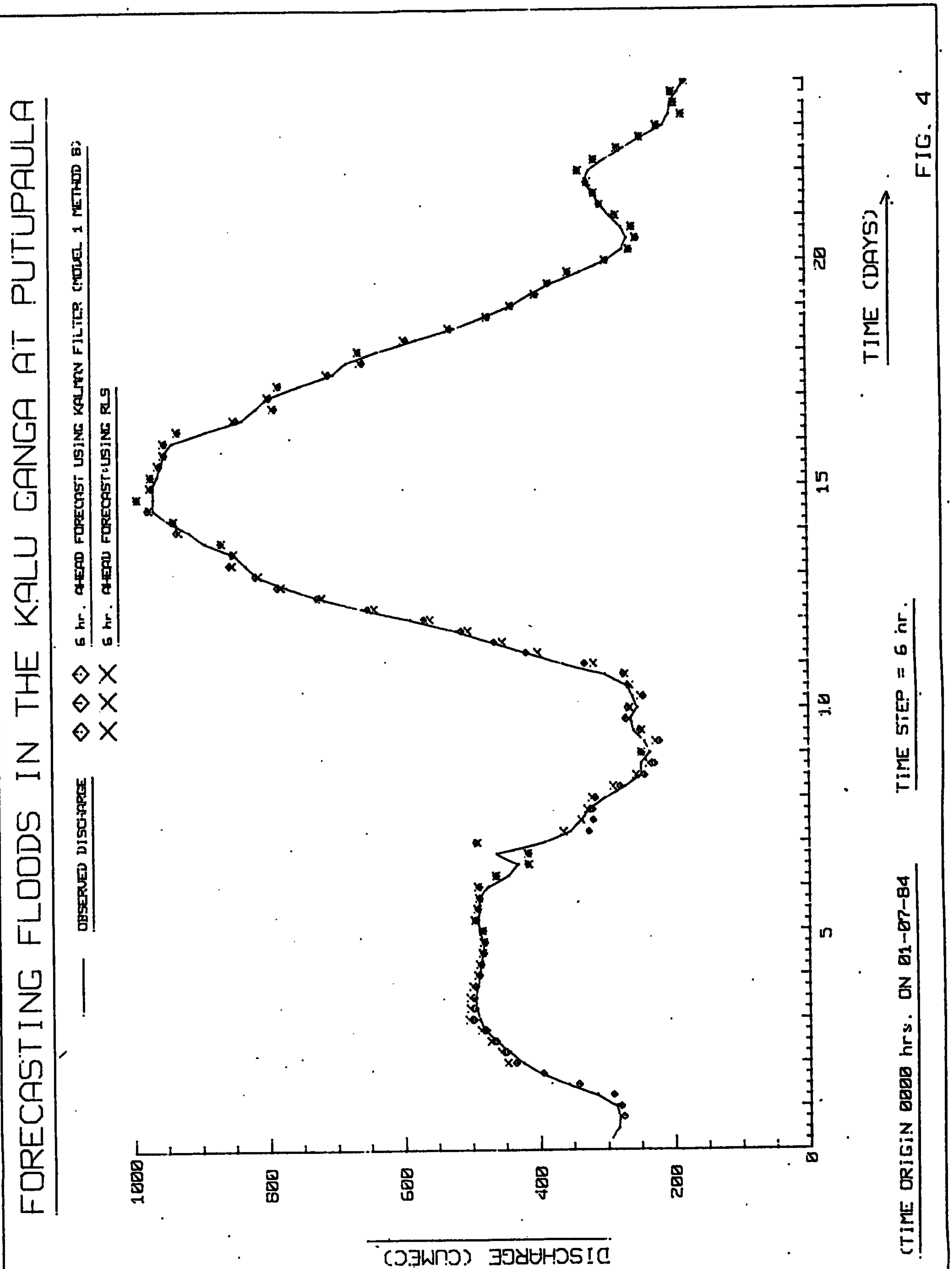

FORECASTING FLOODS IN THE KALU GANGA AT PUTUPAULA
OBSERVED DISCHARGE
6 hr. AHEAD FORECAST USING KALMAN FILTER (MODEL 1 METHOD 6)
6 hr. AHEAD FORECAST USING RLS
DISCHARGE (CUMEC)
1000
800
600
400
200
0
5
10
15
20
TIME (DAYS)
TIME ORIGIN 0000 hrs. ON 01-07-84
TIME STEP = 6 hr.

FIG. 4

FORECASTING FLOODS IN THE KALU GANGA AT PUTUPAULA

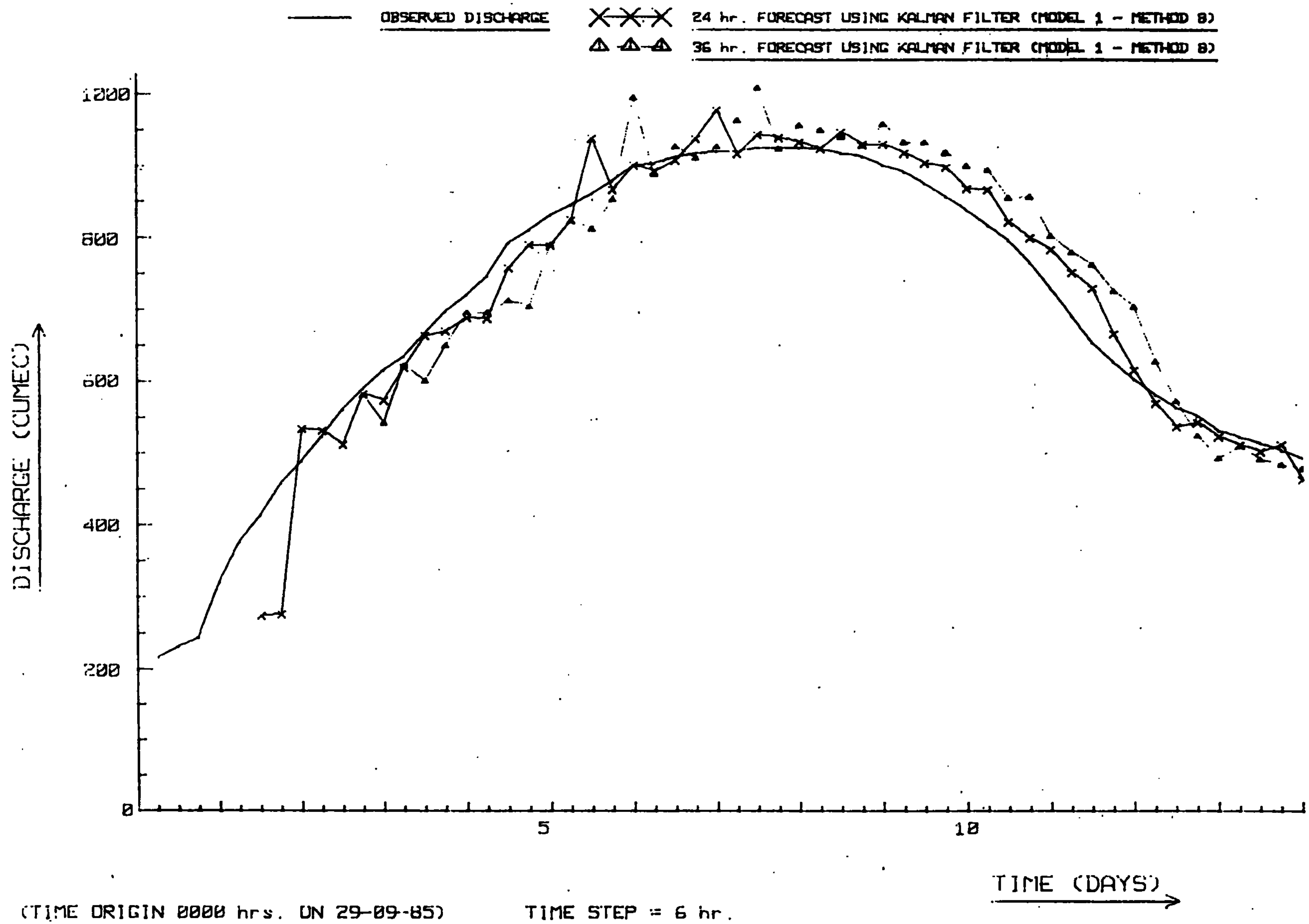



FIG. 5

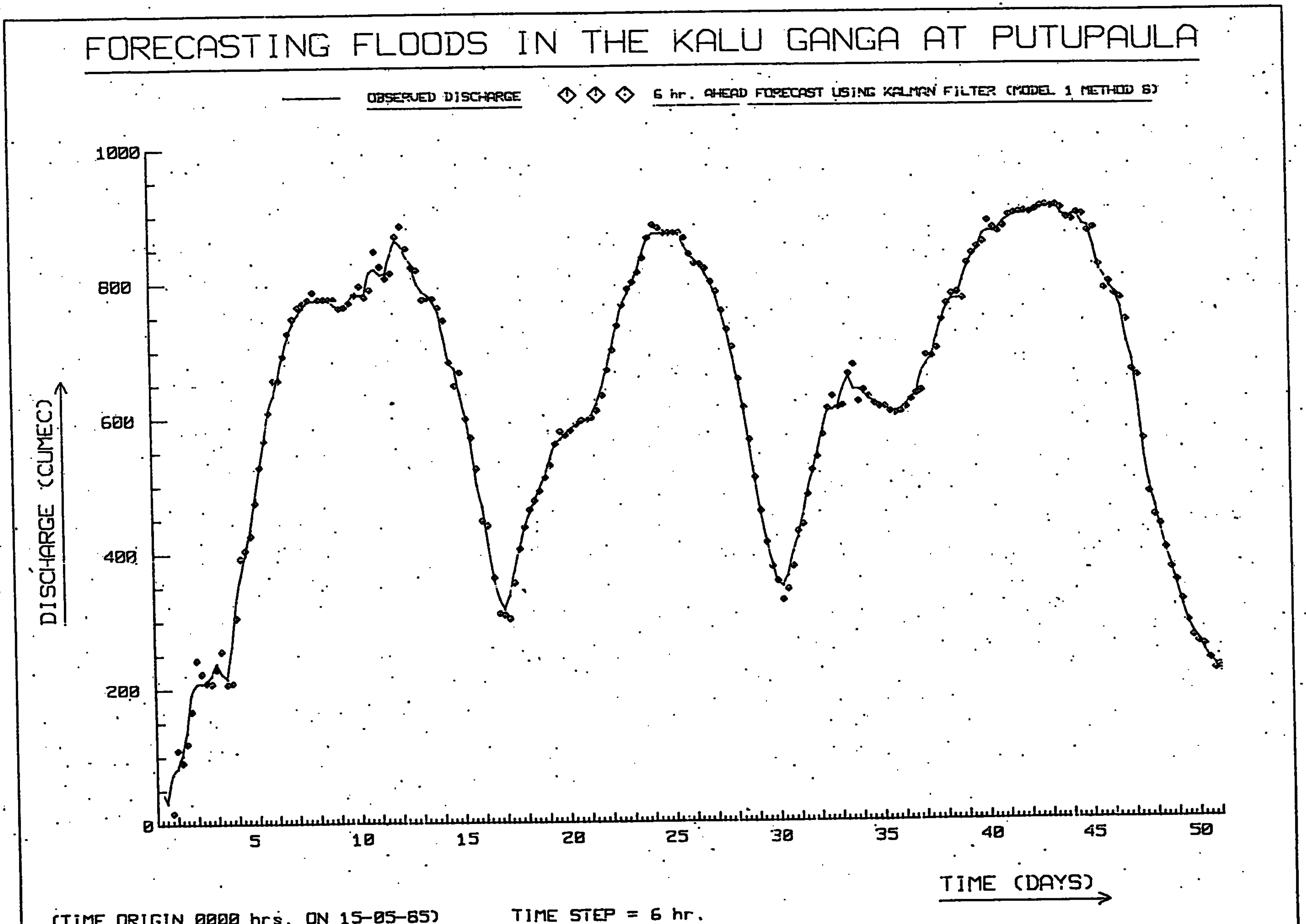

FORECASTING FLOODS IN THE KALU GANGA AT PUTUPAULA
OBSERVED DISCHARGE
6 hr. AHEAD FORECAST USING KALMAN FILTER (MODEL 1 METHOD 6)
DISCHARGE (CUMEC)
1000
800
600
400
200
0
5
10
15
20
25
30
35
40
45
50
TIME (DAYS)
(TIME ORIGIN 0000 hrs. ON 15-05-65)
TIME STEP = 6 hr.

FIG. 6