DEVELOPMENT AND VALIDATION OF A NEW MAXIMUM LIKELIHOOD CRITERION SUITABLE FOR DATA COLLECTED AT UNEQUAL TIME INTERVALS

By

Qingyun Duan

A Thesis Submitted to the Faculty of the

DEPARTMNET OF HYDROLOGY AND WATER RESOURCES

In Partial Fulfilment of the Requirements
For the Degree of

MASTERS OF SCIENCE
WITH A MAJOR IN WATER RESOURCES ADMINISTRATION

In the Graduate College

THE UNIVERSITY OF ARIZONA

1987

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfilment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgement the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: ____________________

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

S. Sorooshian
Professor of Hydrology and Water Resources

9/4/87
Date

ACKNOWLEDGEMENTS

The author would like to express his deep gratitude towards Professor Soroosh Sorooshian, his adviser, and Dr. Richard P. Ibbitt for their guidance and support during the course of this study.

The author also wishes to extend his thanks to the University of Arizona for providing computer facilities which were necessary for the research. Financial support for this research was provided by funds from the National Science Foundation (Grants INT-8413539 and ECE-8610584).

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS--CONTINUED

LIST OF ILLUSTRATIONS--CONTINUED

LIST OF TABLES

ABSTRACT

A new Maximum Likelihood Criterion (MLE) suitable for data which are recorded at unequal time intervals and contain auto-correlated errors is developed. Validation of the new MLE criterion has been carried out both on a simple two-parameter reservoir model using synthetical data and on a more complicated hillslope model using real data from the Pukeiti Catchment in New Zealand. Comparison between the new MLE criterion and the Simple Least Squares (SLS) criterion reveals the superiority of the former over the latter. Comparison made between the new MLE and the MLE for auto-correlated case proposed by Sorooshian in 1978 has shown that both criteria would yield results with no practical difference if equal time interval data were used. However, the new MLE can work on variable time interval data which provide more information than equal time interval data, and therefore produces better visual results in hydrologic simulations.

CHAPTER ONE

INTRODUCTION

The Hydrology Center in New Zealand is developing conceptual models of catchment hydrology for use in river basin management. Primary objectives are to improve flood forecasting by making greater use of rainfall data and to assess the consequences on river flows due to land use changes. The work is motivated by concerns for protection of the environment as well as properties and human lives, and is being made feasible by recent advances in conceptual catchment modeling techniques and improvements in data collection and archival procedures.

For a catchment model to be realistic, units of land with different terrain need to be accounted for individually. Given a spatial distribution of these units within a watershed,

(a) the response to rainfall of each unit can be estimated by using a conceptual model to approximate the chief physical processes taking place;

(b) rainfall can be estimated for each unit and the storm passage across a catchment can be allowed for.

Conceptual models approximating the chief physical processes involved in the conversion of rainfall to runoff and evaporation have been available for many years [Crawford and Linsley, 1966], while published evidence of the computational resources needed to simulate spatially-distributed models appears in Ross et al. [1978]. The model used by Ross and his co-workers is designed for individual floods and uses a spatially-invariant rainfall distribution within a storm, and fixed subsurface conditions to simulate the infiltration of rainfall. The event-based generation of streamflow from surface runoff avoids the need to estimate many of the physical coefficients associated with water transmission through the soil. Thus, by concentrating on flood events that are dominated by surface runoff, Ross et al. avoid a major difficulty encountered in conceptual modeling, i.e., the quantitative estimation of parameters controlling the operation of the individual storage/transfer mechanisms in the model [Ibbitt, 1979]. More general models seeking to estimate a catchment's water resource on a continuing basis have to face this difficulty.

To obtain parameter estimates quickly in operational situations, it is desirable to use an automatic computer-based procedure for fitting models. To obtain reliable parameter estimates, the fitting process must be unbiased and should take into account available physical information about the catchment. Many techniques can be used to fit a conceptual model. Ibbitt [1970] examines nine possible techniques, all of which can be made to ensure physically realistic parameter values by restricting the ranges that each parameter value can take. The narrower the range, the greater is the confidence in the initial estimate. However, before automatically fitting a model, a suitable estimation criterion must also be chosen to quantify the deviation of the modeled flows from those measured. The fitting process then involves searching through different sets of parameter values that minimize (or maximize, if appropriate) the chosen criterion.

Although minimization of any reasonable criterion will usually result in a model being able to satisfactorily mimic the measured flows used in the calculation of the criterion, the results of using the model to simulate flows not used in the fitting process are generally less satisfactory. This problem arises from the parameter values being biased, and in real world situations, biased parameters often result when the data that are used contain measurement errors. Due to the presence of the measurement errors, whenever a model is fitted to data, some discrepancy between model and prototype behavior occurs. The discrepancies take the form of either under- or over-estimates of what was measured. The factor of critical importance in choosing a criterion for fitting a model depends on whether the unders or overs occur for only one or two data points or whether they persist over longer periods. In the former case, the discrepancies are often assumed to result from random measurement error, and either a sum of squares or a heteroscedastic criterion as described by Sorooshian [1978] may be suitable. When a discrepancy of the same sign persists for more than one or two intervals, it is probable that there is serial dependence between successive values. Several reasons can be advanced for the presence of serial dependence in the errors: (1) erroneous changes of stage to discharge rating curve; (2) sub-optimum values for model parameters; (3) smoothing of input data errors by the storage components of catchment models; and (4) the inevitable structural differences between a catchment and its simplified representation by a model. In some cases, the serial dependence can be reduced by measures such as careful data checking or building of more realistic models. Whatever steps are taken are unlikely to completely eliminate the problem. Therefore, criteria that recognize the problems are necessary. Such criteria have been derived by Sorooshian [1978] and Kuczera [1982], and they are based on the Maximum Likelihood Theory.

Sorooshian et al. [1983] tested the Maximum Likelihood approach on the US National Weather Service Soil Moisture Accounting model and reported that conceptually realistic parameters were obtained which gave good forecast performance. In contrast, the use of the popular Simple Least Squares approach gave unrealistic parameter estimates and poor forecasts. Similar results were reported by Ibbitt and Hutchinson [1984].

In the work of Sorooshian [1978], Sorooshian et al. [1983], and Ibbitt and Hutchison [1984] mentioned above, the rainfall and streamflow data used were sequences of values at equally spaced intervals of time. However, to archive hydrological data in this format is not only inefficient on the computer storage space, but often the extremes of hydrological processes are smoothed out. The New Zealand Hydrological Archive, containing some 20000 station-years of continuous streamflow record, exploits the achiving of data at unequal intervals of time, resulting in a reduction of computer storage requirement from 500M bytes to 50M bytes, with attendant savings on input/output operations and the preservation of extreme values. To take advantage of the more informative data, we hope to be able to employ appropriate criteria which can work with variable time interval data. Unfortunately, even though there are models built that work on variable time steps, the criterion necessary for fitting them in an informative manner is unavailable. Hence, the aim of this work is to develop a Maximum Likelihood Criterion for conceptual models using data in the format described above.

In deriving the new Maximum Likelihood Criterion, the differences between measured flows and simulated flows are treated as entirely caused by errors in the measured streamflow data. This is a conceptual convenience that assists with setting up test procedures rather than a necessary assumption. However, it is assumed that streamflow errors are from an auto-correlational process, and the further assumption is made that errors result from a stationary process.

This thesis is organized as follows: Chapter Two discusses the systematical treatment of auto-correlated errors; Chapter Three shows the procedures of deriving the new Maximum Likelihood Criterion (MLE); Chapter Four describes the optimization methods and the model calibration procedures; Chapter Five presents numerical results from the experiments with a simple two-parameter reservoir model.

Comparison is made between the new MLE and the Simple Least Squares (SLS) Criterion; Chapter Six gives the validation results of the new criterion on a hillslope model. Last, conclusions drawn from this study are presented in Chapter Seven.

CHAPTER TWO

TREATMENT OF SYSTEMATIC ERRORS WHEN DATA WERE MEASURED AT UNEQUAL TIME INTERVALS

In Sorooshian's work [1978], it was assumed that the additve errors were auto-correlated to lag-one by a simple linear relationship:

$$e_t = \rho \cdot e_{t-1} + a_t \tag{2.1}$$

where

e_t = additive errors at time t;

a_t = purely random component of measurement error (assumed Gausian distribution with zero mean and constant variance, independently identically distributed for all t s);

ρ = parameter known as the first-lag auto-correlation coefficient, which measures degree of systematic error, $|\rho| \leq 1$.

When the data were collected at unequally spaced time intervals, the above model could not be used in its present form. It should be noted, intuitively, that we expect a greater degree of error correlation in data recorded at closely spaced time intervals than in data recorded at widely spaced time intervals. Hence, any model chosen to represent correlated errors in variable time intervals should have that property.

It is easily shown [Gupta, 1982] that a first order auto-regressive process in fixed-interval-length discrete time steps can be modeled by a discrete time (fixed-interval-length) linear reservoir model, i.e.

$$S_t = (1-K) \cdot S_{t-1} \tag{2.2}$$

where S_t is the storage level at time t and K is the storage constant. If there is a random additive input p_t, then we can write

$$S_t = (1-K) \cdot S_{t-1} + p_t \tag{2.3}$$

Equation (2.3) is isomorphic to the auto-regressive error model, equation (2.1), with $\rho = (1-K)$. The continuous time analog of the discrete time reservoir mentioned above is given by

$$S(t_2) = S(t_1) \cdot \exp\{-\alpha(t_2-t_1)\} \tag{2.4}$$

which indicates that the reservoir storage depletes exponentially with time. As a demonstration, consider time step (t_2-t_1) to be one time unit and let us compute the storage level S(t) at equally spaced time intervals. In this case,

$$S(t+1) = S(t) \cdot \exp\{-\alpha(t+1-t)\} = S(t) \cdot \exp\{-\alpha\} \tag{2.5}$$

Letting $S(t+1) = S_{t+1}$ and $S(t) = S_t$, we get

$$S_{t+1} = S_t \cdot \exp\{-\alpha\} \tag{2.6}$$

Comparing equations (2.6) and (2.2), we see that

$$\exp\{-\alpha\} = (1-K) \tag{2.7}$$

or

$$\alpha = -\ln(1-K) \tag{2.8}$$

or

$$K = 1 - \exp\{-\alpha\} \tag{2.9}$$

Clearly, for a variable time step, the analogous value for the parameter K of the linear reservoir would be a time variable value given by

$$K(t_2, t_1) = 1 - \exp\{-\alpha(t_2-t_1)\} \tag{2.10}$$

If we include consideration of a random additive input p_t in the model, we can write

$$S_{t_2} = S_{t_1} \cdot \exp\{-\alpha(t_2-t_1)\} + p_t \tag{2.11}$$

and, by exploiting the isomorphism between the linear reservoir model and the first-lag auto-regressive error model, we get

$$e_{t_2} = e_{t_1} \cdot \exp\{-\alpha(t_2-t_1)\} + a_t \tag{2.12}$$

Here $\exp\{-\alpha(t_2-t_1)\}$ is the time variable auto-correlation coefficient $\rho(t_2,t_1)$

$$\rho(t_2,t_1) = \exp\{-\alpha(t_2-t_1)\} \quad (2.13)$$

Note that $\rho(t_2,t_1)$ has the desirable property discussed earlier that as (t_2-t_1) increases, the correlation between errors (for fixed value of parameter α) decreases. Hence, measurement errors at widely spaced time intervals have effectively negligible correlation, while closely spaced measurement errors are more significantly correlated. The degree of correlation at a fixed interval is determined by the magnitude of α. If α is small, widely spread measurements will be significantly correlated, while if it is large, even closely spaced measurements may be effectively uncorrelated. If $\alpha = \infty$, then the correlation among measurement errors is zero.

Based on the above presentation, a procedure for the time-variable case when the measurement errors are assumed to be auto-correlated can be developed.

CHAPTER THREE

DEVELOPMENT OF MAXIMUM LIKELIHOOD CRITERION FOR DATA SET MEASURED AT UNEQUAL TIME INTERVALS

Let q_t, $\tilde{q}_t$, and $q_t(\theta)$ be the "true", the observed, and the simulated streamflows, respectively, where θ is the model parameter vector. The measurement error e_t can be calculated by

$$e_t = \tilde{q}_t - q_t \tag{3.1}$$

The model structural error $m_t(\theta)$ which is defined as the difference between the simulated streamflow and the true flow can be obtained by

$$m_t(\theta) = q_t(\theta) - q_t \tag{3.2}$$

The difference between the observed streamflow and the simulated streamflow, known as model residual r_t, can be computed through

$$r_t = \tilde{q}_t - q_t(\theta) \tag{3.3}$$

Adding and substracting q_t on the right side of (3.3), we get

$$\begin{aligned} r_t &= \left[\tilde{q}_t - q_t \right] - (q_t(\theta) - q_t) \\ &= e_t - m_t(\theta) \end{aligned} \tag{3.4}$$

In realistic cases, it is impossible to drive $m_t(\theta)$ to zero for all time intervals. However, for the sake of simplicity, we assume that the model used is perfect, and that $m_t(\theta)$ can be driven to zero for all t s by appropriate selection of parameter set θ. Therefore, we have

$$e_t \cong r_t = \tilde{q}_t - q_t(\theta) \tag{3.5}$$

Rearranging equation (2.12) and replacing e_t by $\left[\tilde{q}_t - q_t(\theta)\right]$, we get the following equation,

$$\begin{aligned} a_{t_2} &= e_{t_2} - e_{t_1} \cdot \exp\{-\alpha(t_2-t_1)\} \\ &= \left[\tilde{q}_{t_2} - q_{t_2}(\theta)\right] - \left[\tilde{q}_{t_1} - q_{t_1}(\theta)\right] \cdot \exp\{-\alpha(t_2-t_1)\} \end{aligned} \tag{3.6}$$

Before deriving the maximum likelihood equation, let s make the following assumptions:

(a) a_t is normally distributed, with zero mean, and constant variance, i.e. $a_t \sim N(0, \sigma^2{}_t)$;

(b) a_{t_i} and a_{t_j} are independent, i.e.

$$E\,[a_{t_i}, a_{t_j}] = \begin{cases} \sigma^2_{t_i} & t_i = t_j \\ 0 & t_i \neq t_j \end{cases} \tag{3.7}$$

(c) a_t is homoscedastic, i.e.

$$\sigma^2_{t_1} = \sigma^2_{t_2} = \ldots = \sigma^2_{t_n} = \sigma^2_a \tag{3.8}$$

Based on assumption (a), the probability density function of the a_t s has the form:

$$P(a_t) = (2\pi)^{-\frac{n}{2}} \, |\Omega| \, \exp\left\{-\frac{1}{2} A^T \Omega^{-1} A\right\} \tag{3.9}$$

where

Ω is the covariance matrix, i.e.

$$\Omega = \begin{bmatrix} E(e_1e_2) & E(e_2e_1) & . & E(e_ne_1) \\ E(e_1e_2) & E(e_2e_2) & . & E(e_ne_2) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ E(e_1e_n) & E(e_2e_n) & . & E(e_ne_n) \end{bmatrix} \tag{3.10}$$

$|\Omega|$ is the determinant of Ω;

Ω^{-1} is the inverse of Ω;

$A = \left[a_{t_1}, a_{t_2}, \dots, a_{t_n} \right]^T$, T denotes the transpose.

Since the joint probability of the series a_t is algebraically equivalent to the likelihood function L of the parameters θ and of the same series, then

$$L(\theta,\Omega) = P(a_t) \tag{3.11}$$

and the log likelihood function is given by

$$\begin{aligned} \ell(\theta,\Omega) &= \ln[L(\theta,\Omega)] \\ &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln|\Omega| - \frac{1}{2} A^T\Omega^{-1}A \end{aligned} \tag{3.12}$$

From assumptions (b) and (c), we get

$$\Omega = \begin{bmatrix} \sigma_a^2 & 0 & \cdot & 0 \\ 0 & \sigma_a^2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \sigma_a^2 \end{bmatrix}$$

$$= \sigma_a^2 \cdot I \tag{3.13}$$

where I is the identity matrix. Therefore, we have

$$|\Omega| = (\sigma_a^2)^n \tag{3.14}$$

$$\Omega^{-1} = \frac{1}{\sigma_a^2} \cdot I \tag{3.15}$$

Replacing equations (3.14) and (3.15) into equation (3.12),

$$\ell(\theta,\Omega) = -\frac{n}{2} \ell n(2\pi) - \frac{1}{2} \ell n(\sigma_a^2)^n - \frac{1}{2\sigma_a^2} \sum_{i=1}^{n} a_{t_i}^2 \tag{3.16}$$

Replacing equation (3.6) into equation (3.16),

$$\ell(\theta, \Omega) = -\frac{n}{2} \ell n(2\pi) - \frac{1}{2} \ell n(\sigma_a^2)^n - \frac{1}{2\sigma^2{}_a} \sum_{i=1}^{n} \left[\left[\tilde{q}_{t_i} - q_{t_i}(\theta)\right] - \exp\{-\alpha(t_i - t_{i-1})\} \left[\tilde{q}_{t_{i-1}} - q_{t_{i-1}}(\theta)\right]\right]^2 \tag{3.17}$$

Here, we assume $q_{t_0} = q_{t_0}(\theta) = 0$.

Equation (3.17) is the new Maximum Likelihood Criterion (MLE) which is suitable for data collected at unequal time intervals. It can be shown that when α is considerably large and σ is known, the parameter values obtained by maximization of $\ell(\theta, \Omega)$ will be identical to those obtained by minimization of the Simple Least Squares (SLS). It can also be seen that the new criterion has a very similar form to that derived by Sorooshian [1978] for the case that auto-correlated errors exist in input data, if $(t_2 - t_1)$ is constant for all time intervals.

CHAPTER FOUR

OPTIMIZATION METHODS AND MODEL CALIBRATION PROCEDURES

To examine the effectiveness of the new MLE criterion, appropriate optimization methods must be chosen in order to estimate all the parameters in equation (3.17) (including catchment model parameters θ, and the error model parameters σ_a^2 and α). It has been suggested that the gradient-based optimization algorithms generally perform better than the direct search algorithms when the optimizing functions display strong nonlinearities [Bard, 1974; Sorooshian and Gupta, 1984; and Gupta and Sorooshian, 1985]. The two models used in this study possess nonlinear characteristics. Gradient search algorithms should be desirable. However, both models have threshold parameters which make the evaluation of the derivatives of objective functions with respect to these parameters a very difficult task. Although Gupta and Sorooshian [1984] have proposed methods to handle the derivatives of threshold parameters, the author has opted to use a revision of the Rosenbrock s algorithm [Rosenbrock, 1960; Ibbitt, 1970]. This is a direct search type algorithm that incorporates the constraints on each parameter into the optimization search procedures. In case of hydrological model calibrations, use of constraints restricts parameter values to physically realistic ranges.

For determination of statistical paramters σ_a and α in equation (3.17), the necessary conditions require that the first derivatives of the optimizing function with respect to the each parameter be zero.

The value for σ_a can thus be obtained directly by

$$\frac{\partial \ell(\theta,\Omega)}{\partial \sigma_a} = -\frac{n}{\sigma_a} + \frac{1}{\sigma_a^3} \sum_{i=1}^{n} [\left[\tilde{q}_{t_i} - q_{t_i}(\theta)\right] - \left[\tilde{q}_{t_{t-1}} - q_{t_{i-1}}(\theta)\right] \cdot \exp\{-\alpha(t_i - t_{i-1})\}]^2 = 0 \tag{4.1}$$

i.e.

$$\sigma_a^2 = \frac{1}{n} \sum_{i=1}^{n} \left[\left[\tilde{q}_{t_i} - q_{t_i}(\theta)\right] - \left[\tilde{q}_{t_{i-1}} - q_{t_{i-1}}(\theta)\right] \cdot \exp\{-\alpha(t_i - t_{i-1})\}\right]^2 \tag{4.2}$$

Substituting equation (4.2) into equation (3.17), we get

$$\ell(\theta, \Omega) = -\frac{n}{2} \ell n(2\pi) - \frac{n}{2} \ell n(\sum_{i=1}^{n} [\left[\tilde{q}_{t_i} - q_{t_i}(\theta)\right] - \left[\tilde{q}_{t_{i-1}} - q_{t_{i-1}}(\theta)\right] \cdot \exp\{-\alpha(t_i - t_{i-1})\}]^2) + \frac{n}{2} \ell n(n) - \frac{n}{2}$$

$$= C - \frac{n}{2} \ell n(\sum_{i=1}^{n} \left[\left[\tilde{q}_{t_i} - q_{t_i}(\theta)\right] - \exp\{-\alpha(t_i - t_{i-1})\} \left[\tilde{q}_{t_{i-1}} - q_{t_{i-1}}(\theta)\right]\right]^2 \tag{4.3}$$

where

$$C = -\frac{n}{2} \ell n(2\pi) - \frac{n}{2} + \frac{n}{2} \ell n(n)$$

To determine the value for α, no explicit expression exists. The derivative of α has the following form

$$\frac{\partial \ell(\theta,\Omega)}{\partial \alpha} = \left[\frac{1}{2\sigma_a^2}\right] \sum_{i=1}^{n} [\left[\tilde{q}_{t_i} - q_{t_i}(\theta)\right]$$

$$- \exp\{-\alpha(t_i - t_{i-1})\} \left[\tilde{q}_{t_{i-1}} - q_{t_{i-1}}(\theta)\right]] \cdot \left[\tilde{q}_{t_{i-1}} - q_{t_{i-1}}(\theta)\right]$$

$$\cdot (t_i - t_{i-1}) \cdot \exp\{-\alpha(t_i - t_{i-1})\}] = 0 \qquad (4.4)$$

Apparently, α cannot be determined directly from equation (4.4). Thus, an iterative procedure is used and the method of false position was chosen.

The entire optimization procedures for determining θ, α, and σ_a can be summarized as:

(1) Choose initial values for model parameters, θ_0;

(2) estimate α from equation (4.4) using iterative procedures (e.g. false position method);

(3) estimate σ_a^2 using equation (4.2);

(4) use iterative procedures (e.g. Rosenbrock s method) to locate a new parameter set (say, θ) , and return to step 2. Continue the optimization procedures until the convergence criteria for all unknown parameters are met and the objective function value is within the assigned tolerance.

CHAPTER FIVE

VALIDATION OF THE NEW CRITERION ON A TWO-PARAMETER RESERVOIR MODEL

A two-parameter reservoir model is chosen for theoretical analysis of the new MLE criterion. The reason for selecting such a simple model is that data used for model calibration can be synthesized in such a way that all the assumptions made earlier on use of the new criterion can be satisfied. In this way, the applicability of the new criterion can be fully studied.

Description of the Model

The two-parameter reservoir model used for testing of the newly-developed estimation criterion is illustrated in Figure 5.1. The model has a threshold parameter C_{max}, the maximum capacity of the reservoir (in milimeters). The other parameter is the storage constant, K (in day^{-1}).

Let S_t = reservoir content at the beginning of time interval t;

p_t = precipitation input during time interval t;

b_t = baseflow from reservoir during time interval t;

r_t = spillflow from reservoir during time interval t;

q_t = total discharge from reservoir during time interval t, $q_t = r_t + b_t$.

Precipitation { p_t } is taken as the basic input of the model and discharge from the reservoir { q_t } as the basic output of the model. The output in time interval t, q_t, is controlled by the reservoir content at the start of that period, S_t, and rainfall input in that time period, p_t.

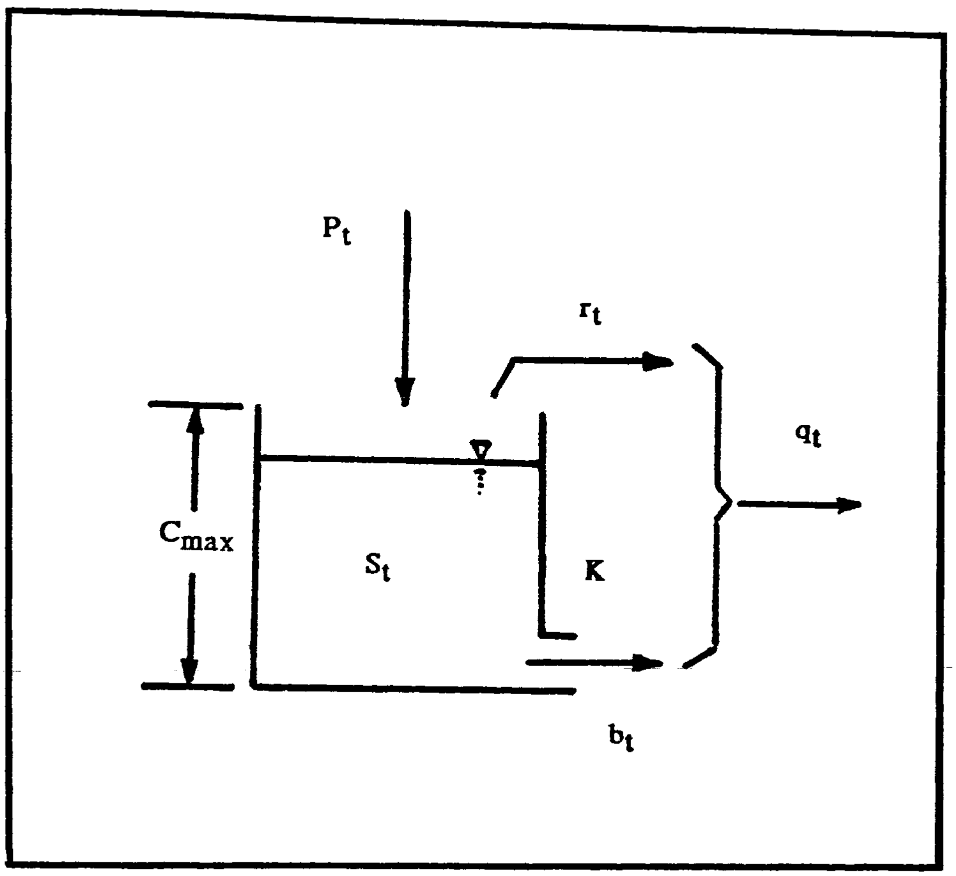

Figure 5.1 Pictorial Representation of a Two-parameter Reservoir Model

The model operates in the following manner: The precipitation input p_t is first added to the reservoir content, S_t, at the beginning of time interval t, i.e. $\tilde{S}_t = S_t + p_t$, where $\tilde{S}$ is interpreted as the intermediate reservoir content. If $\tilde{S}_t \leq C_{max}$, there would no spillflow (i.e. $r_t = 0$), and the output of the model would be equal to base-flow. The equation for the output q_t therefore is

$$q_t = b_t = K \cdot \tilde{S}_t = K \cdot (S_t + p_t) \tag{5.1}$$

If $\tilde{S}_t > C_{max}$, spillflow is resulted, and the output of the model is the sum of base-flow and spillflow. i.e.

$$r_t = S_t + p_t - C_{max}$$

$$b_t = K \cdot C_{max}$$

$$q_t = r_t + b_t = S_t + p_t - (1\text{-}K) \cdot C_{max} \tag{5.2}$$

<u>Calibration of the Model and Test Results</u>

Assuming a set of hypothetically "true" parameters (K^*, C^*_{max}) and an initial reservoir content S_0, a sequence of rainfall data was used by the model (equations (5.1) - (5.3)) to generate a sequence of "true" outflows (q_t s). The rainfall values were assumed to form a series of daily values, i.e., data at equal time intervals. To derive an outflow series at unequal time intervals, a subset of the q_t s values was randomly selected so that the intervals between successive values ranged from one to many days. To this subset of q_t values were added error terms based on equation (2.12). The result of this operation provided an observed series with errors that confirm to those assumed in the development of the criterion given by equation (3.17).

Using the optimization procedure described previously, numerical experiments were performed on the reservoir model using both the new Maximum Likelihood Criterion (MLE) and the more traditional Simple Least Square (SLS) one. In the numerical experiments, 78 values of daily precipitation data were used as input (see

Figure 5.2) and 78 values of daily outflows were calculated. Of the 78 values of daily outflows, the 24 values at times corresponding to the "observed" series were used to calculate the two criteria. The time steps had duration ranging from one day to eight days. The mean duration of one time interval was 3.25 days. The "true" threshold parameter, C^*_{max} was set to 35 mm, while the "true" storage constant was set to 0.2 day^{-1}. The "true" values for α (as defined in equation 2.12) were assumed to be 2.125, 0.495, 0.282, 0.157, and 0.0685, respectively, to reflect various degrees of auto-correlation among measurement errors of outflow data. For instance, if α is equal to 0.0685, the degree of auto-correlation corresponding to the average time duration, 3.25 days, is exp{-0.0685x3.25} = 0.8; if the time duration is one day, the degree of auto-correlation will be 0.934, if 8 days, 0.578. Thus, we observe a variation in degree of auto-correlation from 0.578 to 0.934. By setting the standard deviation σ_a of the a_t's to 15%, 20%, and 30% of the average outflow for the whole series, increasing amounts of random error were introduced into the "observed" outflows. Thus, some 15 combinations of auto-correlation and random error could be tested.

The test results from use of the new criterion and the SLS criterion are recorded in Tables 5.1 to 5.3. Column 1 of the tables gives the α values that are used to synthesize the observed flows. Column 2 lists the degree of auto-correlation of the measurement errors for each α value corresponding to 3.25 days, the average duration of one time interval. Columns 4 and 5 show the parameter estimates. The objective function values, the sum of squared differences of the computed flows from the true flows, $\Sigma SSQ_{com-true}$, and the sum of the squared differences of the computed flows from the observed flows, $\Sigma SSQ_{com-mes}$, are listed in columns 6 to 8. The coefficient of efficiency, E, and the Durbin-Watson d statistic are given in columns 9 and 10.

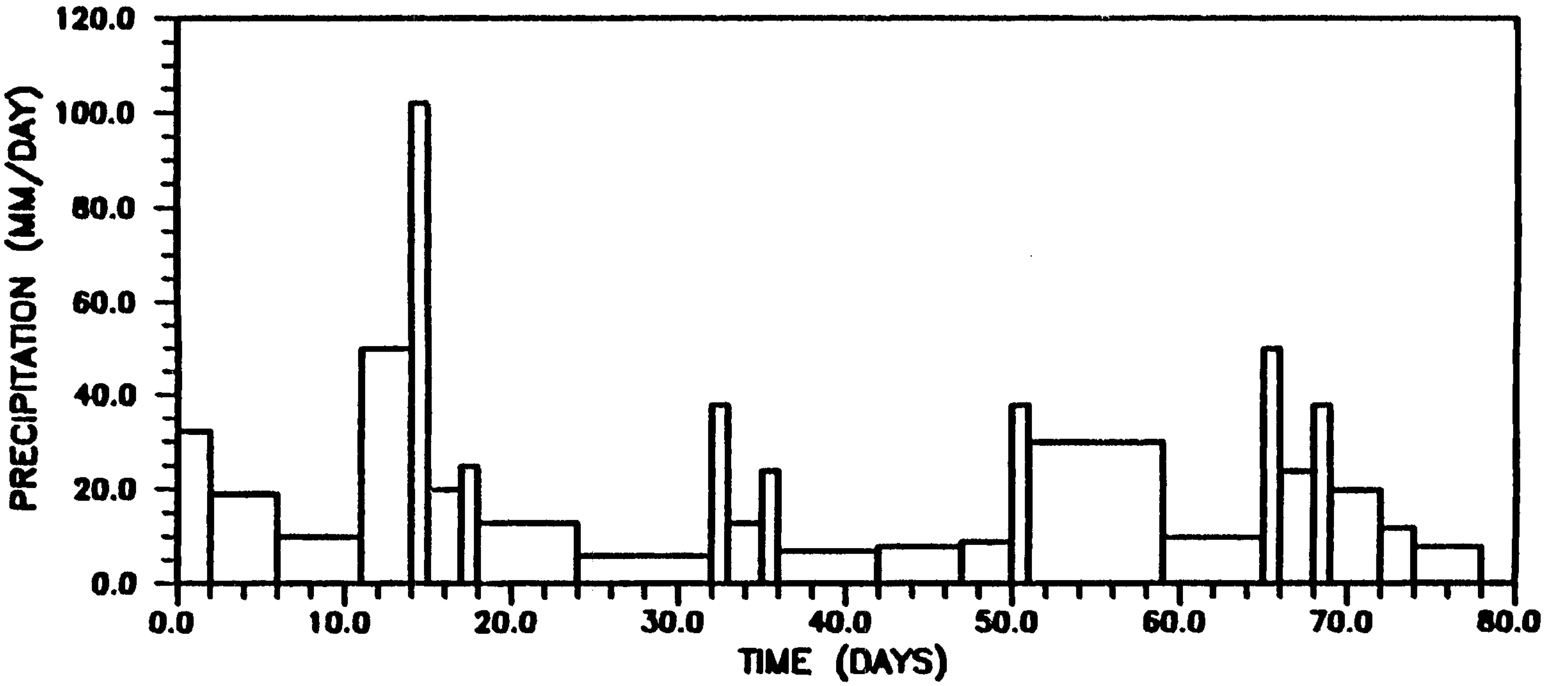

Figure 5.2 Precipitation Data Used for Calibration of the Two-parameter Reservoir Model

Discussion of the Results

For analysis of this model, we have the advantage of being able to compare the parameter estimates against the true parameters. From Tables 4.1 to 4.3, we can see that if there is a correlation existing among the measurement errors (e.g. α = 0.495, 0.282, 0.157, and 0.0695), the parameter estimates obtained by using the new MLE criterion are very close to the true parameter values regardless the amount of errors introduced into data. The SLS estimates, however, are poor approximations of the true parameters. For example, when observed data contain 30 % errors, the parameter K reaches optimum at about 0.65 day^{-1}, far from the true value 0.2, if the SLS criterion is used.

$\Sigma SSQ_{com-true}$ values are a measure of model fit. A small $\Sigma SSQ_{com-true}$ value means that the simulated flows are a close approximation of the true flows. The $\Sigma SSQ_{com-true}$ values for the new MLE are very small compared to the values for the SLS, except for α = 2.125, in which case the auto-correlation among errors are insignificant. The non-correlated case will be addressed later.

Figures 5.3 to 5.7 are the true, observed and simulated hydrographs derived from the use of the new MLE for the case when 30 % errors were introduced into streamflow data. Figures 5.8 to 5.12 are the hydrographs obtained from the use of the SLS criterion. The simulated hydrographs from the use of the MLE parameter estimates match very well against the corresponding true hydrographs except for α = 2.125. In contrast, the simulated hydrographs from use of the SLS parameter estimates deviate from the true hydrographs.

Because the model has only two parameters, we are able to produce the contour maps of the objective function response surface for any given model configuration. Figures 5.13 to 5.22 are the contour maps for the case that 15 % measurement errors

Table 5.1. Estimation results for cases when 15% errors were introduced into observed outflows

	Error Model Configuration $\sigma_{true} = 3.646$					Rainfall-Runoff Model Configuration $C_{max_{true}} = 35$ mm, $K_{true} = 0.2$ day^{-1}				
	1	2	3	4	5	6	7	8	9	10
No.	α	$\bar{\rho}^{\dagger}$	Objective function type	C_{max}	K	$f^{\dagger\dagger}$	$\Sigma SSQ_{com\text{-}true}$	$\Sigma SSQ_{com\text{-}mes}$	E	'd '
1	2.125	0.001	MLE	24.7660	0.3150	196.3	123.51	224.19	0.9792	2.6234
			SLS	30.0018	0.2522	196.3	33.13	196.35	0.9817	2.8086
2	0.495	0.2	MLE	33.4379	0.2234	206.9	6.43	196.10	0.9812	2.0953
			SLS	29.9014	0.2511	179.8	32.64	179.76	0.9827	2.3020
3	0.282	0.4	MLE	34.8351	0.2129	216.7	2.88	234.65	0.9769	1.5603
			SLS	29.8998	0.2499	205.7	32.07	205.69	0.9798	1.8269
4	0.157	0.6	MLE	35.2382	0.2096	238.4	2.47	303.46	0.9697	0.8196
			SLS	29.9093	0.2523	270.3	33.23	270.32	0.9730	1.3617
5	0.0685	0.8	MLE	35.4034	0.2080	254.4	2.27	389.29	0.9615	0.8196
			SLS	30.2829	0.2667	352.8	41.19	352.81	0.9651	1.0053

$\dagger\bar{\rho}$ **--equivalent degree of auto-correlation among measurement errors corresponding to 3.25 days, the average** duration of one time interval.

$\dagger\dagger f$ -- the objective function values.

Table 5.2. Estimation results for cases when 20% errors were introduced into observed outflows

	Error Model Configuration $\sigma_{true} = 4.861$					Rainfall-Runoff Model Configuration $C_{max_{true}} = 35$ mm, $K_{true} = 0.2$ day^{-1}				
	1	2	3	4	5	6	7	8	9	10
No.	α	$\bar{\rho}^{\dagger}$	Objective function type	C_{max}	K	$f^{\dagger\dagger}$	$\Sigma SSQ_{com-true}$	$\Sigma SSQ_{com-mes}$	E	'd'
1	2.125	0.001	MLE	24.9935	0.3121	323.1	118.54	329.70	0.9690	2.8240
			SLS	27.883	0.2797	318.6	65.05	318.62	0.9701	2.8548
2	0.495	0.2	MLE	33.4192	0.2293	334.4	9.84	319.74	0.9690	2.0750
			SLS	28.1787	0.2754	284.8	59.79	315.80	0.9721	2.3409
3	0.282	0.4	MLE	35.0186	0.2167	347.6	5.59	411.41	0.9626	1.5573
			SLS	28.1104	0.2736	319.7	59.29	372.83	0.9679	1.8845
4	0.157	0.6	MLE	35.4717	0.2126	378.9	4.94	411.28	0.9522	1.1712
			SLS	28.2431	0.2751	411.4	59.04	372.83	0.9583	1.4252
5	0.0685	0.8	MLE	35.6087	0.2097	448.0	3.89	695.51	0.9301	0.8061
			SLS	28.8043	0.2918	629.4	68.30	629.36	0.9368	0.9919

$^{\dagger}\bar{\rho}$ --equivalent degree of auto-correlation among measurement errors corresponding to 3.25 days, the average duration of one time interval.

$^{\dagger\dagger}f$ -- the objective function values.

Table 5.3. Estimation results for cases when 30% errors were introduced into observed outflows

	Error Model Configuration $\sigma_{true} = 7.292$				Rainfall-Runoff Model Configuration $C_{max_{true}} = 35$ mm, $K_{true} = 0.2$ day^{-1}					
	1	2	3	4	5	6	7	8	9	10
No.	α	$\bar{\rho}^{\dagger}$	Objective function type	C_{max}	K	$f^{\dagger\dagger}$	$\Sigma SSQ_{com-true}$	$\Sigma SSQ_{com-mes}$	E	'd'
1	2.125	0.001	MLE	28.4538	0.7416	623.6	513.58	701.95	0.9328	2.5917
			SLS	25.0840	0.3121	639.3	117.33	639.26	0.9387	2.8815
2	0.495	0.2	MLE	32.3539	0.2411	699.5	17.70	699.55	0.9346	2.0867
			SLS	25.1863	0.3148	588.7	117.74	588.74	0.9406	2.3647
3	0.282	0.4	MLE	35.0554	0.2225	718.1	9.91	759.50	0.9216	1.5519
			SLS	32.2558	0.6541	638.6	453.67	638.60	0.9341	1.9860
4	0.157	0.6	MLE	36.0756	0.2162	771.9	11.07	948.45	0.9017	1.1677
			SLS	32.5131	0.6490	776.6	451.86	776.57	0.9195	1.0107
5	0.0685	0.8	MLE	36.1060	0.2141	905.7	9.62	1361.00	0.8635	0.8244
			SLS	33.2544	0.6345	1122.3	449.25	1122.31	0.8875	1.1658

$^{\dagger}\bar{\rho}$ **--equivalent degree of auto-correlation among measurement errors corresponding to 3.25 days, the average** duration of one time interval.

$^{\dagger\dagger}f$ -- the objective function values.

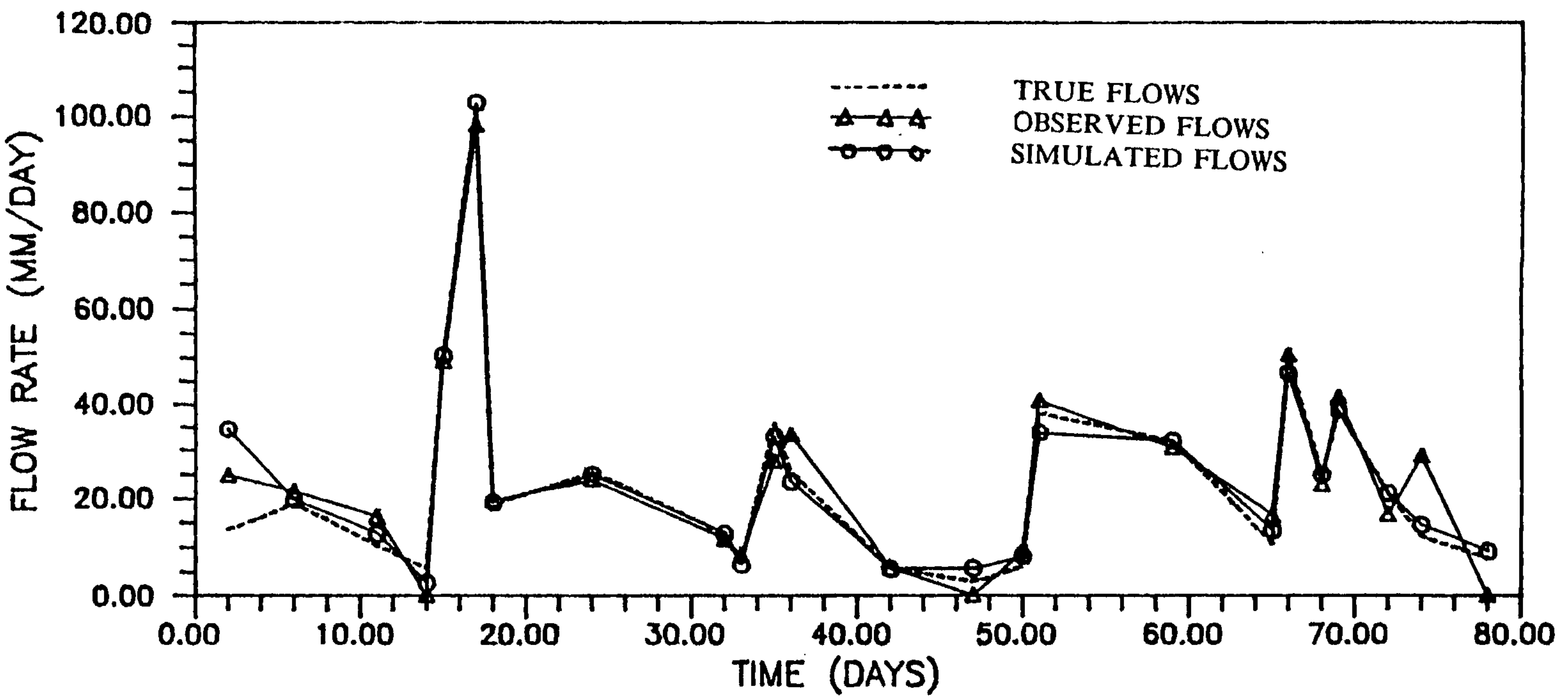

Figure 5.3 Hydrographs for Case C-1 Using MLE:
30% Measurement Error ($\sigma_a = 7.292$)
Degree of Auto-correlation = 0.001 ($\alpha = 2.125$)

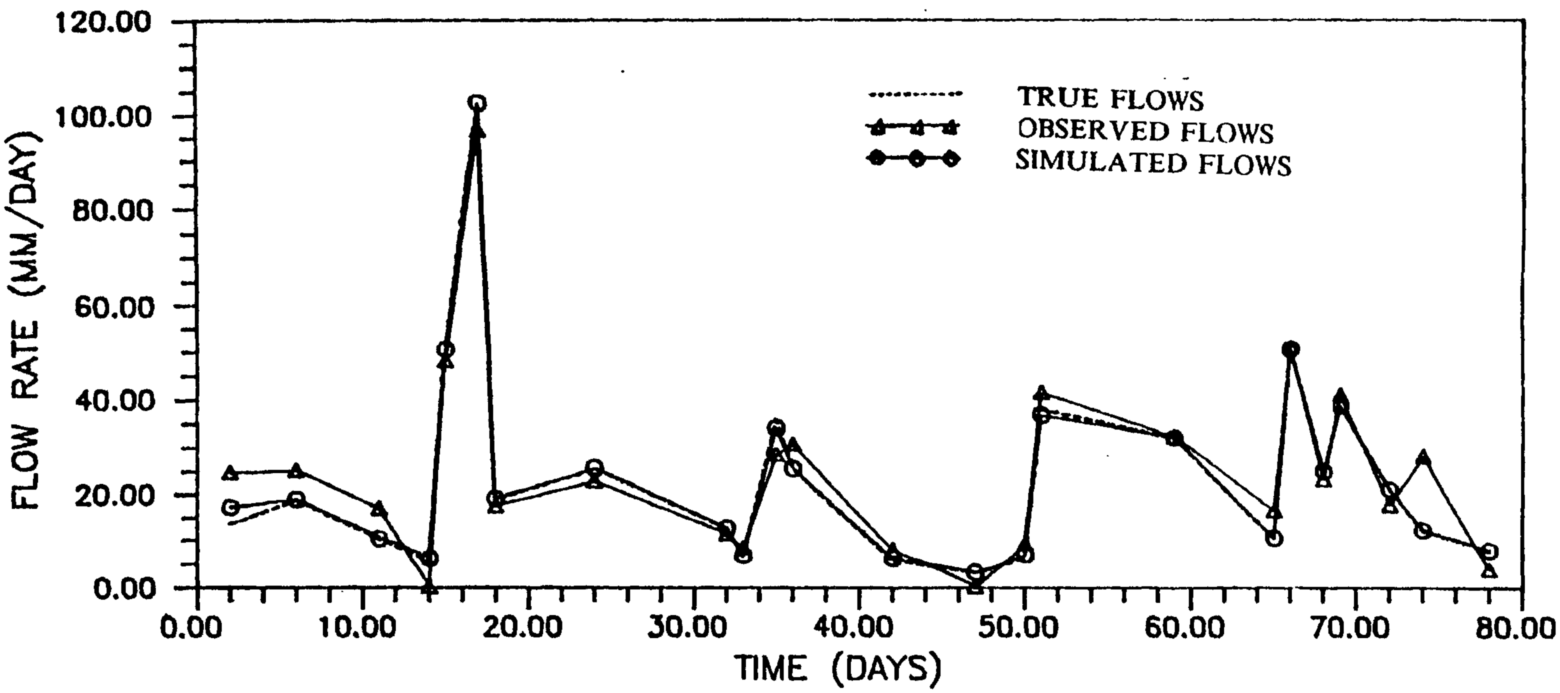

Figure 5.4 Hydrographs for Case C-2 Using MLE:
30% Measurement Error ($\sigma_a = 7.292$)
Degree of Auto-correlation = 0.2 ($\alpha = 0.495$)

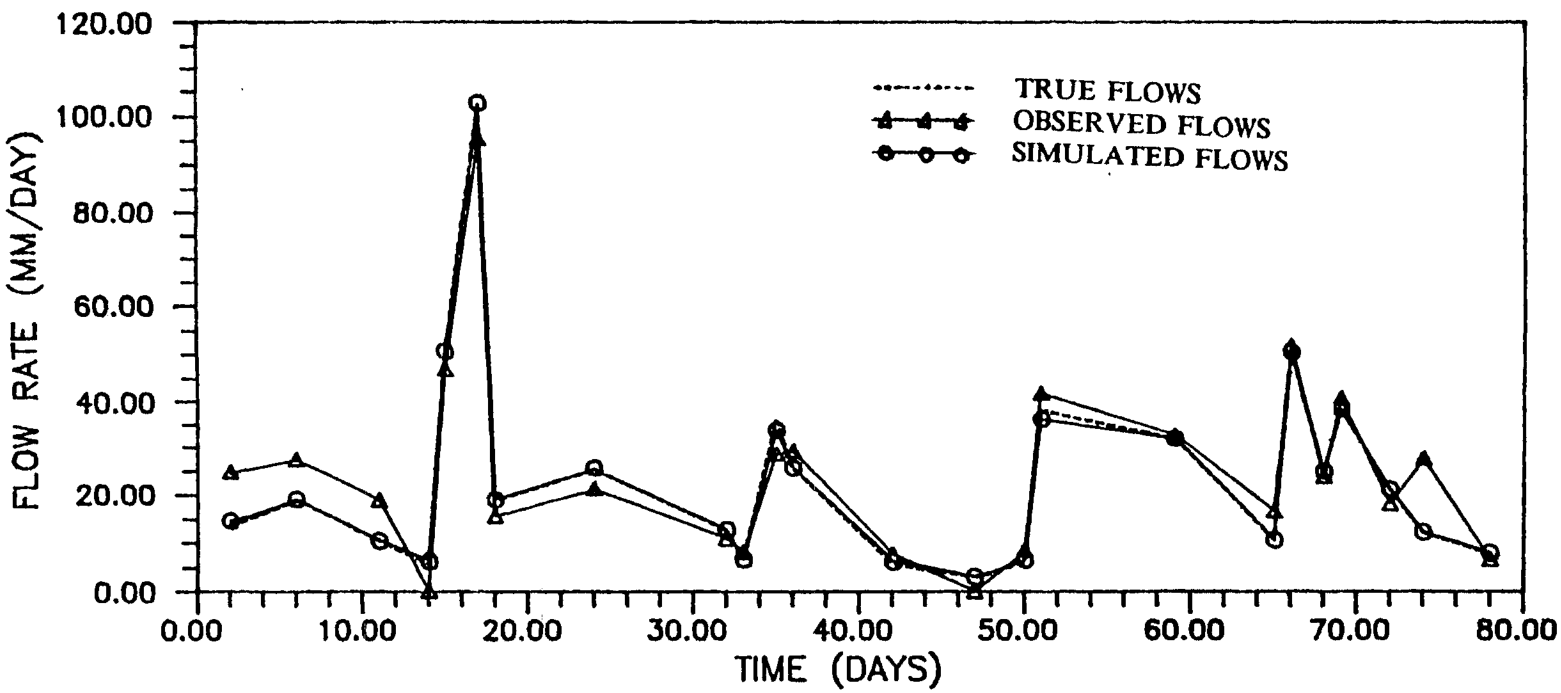

Figure 5.5 Hydrographs for Case C-3 Using MLE:
30% Measurement Error (σ_a = 7.292)
Degree of Auto-correlation = 0.4 (α = 0.282)

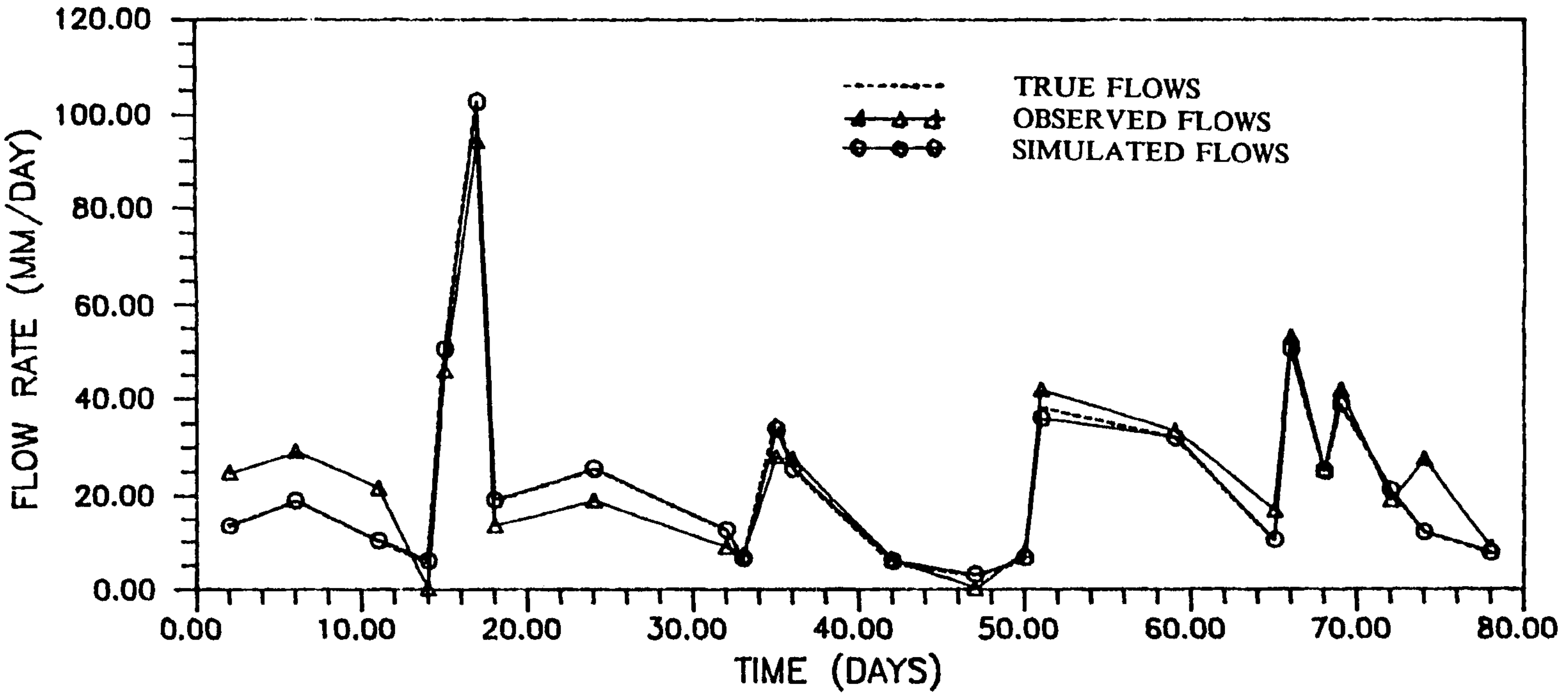

Figure 5.6 Hydrographs for Case C-4 Using MLE:
30% Measurement Error (σ_a = 7.292)
Degree of Auto-correlation = 0.6 (α = 0.157)

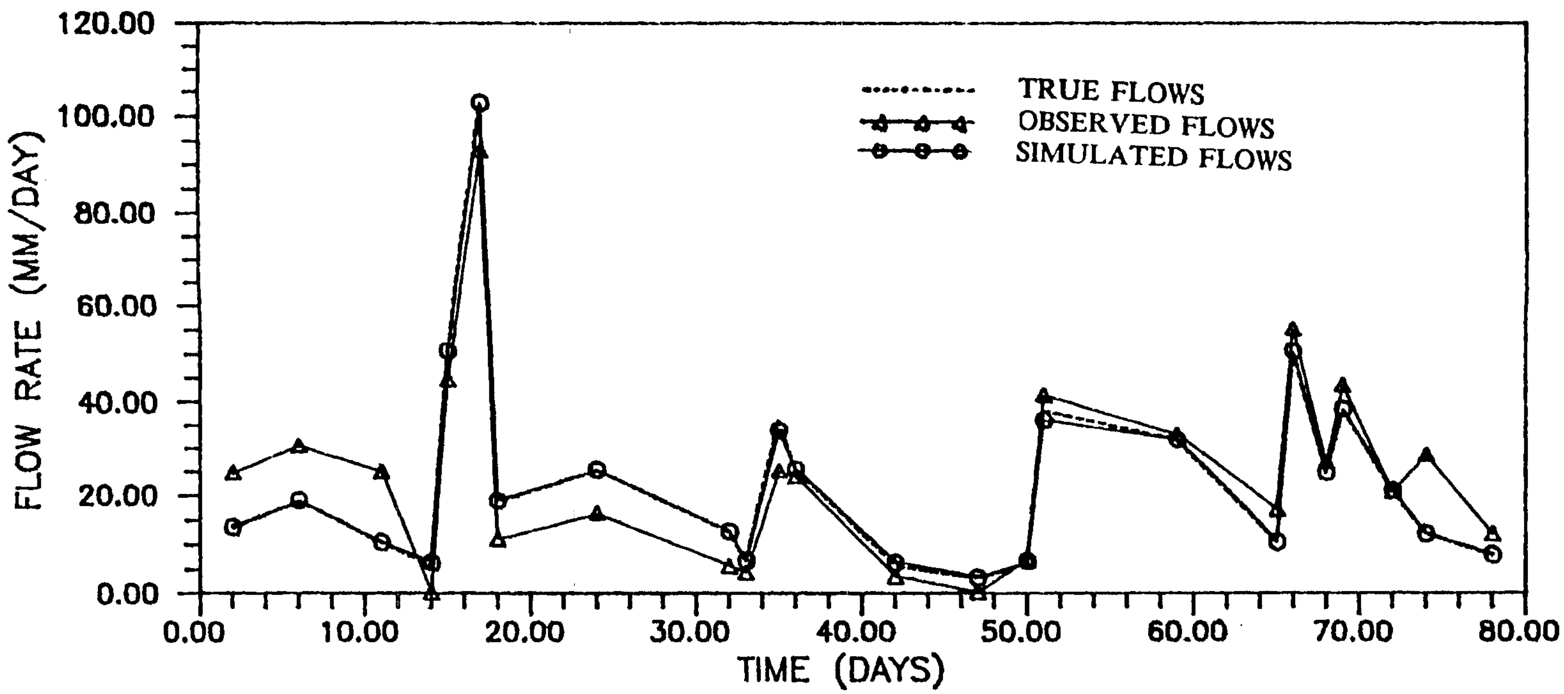

Figure 5.7 Hydrographs for Case C-5 Using MLE:
30% Measurement Error ($\sigma_a = 7.292$)
Degree of Auto-correlation = 0.8 ($\alpha = 0.0685$)

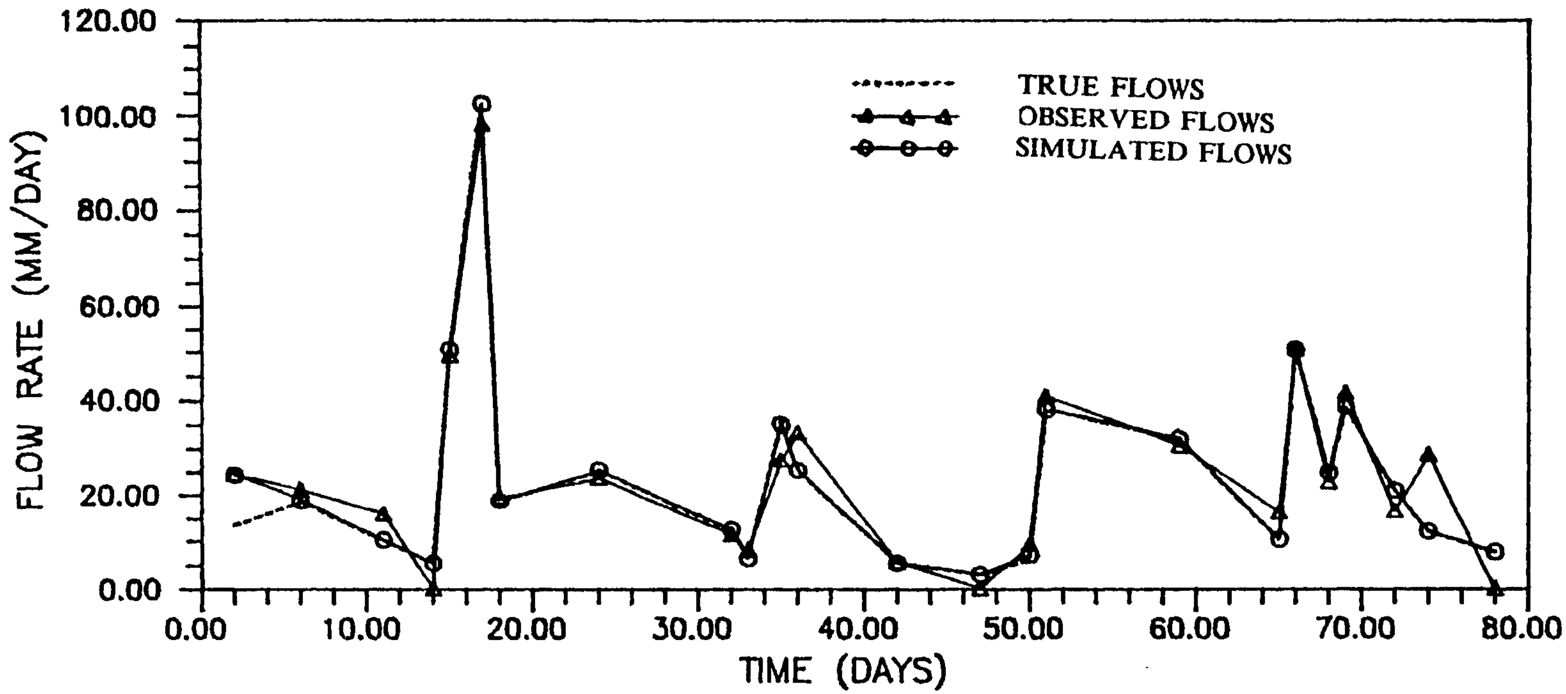

Figure 5.8 Hydrographs for Case C-1 Using SLS: 30% Measurement Error (σ_a = 7.292) Degree of Auto-correlation = 0.001 (α = 2.125)

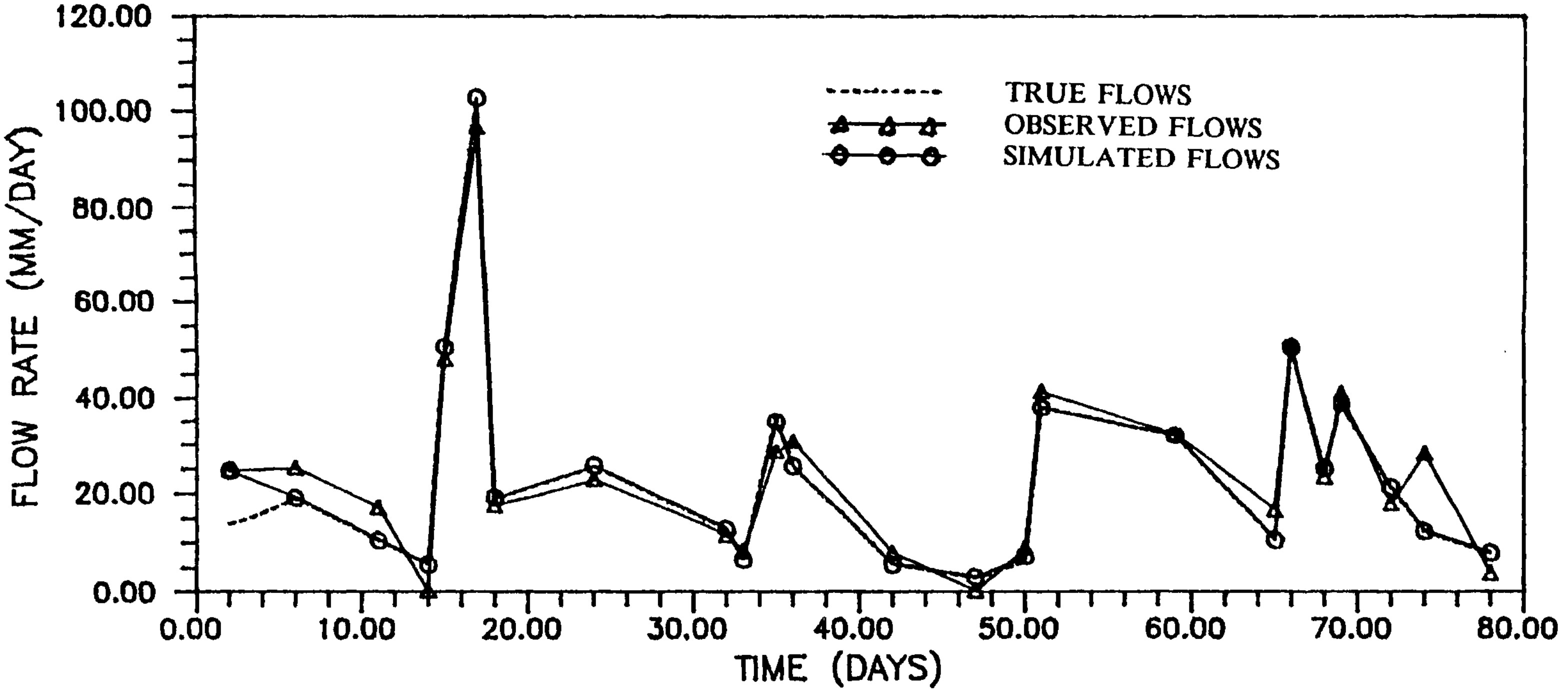

Figure 5.9 Hydrographs for Case C-2 Using SLS: 30% Measurement Error (σ_a = 7.292) Degree of Auto-correlation = 0.2 (α = 0.495)

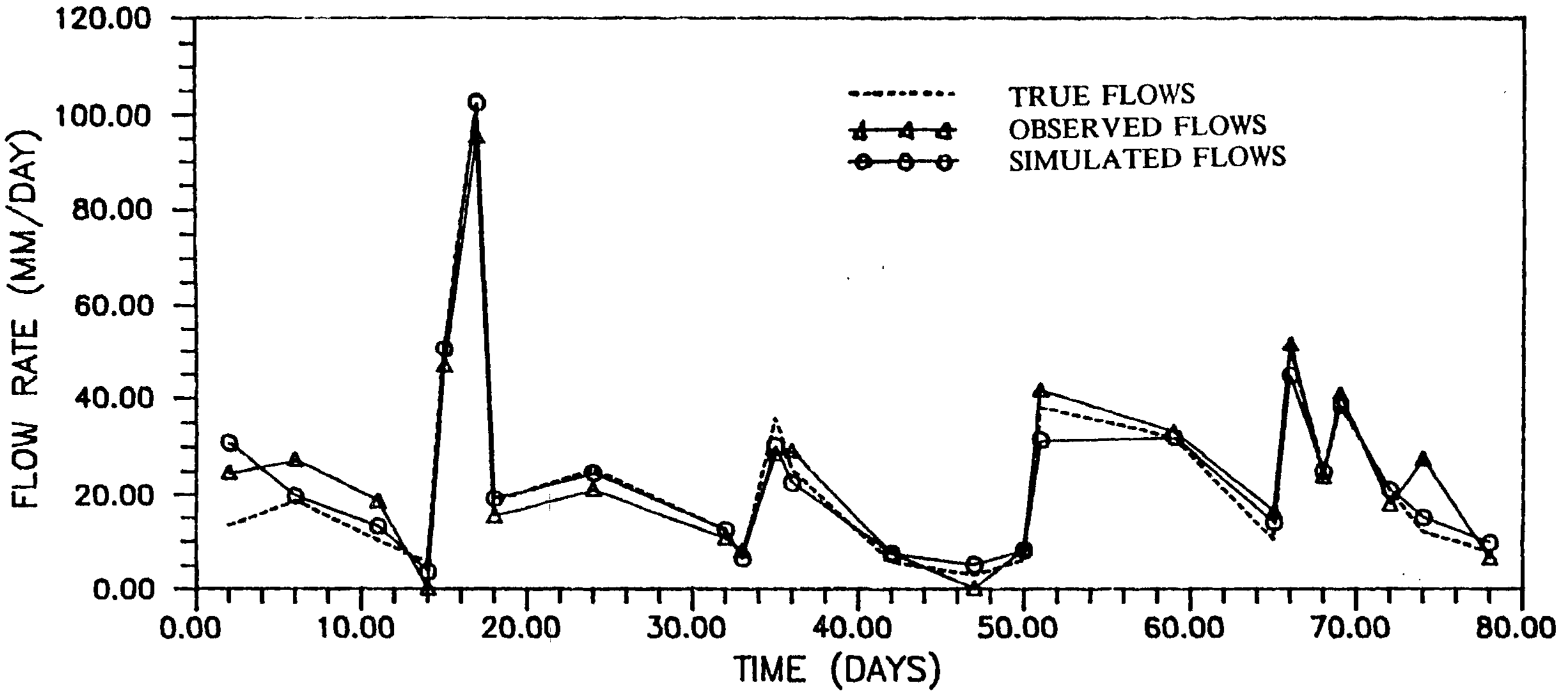

Figure 5.10 Hydrographs for Case C-3 Using SLS: 30% Measurement Error ($\sigma_a = 7.292$) Degree of Auto-correlation = 0.4 ($\alpha = 0.282$)

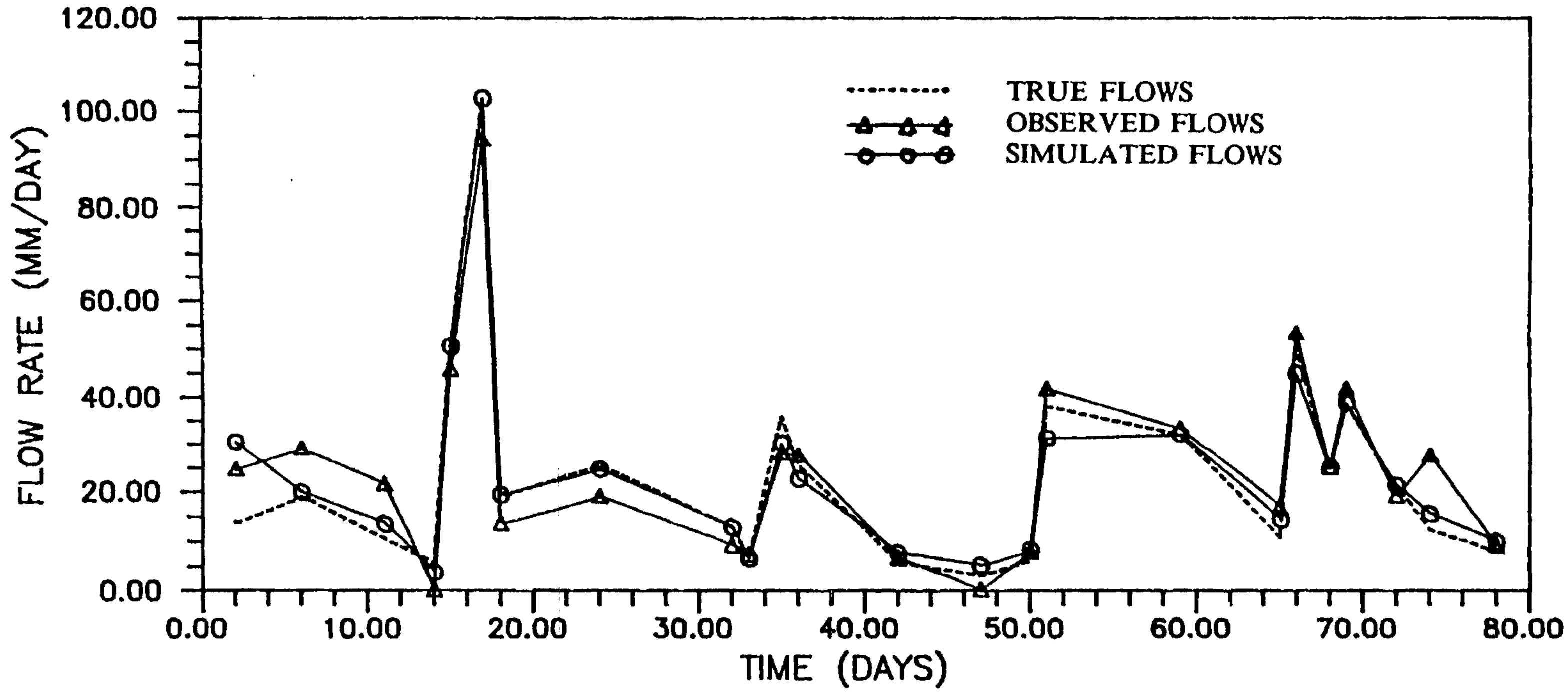

Figure 5.11 Hydrographs for Case C-4 Using SLS:
30% Measurement Error (σ_a = 7.292)
Degree of Auto-correlation = 0.6 (α = 0.157)

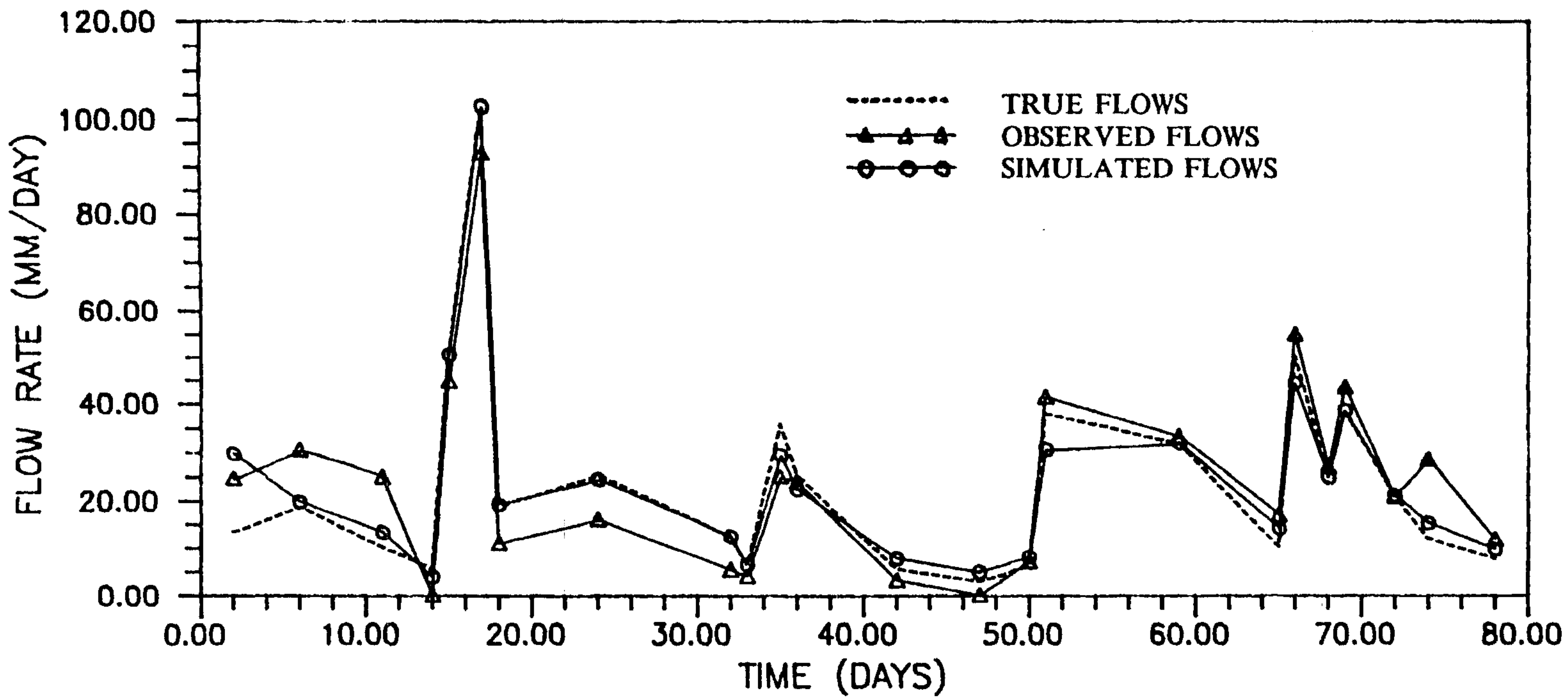

Figure 5.12 Hydrographs for Case C-5 Using SLS:
30% Measurement Error (σ_a = 7.292)
Degree of Auto-correlation = 0.8 (α = 0.0685)

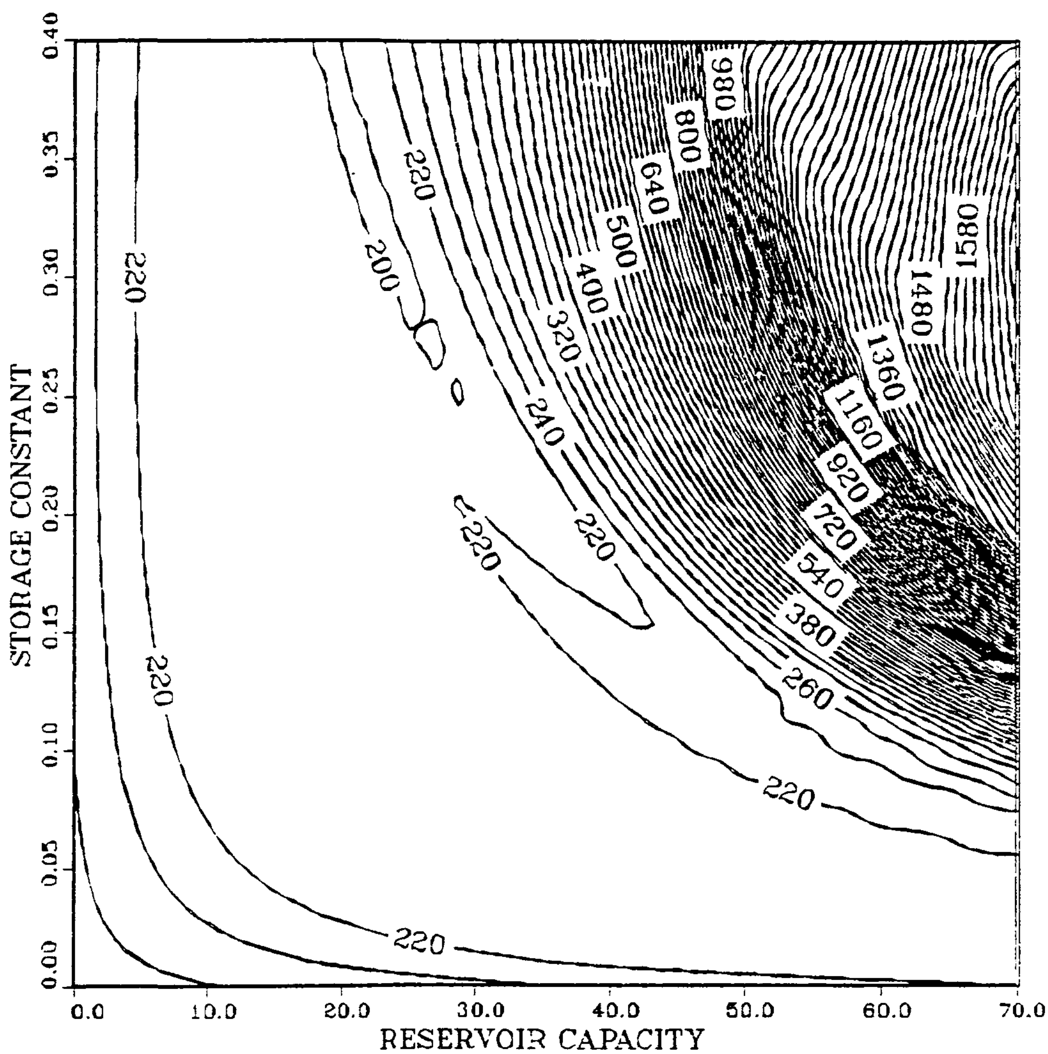

Figure 5.13 Contour Map of MLE for Case A-1:
15% Measurement Error (σ_a = 3.646)
Degree of Auto-correlation = 0.001 (α = 2.125)

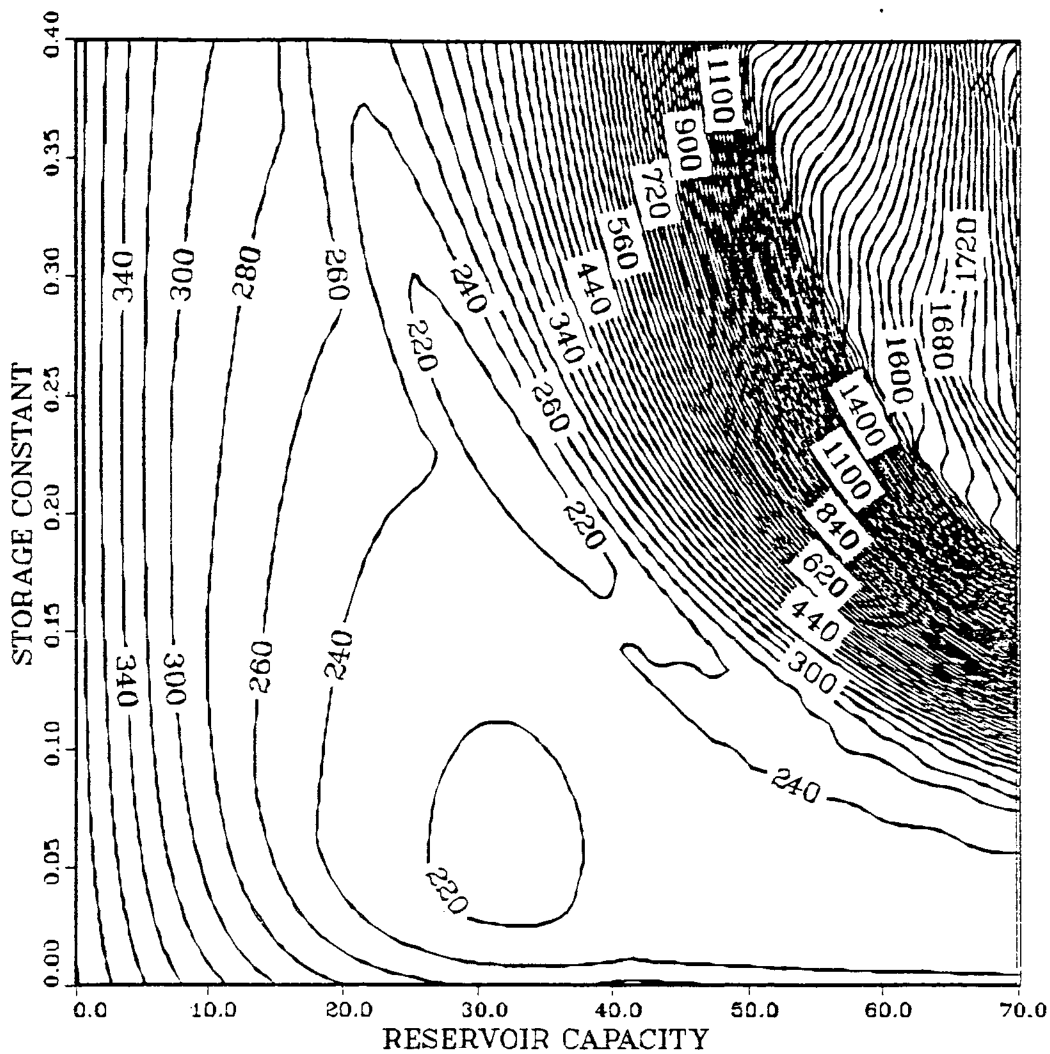

Figure 5.14 Contour Map of MLE for Case A-2:
15% Measurement Error (σ_a = 3.646)
Degree of Auto-correlation = 0.2 (α = 0.495)

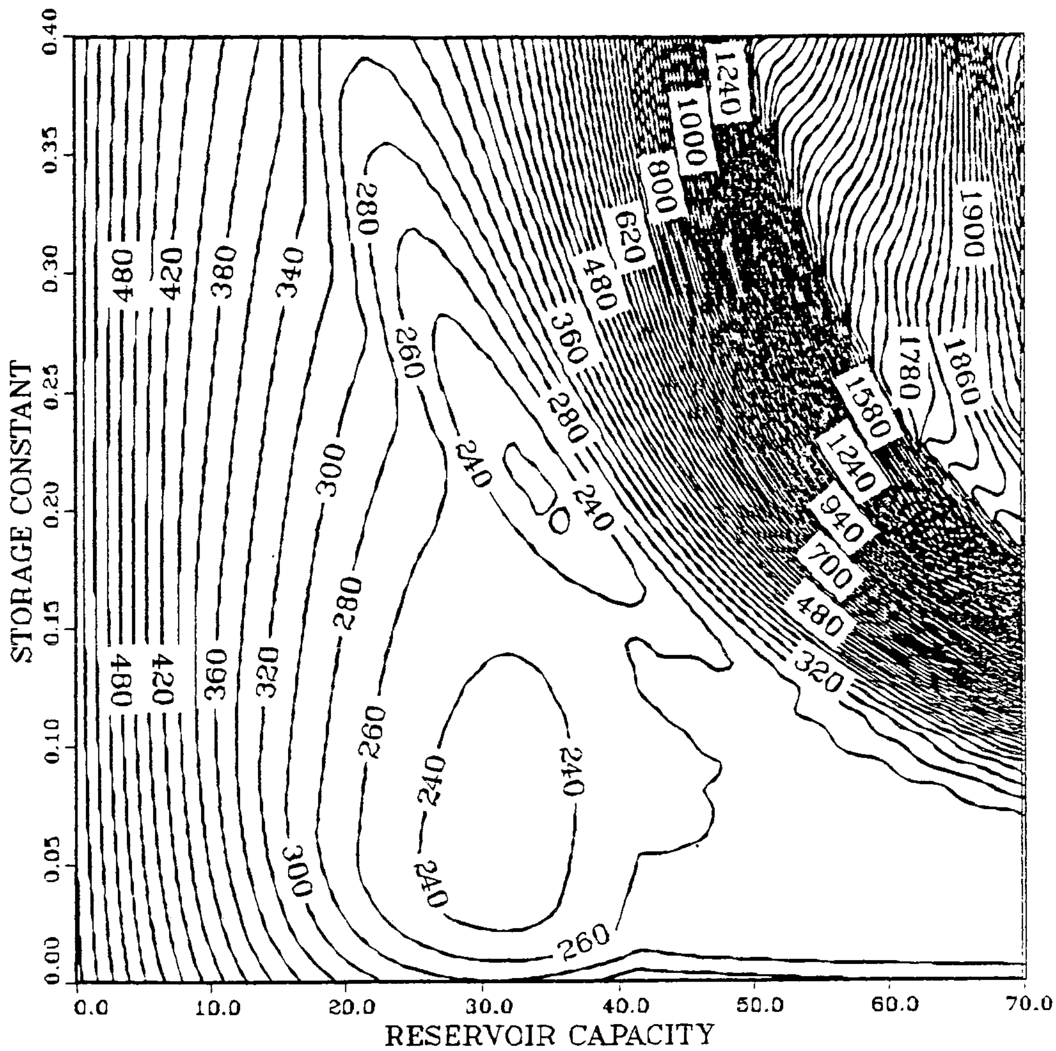

Figure 5.15 Contour Map of MLE for Case A-3:
15% Measurement Error (σ_a = 3.646)
Degree of Auto-correlation = 0.4 (α = 0.282)

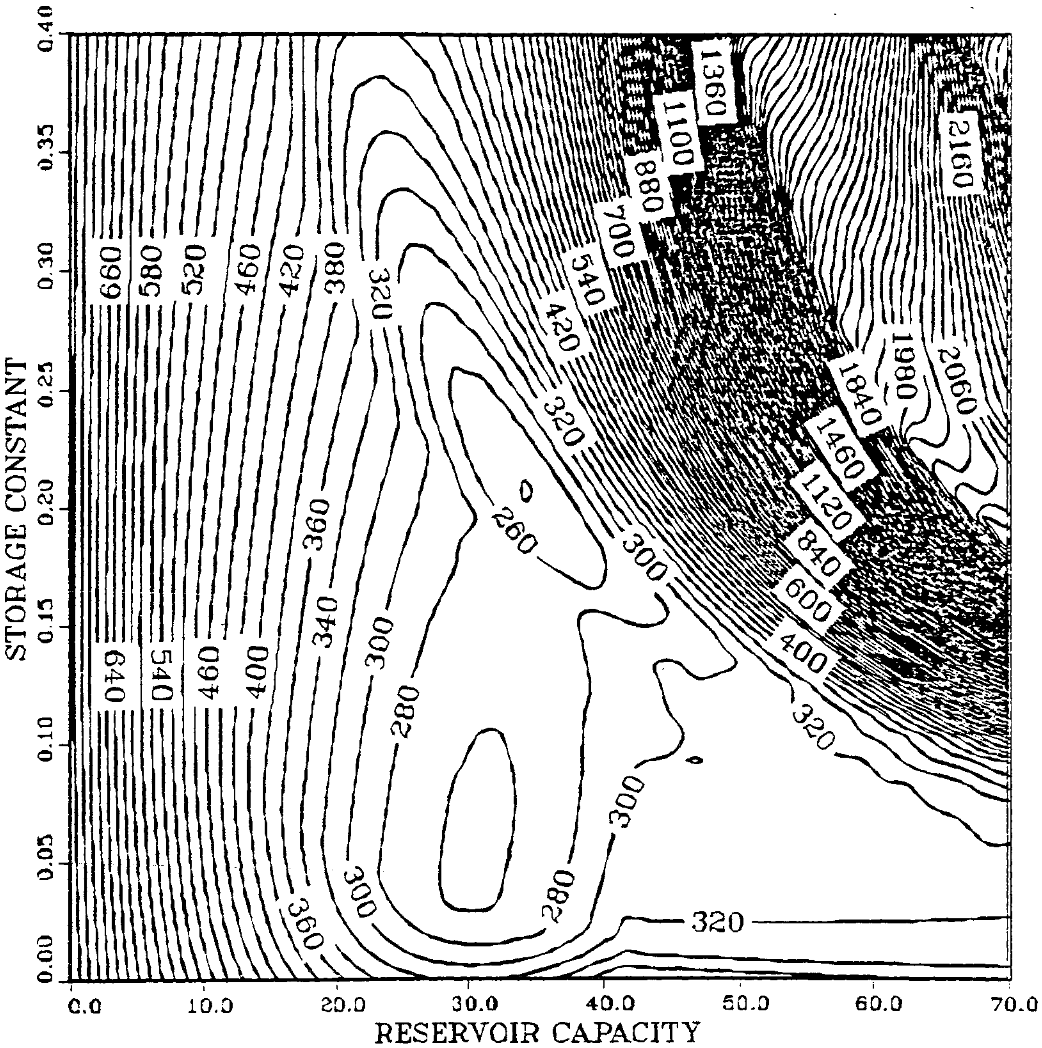

Figure 5.16 Contour Map of MLE for Case A-4:
15% Measurement Error (σ_a = 3.646)
Degree of Auto-correlation = 0.6 (α = 0.157)

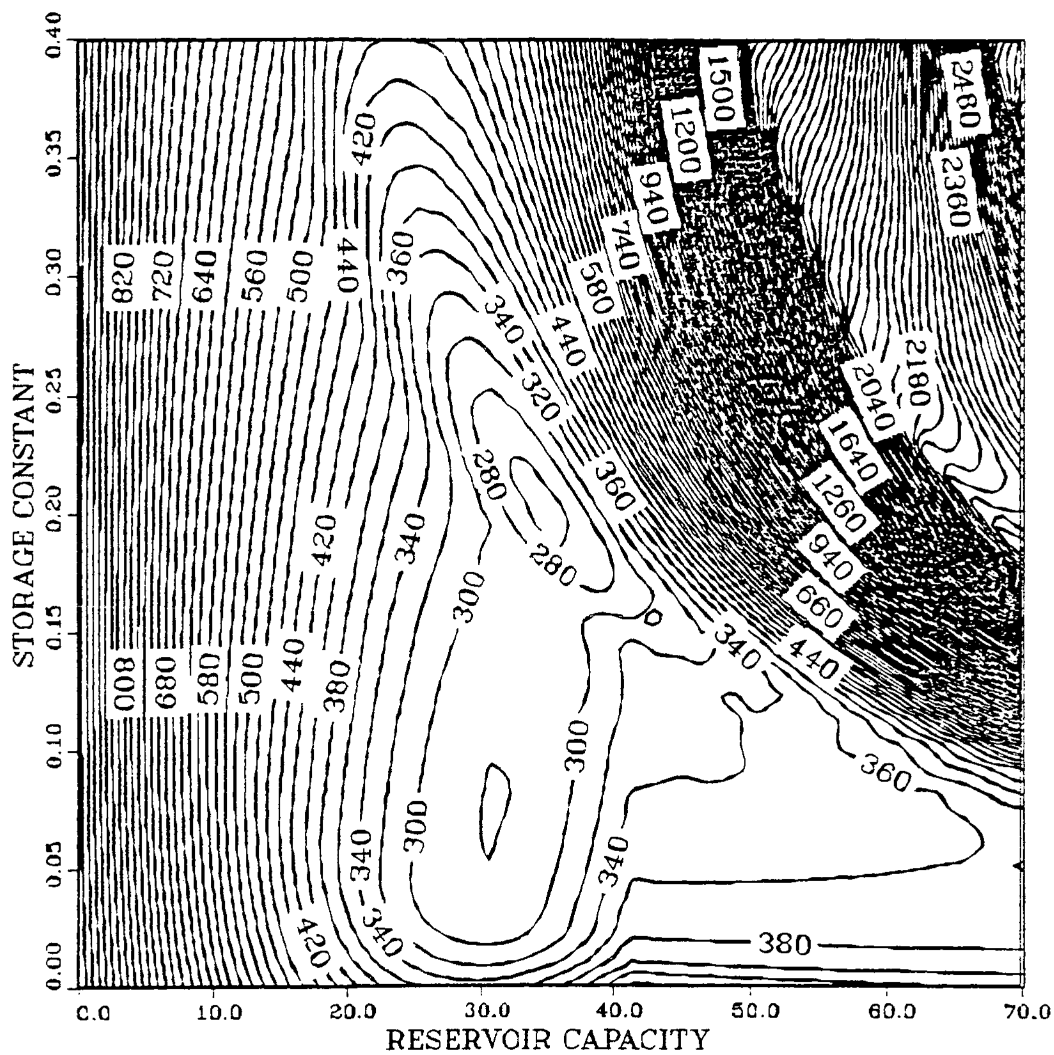

Figure 5.17 Contour Map of MLE for Case A-5:
15% Measurement Error (σ_a = 3.646)
Degree of Auto-correlation = 0.8 (α = 0.0685)

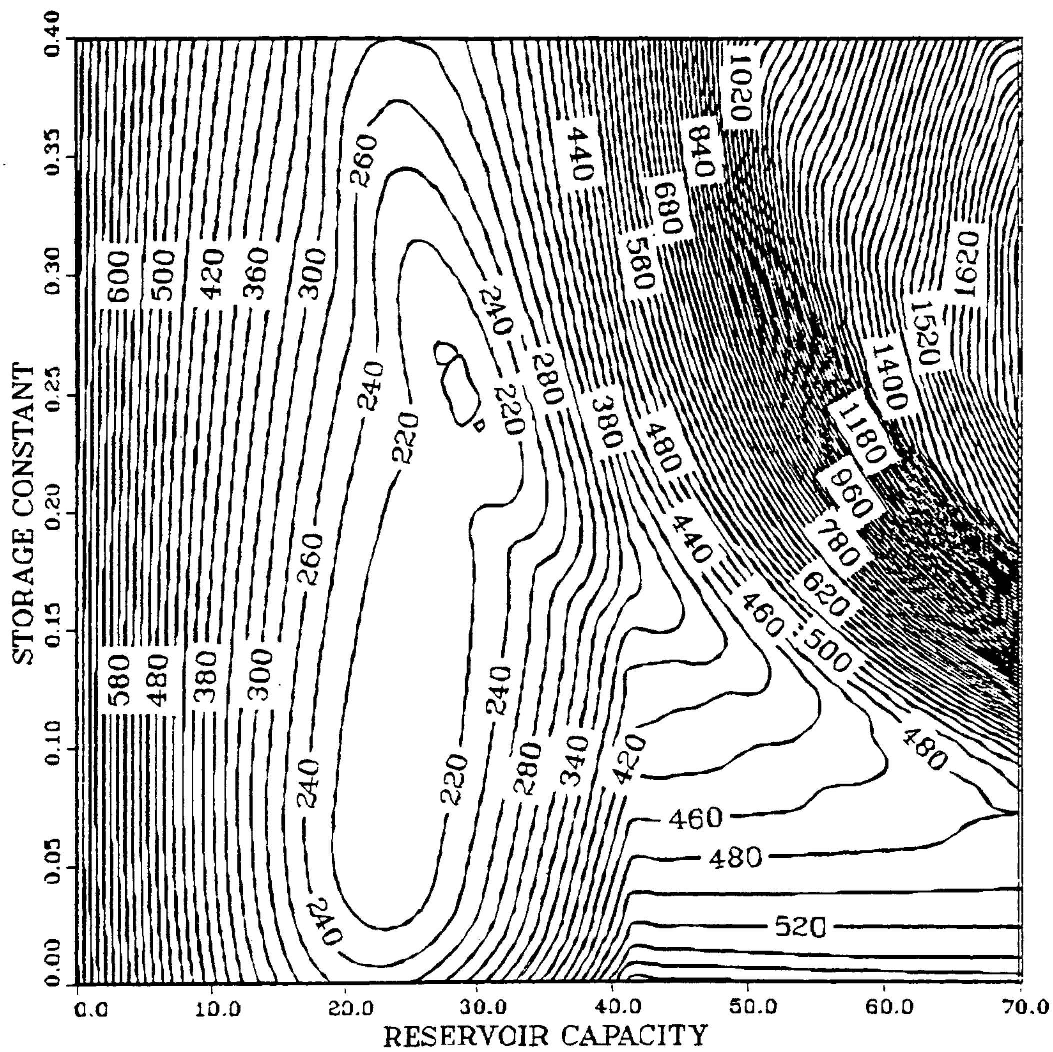

Figure 5.18 Contour Map of SLS for Case A-1:
15% Measurement Error (σ_a = 3.646)
Degree of Auto-correlation = 0.001 (α = 2.125)

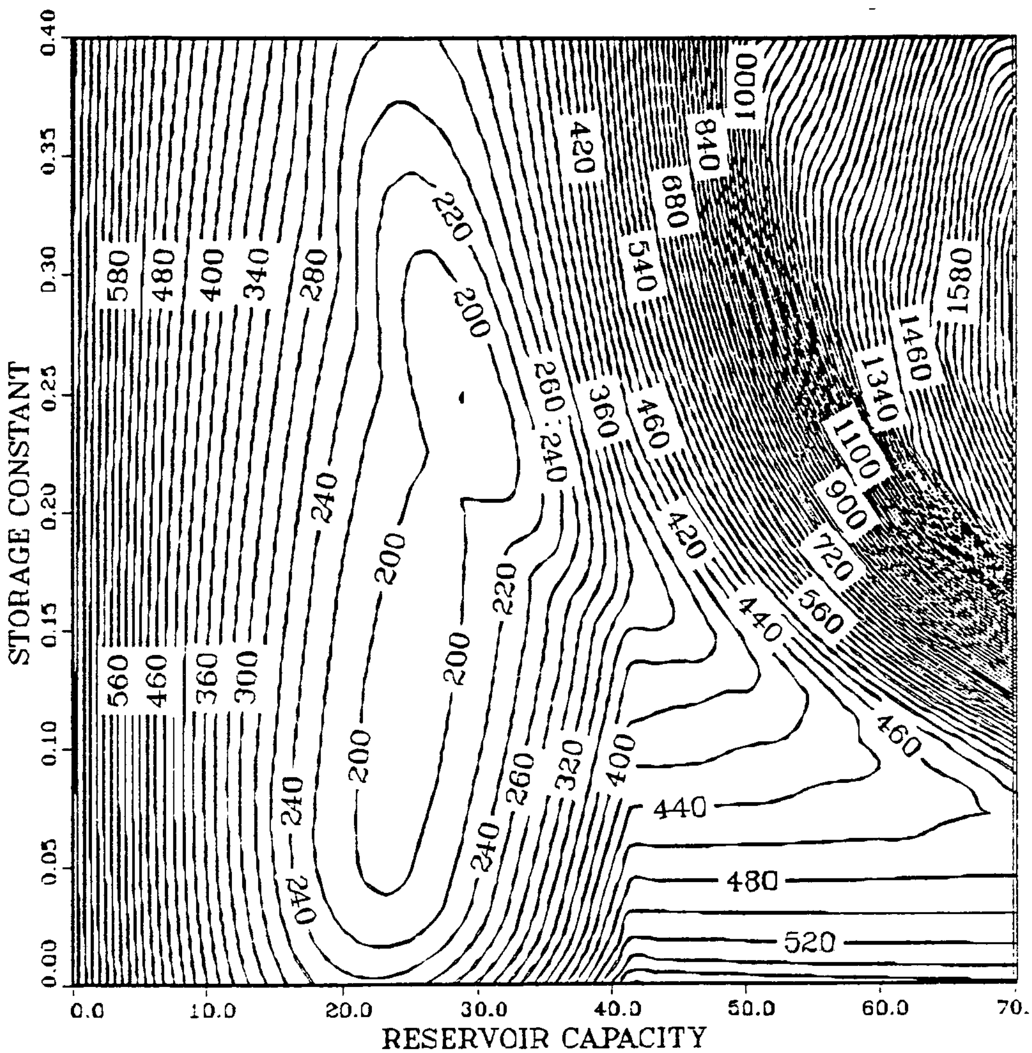

Figure 5.19 Contour Map of SLS for Case A-2:
15% Measurement Error (σ_a = 3.646)
Degree of Auto-correlation = 0.2 (α = 0.495)

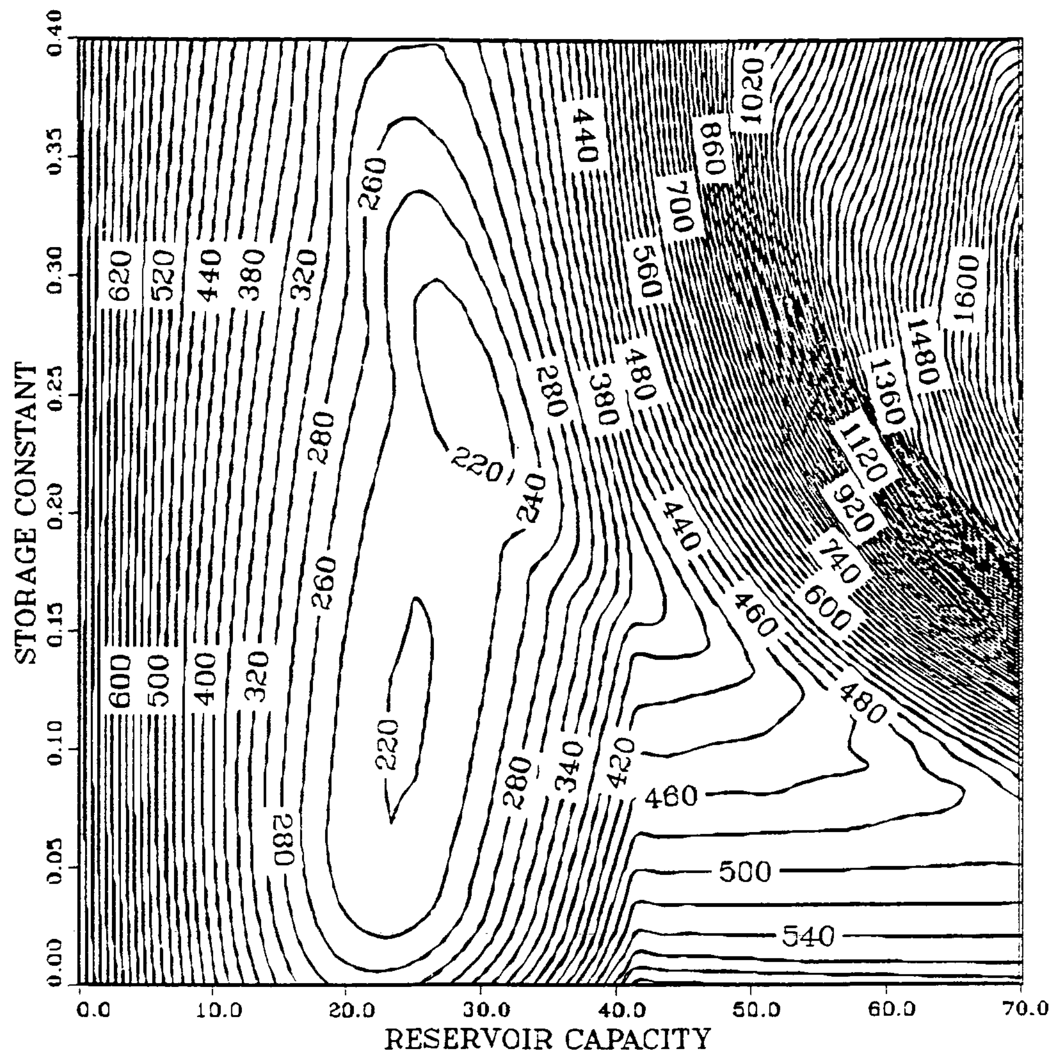

Figure 5.20 Contour Map of SLS for Case A-3:
15% Measurement Error (σ_a = 3.646)
Degree of Auto-correlation = 0.4 (α = 0.282)

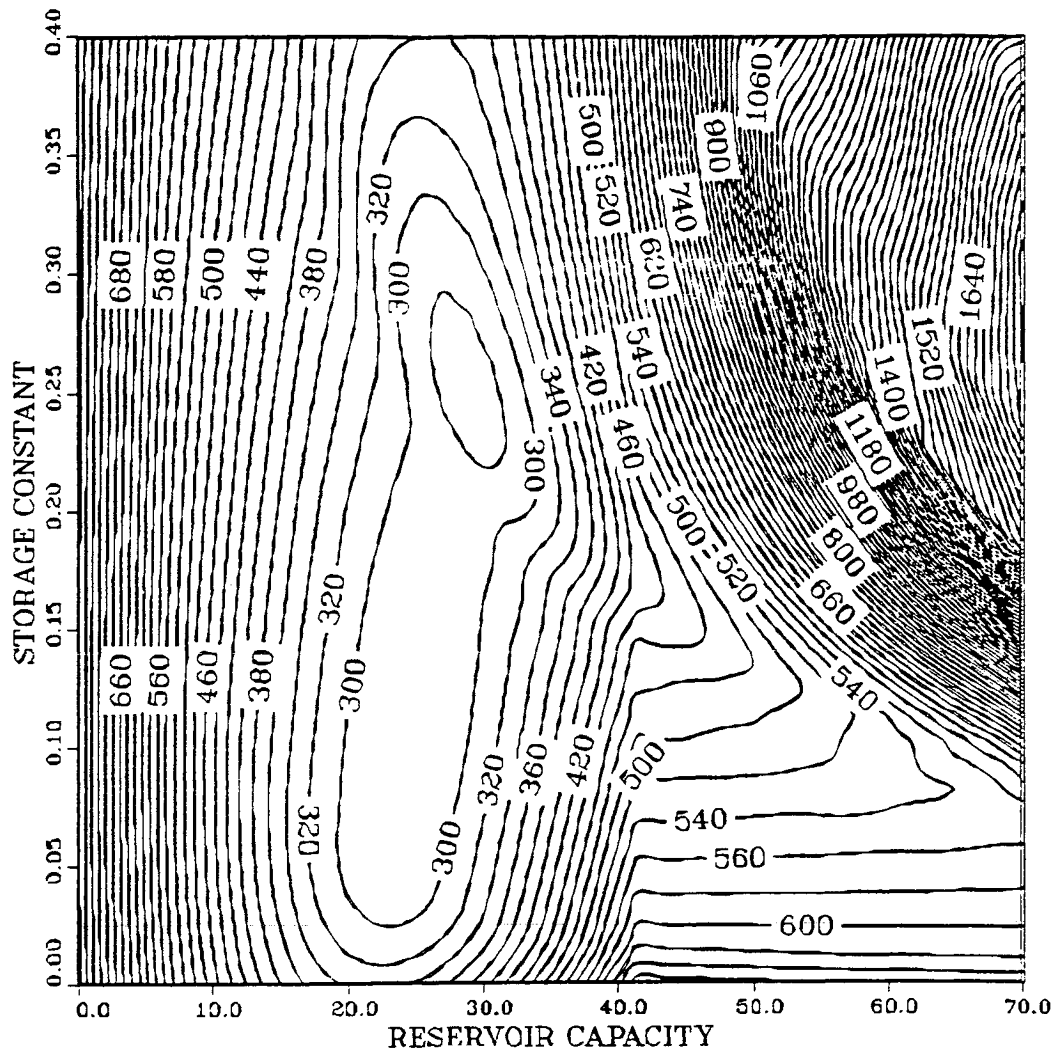

Figure 5.21 Contour Map of SLS for Case A-4: 15% Measurement Error (σ_a = 3.646) Degree of Auto-correlation = 0.6 (α = 0.157)

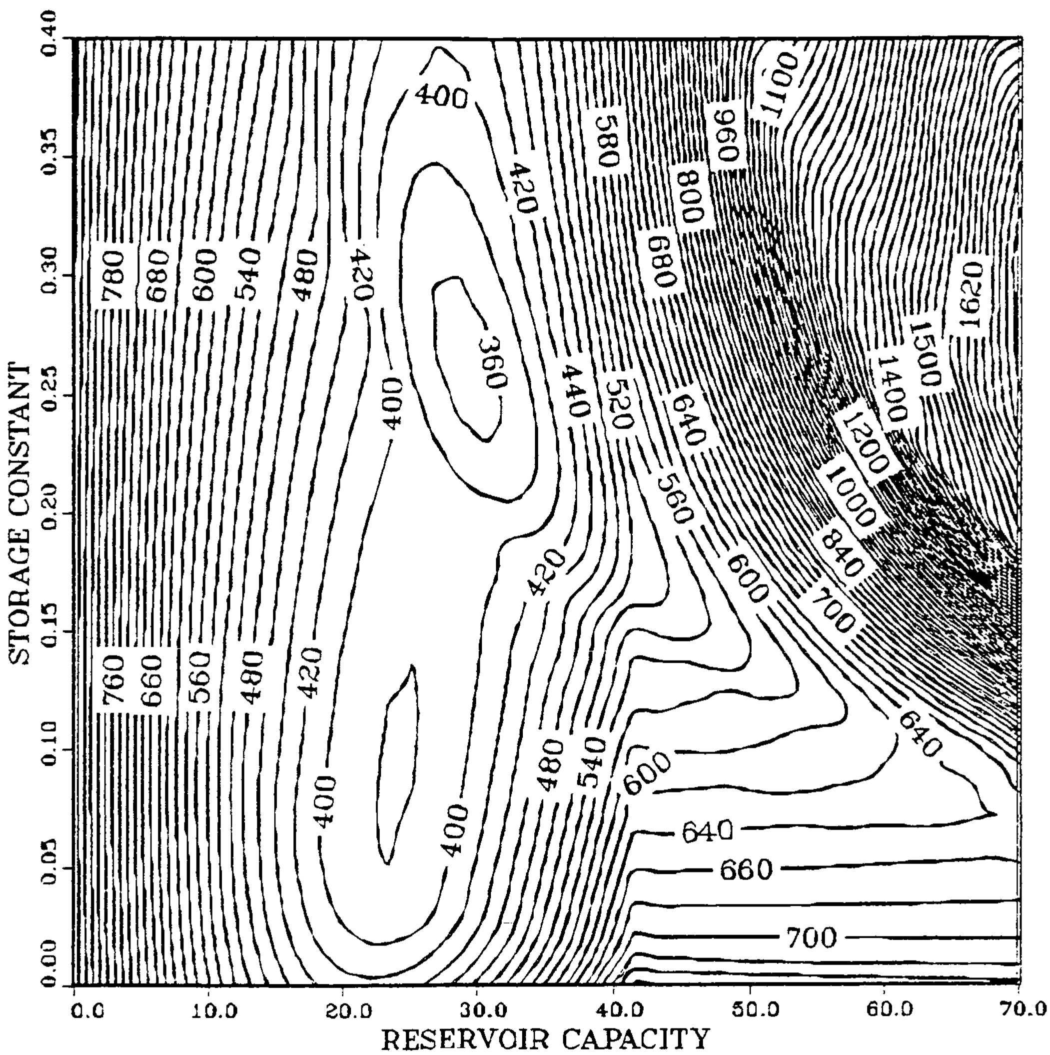

Figure 5.22 Contour Map of SLS for Case A-5:
15% Measurement Error (σ_a = 3.646)
Degree of Auto-correlation = 0.8 (α = 0.0685)

are introduced into data. Figures 5.13 to 5.17 are the contour maps of the new MLE while Figure 5.18 to 5.22 are those of the SLS. We can see that the global minima of the contours for the MLE are observed around the true parameters (K^*, C^*_{max}) = (0.2, 35.0) except for α = 2.125. But the minima of the SLS response functions are located away from the true parameters.

During this study, $\Sigma SSQ_{com-mes}$ values, which are used to indicate the goodness of curve fitting, were also computed. For a better curve fit, $\Sigma SSQ_{com-mes}$ takes on a small value. The $\Sigma SSQ_{com-mes}$ values recorded in Tables 5.1 to 5.3 show that the SLS criterion does a better curve fitting job than the new MLE criterion.

Another parameter E, the coefficient of efficiency, introduced by Nash and Sutcliffe in 1970, describes the degree of association between the observed flows and the computed flows. The coefficient of efficiency E is defined as

$$E = \frac{\sum_{t=1}^{n} \left[q_m - \bar{q}_m\right]^2 - \sum_{t=1}^{n} (q_c - q_m)^2}{\sum_{t=1}^{n} \left[q_m - \bar{q}_m\right]^2} \qquad (5.3)$$

where q_m is the observed flow; $\bar{q}_m$ is the mean of the observed flow; and q_c is the computed flow. A value close to 1 indicates good curve fitting. We observe that E parameter for the MLE is lower than that for SLS. This is another indication that the SLS is superior to the MLE in curve fitting, but not in model fitting.

Also recorded is the d statistic, which is used to test the correlation among measurement errors. It is mathematically defined by

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \qquad (5.4)$$

where e_t is the flow residual at the end of time period t. A large d should indicate a weak correlation among errors while a d close to zero implies that the errors are highly correlated. The last columns of Tables 5.1 to 5.3 demostrate that d statistic decreases with the value of α. This confirms that the degree of correlation increases when α decreases in value.

The above discussion applies for cases where the measurement errors contained in data are auto-correlated. In case that the measurement errors are not auto-correlated or weakly-correlated (e.g. α = 2.125), the new MLE criterion produces less satisfactory results. Various indicators of model fit, such as parameter estimates and $\Sigma SSQ_{com-true}$, show that the SLS criterion gives even relatively better results than the MLE criterion in such cases. Figures 5.3 and 5.8 are the hydrographs for α = 2.125. It can be seen that the simulated hydrograph for the MLE differs farther from the true hydrograph than the SLS simulated hydrograph. Another indication of poor performance in non- or weakly-correlated cases by the MLE is that the function response surface is highly non-convex and the minimum can not be easily identified (see Figure 5.13).

From the above findings, the following conclusions can be drawn:

(1) For cases when data contain auto-correlated errors, the new criterion is able to produce good estimates of the true model parameters. It is true even if a high percentage of measurement errors is introduced in observed outflow data. The performance of the new MLE enhances as the degree of auto-correlation becomes higher. On the other hand, the SLS is unable to give good estimates of the true model parameters. However, if the degree of auto-correlation among measurement errors becomes significantly small, the new criterion will fail to produce satisfactory results. Therefore, to use the new criterion, care

must be taken in ensuring that the measurement errors have the properties discussed in earlier sections.

(2) The traditional tests for goodness of curve fit may not necessarily reflect the goodness of model fit. Previous discussion clearly confirms this point. For this study, comparison of parameter estimates against the true parameters and $\Sigma SSQ_{com-true}$ have been employed for analysis. Since in realistic situations the true parameters and the true flows are unknown, it will not be possible to use $\Sigma SSQ_{com-tru}$ as a measure of fit. However, these tests suggest that the Maximum Likelihood Criterion investigated in this report could be a satisfactory substitute.

(3) One interesting phenomenon observed in the contour maps is that the function response surfaces are non-convex and multiple local optima exist. One possible cause for this phenomenon is related to the model structure. Another possible cause is due to the purely random term a_t in the error model. To resolve this problem, one approach is to use global optimization algorithms [Dixon, 1975; 1978]. An alternative is to use the traditional locally-based optimization algorithms, but starting at different initial points.

CHAPTER SIX

VALIDATION OF THE NEW CRITERION ON THE PUKEITI CATCHMENT OF PUKETURUA BASIN IN NEW ZEALAND

Following the successful application of the Maximum Likelihood Criterion to test problems, it has been used to fit a more realistic model to some measured flows from a catchment in New Zealand. The model simulates outflow from a hillslope by combining the Storage/Discharge model of Sloan and Moore [1984] for the saturated zone with a simple Green and Ampt Piston-flow model for moisture movement in the unsaturated zone. The resultant model was fitted to data from the small Pukeiti subcatchment of the Puketurua Experimental basin in Northland, New Zealand. Ibbitt and Hutchinson [1984], who compared the various Maximum Likelihood Criterion proposed by Sorooshian in 1978 with an empirical one aiming at minimizing water balance errors, concluded that for data recorded in the Puketurua Basin, the Maximum Likelihood approach worked better than the other criterion. Based on their findings, we feel justified to assume that the field data from Pukeiti catchment have the previously described characteristics suited for using the new MLE. That is, the streamflow data contain auto-correlated measurement errors, and that the measurement errors form a stationary time series. In Ibbitt and Hutchinson [1984], the calibration procedures were implemented using data at equal time intervals. As mentioned earlier, the consequences of using the equal time interval data are that the ability of model to estimate runoff from severe events is diminished. This study calibrates the hillslope model on the Pukeiti catchment by taking the advantage of the new criterion which is specially designed for variable time interval data containing auto-correlated measurement errors.

In the following sections, the catchment characteristics are briefly summerized. Then, the catchment model is described. Last, the calibration results are presented along with analysis.

Catchment Characteristics

The Puketurua catchment is an experimental basin established in 1962. The basin has an area of 251 hectares and contains a small sub-catchment, Pukeiti, with an area of 1.4 hectares. The simulation test was done on the Pukeiti catchment. The Puketurua basin sustained a land use change when the vegetation was burned in February,1971, and later was replanted with grass in March of 1972, after which sheep and cattle began grazing the pasture. The numerical simulations were done using data from post-change period 1974.

The soils of the basin are predominantly clay and can develop large cracks during periods of drought when flow at the outlet may cease. During the winter months (April to September), flow is usually continuous, although it drops to low values (0.1 mm/day) if more than several days pass without precipitation.

Description of the Catchment Model and its Fitting Characteristics

In its simplest form the model represents a catchment, as a single rectangular hillslope providing inflow to a channel running along the base of the slope. The saturated behavior of the hillslope is simulated using the 2-dimensional model described by Sloan and Moore [1984], while the behavior of the unsaturated soil on the hillslope is modelled using a 1-dimensional Green and Ampt [1911] piston-flow model based on the work of Mein and Larson [1973] and Clapp et al. [1983].

For the Pukeiti catchment, a single slope unit adequately represents the catchment so that it was unnecessary to use a more complicated model configuration. Where more than one slope unit is needed to adequately represent a catchment, the model can be easily reconfigured to allow this. The model then becomes a spatially distributed model.

In New Zealand, the subdivision of a catchment is based on land resource units. Adjacent land resource units differ because of changes in ground slope, soil type, geology, revegetation type and erosion hazard. Use of the model to date has used ground slope as major discriminator between areas.

Figure 6.1 shows a pictorial representation of the model. The length of slope is L, the depth of the soil is d, and the angle of the slope is β. The water table is assumed to behave like a straight rod pivoted at D, and discharge from the hillslope is given by $h \cdot Ks_h \cdot \sin\beta$ where h is the height of the water table at the bottom end of the slope, and Ks_h is the saturated hydraulic conductivity in the horizontal direction. If the water table is high enough to intersect the slope AB, precipitation falling on the saturated part of the slope is immediately added to the channel. The saturated zone is supplied by the vertical movement of the moisture from the unsaturated zone. This is estimated using a 1-dimensional Green-Ampt piston type model. The depth of the profile is taken to be the mean thickness of the 2-dimensional unsaturated zone in the figure. Infiltration rate is calculated using the moisture content of the top block of the profile, and if this is exceeded by the precipitation, the infiltration excess is added directly to the channel contents. For small catchments, water entering the channel system is assumed to immediately reach the catchment outlet.

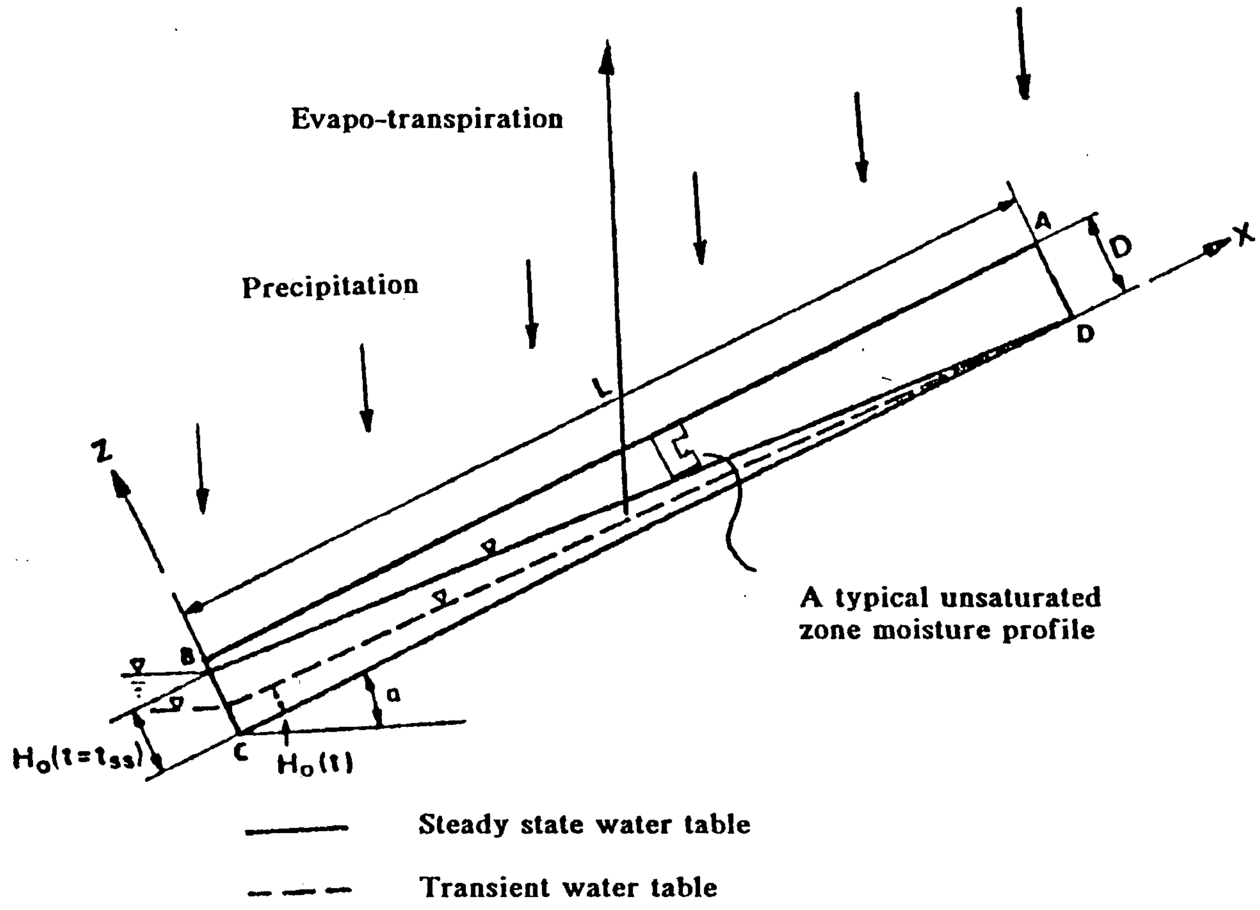

Figure 6.1 Pictorial Representation of the Hillslpoe Model

In its single hill slope configuration, the model uses parameters given in Table 6.1. Many of these can and should be estimated from survey data and maps. Where considerable uncertainty exists, then optimization using precipitation, potential evaporation, and runoff data is used to better define the values that are appropriate.

When the initial choice of parameters fails to provide a satisfactory simulation, "a typical reaction is to bring in new variables, to increase the dimensionality of the system. Invariably, the urge is to manipulate more rather than to comprehend the interaction of fewer" [Kane et al., 1973]. Never was a truer statement made. In the present model, several pitfalls await the unwary. For example, with the hillslope model of Sloan and Moore, the ground slope β always appears in the algebra of the equations with horizontal saturated hydraulic conductivity Ks_h in the combination $Ks_h \cdot \sin \beta$. Consequently, Ks_h and β cannot be separately estimated. Similar but less obvious correlations exist. For instance, if the length of the slope is over-estimated, then to a first approximation, this can be compensated for by increasing the horizontal saturated hydraulic conductivity to preserve the length of time it takes the water to traverse the slope. A similar situation exists between the depth of the soil and the vertical saturated hydraulic conductivity. Consequently, only a subset of the 11 parameters given in Table 6.1 should be optimized on.

Model Calibration and Results

The first step in fitting the model is to identify those model parameters that can be estimated from the catchment's physiography. All New Zealand catchments are documented in the National Land Resources Inventory (LRI) so that information is readily available on catchment area and surface slopes. Also, using soil data in the LRI, realistic estimates can be deduced for hillslope length and soil depth. Of the 11 potential parameter values to be found by optimization, those for area and slope

Table 6.1. Physical parameters used by the hillsope model.

Parameter Number	Description	Units
1	Area of hillslope unit	mm^2
2	Length of the slope	mm
3	Angle of the slope	radians
4	Horizontal saturated hydraulic conductivity of the soil	mm/sec
5	Depth of the soil on the hillslope	mm
6	Saturated soil moisture capacity of the soil	-
7	Exponent in Campbells equation for the unsaturated hydraulic conductivity	-
8	Initial soil moisture content of the unsaturated zone	-
9	Moisture content at wilting point	-
10	Vertical saturated hydraulic conductivity of the soil	mm/sec
11	Initial position of the water table	mm

were excluded while those for length and depth were constrained to lie in narrow ranges. Of the remaining 9 parameters, two represent initial conditions and do not have a significant effect upon the simulated hydrographs after the first few time intervals of non-zero precipitation. They can be handled in one of the two ways: (1) Fix them at realsitic values and ignore the initial discrepancies in the hydrograph by having a "warm-up" period not used in the calculation of the criterion; or (2) Include them as parameters to be found by optimization. The first approach is useful when some experience has been gained with model since it can reduce the cost of running the model. However, since our experience with the model was initially limited we chose to treat both initial conditions as parameters to be found by optimization.

The calibration results are listed in Table 6.2. Three schemes have been used in the calibration process. Scheme One is to test the new MLE using equal time interval data. In this particular test, data used are from 0:00:00 am April 6,1974 to 4:00:00 am of April 7, 1974. The duration of one time interval is fixed to 15 minutes, resulting in 112 time steps. Scheme Two uses the same data with the Maximum Likelihood Criterion for stationary auto-correlated errors presented in Sorooshian [1978]. Scheme Three is designed to test the effectiveness of the new procedures using data collected at unequal time intervals. The data used cover a time span from 7:54:12 am to 23:41:33 pm of April 6, 1974. The major storm events during that day are included for model calibration, so that comparison between Schemes One and Two and Scheme Three can be made.

Figure 6.2 displays the simulation results using Scheme One. Figure 6.2 (a) shows the rainfall process, and Figure 6.2 (b) shows the corresponding observed and simulated hydrographs.

Table 6.2. Calibration results of the hillslope model on the Pukeiti catchment.

Parameter Number	Scheme One	Scheme Two	Scheme Three
1*	1.400×10^{10}**	1.400×10^{10}	1.4×10^{10}
2	0.249×10^{2}	0.249×10^{2}	0.455×10^{2}
3*	0.100	0.100	0.100
4	0.549×10^{-1}	0.549×10^{-1}	0.853×10^{-1}
5	0.111×10^{3}	0.111×10^{3}	0.135×10^{3}
6	0.543	0.543	0.435
7	0.134×10	0.134×10	0.557
8	0.570×10^{-1}	0.570×10^{-1}	0.531×10^{-1}
9	0.534×10^{-1}	0.534×10^{-1}	0.376×10^{-1}
10	0.102	0.102	0.462×10^{-1}
11	0.922×10^{-1}	0.991×10^{-1}	0.996×10^{-1}

*Parameter obtained not from optimization
**For units, see Table 6.1.

Figure 6.3 gives the simulation results from using Scheme Three. Figures 6.4 - 6.6 show some streamflow forecasts using parameters derived from Scheme Three.

Finally, Figure 6.7 displays what happens when the parameter values derived using unequal time intervals are used at equal time intervals.

Discussion of the Results

From Table 6.2, we see that the parameter estimates from Scheme One and Scheme Two have the same values except a minor difference in the initial position of the water table. This illustrates that when the new criterion is applied to equal time interval data, it behaves in the same way as the criterion proposed by Sorooshian in 1978 for handling the case when data contain stationary auto-correlated mea-surement errors.

From the hydrographs, we observe that whatever scheme is used, there exists a discrepancy between the simulated hydrographs and the observed hydrographs. Comparing the observed hydrographs in Figures 6.2 (b) and 6.3 (b), we notice a difference in the magnitude of the peak flow rates. This difference is the result of recording streamflows at equal time intervals. The simulated hydrographs in Figures 6.2 (b) and 6.3 (b) show that in both tests, the highest peak is badly missed. However, we can notice that the peak value of the simulated flows in Figure 6.3 (b) takes on a higher value than that in Figure 6.2 (b). This is an indication of the superiority of using variable time interval data.

In Figure 6.4, the parameter estimates derived from Scheme Three are used to forecast the streamflows between 15:40:51 pm of June 9, 1974 to 21:04:30 pm of June 10, 1974. We notice that the very first peak is missed in the simulated hydrograph. This feature was also evident in the fitting and implied a structural error in the model. But overall, the simulated hydrograph matches very well against the observed hydrograph. Another simulation, Figure 6.5, for the period between 4:06:38 am and 8:24:46 am of August 5, 1974, also shows the simulated hydrograph closely approximates the observed hydrograph.

Figure 6.6 shows the forecast of the hydrological event that occurred between 11:14:00 am and 20:52:47 pm of June 27, 1974. We see that the simulated hydrograph overestimates the observed hydrograph. A probable explanation for the overestimation would be if the initial moisture contents of the soil were too great.

Comparison between Figure 6.7 and Figure 6.2 shows that the unequal time interval parameters were marginally better and implies that the interval length is not implicitly buried in the final parameter values.

The results show that for data at equal time intervals, both the new criterion and that proposed by Sorooshian give comparable results. This implies that the different assumptions made in the derivation of the two criteria are of little practical difference. On variable interval data, the new criterion allows better use of the information contained in the record and better visual simulation results. When used to predict the runoff for data not used in the fitting, the results are visually comparable with those obtained during fitting. Some discrepancies are, however, apparent and these are attributed to possible errors in the model's structure and the initial conditions used for simulations and predictions.

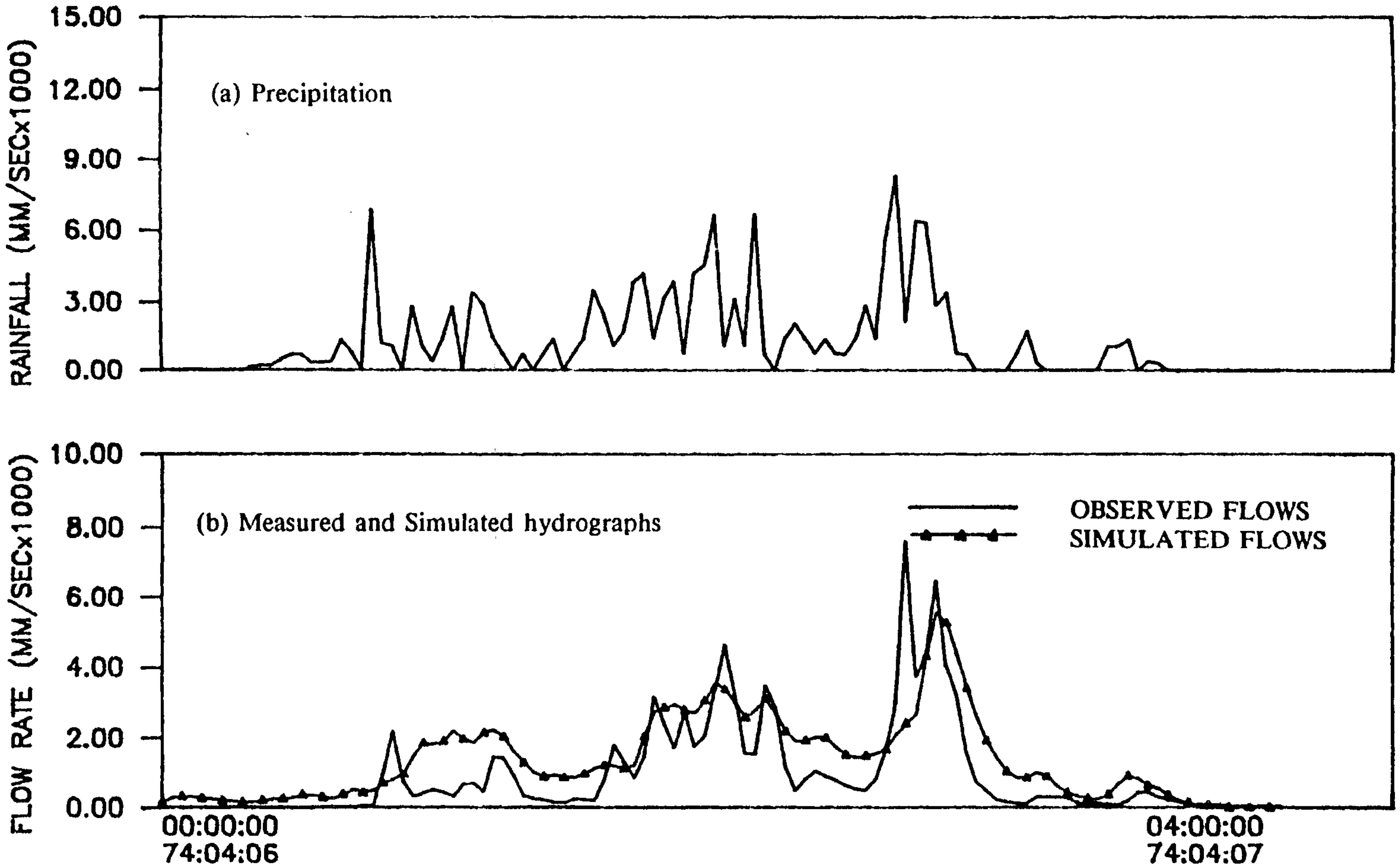

Figure 6.2 Calibration of the Hillslope Model Using Scheme One

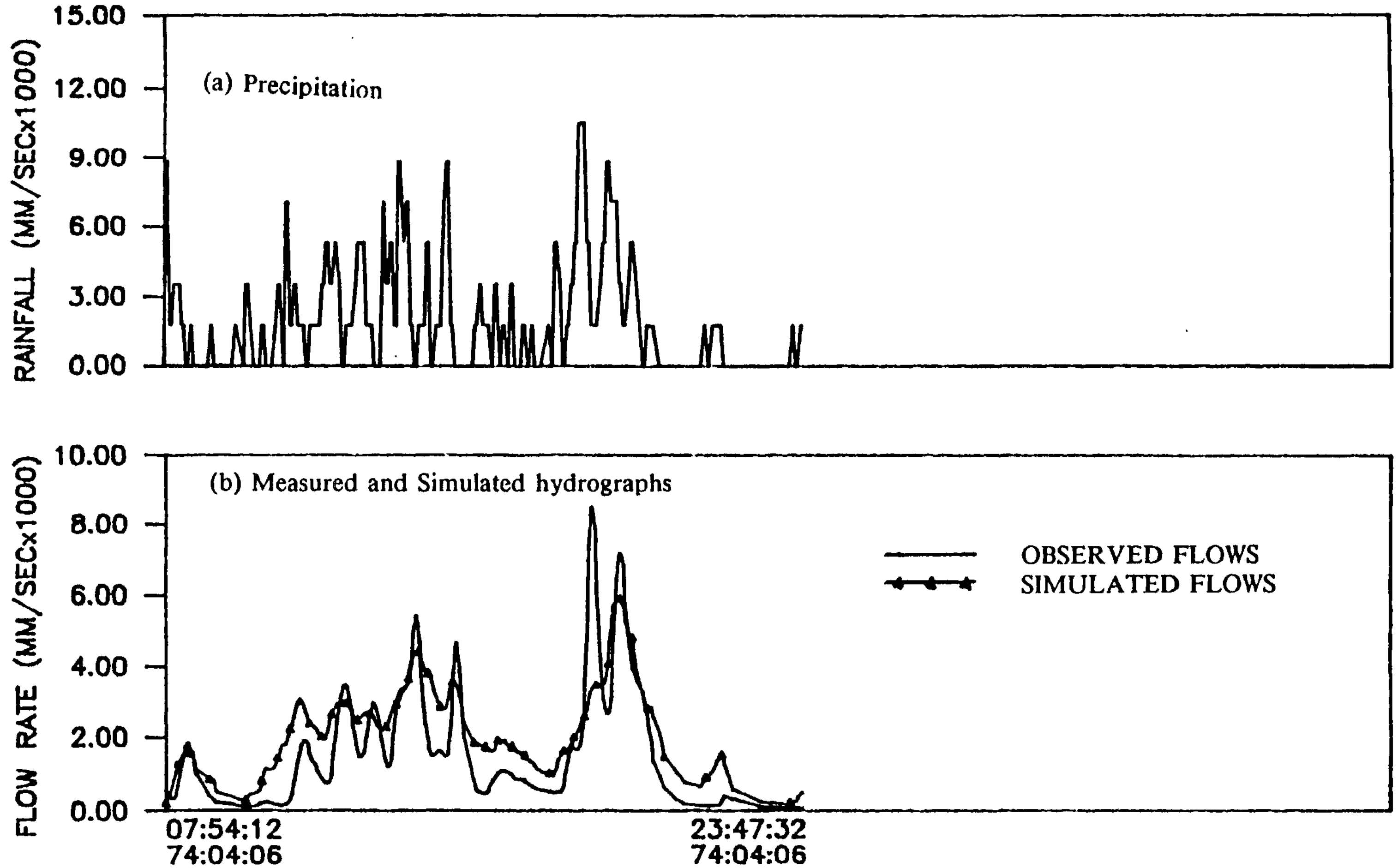

Figure 6.3 Calibration of the Hillslope Model Using Scheme Three

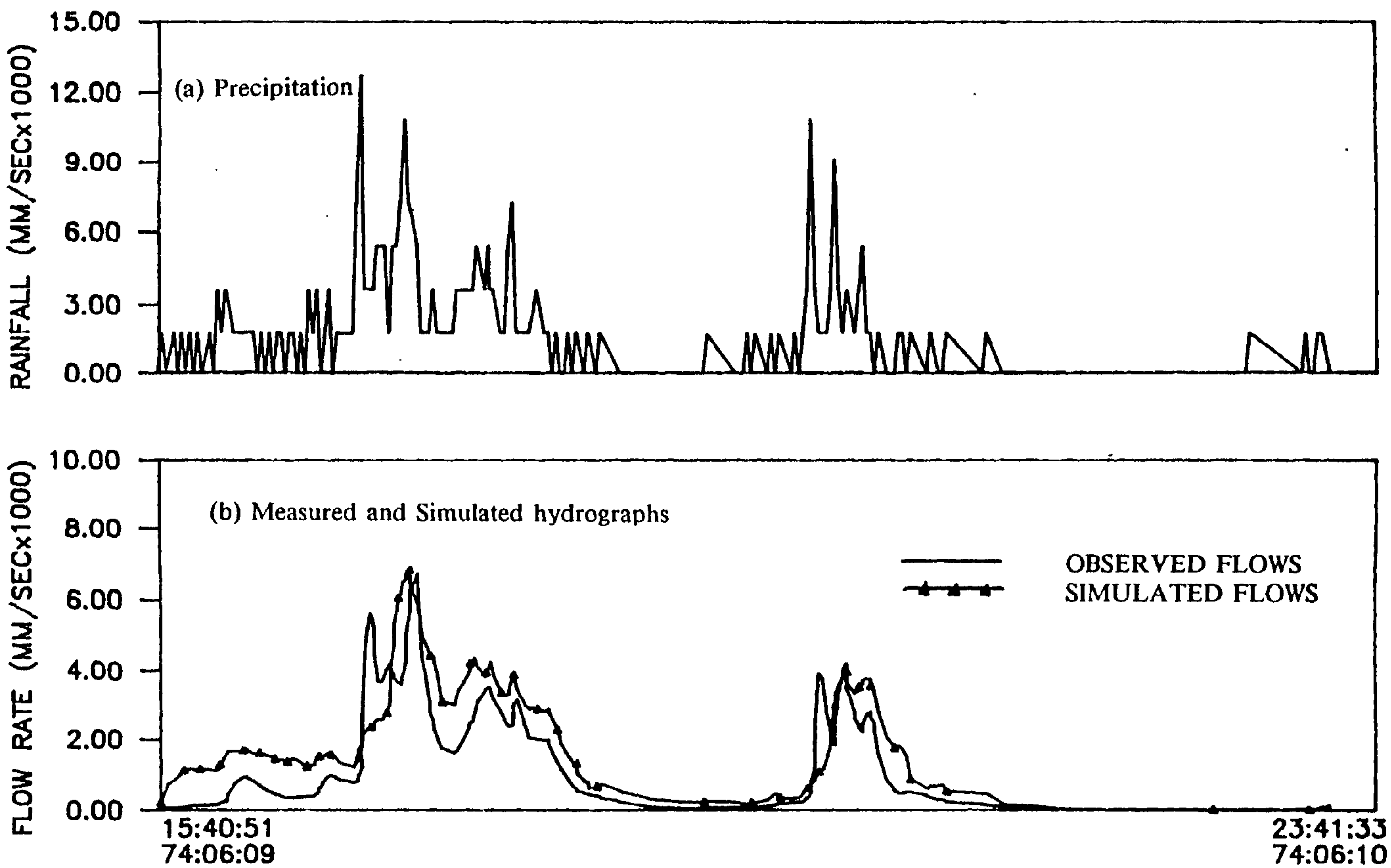

Figure 6.4 Prediction I Using Scheme Three

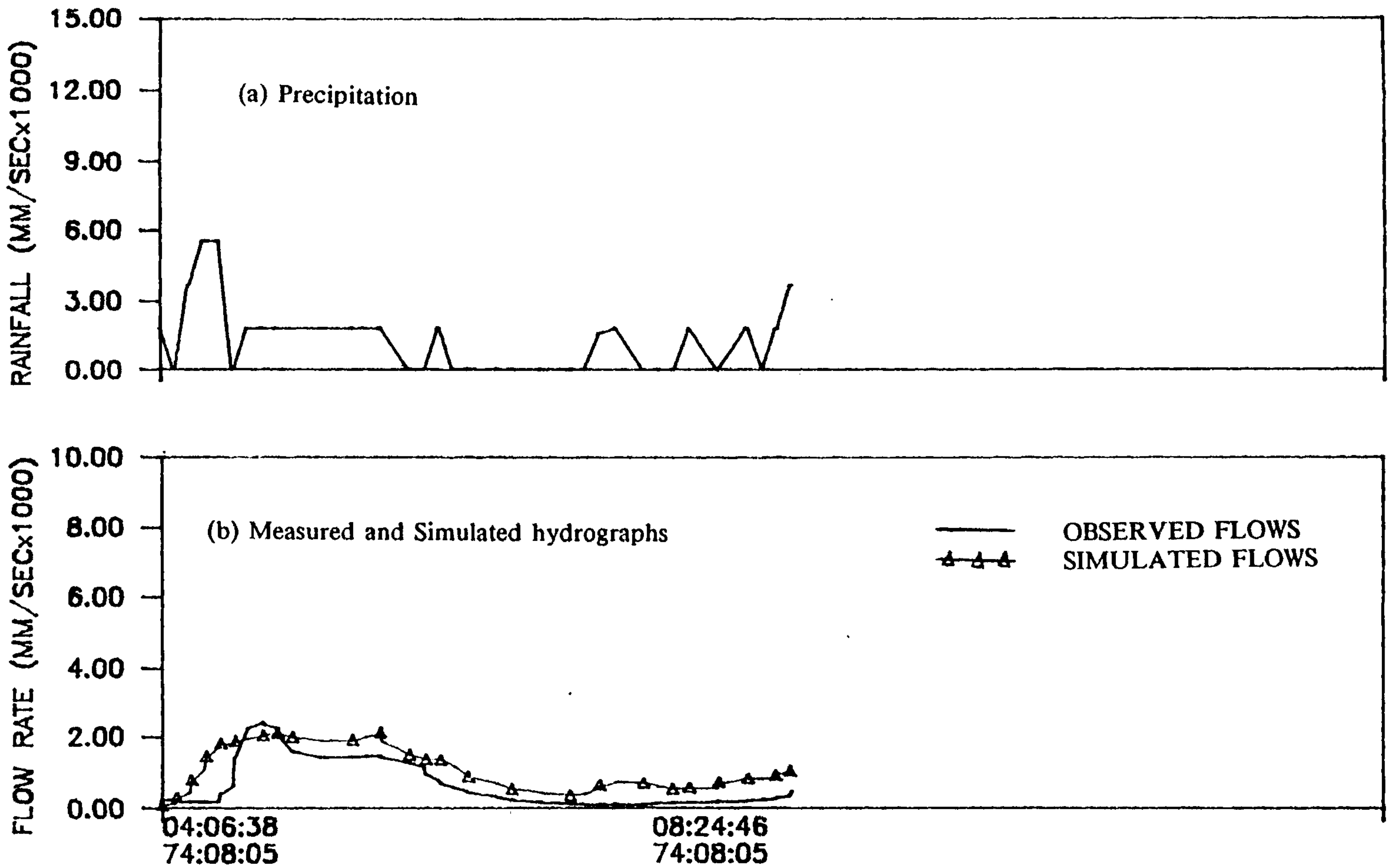

Figure 6.5 Prediction II Using Scheme Three

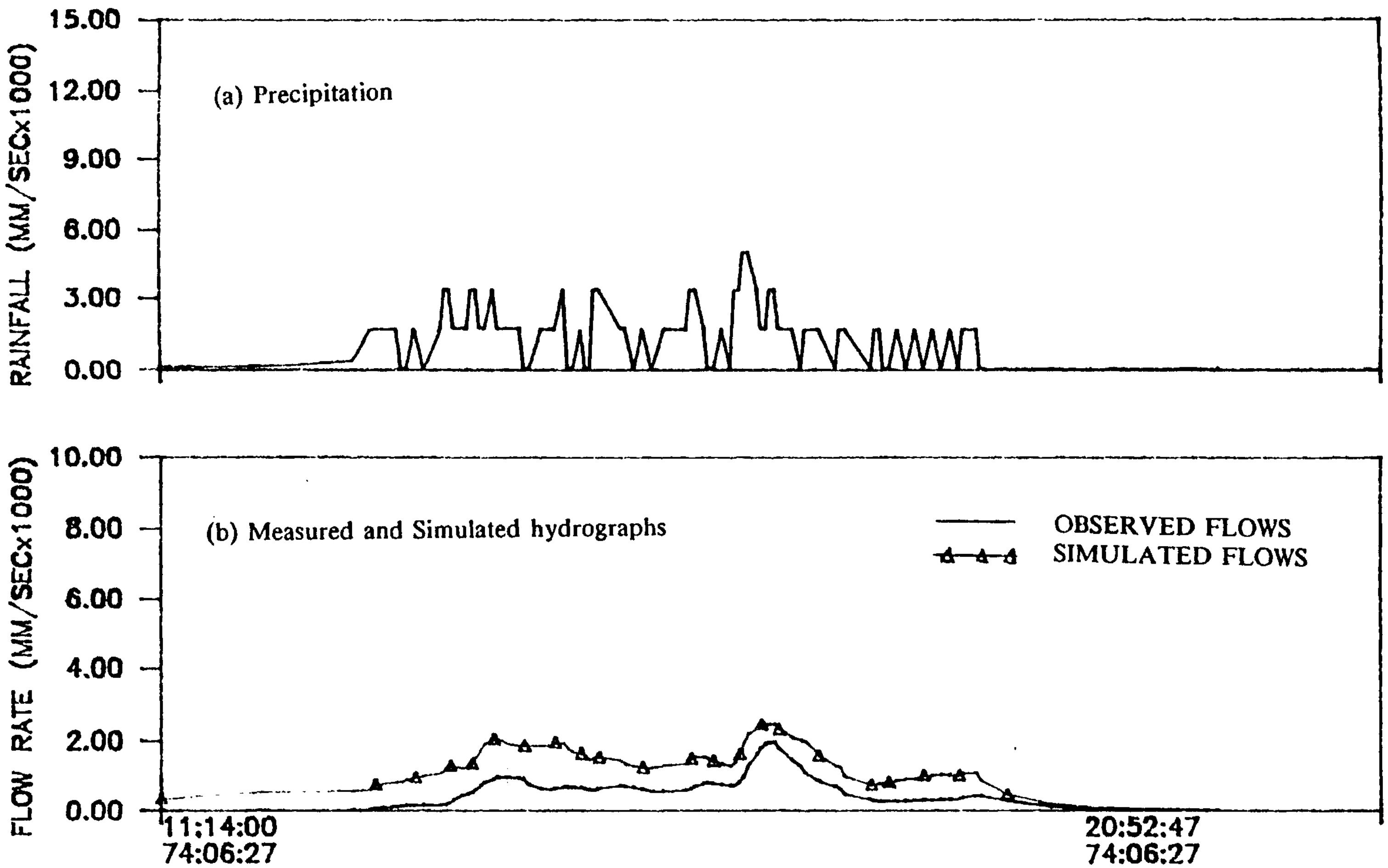

Figure 6.6 Prediction III Using Scheme Three

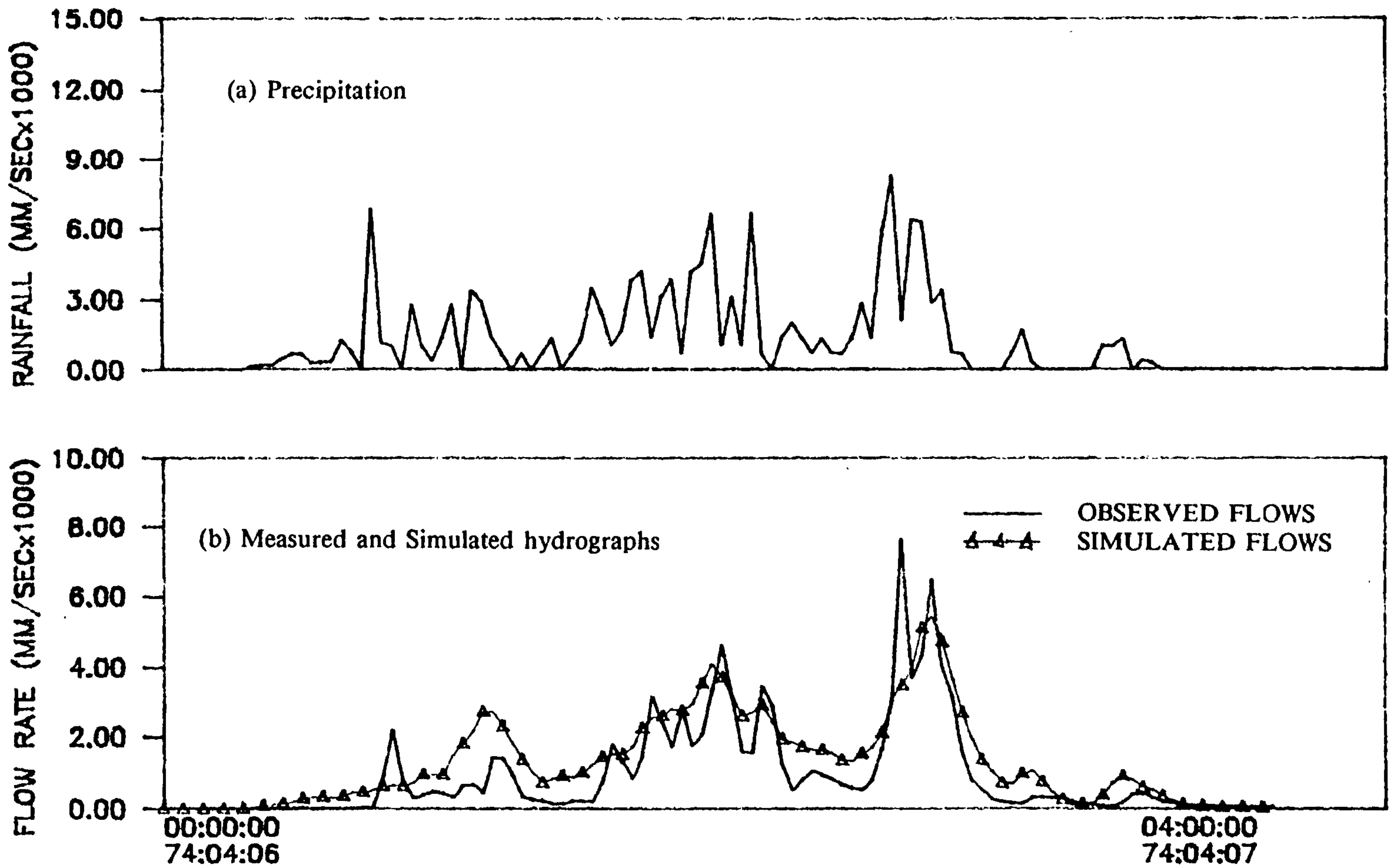

Figure 6.7 Prediction of Equal Time Interval Hydrograph Using Parameters from Scheme Three

CHAPTER SEVEN

CONCLUSIONS

A new Maximum Likelihood Criterion has been successfully derived for use in fitting models to data recorded at unequal time intervals. Initial testing used a simple two-parameter model and synthetic data. It showed that when the observed flows contain serially correlated errors, the fitting routine was able to find the "true" parameter values and generate the "true" hydrograph, whereas use of Simple Least Squares resulted in a curve fitting exercised that largely ignored the underlying "truth" in the data.

One interesting problem was revealed when the new criterion was used on data that did not contain serially correlated errors. In this case, the model fits were inferior to those obtained with the Simple Least Squares criterion. Examination of the model response surfaces showed it to be poorly conditioned when the fitting used the new criterion. The reasons for this unexpected behavior have not been resolved.

The criterion has also been tested with a more complicated model fitted to real data. Initially, the new criterion and that proposed by Sorooshian in 1978 were used with data at equal time intervals. No practical differences in the fits were found.

The next series of tests investigated comparative model performance when the parameters were estimated from the same storm but using data at equal and unequal time intervals. The results indicate that the unequal time interval method of fitting gives better visual results. The evaluation of the parameter values from the fitting made using unequal time intervals was extended by predicting the hydrographs for some previously unsimulated storms. The results show good peak predictions, although some unusual features indicate poor estimates of initial conditions in some cases, and possible structural errors in the models in others.

In the final test, the parameters from the unequal time interval fitting were used reconstruct the fitted storm but at equal time intervals. The results are at least as good as when the model was fitted to equal time interval data and indicate that the simulation time interval has not been implicitly hidden in the parameter values.

The work reported on here shows some encouraging developments that further research in the area of hydrlogic simulation may based on.

REFERENCES

Agiralioglu, N., Water routing on diverging-converging watersheds, J. Hydraul., Div. Amer. Soc. Civil Eng., 107(HY8), 1003-1017, 1981.

Aitken, A.P., Assessing systematic errors in rainfall-runoff models, J. of Hydrology, 20(2), 121-136, 1973.

Alley, W.M., D.R. Dawdy, and J.C. Schaake, Parametric-deterministic urban watershed model, J. Hydraul., Div. Amer. Soc. Civil Eng., 106(HY5), 679-690, 1980.

Bard, Y., Nonlinear parameter estimation, Academic Press, New York, 1974.

Beable, M.E., A simulation method for predicting hydrological effects of land-use changes, Ph.D. Dissertation, Univ. of Canterbury, New Zealand, 1976.

Beven, K. and M.J. Kirkby, Considerations in the development and validation of a simple physically-based, variable contributing area model of catchment hydrology, Surface and subsurface hydrology, 37-51, 1978.

Burnash, R.J.C., K.L. Ferral, and R.A. McGuire, A generalized streamflow system--conceptual modelling for digital computers, Report, Joint Fed. State River Fore-cast Center, U.S. National Weather Service, and Calif. Dept. of Water Resour., Sacramento, 1973.

Chanasyk, D.S., A model to evaluate the hydrologic response to land use changes, Ph.D. Dissertation, Univ. of Alberta, Canada, 1980.

Clapp, R.B., G.M. Hornberger, and B.J. Cosby, Estimating spatial variability in soil moisture with a simplified dynamic model, Water Resour. Res., 19(3), 739-745, 1983.

Clarke, R.T., A review of some mathematical models used in hydrology with obsevations on their calibration and use, J. of Hydrology, 19(1), 1-20, 1973.

Crawford, N.H. and R.K. Linsley, Digital simulation in hydrology: Stanford Watershed Model IV, Stanford Univ. Civil Eng. Tech. Rept. 39, 1966.

Dawdy, D.R., T. O'Donnell, and M. Asce, Mathematical models of catchment behavior, J. Hydraul., Div. Amer. Soc. Civil Eng., 91(HY4), 123-137, 1965.

Dawdy, D.R., J.C. Schaake, and W.M. Alley, Distributed routing rainfall-runoff model, U.S. Geological Survey, Water Resour. Investigation 78-90, 1978.

Diskin, M.H., and E. Simon, A procedure for the selection of objective functions for hydrologic simulation models, J. of Hydrology, 34, 129-149, 1977.

Diskin, M.H. and E.S. Simpson, A quasi-linear spatially distributed cell model for the surface runoff system, Water Resour. Bull., 14, 903-918, 1978

Dixon, I.C.W., and G.P. Szegö, Towards global optimization, North-Holland, Amsterdam, 1975.

Dixon, I.C.W., and G.P. Szegö, Towards global optimization 2, North-Holland, Amsterdam, 1978.

Durbin, J., and G.S. Watson, Testing of serial correlation in Least squares regression, Biometrika, 59(1), 1-19, 1971.

Engman, E.T. and A.S. Rogowski, A partial area model for storm flow synthesis, Water Resour. Res., 10, 464-472, 1974.

Green, W.H., and G.A. Ampt, Studies in soil physics: 1. The flow of air and water through soils, J. Agric. Soils, 4, 1-24, 1911.

Gupta, V.K., Parameter estimation problems caused by model structure: case of conceptual rainfall-runoff models, M.S. Thesis, Case Western Reserve Univ., Jan. 1982.

Gupta, V.K., The identification of conceptual watershed models, Ph.D. Dissertation, Case Western Reserve University, 1984.

Gupta, V.K., and S. Sorooshian, The automatic calibration of conceptual catchment models using derivative-based optimization algorithms, Water Resour. Res., 21(4), 473-485, 1985.

Holtan, H.N. and N.C. Lopez, USDAHL-70 model of watershed hydrology, U.S. Dept of Agric., Agric. Res. Service Tech. Bull., 1435, 1971.

Ibbitt, R.P., Systematic parameter fitting for conceptual models of catchment hydrology, Ph.D. Dissertation, University of London, 1970.

Ibbitt, R.P., Effect of random data error on the parameter values for a conceptual model, Water Resour. Res., 8(1), 70-78, 1971.

Ibbitt, R.P., and T. O'Donnell, Fitting methods for conceptual catchment models, J. Hydraul. Div. Amer. Soc. Civil Eng., 94(HY9), 1331-1342, 1971.

Ibbitt, R.P. and T. O'Donnell, Designing conceptual catchment models for automatic fitting methods, Int. Ass. Hydrol. Soc. Publ., 101, 461-475, 1974.

Ibbitt, R.P., Automation in hydrology with particular reference to data processing and conceptual catchment models, in: Physical Hydrology, New Zealand Experience, Ed(s) Murray, D.L. and Ackroyd, P., 1979.

Ibbitt, R.P. and P.D. Hutchinson, Model parameter consistency and fitting criteria, paper accepted for presentation at the 1984 IFAC World Congress, Budapest, Hungary, 1984.

Kane, J., I. Vertinsky, and W. Thomson, KSIM: A methodology for interactive resource policy simulation, Water Resour. Res., 9(1), 65-79, 1973.

Kuczera, G., Improved parameter inference in catchment models, 1, evaluating parameter uncertainty, Water Resour. Res., 19, 1151-1162, 1983.

Laurenson. E.M., A catchment storage model for runoff routing, J. of Hydrology, 2, 141-163, 1964.

Law, A.M., and W.D. Kelton, Simulation modeling and analysis, McGraw-Hill, New York, 1982.

Mathur, B.S., Natural catchment representation by a series model of linear channels: Part(1), Int. Ass. Hydrol. Sci. Publ., 101, 634-642, 1974a.

Mathur, B.S., Natural catchment representation by a series model of linear channels: Part(2), Int. Ass. Hydrol. Sci. Publ., 101, 643-656, 1974b.

Mein, R.G., and C.L. Larson, Modeling infiltration during a steady rain, Water Resour. Res., 9(2), 384-394, 1973.

Nash, J.E., and J.V. Sutcliffe, River flow forecasting through conceptual models, 1. A discussion of principles, J. of Hydrology, 282-290, 1970.

Rosenbrock, H.H., An automatic method of finding the greatest or least value of a function, The Computer Journal, 3(*), 175-184, 1960.

Ross, B.B., V.O. Shanholtz, D.N. Contractor, and J.C. Carr, A model for evaluating the effects of land use on flood flows, Virginia Polytechnic Inst. and State Univ. Bull., 85, 1978.

Shanholtz, V.O., B.B. Ross, and J.C. Carr, Effect of spatial variability on the simulation of overland and channel flow, Trans. Amer. Soc. Agri. Eng., 24(1), 124-133, 1981.

Sloan, P.G., and I.D. Moore, Modeling subsurface stormflow on steeply sloping forested watersheds, Water Resour. Res., 20(12), 1815-1822, 1984.

Sorooshian, S., Considerations of stochastic properties in parameter estimation of hydrologic rainfall-runoff models, Ph.D. Dissertation, Univ. of Calif., Los Angelos, 1978.

Sorooshian, S., and J.A. Dracup, Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases, Water Resour. Res., 16, 430-442, 1980.

Sorooshian, S., V.K. Gupta, and J.L. Fulton, Evaluation of maximum likelihood estimation techniques for conceptual rainfall-runoff models: influence of calibration data variability and length on model credibility, Water Resour. Res., 19, 251-259, 1983.

Sorooshian, S., and V.K. Gupta, Study and modification of SMA-NWSRFS model: Automatic calibration Aspect, Final Report, Dept. of Hydrol. and Water Resour., Univ. of Ariz., Tucson, Arizona, 1984.

Sutcliffe, J.V., The assessment of random errors in areal rainfall estimation, Bull. Inst. Assoc. Sci. Hydrol., 11, 35-42, 1966.

White, J.K., and A.W. Jaywardene, A distributed and deterministic catchment model using finite elements, Proc. 2nd World Congress on Water Resour., New Delhi, Water for Human Needs, 5, 251-259, 1980.

Wooding, R.A., A hydraulic model for the catchment stream problem, Part I: the kinematic wave theory, J. of Hydrology, 3, 254-267, 1965a.

Wooding, R.A., A hydraulic model for the catchment stream problem, Part II: Numerical solutions, J. of Hydrology, 3, 268-282, 1965b.

Wooding, R.A., A hydraulic model for the catchment stream problem, Part III: Comparison with runoff observations, J. of Hydrology, 4, 21-37, 1966.

Woolhiser, D.A., Deterministic approach to watershed modelling, Nordic Hydrology, 2, 146-166, 1971.

Zaslavsky, D., and G. Sinai, Surface hydrology IV - flow in sloping layered soil, J Amer. Soc. Civil Eng., 107(hy1), 53-64, 1981.